

GERMANISTISCHE LINGUISTIK

Herausgegeben vom
Forschungszentrum
Deutscher
Sprachatlas

223-224 | 2014

Michael Mann (Hrsg.)

Digitale Lexikographie

Ein- und mehrsprachige elektronische
Wörterbücher mit Deutsch:
aktuelle Entwicklungen und Analysen

OLMS



Begründet von Ludwig Erich Schmitt
Friedhelm Debus (Kiel)
Peter O. Müller (Erlangen)
Damaris Nübling (Mainz)
Jürgen Erich Schmidt (Marburg)
Herbert Ernst Wiegand (Heidelberg)

Redaktion: Hanna Fischer

Redaktionsanschrift: 35032 Marburg/Lahn, Hermann-Jacobsohn-Weg 3

E-Mail: gl@deutscher-sprachatlas.de



Georg Olms Verlag
Hildesheim · Zürich · New York
2014

Michael Mann (Hrsg.)

Digitale Lexikographie

Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch:
aktuelle Entwicklungen und Analysen



Georg Olms Verlag
Hildesheim · Zürich · New York
2014

Die Zeitschrift und alle in ihr enthaltenen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Germanistische Linguistik erscheint 4-6 mal jährlich.
Ab 1985 werden die Hefte unter Berücksichtigung der bisher erschienenen fortlaufend durchnummeriert. Vorschlag für die Zitierweise:
GL Heft-Nummer, Jahr, Seite (z. B. GL 79-80. 1985, ...).

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

 ISO 9706

Gedruckt auf säurefreiem und alterungsbeständigem Papier

Umschlagentwurf: Anna Braungart, Tübingen

Herstellung: KM-Druck, 64823 Groß-Umstadt

Alle Rechte vorbehalten

Printed in Germany

© Georg Olms Verlag AG, Hildesheim 2014

www.olms.de

ISBN 978-3-487-15191-5

ISSN 0072-1492

Inhalt

MICHAEL MANN Digitale Lexikographie. Einführung in Thema und Beiträge dieses Bandes	1
CAROLINA FLINZ Wörterbuchbenutzung: Ergebnisse einer Umfrage unter Studenten der Tourismuswissenschaft	13
CHRISTIANE ZEHRER Kriterien für die Terminus-Auswahl aus prozessbezogener Perspektive – ein Beitrag zur Meta-Lexikographie	35
SANDRA DENZER / FRANZISKA HORN Die Arbeitsumgebung des Digitalen Familiennamenwörterbuch Deutschlands. Ein XML-basiertes Redaktionssystem	67
PETER MEYER Entlehnungsketten in einem Internetportal für Lehnwörterbücher. IT-Infrastruktur und computerlexikographischer Prozess in einem Projekt zu polnisch vermittelten Germanismen im Ostslavischen	97
MÓNICA MIRAZO Balsa Überlegungen und Vorschläge zur Strukturierung und Darstellung lexikographischer Information in kontrastiven Onlinewörterbüchern	133

AXEL HEROLD	
Das Wörterbuch als zentrale Komponente eines digitalen lexikalischen Systems	155
ANTJE TÖPEL	
Die Beständigkeit von Wortartikeln im Onlinewörterbuch am Beispiel von <i>elexiko</i>	169
JULIA STEUBE	
Wörterbücher und lexikografisches Wissen von Amateurlexikografen im Internet	201
NATHALIE MEDERAKE	
Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive. Oder: Wenn das Leben dir Zitronen gibt	229
SUSANNE DYKA / EVA SCHARF	
Vergleichende Analyse von Artikelstrukturen deutsch-englischer Onlinewörterbücher	251

MICHAEL MANN

Digitale Lexikographie

Einführung in Thema und Beiträge dieses Bandes

1. Digitale Lexikographie

Der Titelausdruck *Digitale Lexikographie* ist mehrdeutig:¹ Er kann sich zum einen auf den unter Verwendung des Computers vollzogenen **Herstellungsprozess** beziehen, der zu einem Wörterbuch führt („computerunterstützter lexikographischer Prozess“ oder „computerlexikographischer Prozess“ nach WIEGAND (1998, 153ff.)). Er kann zum anderen das Produkt, das **elektronische Wörterbuch**, mit seinen Funktionalitäten und Inhalten in den Fokus rücken.² Hier kann wiederum zwischen dem Wörterbuch als Software, das auf dem Computer des Benutzers installiert oder mittels eines Browsers von einem Server abgerufen und angezeigt wird (ersteres üblicherweise als *Offline-*, letzteres als *Online-Wörterbuch* bezeichnet), unterschieden werden, und dem Wörterbuch als relativ untrennbarem Hardware-Software-Fabrikat, wie es etwa die tragbaren sog. *elektronischen Taschenübersetzer* darstellen. Außerdem kann das Produkt ein sog. *Maschinenwörterbuch* sein (vgl. ebd., 240f.), das in seiner Funktion als Wörterbuch nicht durch Menschen genutzt wird, sondern dessen Daten von sprachverarbeitenden Systemen mit unterschiedlichen Zielsetzungen wie Spracherkennung, Textkorrektur

¹ Für die Einladung, den vorliegenden Band dieses Titels und Themas herauszugeben, möchte ich mich bei den Reihenherausgebern der *Germanistischen Linguistik* bedanken.

² Die Attribute *digital* und *elektronisch* in Ausdrücken wie *digitale Lexikographie/elektronische Lexikographie*, *digitales Wörterbuch/elektronisches Wörterbuch* usw. werden synonym verwendet.

etc. abgerufen werden (vgl. etwa die Beiträge in Kapitel XV zu „Electronic dictionaries for machine use“ in GOUWS u. a. (Hrsg.) (2013)).

In diesem Band werden sowohl Aspekte des digitalen lexikographischen Prozesses behandelt als auch digitale Wörterbücher als Produkte vorgestellt und analysiert, wobei sich alle konkret behandelten Wörterbücher der Kategorie der *Online-Wörterbücher* (= *Internet-Wörterbücher*)³ im weiteren Sinne, d. h. inklusive enzyklopädisch ausgerichteter Nachschlagewerke, zurechnen lassen. Dies kann die Einschätzung aus TARP (2012, 253) unterstützen, dass „the Internet seems to have consolidated itself as the dominant electronic platform for dictionaries“; nach FUERTES-OLIVERA (2013) ist die Situation für (elektronische) Enzyklopädien vergleichbar. *Digitale Lexikographie* (für menschliche Benutzerinnen) ist demnach zum heutigen Stand weitgehend gleichbedeutend mit *Online-Lexikographie*. Nachschlagewerke, die ausschließlich auf CD/DVD oder als Download angeboten werden, wie ⁴DUDEN-GWDS (2012), stellen im PC-Bereich derzeit die sprichwörtlichen Ausnahmen dar.

Der Typ der Online-Wörterbücher wird bereits in der häufig zitierten, frühen Typologie für elektronische Wörterbücher von LEHR (1996) genannt und ist heute etabliert; in LEW (2011) stellt er den Ausgangspunkt weiterer Unterscheidungen dar. Dieser Typ kann, wie oben angedeutet, auch für den vorliegenden Band übernommen werden, wenn auch generell die strikte Trennung von Online- und Offline-Wörterbüchern aufgebrochen und die Typologie durch einen hybriden Subtyp ergänzt werden muss: Bereits heute liegen Wörterbuch-Apps für mobile Endgeräte vor, die Charakteristika sowohl von Online- wie von Offline-

³ *Online-Wörterbuch* wird meist – und auch in diesem Band – gleichbedeutend mit *Internet-Wörterbuch* verwendet, obwohl auch die Intranet-Wörterbücher (dazu etwa SIMONSEN 2004) in Hinblick auf ihre Client-Server-Architektur und Funktionsweise Online-Wörterbücher sind, auf die jedoch nicht öffentlich aus dem WWW zugegriffen werden kann. Hinsichtlich ihrer technisch-lexikographischen Möglichkeiten bestehen zwischen Internet- und Intranet-Wörterbüchern jedoch so wenige Unterschiede, dass Aussagen über Internet-Wörterbücher üblicherweise auch auf Intranet-Wörterbücher übertragbar sind.

Wörterbüchern aufweisen. Die spezifische App muss in diesen Fällen heruntergeladen und auf dem Endgerät installiert werden wie bei einem Offline-Wörterbuch; die lexikographischen Daten werden jedoch beim Aufruf über das Internet geladen wie bei einem Online-Wörterbuch. Da Wörterbuch-Apps in diesem Band nicht explizit thematisiert werden – die eingehende Analyse der Inhalte, Strukturen und der Benutzung von Wörterbuch-Apps bleibt ein Desideratum der Wörterbuchforschung –, wird diese Problematik jedoch an dieser Stelle nicht weiter vertieft. Ebenso soll hier nur cursorisch erwähnt werden, dass die oben genannte Dominanz von Online-Wörterbüchern in der digitalen Lexikographie zumindest im mobilen Bereich noch nicht zementiert zu sein scheint. Es dürfte interessant werden, die weitere Entwicklung in diesem Marktsegment zu verfolgen, wo digitale Wörterbücher auch als Apps verfügbar sind, die komplett auf Smartphone oder Tablet heruntergeladen und somit anschließend offline genutzt werden können. Bei tragbaren Geräten ist ein solches Verfahren sinnvoll, da eine drahtlose Internetverbindung nicht überall zur Verfügung steht bzw. mit Folgekosten verbunden sein kann. Sollten sich derartige Angebote durchsetzen, würden Offline-Wörterbücher wieder an Boden gewinnen.

Der Erfolg der Online-Produkte im PC-Bereich ist jedoch auch relativ einfach zu erklären: Aus Sicht der Benutzer sind sie direkt an jedem modernen Computerarbeitsplatz abrufbar und sie sind meist kostenlos; aus Sicht der Lexikographinnen garantieren sie eine weite Verbreitung des Produkts, sind mit relativ einfachen Mitteln und mittels Standardlösungen realisierbar, bieten jedoch auch das Potenzial vieler technischer Möglichkeiten. Diese Möglichkeiten sind bereits an anderer Stelle ausführlich erörtert worden (vgl. etwa STORRER/FREESE 1996; STORRER 1998; LEMBERG 2001; DE SCHRYVER 2003; ENGELBERG/LEMNITZER 2009, 166ff.; SCHMITZ 2013; STORRER 2013); nur stichwortartig und beispielhaft genannt werden sollen daher die zentralen und medientypischen Aspekte der Offenheit für Aktualisierungen, der Vernetzung mit anderen Internetseiten, der Anbindung an Textcorpora, der Multimedialität, der Variabilität und Interaktivität, der Benutzeradaptivität und der Kollaborativität.

Die Einschätzungen darüber, ob Online-Wörterbücher zum jetzigen Stand noch immer überwiegend als digitale Reproduktionen von zugrunde liegenden Printwörterbüchern einzustufen sind (so vor über zehn Jahren das Urteil in HERBST/KLOTZ (2003, 251)) oder ob sie inzwischen direkt für das elektronische Medium konzipiert und erarbeitet werden, somit das technische Potenzial besser ausgeschöpft wird und sich auch ihr Aufbau deutlicher von einem gedruckten Exemplar unterscheidet, gehen auseinander. So greift etwa MÜLLER-SPITZER (2013, 367) eine Zustandsbeschreibung aus TARP (2009, 21) auf, der zufolge digitale Wörterbücher weiterhin „are more or less produced as exact copies of their printed counterparts“ – dies ist einer von mehreren Punkten, aufgrund derer TARP eine *Identitätskrise der Lexikographie* konstatiert (vgl. ebd.). Auch Analysen wie MANN (2010) haben gezeigt, dass zahlreiche bekannte und vielgenutzte Online-Angebote zur Zeit der Untersuchung noch wenige der einschlägigen Möglichkeiten des elektronischen Mediums ausnutzen. Andererseits stellen GOUWS/SCHWEICKARD/WIEGAND (2013, 10) fest: „There are fewer digitalized Internet dictionaries with printed dictionaries as a basis [...] than Internet dictionaries conceptualized for on-line use“. Diese unterschiedlichen Aussagen deuten darauf hin, dass sich die digitale Lexikographie in einer Umbruchphase befindet und dass sie das Print-Erbe tatsächlich langsam hinter sich lässt. Die negative Interpretation von TARP (s. o.) wird hier bewusst nicht aufgegriffen, zumal auch in den Beiträgen dieses Bandes Beispiele für innovative und darüber hinaus in der praktischen Umsetzung befindliche lexikographische Lösungen präsentiert werden.

2. Überblick über die Beiträge dieses Bandes

Alle Beiträge dieses Bandes beschäftigen sich mit Aspekten der digitalen Lexikographie mit Deutsch als beteiligter Sprache. In diesem Bereich herrscht eine große Vielfalt in den lexikographischen Projekten und Unternehmungen vor: Es gibt Wörterbücher, die ganz am Anfang des lexikographischen Prozesses stehen (zu diesem WIEGAND 1998, 134ff.; KLOSA 2013), und solche, die relativ weit fortgeschritten sind;

die Wörterbücher lassen sich ganz unterschiedlichen Wörterbuchtypen mit unterschiedlichem Inhalt, Aufbau, unterschiedlicher Funktion und unterschiedlichen Adressatengruppen zuordnen; und es liegen hinsichtlich der Zahl der beteiligten Lexikographen und hinsichtlich der materiellen Ausstattung ganz unterschiedliche Voraussetzungen vor. Ziel dieses Bandes ist es daher nicht, im Gesamt aller Beiträge „den“ digitalen lexikographischen Prozess abzubilden, der in der einheitlichen Form, die solch eine Formulierung nahelegt, in der lexikographischen Realität gar nicht existiert.⁴ Vielmehr erlauben die hier versammelten Beiträge Einblicke in unterschiedliche Herangehensweisen zu aktuellen Themen, Problemfeldern und Herausforderungen während unterschiedlicher Projektphasen bei den verschiedenen behandelten Online-Wörterbüchern; daneben werden aktuelle Analysen der Wörterbuchforschung zur digitalen Lexikographie vorgestellt. Alle Beiträgerinnen und Beiträger gehören zu dem großen akademischen Segment der Nachwuchswissenschaftler in der Promotions- oder Postdoc-Phase und präsentieren hier aktuelle Forschungsergebnisse.⁵

Die Beiträge lassen sich inhaltlich in drei Großbereiche gliedern: Beiträge (i) zur Planung und Vorbereitung von Wörterbüchern, (ii) zur Erstellung und (Weiter-)Entwicklung von digitalen Wörterbüchern, sowie (iii) zur Analyse solcher Wörterbücher, die bereits publiziert vorliegen; dabei ist die teilweise fließende Grenze zwischen (i) und (ii) zu berücksichtigen.

⁴ Eine stärker linear ausgerichtete, schematischere Darstellung ist in der Abschlusspublikation des Wissenschaftlichen DFG-Netzwerks „Internetlexikografie“ (<http://www.internetlexikografie.de>) zu erwarten, die derzeit erarbeitet wird.

⁵ Die Anordnung der Beiträge folgt grob dem in ihnen jeweils beschriebenen Stand im Verlauf des lexikographischen Prozesses. Alle Beiträge sind jedoch für sich eigenständig, was sich auch in individuellen Ausdrucksweisen etwa hinsichtlich der unterschiedlichen Verwendung immer generisch zu verstehender männlicher/weiblicher Personenbezeichnungen oder bestimmter Schreibvarianten (z. B. *Lexikographie* vs. *Lexikografie*) ausdrückt. An dieser Stelle sei auch den Beiträgerinnen und Beiträgern sehr herzlich für ihre Mitwirkung am vorliegenden Band gedankt.

(i) *Beiträge zur Planung und Vorbereitung von Wörterbüchern:*

CAROLINA FLINZ (Università di Pisa) eröffnet diesen Großbereich mit ihrem Beitrag „**Wörterbuchbenutzung: Ergebnisse einer Umfrage unter Studenten der Tourismuswissenschaft**“. Die hier vorgestellte Internet-Umfrage und ihre Ergebnisse dienen zur Vorbereitung des deutsch-italienischen Online-Fachwörterbuchs *Tourlex* über die Fachsprache des Tourismus. Abgedeckt werden Aspekte der Wörterbuchbenutzer, der Benutzungssituation, des Wörterbuchformats sowie des speziellen fachlichen Zusammenhangs. Neben Ergebnissen, die im Einklang mit anderen Untersuchungen zur Wörterbuchbenutzung stehen – etwa der Bevorzugung zweisprachiger gegenüber einsprachigen Wörterbüchern – liefert die Befragung auch überraschende Antworten, so etwa die hohe Affinität der Teilnehmenden zu gedruckten Wörterbüchern.

CHRISTIANE ZEHRER (Universität Hildesheim) behandelt in „**Kriterien für die Terminus-Auswahl aus prozessbezogener Perspektive – ein Beitrag zur Meta-Lexikographie**“ den Aspekt der Ermittlung von Termkandidaten (und damit Lemmakandidaten) in der Gegenüberstellung Mensch vs. Maschine. Der Beitrag, der seine Wurzeln in der Terminologielehre hat, diskutiert zunächst den traditionellen *Terminus*-Begriff und erweitert das Konzept *Terminus* um einen funktionalen, prozessorientierten Aspekt, indem es in einen Arbeitszusammenhang eingebettet und die handelnde Person einbezogen wird. Auf dem Hintergrund dieses erweiterten *Terminus*-Konzepts werden die Ergebnisse einer vergleichenden Studie vorgestellt, in der zum einen von studentischen Studienteilnehmern und zum anderen von einem computerlinguistischen Algorithmus solche Ausdrücke aus einem Fachtextcorpus extrahiert worden sind, die als Termini aufgefasst werden und damit potenzielle Lemmata in einem Fachwörterbuch darstellen. Die Studie zeigt Unterschiede der intellektuellen gegenüber der maschinellen Termkandidatenselektion auf.

(ii) *Beiträge zur Erstellung und (Weiter-)Entwicklung digitaler Wörterbücher:*

SANDRA DENZER und FRANZISKA HORN (Technische Universität Darmstadt) erläutern „**Die Arbeitsumgebung des Digitalen Familiennamenwörterbuch Deutschlands. Ein XML-basiertes Redaktionssystem**“. Dieses große Online-Spezialwörterbuchprojekt stellt aufgrund seines Gegenstandsbereichs und dessen Eigenschaften – Familiennamen in Deutschland mit unterschiedlicher Benennungsmotivik, Herkunft etc. – sowie aufgrund seiner Erarbeitungsbedingungen – insbesondere der großen Datenmenge bei rund 200.000 im Rahmen des XML/TEI-Formats zu bearbeitenden Namen und mehreren dafür zuständigen Lexikographinnen – hohe Anforderungen an die Bearbeitungswerkzeuge. Der Beitrag gibt einen Einblick in die Hintergründe, Notwendigkeiten und Funktionalitäten des verwendeten Editors und der Datenbank sowie in die Mechanismen zur Projektverwaltung wie Rechtemanagement, Suchmöglichkeiten u. a. m.

PETER MEYER (Institut für Deutsche Sprache, Mannheim) gibt in seinem Beitrag „**Entlehnungsketten in einem Internetportal für Lehnwörterbücher. IT-Infrastruktur und computerlexikographischer Prozess in einem Projekt zu polnisch vermittelten Germanismen im Ostslavischen**“ Einblicke in die Arbeit am *Lehnwortportal Deutsch*, dessen Datenmodellierung und -haltung am Institut für Deutsche Sprache durchgeführt wird. Wörterbuchdaten zu deutschen Lehnwörtern in den drei Sprachen Russisch, Weißrussisch und Ukrainisch werden hier derart vernetzt, dass eine Suche nach Entlehnungsketten von Sprache zu Sprache ermöglicht wird. Ein Schwerpunkt des Beitrags liegt auf der Erläuterung der Modellierung und Weiterverarbeitung der lexikalischlexikographischen Daten (Angaben, Mikrostrukturausschnitte), die im Sinne der objektorientierten Programmierung als Objekte mit Attributen und Methoden implementiert sind und im XML-Format gespeichert werden. Die Entlehnungsverhältnisse werden in der Form von gerichteten azyklischen Graphen mit Knoten (Wörter) und Kanten (Beziehungen zwischen den Wörtern) dargestellt, deren Eigenschaften anschaulich dargelegt werden und deren Funktionsweise anhand eines Recher-

chebeispiels demonstriert wird. Einen weiteren Schwerpunkt bildet die Darstellung des lexikographischen Prozesses hinsichtlich der Exzerpierungs- und Endredaktionsphase, wobei insbesondere zu berücksichtigen ist, dass die auszuwertenden Wörterbücher teilweise noch nicht abgeschlossen sind.

MÓNICA MIRAZO Balsa (Universidade de Santiago de Compostela) präsentiert „**Überlegungen und Vorschläge zur Strukturierung und Darstellung lexikographischer Information in kontrastiven Onlinewörterbüchern**“ anhand eines in Arbeit befindlichen, bidirektional spanisch-deutschen Wörterbuchs zur Substantivvalenz (*CSVEA*). Dieses auf bestehenden Wörterbüchern und eigens durchgeführten Corpusanalysen basierende Wörterbuch wird eine Lücke in der spanischen und kontrastiven Lexikographie schließen und soll dafür sorgen, dass der Typ des Valenzwörterbuchs beim spanischsprachigen Publikum bekannter wird und häufiger genutzt wird. Die Daten (syntaktische, semantische und morphologische Information, Belege, Äquivalente, Kommentare und Hinweise zu Testverfahren) werden nach Benutzerinteressen und Benutzungssituationen modelliert. Im Beitrag wird die Benutzeroberfläche des Online-Wörterbuchs skizziert, die modular einen Überblick über die unterschiedlichen Bedeutungsvarianten eines Substantivs, die Beschreibung der ausgewählten Bedeutungsvariante, die Beschreibung der spezifischen Aktanten, die Kombinationsmöglichkeiten dieser Aktanten sowie Kommentare zum kontrastiven Vergleich anzeigen kann. Aufgrund der modularen und datenbankbasierten Konzeption ist auch eine spätere Erweiterung auf andere Sprachen möglich und angedacht.

AXEL HEROLD (Berlin-Brandenburgische Akademie der Wissenschaften, Berlin) stellt in seinem Beitrag „**Das Wörterbuch als zentrale Komponente eines digitalen lexikalischen Systems**“ grundlegende Eigenschaften solcher digitaler lexikalischer Systeme (DLS) vor und veranschaulicht diese am Beispiel des *Digitalen Wörterbuchs der deutschen Sprache (DWDS)*. Dieses integriert sowohl mehrere (potenziell) eigenständige Wörterbücher als auch Corpora und lexikographische Werkzeuge, wobei letztere wiederum zur Weiterbearbeitung der Wör-

terbücher genutzt werden: „Der lexikografische Arbeitsplatz und das DLS gehen hier – jedenfalls zum Teil – ineinander über“ (HEROLD, S. 163). Die Weiterbearbeitung, aber besonders auch die automatische Vernetzung von Artikeln innerhalb des DLS erfordert eine konsistente Bearbeitung und Datenvorhaltung – diesen Aspekten widmet sich der Beitrag abschließend.

ANTJE TÖPEL (Institut für Deutsche Sprache, Mannheim) behandelt „**Die Beständigkeit von Wortartikeln im Onlinewörterbuch am Beispiel von *elexiko***“, und diesbezüglich insbesondere den Aspekt der Überarbeitung einmal veröffentlichter Wörterbuchartikel. Obwohl eine Überarbeitung und Aktualisierung der Inhalte, wie sie in der Literatur zu Internet-Wörterbüchern häufig als Vorteil des Mediums genannt und entsprechend gefordert wird, auch bei *elexiko* angestrebt war, sind entsprechende Pläne aufgrund begrenzter personeller Ressourcen und anderer Schwerpunktsetzung (noch) nicht umgesetzt worden. Nach einem Hinweis auf die notwendige Versionierung und Archivierung früherer Inhalte im Falle einer Überarbeitung werden Gründe für Veränderungen an Wörterbuchartikeln erläutert: sich verändernde (wachsende) Erfahrungen der bearbeitenden Lexikographinnen, ein veränderliches Textcorpus als Grundlage, sowie der Sprachwandel (Mikrodiachronie). Die Auswirkungen der beiden letztgenannten Aspekte werden an frühen Beispielen aus *elexiko* demonstriert, zu denen auf der Basis aktueller Corpusdaten möglicher Überarbeitungsbedarf festgestellt wird.

(iii) *Analysen veröffentlichter digitaler Nachschlagewerke:*

JULIA STEUBE (Pädagogische Hochschule Luzern) stellt in ihrem Beitrag „**Wörterbücher und lexikografisches Wissen von Amateurlexikografen im Internet**“ eine Untersuchung zu sachbezogenen Online-Wörterbüchern und deren nicht professionell (beruflich) lexikografisch arbeitenden Autorinnen und Autoren vor. Nach einer terminologischen Diskussion, die die Wahl des Ausdrucks (*lexikographischer*) *Amateur* für diese Autoren begründet, werden Ergebnisse einer Analyse amateurlexikographischer Nachschlagewerke vorgestellt, wobei u. a. das Zugriffsangebot der Nachschlagewerke, die Struktur des Datenan-

gebots, Verweise und deren Hypertextualisierung, Multimedialität, das Vorhandensein von Web-2.0-Elementen und die Abgeschlossenheit des Wörterbuchs in quantitativer und qualitativer Hinsicht in Betracht gezogen werden. Anschließend werden Antworten einer schriftlichen Befragung der Amateurlexikographen bzgl. der Quellen ihres lexikographischen Wissens und ihres Wissens über die Lexikographie als professionelle Tätigkeit präsentiert und erläutert. Die Befragung ergab beispielsweise, dass Amateurlexikographen so gut wie keine Vorstellung von professioneller Lexikographie und von deren wissenschaftlicher Untersuchung haben, was auch als Aufforderung an die professionelle Lexikographie und die akademische Wörterbuchforschung zu einer besseren Kommunikation ihrer Tätigkeiten nach außen hin interpretiert werden kann.

NATHALIE MEDERAKE (Akademie der Wissenschaften zu Göttingen) gibt in ihrem Beitrag „**Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive. Oder: Wenn das Leben dir Zitronen gibt ...**“ Einblicke in die textuelle Entwicklung von Artikeln dieser kollaborativ erstellten Internet-Enzyklopädie. Nach einer Einordnung von Wikipedia-Artikeltexten relativ zu Textmodellen der Textlinguistik liegt der Fokus auf der Thematik der *Aushandlungskohärenz* bei der Artikelbearbeitung im Laufe der Zeit und auf einer Methode der Erfassung der Textdynamik in Anlehnung an das Matrixframe-Konzept. Diese Methode wird im Rahmen einer Fallbeispielanalyse zum Artikel *Zitronenpresse* angewendet und so u. a. veranschaulicht, dass verschiedene Themenaspekte und die entsprechenden Artikelabschnitte eine unterschiedliche Bearbeitungsintensität zeigen, bevor ein Konsens hergestellt ist.

SUSANNE DYKA (Universität Bremen/Friedrich-Alexander-Universität Erlangen-Nürnberg) und **EVA SCHARF** (Friedrich-Alexander-Universität Erlangen-Nürnberg) untersuchen in ihrem Beitrag „**Vergleichende Analyse von Artikelstrukturen deutsch-englischer Onlinewörterbücher**“ diesen wahrscheinlich mit am häufigsten benutzten Wörterbuchtyp. Für fünf bekannte Exemplare (*Leo*, *dict.cc*, *Beolingus*, *Linguee* und *Pons*) werden Makrostruktur bzw. Zugriffsmöglichkeiten

und Mikrostruktur formal verglichen und im Hauptteil inhaltlich geprüft: Die Analyse der Suchergebnisse berücksichtigt insbesondere Kollokationen sowie pragmatische Aspekte des Sprachgebrauchs und entsprechende Äquivalentdifferenzierung, aber etwa auch Aspekte der lexikographischen Beschreibungssprache wie verwendete grammatische Bezeichnungen. Ein eigener Abschnitt beschäftigt sich mit der Suche nach phraseologischen Ausdrücken. Die Analyse zeigt Gemeinsamkeiten und Unterschiede zwischen den Wörterbüchern auf und benennt systematische Fehler in der lexikographischen Behandlung sowie inadäquate Mechanismen beim Einsatz der medialen Möglichkeiten.

Literatur

- DE SCHRYVER, GILLES-MAURICE (2003): Lexicographers' Dreams in the Electronic-Dictionary Age. In: *International Journal of Lexicography* 16/2, 143–199.
- ⁴DUDEN-GWDS (2012): Duden. Das große Wörterbuch der deutschen Sprache. 4., vollständig überarbeitete Auflage. Hrsg. von der Dudenredaktion. Mannheim [CD-ROM].
- ENGLBERG, STEFAN/LEMNITZER, LOTHAR (2009): Lexikographie und Wörterbuchbenutzung. 4., überarbeitete und erweiterte Auflage. Tübingen (Stauffenburg Einführungen. 14). [1. Auflage 2001.]
- FUERTES-OLIVERA, PEDRO A. (2013): Electronic encyclopedias. In: GOUWS u. a. (Hrsg.), 1069–1081.
- GOUWS, RUFUS H./HEID, ULRICH/SCHWEICKARD, WOLFGANG/WIEGAND, HERBERT ERNST (Hrsg.) (2013): Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/Boston (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.4).
- GOUWS, RUFUS H./SCHWEICKARD, WOLFGANG/WIEGAND, HERBERT ERNST (2013): Lexicography through the ages: From the beginnings to the electronic age. In: GOUWS u. a. (Hrsg.), 1–24
- HERBST, THOMAS/KLOTZ, MICHAEL (2003): Lexikografie. Paderborn [u. a.].
- KLOSA, ANNETTE (2013): The lexicographical process (with special focus on online dictionaries). In: GOUWS u. a. (Hrsg.), 517–524.
- LEHR, ANDREA (1996): Zur neuen Lexicographica-Rubrik „Electronic Dictionaries“. In: *Lexicographica* 12, 310–317.
- LEMBERG, INGRID (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: LEMBERG, INGRID/SCHRÖDER, BERNHARD/

- STORRER, ANGELIKA (Hrsg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen (Lexicographica, Series Maior. 107), 71–91.
- LEW, ROBERT (2011): Online Dictionaries of English. In: FUERTES-OLIVERA, PEDRO A./BERGENHOLTZ, HENNING (Hrsg.): e-Lexicography. The Internet, Digital Initiatives and Lexicography. London/New York, 230–250.
- MANN, MICHAEL (2010): Internet-Wörterbücher am Ende der „Nullerjahre“: Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien unter besonderer Berücksichtigung der Fachlexikographie. In: Lexicographica 26, 19–45.
- MÜLLER-SPITZER, CAROLIN (2013): Textual structures in electronic dictionaries compared with printed dictionaries: A short general survey. In: GOUWS u. a. (Hrsg.), 367–381.
- SCHMITZ, ULRICH (2013): Monolingual and bilingual electronic dictionaries on the Internet. In: GOUWS u. a. (Hrsg.), 1013–1023.
- SIMONSEN, HENRIK KØHLER (2004): Nine Key Principles on Corporate LSP Intranet Lexicography. In: WILLIAMS, GEOFFREY/VESSIER, SANDRA (Hrsg.): Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004. Lorient, France, July 6–10, 2004. Lorient, 603–613. Auch online, Zugriff via http://www.euralex.org/elx_proceedings/Euralex2004/ [Abruf: 2014-08-26].
- STORRER, ANGELIKA (1998): Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher. In: Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Hrsg. von HERBERT ERNST WIEGAND. Tübingen (Lexicographica, Series Maior. 84), 106–131.
- STORRER, ANGELIKA (2013): Representing (computational) dictionaries in hypertextual form. In: GOUWS u. a. (Hrsg.), 1244–1253.
- STORRER, ANGELIKA/FREESE, KATRIN (1996): Wörterbücher im Internet. In: Deutsche Sprache 24, 97–153.
- TARP, SVEN (2009): Beyond Lexicography: New Visions and Challenges in the Information Age. In: BERGENHOLTZ, HENNING/NIELSEN, SANDRO/TARP, SVEN (Hrsg.): Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow. Frankfurt am Main [u. a.] (Linguistic Insights. Studies in Language and Communication. 90), 17–32.
- TARP, SVEN (2012): Online dictionaries: today and tomorrow. In: Lexicographica 28, 253–267.
- WIEGAND, HERBERT ERNST (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Berlin/New York.

CAROLINA FLINZ

Wörterbuchbenutzung: Ergebnisse einer Umfrage unter Studenten der Tourismuswissenschaft

1. Einleitung

Nach WIEGAND (1987, 179) sollten neue Wörterbücher und Neuauflagen aus der wissenschaftlichen Erkenntnis der Wörterbuchbenutzung entstehen. Diese Anweisung wurde für das Projekt *Tourlex*, ein zweisprachiges (deutsch-italienisches) Online-Fachwörterbuch der Tourismusfachsprache, das an der Universität Pisa in Planung ist, befolgt, denn die Erkundung, in welchen Situationen die Studenten ein Fachwörterbuch benutzen und wozu sie ein Wörterbuch zur Tourismusfachsprache verwenden würden, war von großer Bedeutung für bestimmte Entscheidungen zur Makro- und Mikrostruktur.

Zu dem obengenannten Zweck wurde eine Online-Umfrage realisiert, die ca. 50 Studenten des ersten und zweiten Jahres der Fakultät Tourismuswissenschaft der Universität Pisa unterbreitet wurde.¹ Da *Tourlex* ein Online-Fachwörterbuch sein wird, wurde auch untersucht, wie Studenten Print- und Online-Produkte beurteilen, welche Eigenschaften sie für positiv und negativ halten und welche Präferenzen sie haben.

¹ In einem weiteren Verlauf der Untersuchung werden die Ergebnisse der Studenten des dritten Jahres und des Magisterstudienganges zusätzlich in Betracht gezogen werden, die zum Zeitpunkt des Fragebogens wegen Praktikumsaktivitäten nicht vor Ort waren.

2. Stand der Forschung

Die „Wörterbuchbenutzung“ hat sich im Rahmen der Lexikographie in den 80er Jahren entwickelt und wird als gleichwertiger Forschungsgegenstand neben der Erforschung der Struktur, Erstellung und Geschichte von Wörterbüchern betrachtet (ENGELBERG/LEMNITZER 2009, 83). Sie hat die Erschaffung einer größeren Benutzereffizienz zum Ziel (WIEGAND 1987, 179) und setzt sich mit folgenden Fragen auseinander: Wer ist der Benutzer und wie wird das Wörterbuch unter welchen Umständen zu welchem Zeitpunkt für welche Dauer an welchem Ort benutzt? Und zusätzlich warum, aus welchem Anlass und mit welchem Ziel? Und natürlich ob mit Erfolg und mit welchen Konsequenzen für die Benutzung? (vgl. WIEGAND 1987).

Auch wenn die Wörterbuchbenutzung nach WIEGAND der am wenigsten erforschte Bereich der Lexikographie bleibt (2008, 1) und der Mangel an empirischen Untersuchungen weiterhin von einigen Arbeiten hervorgehoben wird (WIEGAND 1977; HARTMANN 2001; KROMANN 1995; WIEGAND 2008), ist das wachsende Interesse an der Disziplin von mehreren Autoren bezeugt (vgl. u. a. HAUSMANN u. a. 1989; RIPFEL/WIEGAND 1988; ROSSENBECK 2005; WELKER 2010; WIEGAND 1987; 1998; 2008; 2010).

Die Veröffentlichungen haben unterschiedliche thematische Schwerpunkte:

- Erstellung von Wörterbüchertypologien nach Benutzungsmöglichkeiten (u. a. ENGELBERG/LEMNITZER 2009; KÜHN 1989);
- Problemlösung und Optimierungsvorschläge (u. a. WIEGAND 1995; RIPFEL 1989a; DOMÍNGUEZ 2006; KEMMER 2010);
- Erkundung typischer Benutzungssituationen (u. a. KROMANN 1995; TARP 2008) auch im Fachsprachenbereich (u. a. ENGELBERG/LEMNITZER 2009).

Auch zur Online-Wörterbuchbenutzung, die sich der Erkundung unterschiedlicher Gebrauchssituationen und spezifischer Bedürfnisse des Online-Mediums widmet, sind mehrere Veröffentlichungen getätigt worden: mehrere davon haben die tatsächliche Benutzung als Kern-

punkt (vgl. ATKINS/VARANTOLA 1998; HARTMANN 2000; MÜLLER-SPITZER/KOPLINIG/TÖPEL 2012)², auch wenn diese Forschungsrichtung noch in ihren Anfängen ist (vgl. NESI 2000, 845; SIMONSEN 2011, 7).

2012/2014 sind sehr interessante Studien durchgeführt worden:

- die obenerwähnte Studie am IDS (im deutschen Raum);
- die UDALPE-Befragung³ (im spanischen Raum).

Im italienischen Raum wurde eine kleine Untersuchung getätigt, die sich mit der Fragestellung beschäftigt, in welchen Situationen, mit welchen Zwecken und bei welchen Suchanfragen italienische Deutschlerner an der Sprachwissenschaftlichen Fakultät deutsche und italienisch-deutsche Wörterbücher verwenden. Besonderes Augenmerk wurde auf Online-Produkte und sprachwissenschaftliche Wörterbücher gelegt (FLINZ 2014, 213ff.).

Die getätigten Studien bedienen sich sowohl einzelner unterschiedlicher empirischer Methoden⁴, wie der Beobachtung, der Befragung⁵ (schriftliche oder mündliche), des Protokolls (auch Log-File-Protokolls⁶ und Think-Aloud-Protokolls), des Tests, des Experiments, des interpretativen Verfahrens, des „Simultaneous Feedbacks“ (DE SCHRYVER/PRINSLOO 2000) als auch der Verbindung mehrerer⁷. Trotz der anfänglichen Kritiken, die die Wahl, Anzahl der Probanden, das Design der empirischen Methode und die statistische Auswertung betrafen (ENGELBERG/LEMNITZER 2009, 86) hat sich die Situation heute geändert.

² Vgl. die am IDS durchgeführte Studie „User-adaptive access and cross-references in elexiko (BZV elexiko)“.

³ Sie beschäftigt sich mit der Fragestellung, in welchen Situationen, mit welchen Zwecken und bei welchen Suchanfragen spanische Deutschlernende deutsche und spanisch-deutsche Wörterbücher verwenden (DOMÍNGUEZ/MIRAZO/VIDAL 2013).

⁴ Einen Überblick über die Methoden bieten u. a. WIEGAND (1987; 2008); WELKER (2010); HARTMANN (2001).

⁵ Vgl. HATHERALL (1984); WIEGAND (1998, 589ff.).

⁶ Vgl. LEMNITZER (2001); DE SCHRYVER u. a. (2006).

⁷ Vgl. DOMÍNGUEZ/MIRAZO/VIDAL (2013, 139).

Die Umfrage erweist sich durch ihre mehrmalige Verwendung als eine der meistbenutzten Methoden (vgl. DOMÍNGUEZ/MIRAZO/VIDAL 2013, 139), um die Benutzungssituationen unterschiedlicher Benutzer (Muttersprachler⁸, Fremdsprachenlerner⁹) und unterschiedlicher Wörterbuchtypen (Spezialwörterbücher¹⁰, Online-Wörterbücher) zu erkunden.

Auch die Inhalte der Untersuchungen stellen unterschiedliche Aspekte in den Vordergrund wie u. a. Auswahl und Besitz von Wörterbüchern, Angaben und Angabeklassen, Adäquatheit der Makrostruktur, Nachschlagekompetenz, Vor- und Nachteile elektronischer Wörterbücher und Wörterbuchdidaktik (vgl. ENGELBERG/LEMNITZER 2009, 87f.).

3. Die Untersuchung

Die vorliegende Untersuchung, deren Hauptanliegen die Erkundung der Wörterbuchbenutzung seitens der Studenten des ersten und zweiten Jahrgangs der Fakultät Tourismuswissenschaft ist, wurde in Anlehnung an HARTMANNs 5. Hypothese „*An analysis of users' needs should precede dictionary design*“ (HARTMANN 1989) durchgeführt, da sie grundlegend für das Projekt *Tourlex*, ein zweisprachiges Fachwörterbuch der Tourismusfachsprache, sein wird.

Die Umfrage befasst sich mit der Auswertung der Ergebnisse einer Online-Umfrage von 56 italophonen Studierenden (davon lernen 29 Deutsch als Fremdsprache). Der Fragebogen in italienischer Sprache wurde per Mail an die Studierenden gesendet, bestand aus 30 Fragen und hatte eine Ausfüllzeit von ca. 20 Minuten.¹¹

Die 30 Fragen konzentrierten sich auf folgende Kernfrageblöcke:

1. Wörterbuchbenutzer (auch Soziodemografie): Die Fragen betreffen die persönlichen Daten der Befragten (Alter, Geschlecht, Muttersprache, Niveau der Deutschkenntnisse und die Erlangung von Zertifi-

⁸ Vgl. u. a. RETTI (2004); EKWA EBANÉGA/TOMBA MOUSSAVOU (2008).

⁹ Vgl. u. a. DE SCHRYVER/PRINSLOO (2011).

¹⁰ Vgl. WANG (2001); MURÁTH (2005); TALJARD/PRINSLOO/FRICKE (2011).

¹¹ Die deutsche Übersetzung des Fragebogens ist als Anhang beigefügt.

katens, Motivation zur Auswahl des Deutschen als Fremdsprache, Kenntnis anderer Fremdsprachen und Reihenfolge des Erwerbs);

2. Wörterbuchtyp, Wörterbuchbenutzungssituation, Bedürfnisse: die Gründe zur Benutzung des Wörterbuchs, die typische Benutzungssituation, die Benutzungsfrequenz und die zu erfüllenden Bedürfnisse wurden erforscht. Unterschiedliche Wörterbuchtypen (einsprachige, zweisprachige, fachsprachliche und weitere) und die jeweilige Benutzungsfrequenz wurden untersucht;

3. Wörterbuchformat: Format, Vor- und Nachteile eines Online-Wörterbuchs, die verwendeten dynamischen Elemente und die Art der Verlinkung wurden untersucht. Auch die Informationsstruktur, die vorkommenden Probleme und Optimierungsvorschläge wurden analysiert. Es wurde außerdem auf Copyright und Aktualisierung der Wörterbücher eingegangen. Die traditionellen Umtexte und deren Benutzungsfrequenz, sowie ihre Hauptmerkmale standen ebenfalls im Mittelpunkt des Interesses;

4. Das Fachwörterbuch (Tourismusfachsprache): Auf die Funktionen und Bedürfnisse des Fachwörterbuchs (Tourismusfachsprache) wurde spezifisch und detailliert eingegangen.

4. Die Ergebnisse der Untersuchung

Die Ergebnisse der Untersuchung werden entsprechend ihrer Aufteilung in Kernblöcke präsentiert und reflektiert: 4.1 – der Wörterbuchbenutzer und seine Personalien, 4.2 – der Wörterbuchtyp und die Benutzungssituationen, 4.3 – das Wörterbuchformat; 4.4 – die Funktionen und Merkmale des Fachwörterbuchs zur Tourismusfachsprache.

4.1 Der Wörterbuchbenutzer

Die ersten sieben Fragen haben sich mit den persönlichen Angaben der Befragten befasst, d. h. der Erkundung von Geschlecht, Alter, universitärem Jahrgang, Muttersprache und dem Erlernen von Fremdsprachen

(insbesondere des Deutschen) und deren Reihenfolge, Zertifizierungen der fremdsprachlichen Kompetenzen, dem erreichten Niveau.

Das Geschlecht der Testgruppe kann folgendermaßen zusammengefasst werden:

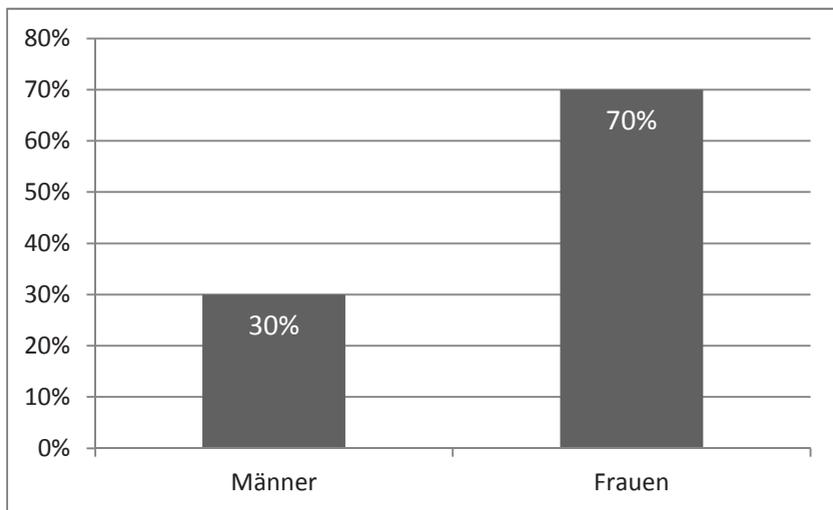


Abb. 1: Geschlechtsaufteilung der Probanden

Die Altersgruppe ist ziemlich homogen, denn 80 % der Befragten sind zwischen 18 und 21 Jahren, während nur 7 % zwischen 22 und 24 sind. Interessant ist die Feststellung, dass 13 % der Studierenden mehr als 25 Jahre alt sind. Dies bezeugt das wachsende Interesse für die Fakultät auch von Erwachsenen, die bereits im Tourismus tätig sind und einen Titel erwerben möchten, um bessere Berufschancen zu erreichen.

Alle Probanden haben Italienisch als Muttersprache und sie besuchen entweder das erste oder das zweite Jahr der Universität.

Sie zeigen ein starkes Interesse für die deutsche Sprache, da 51 % Deutsch als Fremdsprache gelernt haben und weiterhin lernen. Leider haben nur 27 % der Studierenden eine Zertifizierung der fremdsprachlichen Kompetenzen gemacht, auch wenn 87 % ein Mittelstufenniveau erreicht haben (B1–B2). Hinsichtlich des schulischen Backgrounds

wurde festgestellt, dass Deutsch in den folgenden Schultypen gelernt worden ist:

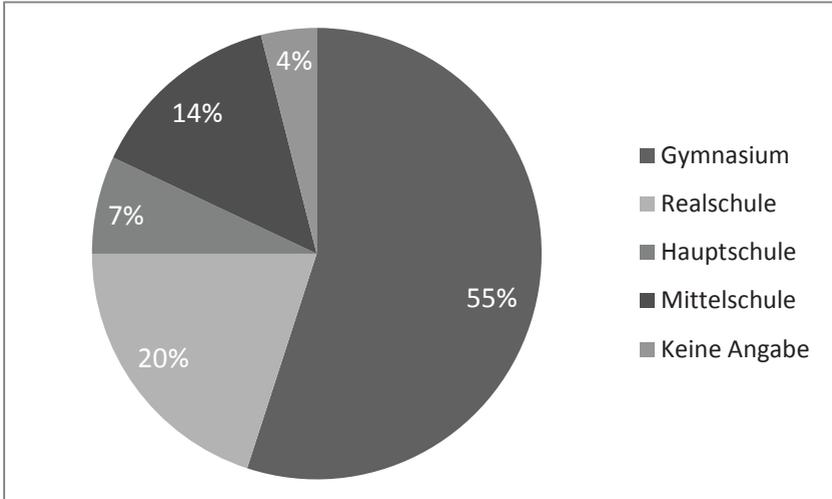


Abb. 2: Schulischer Werdegang der Probanden

Das Gymnasium ist die Schulordnung, in der Deutsch vom größten Teil der Studenten gelernt worden ist; ein kleiner Teil hat es auch in der Realschule erworben. Niedriger sind die Daten zur Mittelschule (6.–8. Klasse) und zur Berufsschule, wo nur am „Istituto Alberghiero“ (Hotelfachschule) Deutsch gelernt wird.

Die Motivationen zur Auswahl des Deutschen sind hingegen sehr homogen, auch wenn 28 % der Befragten keine Angaben dazu gegeben hat: Höhere Arbeitschancen (46 %) und Vorliebe für die deutsche Sprache und Kultur (20 %). 6 % haben Deutsch nicht freiwillig gewählt, sondern es war ein Pflichtfach.

Was die erste erlernte Fremdsprache betrifft, haben 98 % der Testpersonen Englisch gelernt, während davon nur 38 % ihre Kompetenzen zertifiziert haben. Die erreichten Niveaus sind jedoch sehr hoch: Mittelstufenniveau (85 %) und Hochstufenniveau (15 %).

Die zweite erlernte Fremdsprache der Testpersonen kann der folgenden Tabelle entnommen werden:

Französisch	Spanisch	Deutsch
69 %	22 %	9 %

Tab. 1: Zweite erlernte Fremdsprache nach Englisch

Wie zu erwarten war, ist, der italienischen Tradition folgend, das Französische noch sehr stark. Deutsch ist sehr oft die dritte erlernte Sprache, 70 % nach Französisch und 30 % nach Spanisch.

Zusammenfassend kann festgestellt werden, dass die Testgruppe in großem Maße aus Frauen besteht, die ein Alter zwischen 19 und 21 Jahren haben. Sie haben das Italienische als Muttersprache und Englisch als erste Fremdsprache, Französisch als zweite Fremdsprache und Deutsch als dritte Fremdsprache erlernt. Die Zertifizierung der fremdsprachlichen Kompetenzen ist sowohl für die erste, zweite und dritte Fremdsprache niedrig (ca. 30 %). Deutsch wurde größtenteils am Gymnasium gelernt und die Motivationen sind sehr stark mit höheren und besseren Arbeitschancen verbunden.

4.2 Wörterbuchtyp und Wörterbuchbenutzungssituation

Aus der Analyse der Items 8, 9, 10, 11, 12, 13 kann ein Profil der verwendeten Wörterbücher und Fachwörterbücher seitens der Testpersonen geschaffen werden.

Die in Abb. 3 vorgestellten Daten zeigen, dass das bilinguale Wörterbuch öfter als das monolinguale Wörterbuch seitens der Studenten verwendet wird.

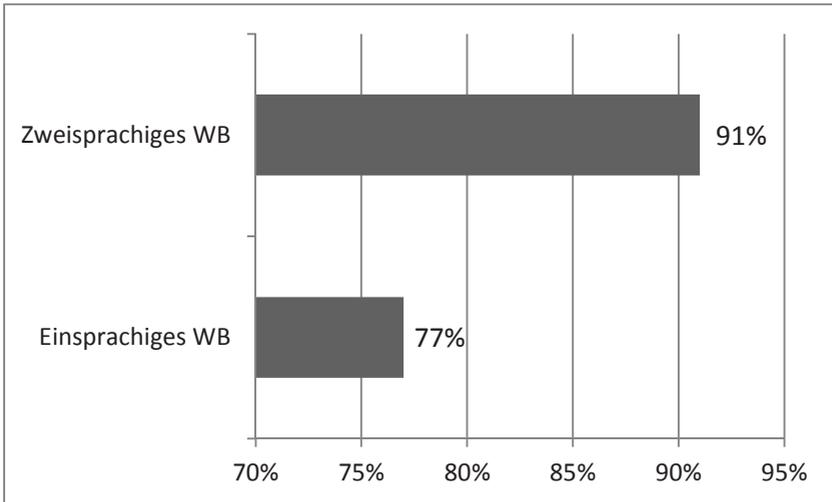


Abb. 3: Von den Probanden benutzte Wörterbuchtypen

Die Frequenz der Verwendung der jeweiligen Wörterbücher kann aus Tab. 2 entnommen werden:

	selten	manchmal	oft	Keine Angabe
einsprachiges Wb.	40 %	30 %	10 %	20 %
zweisprachiges Wb.	18 %	33 %	32 %	17 %

Tab. 2: Frequenz der Verwendung des jeweiligen Wörterbuchtypen

Zweisprachige Wörterbücher werden nicht nur häufiger als einsprachige verwendet, sondern ihre Verwendungsfrequenz ist höher, denn 32 % der Befragten benutzen sie „oft“.

Es wurde auch die Benutzung von anderen Typen von Wörterbüchern untersucht, wodurch deutlich geworden ist, dass nach den oben-

genannten Typen am meisten Synonymwörterbücher (44 %) verwendet werden; hingegen fachsprachliche Wörterbücher (8 %) und Rechtsschreibwörterbücher (4 %) viel seltener. In einer Tourismuswissenschaftlichen Fakultät hätte man einen höheren Anteil an Fachwörterbüchern erwartet – interessant wäre, zu erkunden, warum die Studenten so selten Fachwörterbücher benutzen.

Die Ergebnisse auf die Frage, wie die Testpersonen auf Wörterbücher aufmerksam geworden sind, zeigen, wie hoch der Einfluss des Schulsystems ist (57 %); die Eigeninitiative hat jedoch auch einen großen Einfluss gehabt (14 %), während die Familie eine viel geringere Rolle spielt (4 %).

4.3 Das Wörterbuchformat

Die Fragen 16, 17, 18 konzentrieren sich auf das Wörterbuchformat und zeigen unerwartete Ergebnisse: 41 % der Befragten bevorzugen noch die traditionellen Printwörterbücher, nur 2 % sind an elektronischen Wörterbüchern (d. h. CD-ROM-Wörterbüchern) interessiert und 43 % plädieren für Online-Wörterbücher. Der Unterschied zwischen der Verwendung von Print- und Onlinewörterbüchern ist sehr gering und es gibt dementsprechend auch eine kleine Gruppe von Testpersonen, die beide ohne Bevorzugung verwendet.

Als Vorteile des Printmediums werden folgende Eigenschaften genannt: Die Zuverlässigkeit der Informationen wird an erster Stelle genannt; es werden aber Aspekte wie Vollständigkeit der Lemmata und Detailliertheit der Angaben betont. Die kontinuierliche Verfügbarkeit wird auch als positives Kriterium bewertet (vgl. Abb. 4).

Nachteile des Formats sind jedoch für die Probanden die Schwere und die Handlichkeit des Werkzeugs (52 %), sowie der Zeitfaktor, im Sinne von einer längeren Benutzungsdauer (37 %) und das Fehlen von einer zeitnahen Aktualisierung (5 %).

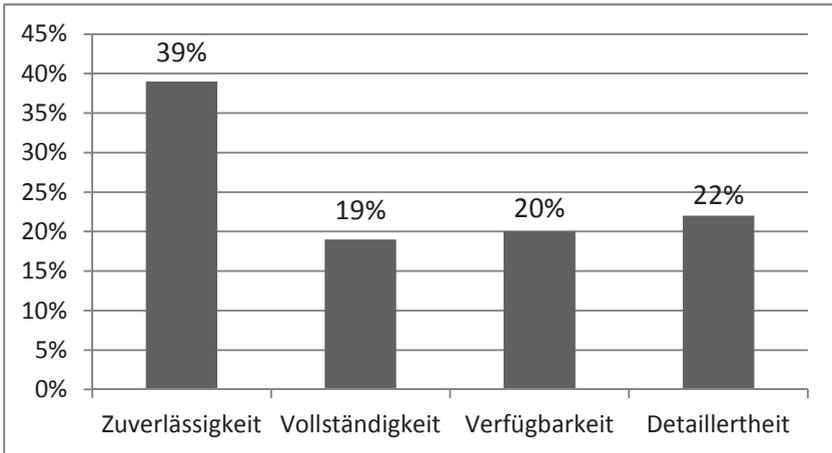


Abb. 4: Vorteile des Printformats

Die Vorteile des Online-Mediums können folgendermaßen zusammengefasst werden:

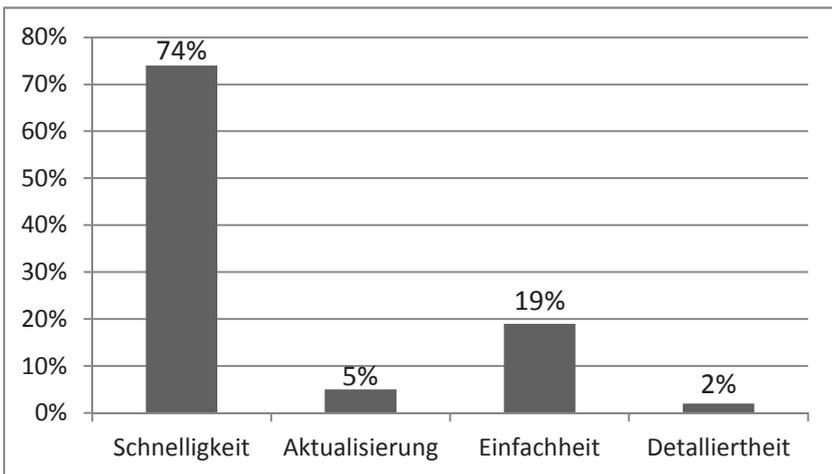


Abb. 5: Vorteile des Online-Formats

Die Schnelligkeit in der Auffindung der Informationen sowie der höhere Erfolg in der Suche sind am meisten genannt; die Einfachheit, im Sinne des Auswählens der richtigen Informationen sowie die ständige Aktualisierung des Produkts und die höhere Detailliertheit der Informationen werden weniger geschätzt, wenn auch für wichtig gehalten.

Als negative Aspekte werden jedoch die möglichen Fehler (57 %) und die Abhängigkeit von einer Internetverbindung (43 %) genannt. Die Probanden beurteilen die Internetprodukte negativ, in dem sie ihre wissenschaftliche Zuverlässigkeit leugnen.

Zu den bevorzugten Instrumenten eines Online-Wörterbuchs gehört die Suchmaschine: die einfache Suchmaschine wird nur von 31 % der Testpersonen bevorzugt, die Suchmaschine mit zusätzlichen Funktionen dagegen von 65 %. Die alphabetische Leiste wird hingegen nur von 7 % benutzt.

Die Schnelligkeit eines Online-Wörterbuchs wird von 86 % der Befragten als positiv beurteilt; das Vorhandensein von Links von 92 %. Nur eine kleine Gruppe von Studenten kritisiert das zu häufige Verwenden von Links in einer Homepage.

Interne Links werden bevorzugt, wie die folgende Graphik zeigt:

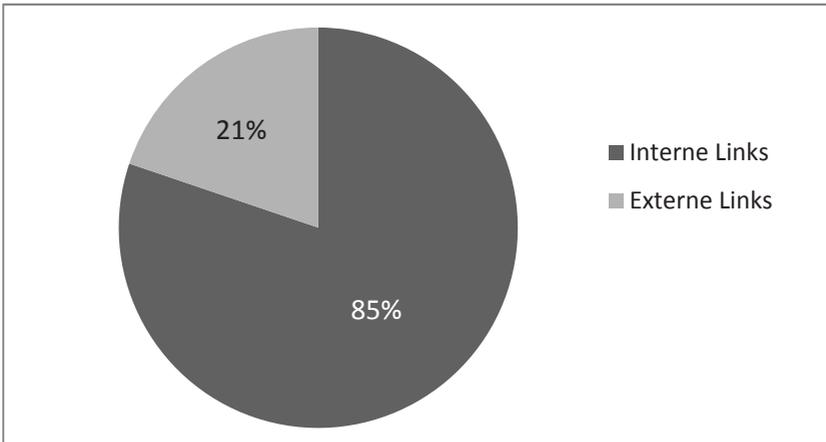


Abb. 6: Bevorzugte Links

Hinsichtlich der anderen Werkzeuge, die in einem Online-Wörterbuch vorhanden sind, wird die Möglichkeit, einen Eintrag auszudrucken, von 18 % der Testpersonen verwendet und die Kontaktfunktion zu dem Wörterbuchteam, um Lücken oder Fehler zu signalisieren, von 14 % geschätzt. Leider sehen nur 18 % der Testpersonen die Wichtigkeit des Copyrights ein, während jedoch die zeitnahe Aktualisierung eines Wörterbuchs von 89 % als positiv beurteilt wird.

Die Verwendung der Umtexte seitens der Studenten wurde auch analysiert, da Umtexte ein wichtiger Teil eines Wörterbuchs sind, aber in Tat wenig benutzt werden.

Die Ergebnisse können folgendermaßen zusammengefasst werden:

- Nur 19 % der Probanden lesen die Einleitung eines lexikographischen Produkts;
- nur 19 % verwenden die Benutzungshinweise;
- 53 % konsultieren das Abkürzungsverzeichnis;
- 61 % benutzen das Register der Einträge.

4.4 Funktion und Bedürfnisse des fachsprachlichen Wörterbuchs

Der fünfte Block der Befragung beschäftigt sich mit den Motivationen und Gründen, die zur Verwendung eines Fachwörterbuchs und insbesondere eines zur Tourismusfachsprache führen könnten. Die Ergebnisse der Untersuchung deuten darauf hin, dass die Testpersonen ein solches Wörterbuch zur Rate ziehen würden, wenn sie einen Fachterminus in der Fremdsprache verstehen möchten und Zweifel über seine Verwendung hätten (60 %) oder wenn sie nach der Übersetzung in die italienische Sprache suchen würden (40 %). Sie würden es auch benutzen, wenn sie einen Text in der Fremdsprache produzieren müssten (46 %) oder um Informationen zum Eintrag zu finden (Grammatik, Verwendung etc.). Eine kleine Gruppe würde hingegen das Wörterbuch verwenden, um ein Problem in der Kommunikation zu lösen.

Wie man aus den Daten entnehmen kann, müsste ein Fachwörterbuch zur Tourismusfachsprache sowohl aktive, bzw. produktive, als auch passive, bzw. rezeptive, Funktionen haben.

Was die erwünschten Angaben betrifft, würden die Probanden das Wörterbuch verwenden, um Folgendes zu suchen:

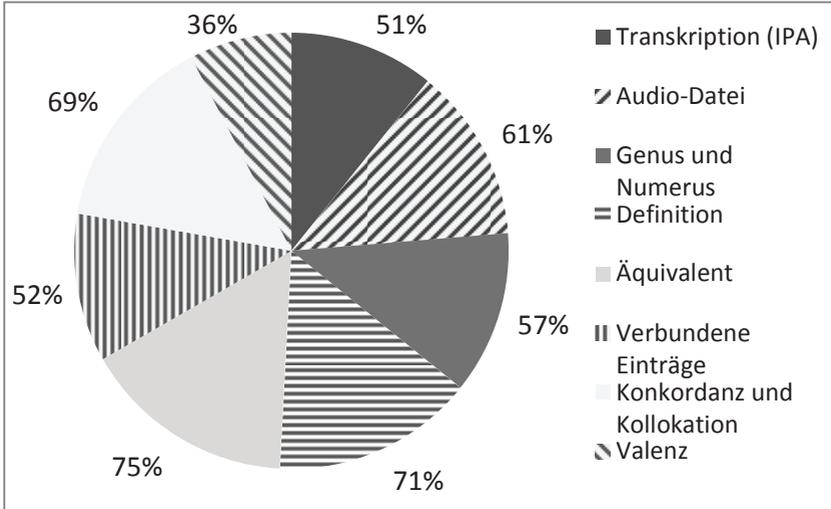


Abb. 7: Diagramm zu den mikrostrukturellen Angaben eines Fachwörterbuchs

Das Äquivalent in der Muttersprache ist die meistgesuchte Information (75 %) vor der fachsprachlichen Definition des Terminus (71 %). Von großem Wert sind auch die Gebrauchsbeispiele des Terminus, mit Exemplifikation zur Konkordanz und zu Kollokationen (69 %), und die Audio-Datei (61 %), um eine normkonforme Aussprache abzuhören. Grammatische Informationen (wie Genus und Numerus) sind auch für einen großen Teil der Gruppe ein Desideratum, und das betrifft sowohl die des Eintrags (57 %) als auch die des Äquivalents (52 %). Die Information zu „verwandten Einträgen“ (52 %) und die Transkription mittels IPA-Zeichen (51 %) sind auch geschätzt, während die Suche nach Angaben zur Valenz nur von 36 % der Studenten angegeben wird.

Es wurde auch zusätzlich gefragt, welche anderen zusätzlichen Informationen sie für nützlich gehalten hätten, daraufhin wurden Informationen zum Sektor (44 %), etymologische Angaben (23 %) und Synonyme und Antonyme (38 %) genannt.

4.5 Schlussbemerkungen

Aus der empirischen Untersuchung geht hervor, welche Einstellung die Probanden gegenüber unterschiedlichen Wörterbuchtypen und Formaten haben und welche Eigenschaften der Makro-, Medio- und Mikrostruktur sie bevorzugen.

Die besondere Vorliebe für zweisprachige Wörterbücher war abzu- sehen, während der seltene Gebrauch von Fachwörterbüchern in einer Fakultät für Tourismuswissenschaft überraschend war.

Auch die Tatsache, dass die Testpersonen gleichermaßen Print- und Online-Wörterbücher verwenden, war in einem Zeitraum, wo Smart- phones und Tablets den studentischen Alltag charakterisieren, unerwar- tet. Die Gründe zu den Vorteilen und Nachteilen der jeweiligen Formate sind jedoch übereinstimmend mit der Literatur, auch wenn es vorge- kommen ist, dass die Internetlexikographie noch einiges bezüglich der Wissenschaftlichkeit der Online-Produkte zu tun hat. Die negative Ein- schätzung der Online-Wörterbücher kann jedoch mit einem stärkeren Einbau der theoretischen Erkenntnisse verbessert werden.

Die Analyse hat auch ergeben, dass nur einige Umtexte (wie Regis- ter und Abkürzungsverzeichnis) für den Benutzer wichtig sind, der auf eine ständige Aktualisierung achtet, aber keinen Blick auf Copyright oder Eigentum des Produkts wirft. Die Wichtigkeit von Suchmaschinen und Links, insbesondere von internen Links, ist auch hervorgehoben worden.

Interessant waren auch die Ergebnisse zu den mikrostrukturellen Angaben eines Fachwörterbuchs: eine Fülle von Informationen (phone- tischen, grammatischen, syntaktischen etc.) wird erwünscht, so dass der Benutzer nicht nur das Lemma verstehen kann, sondern es auch norm- gerecht aussprechen und verwenden kann. Die Bedeutung der Äquiva- lenzangabe sowie der Gebrauch im Kontext wird mehrmals hervorge- hoben und der Bedarf an spezifischen Kollokationen und Phraseolo- gismen von einem großen Teil der Probanden angesprochen.

Ob diese Ergebnisse seitens der Studenten des dritten Jahres und des Masterstudiengangs bestätigt werden ist noch abzuwarten, auch wenn sie in die Untersuchung einfließen werden, so dass die wichtigs-

ten Entscheidungen zur Makro- und Mikrostruktur des Projektes *Tourlex* sowohl auf theoretischen Erkenntnissen als auch auf der empirischen Analyse gründen werden.

Literatur

- ATKINS, B. T. SUE/VARANTOLA, KRISTA (1998): Language learners using dictionaries: The final report on the EURALEX/AILA research project on dictionary use. In: ATKINS, B. T. SUE (Hrsg.): *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. Tübingen: Niemeyer (Lexicographica, Series Maior. 88), 83–122.
- DE SCHRYVER, GILLES-MAURICE/JOFFE, DAVID/JOFFE, PITTA/HILLEWAERT, SARAH (2006): Do Dictionary Users Really Look Up Frequent Words? – On the Overestimation of the Value of Corpus-based Lexicography. In: *Lexikos* 16, 67–83.
- DE SCHRYVER, GILLES-MAURICE/PRINSLOO, DAAN J. (2000): Dictionary-Making Process with ‘Simultaneous Feedback’ from the target Users to the Compilers. In: HEID u. a. (Hrsg.), 197–209.
- DE SCHRYVER, GILLES-MAURICE/PRINSLOO, D. J. (2011): Do Dictionaries Define on the Level of Their Target Users? A Case Study for Three Dutch Dictionaries. In: *International Journal of Lexicography* 24/1, 5–28.
- DOMÍNGUEZ VÁZQUEZ, MARÍA JOSÉ (2006): Von monolingualen Wörterbüchern zu kontrastiven Valenzwörterbüchern. Die Valenzwiedergabe unter der Lupe. In: *Jahrbuch für Deutsch als Fremdsprache. Intercultural German Studies* 1/32, 231–241.
- DOMÍNGUEZ VÁZQUEZ, MARÍA JOSÉ/MIRAZO BALSÀ, MÓNICA/VIDAL PÉREZ, VANESSA (2013): Wörterbuchbenutzung: Erwartungen und Bedürfnisse. Ergebnisse einer Umfrage bei Deutsch lernenden Hispanophonen. Online: http://www.academia.edu/4640188/Worterbuchbenutzung_Erwartungen_und_Bedurfnisse_Ergebnisse_einer_Umfrage_bei_Deutsch_lernenden_Hispanophonen [Abruf: 2014-06-25].
- EKWA EBANÉGA, GUY-MODESTE/TOMBA MOUSSAVOU, FATIMA (2008): A Survey of the Dictionary Use of Gabonese Students at Two South African Universities. In: *Lexikos* 18, 349–365.
- ENGELBERG, STEFAN/LEMNITZER, LOTHAR (2009): *Lexikographie und Wörterbuchbenutzung*. 4., überarbeitete und erweiterte Auflage. Tübingen: Stauffenburg.
- FLINZ, CAROLINA (2014): Wörterbuchbenutzung: Ergebnisse einer Umfrage bei italienischen DaF-Lernern. In: ABEL, ANDREA/VETTORI, CHIARA/

- RALLI, NATASCIA (Hrsg.): Proceedings of the XVI EURALEX International Congress: The User in Focus. 15–19 July 2014, Bolzano/Bozen. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism, 213–224. Online: http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX_Part_1.pdf [Abruf: 2014-08-04].
- HARTMANN, REINHARD RUDOLF KARL (1989): Sociology of The Dictionary User: Hypotheses and Empirical Studies. In: HAUSMANN u. a. (Hrsg.), 649–657.
- HARTMANN, REINHARD (2000): European Dictionary Culture. The Exeter Case Study of Dictionary Use among University Students, against the Wider Context of the Reports and Recommendations of the Thematic Network Project in the Area of Languages (1996–1999). In: HEID u. a. (Hrsg.), 385–391.
- HARTMANN, R. R. K. (2001): Teaching and Researching Lexicography. London: Pearson.
- HATHERALL, GLYN (1984): Studying Dictionary Use: Some Findings and Proposals. In: HARTMANN, REINHARD R. K. (Hrsg.): LEXeter '83 Proceedings: Papers from the International Conference on Lexicography at Exeter, 9–12 September 1983. Tübingen: Niemeyer (Lexicographica, Series Maior. 1), 183–189.
- HAUSMANN, FRANZ JOSEF/REICHMANN, OSKAR/WIEGAND, HERBERT E./ZGUSTA, LADISLAV (Hrsg.) (1989): Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie [...]. Erster Teilband. Berlin/New York: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.1).
- HEID, ULRICH/EVERT, STEFAN/LEHMANN, EGBERT/ROHRER, CHRISTIAN (Hrsg.) (2000): Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000. Stuttgart: Niemeyer.
- KEMMER, KATHARINA (2010): Onlinewörterbücher in der Wörterbuchkritik. Mannheim: Institut für Deutsche Sprache (OPAL – Online publizierte Arbeiten zur Linguistik 2/2010). Online: <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2010-2.pdf> [Abruf: 2014-07-01].
- KROMANN, HANS-PEDER (1995): Deutsche Wörterbücher aus der Perspektive eines fremdsprachigen Benutzers. In: POPP (Hrsg.), 501–512.
- KÜHN, PETER (1989): Typologie der Wörterbücher nach Benutzungsmöglichkeiten. In: HAUSMANN u. a. (Hrsg.), 111–127.
- LEMNITZER, LOTHAR (2001): Das Internet als Medium für die Wörterbuchbenutzungsforschung. In: LEMBERG, INGRID/SCHRÖDER, BERNHARD/STORER, ANGELIKA (Hrsg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion

- und Publikation digitaler Wörterbücher. Tübingen: Niemeyer (Lexicographica, Series Maior. 107), 247–254.
- MÜLLER-SPITZER, CAROLIN/KOPLINIG, ALEXANDER/TÖPEL, ANTJE (2012): Online dictionary use: Key findings from an empirical research project. In: GRANGER, SYLVIANE/PAQUOT, MAGALI (Hrsg.): *Electronic Lexicography*. Oxford: Oxford University Press, 425–457.
- MURÁTH, JUDITH (2005): Wörterbuchbenutzung von Fachübersetzerstudenten Ihre Erwartungen an ein Fachwörterbuch. In: GOTTLIEB, HENRIK/MOGENSEN, JENS ERIK/ZETTERSTEN, ARNE (Hrsg.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002, at the University of Copenhagen*. Tübingen: Niemeyer (Lexicographica, Series Maior. 115), 401–415.
- NESE, HILARY (2000): Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art. In: HEID u. a. (Hrsg.), 839–847.
- POPP, HEIDRUN (Hrsg.) (1995): *Deutsch als Fremdsprache. An den Quellen eines Faches. Festschrift für Gerhard Helbig zum 65. Geburtstag*. München: Iudicium.
- RETTI, GREGOR (2004): *Österreichisches Deutsch und Österreichisches Wörterbuch*. Online: <http://oewb.retti.info/oewb/index.html> und <http://gregor.retti.info/docs/retti1991/4.pdf> [Abruf beide: 2014-03-07].
- RIPFEL, MARTHA (1989): *Wörterbuchkritik. Eine empirische Analyse von Wörterbuchrezensionen*. Tübingen: Niemeyer (Lexicographica, Series Maior. 29).
- RIPFEL, MARTHA/WIEGAND, HERBERT ERNST (1988): Empirische Wörterbuchbenutzungsforschung. In: WIEGAND, HERBERT ERNST (Hrsg.): *Studien zur neuhochdeutschen Lexikographie VI. 2. Teilband*. Hildesheim/Zürich/New York: Olms (Germanistische Linguistik. 87–90/1986), 491–520.
- ROSSENBECK, KLAUS (2005): Die zweisprachige Fachlexikographie in der neueren und neuesten Wörterbuchforschung. In: *Lexicographica* 21, 179–201.
- SIMONSEN, HENRIK KØHLER (2011): User Consultation Behaviour in Internet Dictionaries: An Eye-Tracking Study. In: *Hermes. Journal of Language and Communication Studies* 46, 75–101.
- TALJARD, E./PRINSLOO, D. J./FRICKE, I. (2011): The use of LSP dictionaries in secondary schools? A South African case study. In: *South African Journal of African Languages* 31/1, 87–109.
- TARP, SVEN (2008): *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Niemeyer (Lexicographica, Series Maior. 134).

- WANG, WEIWEI (2001): Zweisprachige Fachlexikographie. Benutzungsforschung, Typologie und mikrostrukturelle Konzeption. Frankfurt am Main [u. a.]: Peter Lang.
- WELKER, HERBERT ANDREAS (2010): Dictionary Use. A General Survey of Empirical Studies. Brasilia: Author's Edition.
- WIEGAND, HERBERT ERNST (1977): Nachdenken über Wörterbücher: Aktuelle Probleme. In: DROSDOWSKI, GÜNTHER/HENNE, HELMUT/WIEGAND, HERBERT ERNST: Nachdenken über Wörterbücher. Mannheim/Wien/Zürich: Bibliographisches Institut/Dudenverlag, 51–102.
- WIEGAND, HERBERT ERNST (1987): Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. In: *Lexicographica* 3, 178–227.
- WIEGAND, HERBERT ERNST (1995): Lexikographische Texte in einsprachigen Lernerwörterbüchern. Kritische Überlegungen anlässlich des Erscheinens von Langenscheidts „Großwörterbuch Deutsch als Fremdsprache“. In: POPP (Hrsg.), 463–499.
- WIEGAND, HERBERT ERNST (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Berlin/New York: De Gruyter.
- WIEGAND, HERBERT ERNST (2008): Wörterbuchbenutzung bei der Übersetzung. Möglichkeiten ihrer Erforschung. In: JESENŠEK, VIDA/LIPAVIC OŠTIR, ALJA (Hrsg.): Wörterbuch und Übersetzung. 4. Internationales Kolloquium zur Lexikographie und Wörterbuchforschung. Universität Maribor. 20. bis 22. Oktober 2006. Hildesheim/Zürich/New York: Olms (Germanistische Linguistik. 195–196), 1–43.
- WIEGAND, HERBERT ERNST (2010): Zur Methodologie der Systematischen Wörterbuchforschung: Ausgewählte Untersuchungs- und Darstellungsmethoden für die Wörterbuchform. In: *Lexicographica* 26, 249–330.

Anhang: Fragebogen (übers. aus dem Italienischen)

TOURLEX

PERSÖNLICHE ANGABEN	
1) GESCHLECHT:	<input type="checkbox"/> M <input type="checkbox"/> W 2) ALTER: _____ 3) MUTTERSPRACHE: <input type="checkbox"/> Deutsch <input type="checkbox"/> Italienisch <input type="checkbox"/> _____
4) UNIVERSITÄT:	Pisa FAKULTÄT: TOURISMUSWISSENSCHAFT JAHRGANG: _____
5) DEUTSCHNIVEAU:	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2
ZERTIFIKATE:	<input type="checkbox"/> Ja (welche? _____) <input type="checkbox"/> Nein
WIE LANGE LERNST DU SCHON DEUTSCH?:	_____ (Jahre / Monate)
SCHULE (ODER KURS), IN DER/DEM DU DEUTSCH GELERNT HAST:	<input type="checkbox"/> Gymnasium (Jahre: _____) <input type="checkbox"/> Realschule (Jahre: _____) <input type="checkbox"/> Hauptschule (Jahre: _____)
	<input type="checkbox"/> Sprachschule (Jahre: _____) <input type="checkbox"/> _____ (Jahre: _____)
6) GRÜNDE ZUR AUSWAHL DES DEUTSCHEN ALS FREMDSPRACHE:	_____
7) WELCHE ANDEREN FREMDSPRACHEN HAST DU VOR DEUTSCH GELERNT? (1. FREMDSPRACHE, 2. FREMDSPRACHE ETC.)	_____
1. FREMDSPRACHE:	_____ WIE LANGE?: _____
NIVEAU:	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2
ZERTIFIKATE:	Ja (welche? _____) <input type="checkbox"/> Nein
SCHULE (ODER KURS), IN DER/DEM DU DIE SPRACHE GELERNT HAST:	<input type="checkbox"/> Gymnasium (Jahre: _____) <input type="checkbox"/> Realschule (Jahre: _____) <input type="checkbox"/> Hauptschule (Jahre: _____)
	<input type="checkbox"/> Sprachschule (Jahre: _____) <input type="checkbox"/> _____ (Jahre: _____)
2. FREMDSPRACHE:	_____ WIE LANGE?: _____
NIVEAU:	<input type="checkbox"/> A1 <input type="checkbox"/> A2 <input type="checkbox"/> B1 <input type="checkbox"/> B2 <input type="checkbox"/> C1 <input type="checkbox"/> C2
ZERTIFIKATE:	Ja (welche? _____) <input type="checkbox"/> Nein
SCHULE/KURS, IN DER/DEM DU DIE SPRACHE GELERNT HAST:	<input type="checkbox"/> Gymnasium (Jahre: _____) <input type="checkbox"/> Realschule (Jahre: _____) <input type="checkbox"/> Hauptschule (Jahre: _____)
	<input type="checkbox"/> Sprachschule (Jahre: _____) <input type="checkbox"/> _____ (Jahre: _____)
WELCHE WÖRTERBUCHTYPEN BENUTZT DU?*	
8) EINSPRACHIGE WÖRTERBÜCHER:	<input type="checkbox"/> Ja (Welche? _____) <input type="checkbox"/> nein Wie oft?: <input type="checkbox"/> oft <input type="checkbox"/> manchmal <input type="checkbox"/> selten <input type="checkbox"/> nie (warum?: _____)
9) ZWEISPRACHIGE WÖRTERBÜCHER:	<input type="checkbox"/> Ja (Welche? _____) <input type="checkbox"/> nein Wie oft?: <input type="checkbox"/> oft <input type="checkbox"/> manchmal <input type="checkbox"/> selten <input type="checkbox"/> nie (warum?: _____)
10) FACHWÖRTERBÜCHER:	<input type="checkbox"/> Ja (Welche? _____) <input type="checkbox"/> nein Wie oft?: <input type="checkbox"/> oft <input type="checkbox"/> manchmal <input type="checkbox"/> selten <input type="checkbox"/> nie (warum?: _____)
11) WÖRTERBUCHER ZUR RECHTSCHREIBUNG:	<input type="checkbox"/> Ja (Welche? _____) <input type="checkbox"/> nein Wie oft?: <input type="checkbox"/> oft <input type="checkbox"/> manchmal <input type="checkbox"/> selten <input type="checkbox"/> nie (warum?: _____)
12) SYNONYMWÖRTERBÜCHER:	<input type="checkbox"/> Ja (Welche? _____) <input type="checkbox"/> nein Wie oft?: <input type="checkbox"/> oft <input type="checkbox"/> manchmal <input type="checkbox"/> selten <input type="checkbox"/> nie (warum?: _____)
13) ANDERE TYPEN?	<input type="checkbox"/> Kollokationsw. <input type="checkbox"/> Valenzw. <input type="checkbox"/> andere (Bitte die jeweilige Frequenz angeben)
14) WIE BIST DU ZUR VERWENDUNG EINES WÖRTERBUCHES GEKOMMEN? WER HAT DICH DAZU GEFÜHRT?	_____
15) WELCHES WÖRTERBUCH WURDE DIR ZULETZT EMPFOHLEN?	_____
FORMAT	
16) WELCHE WÖRTERBÜCHER BEVORZUGST DU?	<input type="checkbox"/> PRINTWÖRTERBÜCHER: WARUM? _____
	<input type="checkbox"/> ELEKTRONISCHE W. (CD-ROM ETC.): WARUM? _____
	<input type="checkbox"/> ONLINE-WÖRTERBÜCHER: WARUM? _____
17) WELCHE VORTEILE HABEN DIE JEWEILIGEN FORMATE?	<input type="checkbox"/> PRINTWÖRTERBÜCHER: _____
	<input type="checkbox"/> ELEKTRONISCHE W. (CD-ROM ETC.): _____
	<input type="checkbox"/> ONLINE-WÖRTERBÜCHER: _____
18) WELCHE NACHTEILE HABEN DIE JEWEILIGEN FORMATE?	<input type="checkbox"/> PRINTWÖRTERBÜCHER: _____
	<input type="checkbox"/> ELEKTRONISCHE W. (CD-ROM ETC.): _____
	<input type="checkbox"/> ONLINE-WÖRTERBÜCHER: _____
ONLINE- WÖRTERBÜCHER	
19) WELCHE DYNAMISCHEN INSTRUMENTE BENUTZT DU?	<input type="checkbox"/> Alphabetische Leiste

VIELEN DANK!!!

TOURLEX

Suchmaschine (elementare Suchmaschine)
 Suchmaschine (fortgeschrittene Suchmaschine)
 Andere: _____

20) IST DIE SCHNELLIGKEIT FÜR DICH WICHTIG? Ja Nein Ich weiß nicht

21) WIE BEURTEILTST DU DAS VORHANDENSEIN VON LINKS IN DEN DEFINITIONEN?
 Positiv (Warum?: _____)
 Negativ (Warum?: _____)

22) WELCHE LINKS BENUTZT DU AM HÄUFIGSTEN?
 Interne Links (zwischen Einträgen etc.)
 Externe Links (zu anderen Wörterbüchern, Ressourcen etc.)

23) LAYOUT: WELCHE DIESER TEXTE / FUNKTIONEN BENUTZT DU IN EINEM ONLINE-WÖRTERBUCH?
 Vorwort oft manchmal nie
 Benutzeranleitung oft manchmal nie
 Abkürzungsverzeichnis oft manchmal nie
 Register der Einträge oft manchmal nie
 Kontaktfunktion (Team, Autoren etc.) oft manchmal nie
 Druckfunktion (Einträge etc.) oft manchmal nie

24) COPYRIGHT: KONTROLLIERST DU DIE ANGABEN ZUM COPYRIGHT UND HÄLTST DU SIE FÜR WICHTIG?
 Ja (Warum?: _____)
 Nein (Warum?: _____)

25) AKTUALISIERUNG: KONTROLLIERST DU, OB SIE VORHANDEN IST, UND HÄLTST DU SIE FÜR WICHTIG?
 Ja (Warum?: _____)
 Nein (Warum?: _____)

DAS FACHWÖRTERBUCH ZUR TOURISMUSSPRACHE

26) WARUM BENUTZT DU EIN FACHWÖRTERBUCH ZUR TOURISMUSSPRACHE?
 um einen Text zu verstehen um einen Text zu übersetzen um einen Text zu schreiben um dich über ein Thema zu informieren um bibliographische Informationen zu bekommen um ein Problem zu lösen (Verwendung eines Wortes etc.) andere Gründe: _____

27) BEDÜRFNISSE: WAS SUCHST DU IN EINEM FACHWÖRTERBUCH ZUR TOURISMUSFACHSPRACHE?
 Angaben zur Phonetik (Transkription) oft manchmal nie
 Audio-Datei zur Aussprache oft manchmal nie
 Grammatische Angaben
 Genus oft manchmal nie
 Numerus oft manchmal nie
 Weitere: _____ oft manchmal nie
 Definition oft manchmal nie
 Übersetzung oft manchmal nie
 Grammatische Angaben zum Äquivalent
 Genus oft manchmal nie
 Numerus oft manchmal nie
 Weitere: _____ oft manchmal nie
 Beispiele oft manchmal nie
 Kollokationen oft manchmal nie
 Valenz oft manchmal nie
 Bibliographische Informationen oft manchmal nie
 Verbundene Einträge oft manchmal nie
 weitere Angaben: _____ oft manchmal nie

28) WELCHE ANGABEN WÄREN FÜR DICH NOCH WICHTIG?:
 Spezifische Informationen zum Sektor Etymologie
 Andere grammatische Informationen (Welche? _____) Synonyme

*Kannst die Unterschiede zwischen Wörterbuchtypen? Schreib deine Definition und mach Beispiele:

Einsprachiges Wörterbuch: _____

Beispiel (Name, Verlag): _____

Zweisprachiges Wörterbuch: _____

Beispiel (Name, Verlag): _____

Fachwörterbuch: _____

Beispiel (Name, Verlag): _____

VIELEN DANK!!!

CHRISTIANE ZEHRER

Kriterien für die Terminus-Auswahl aus prozessbezogener Perspektive – ein Beitrag zur Meta-Lexikographie

1. Einleitung

Die Verwendung korrekter Bezeichnungen für Sachverhalte ist ein zentrales Anliegen von Sprachforschern und Sprachverwendern. Die zunehmende Differenzierung von Adressatengruppen stellt jedoch nicht nur eine starre Bewertung als „korrekt“ oder „inkorrekt“ zugunsten von „adäquat“ in Frage, sondern erfordert auch die Berücksichtigung zieltext- und zielgruppenbezogener Kriterien für Übersetzungsvorschläge. Vor diesem Hintergrund untersucht der vorliegende Beitrag die Terminus-Auswahl von Übersetzungsstudierenden im Vergleich mit einem maschinellen Verfahren. Im Ergebnis zeigen sich systematische Unterschiede, die für die stärkere Beachtung verwendungsbezogener Aspekte bei der Auswahl von Einträgen in Nachschlagewerken sprechen.

2. Terminologiearbeit

Dieser Abschnitt erklärt Zweck und Vorgehensweise der Terminologiearbeit. Dabei wird insbesondere auf die Auswirkungen kooperativer und kognitiver Paradigmen auf den zentralen Begriff des Terminus eingegangen. Als Ergebnis wird der Terminus-Begriff der klassischen Terminologielehre um kognitive und prozessbezogene Aspekte erweitert.

2.1 Klassische Terminologielehre

Die klassische Terminologielehre stellt eine Hilfsdisziplin für die praktische Arbeit von Übersetzern und technischen Redakteuren dar. Als solche ist sie positivistisch ausgerichtet. Dies wird besonders deutlich anhand der einschlägigen Normen. So definiert DIN EN 2342:2011-08 *Terminologie* als

Gesamtheit der Begriffe und ihrer Benennungen in einem Fachgebiet.
(ebd.)

Unter *Terminologearbeit* versteht DIN EN 2342 die

auf der Terminologielehre aufbauende Planung, Erarbeitung, Bearbeitung oder Verarbeitung, Darstellung oder Verbreitung von Terminologie. (ebd.)

Dabei unterscheidet man zwischen punktueller und systematischer Terminologearbeit (ARNTZ/PICHT/MAYER 2004, 216). Während punktuelle Terminologearbeit ein aktuelles Bezeichnungs- oder Übersetzungsproblem zum Anlass nimmt, um eine Lösung für genau dieses Problem zu erarbeiten (ebd., 217), geht systematische Terminologearbeit ihren Gegenstand strategisch an. Das heißt, hier werden in sprachlichem Material potenzielle Fachbegriffe gesucht, vereinheitlicht, definiert und zur Verfügung gestellt (ebd., 219ff.). In ihrer systematischen Variante ist Terminologearbeit demnach Teil eines professionellen Texterstellungs- und Übersetzungsprozesses (DREWER/ZIEGLER 2011, s. a. Abbildung 1).

Auf die Nutzung von Terminologie in einem Arbeitsprozess werde ich in Abschnitt 2.3 noch ausführlich eingehen. Zuvor soll noch geklärt werden, um welche Art sprachlicher Einheiten es sich bei den Termini handelt. Dies dient dazu, die Besonderheit einer prozessorientierten Sicht herauszuarbeiten.

Die aktuell geltende Norm DIN EN 2342:2011-08 definiert den *Terminus* (nachfolgend: Terminus_a) als

sprachliche Bezeichnung eines Allgemeinbegriffs aus einem Fachgebiet.
(ebd.)

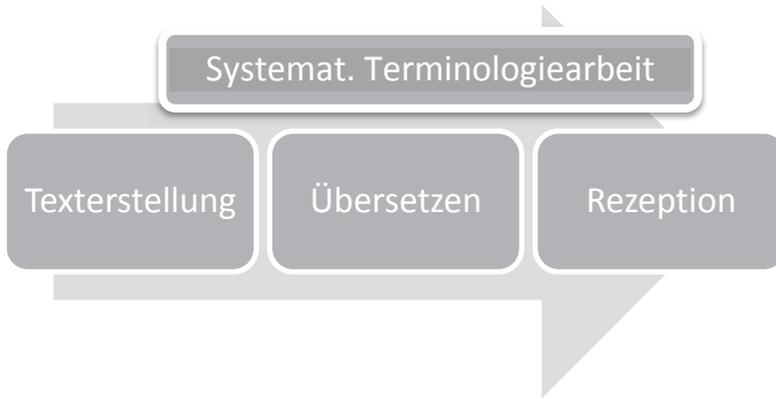


Abb. 1: Systematische Terminologiearbeit: Teil eines Arbeits- und Übersetzungsprozesses

Terminus ist ein Synonym für *Benennung* und bewegt sich somit rein auf der Ausdrucksseite des sprachlichen Zeichens. Hierdurch fehlt jede Bezugnahme zur genauen Identität des mit einem oder mehreren Wörtern Bezeichneten (Saussures „signifié“). Folglich bietet der Fachbegriff *Terminus* keinen Zugriff auf die Sachangemessenheit des Zeichens, die folglich nicht beurteilt werden kann.

Demgegenüber bezog die jetzt überholte Norm DIN EN 2342: 1992-10 die beiden genannten Aspekte in sinnvoller Weise ein. *Terminus* (nachfolgend: *Terminus_b*) war dort definiert als

das zusammengehörige Paar aus Begriff und Benennung als Teil einer Terminologie. (ebd.)

Hier wird die Korrektheit der Bezeichnungsleistung (Bühlers Symbolfunktion) durch die Einordnung in eine Terminologie (Bühlers Symptomfunktion) legitimiert. Dies erscheint stringent vor dem Hintergrund arbiträrer Zeichen, die nur jeweils durch festgelegten Gebrauch – und ebendieser durch gegenseitige Abgrenzung der Zeichen und Bezeichneten – Bedeutung erlangen können.

CABRÉ/VIVALDI PALATRESI (2014, 1488) nennen als Mindestkriterien für einen *Terminus* „unithood (internal coherence of the unit)“,

„termhood (unit usage in one or more domains)“ und „specialised usage (unit usage in general language and/or specialised domains)“.

Der soeben nachgezeichnete Bezug zu einem unterliegenden Begriffssystem und die Regel der Ein-Eindeutigkeit – durch Einführung von Vorzugs- und unerwünschten Benennungen – stellen die Konsistenz einer Terminologie sicher (ARNTZ/PICHT/MAYER 2004, 260). Eine derartige Terminologie hilft nach allgemeiner Ansicht dabei, Texte mit eindeutigen Aussagen zu verfassen (vgl. DREWER/ZIEGLER 2011).

Die erwähnte Einbettung der Terminologienutzung oder auch konkreter Termini in textproduktive Arbeitszusammenhänge wird in der Norm jedoch lediglich benannt (vgl. oben), weiter gehende Überlegungen geschehen aus praktischer Perspektive (ARNTZ/PICHT/MAYER 2004, 219). Der Frage, wie Kriterien der Arbeitsprozess-Perspektive in ein Terminus-Modell einfließen können, geht der vorliegende Beitrag nach.

2.2 Kritik am Terminus-Begriff

Im vorangehenden Unterabschnitt wurde der Terminusbegriff der klassischen Terminologielehre vorgestellt. Dies geschah unter Berücksichtigung von dessen ursprünglichem Zweck, und zeigte seine Grenzen auf. In diesem Unterabschnitt sollen diese Überlegungen mit den Argumenten der kognitiven Wende präzisiert werden.

Die kognitive Wende der Terminologie-Lehre brachte schwerwiegende Kritik am Terminus_b-Begriff der klassischen Terminologielehre hervor. Diese trifft deren Kern in dem Punkt, eine Ein-Eindeutigkeit von Bezeichnungen wäre möglich und gar erstrebenswert. Dem hält die kognitiv orientierte Terminologielehre empirische Untersuchungen entgegen, denen zufolge sich Termini im Gebrauch ebenso entwickeln wie andere sprachliche Einheiten. Die Verwendung von Termini und ihre empfundene Korrektheit hängen stark vom jeweiligen Kontext des Gebrauchs ab (FABER 2009, 113f.).

Auch die Anwendung des Natural Language Processing (NLP) für Zwecke der Terminologiearbeit trug dazu bei, dass der lange Zeit so

hermetisch erscheinende Terminus-Begriff in seinem Kern ausgehöhlt wurde. Sie zeigte, dass die Grenzen zwischen Terminus und Nicht-Terminus aus statistischer Sicht fließend sind – und aus Expertensicht umstritten sein können. Letzteres gilt insbesondere für thematische Experten im Vergleich zu professionellen Terminologen (CABRÉ/VIVALDI PALATRESI 2013, 1495; s. a. KIT 2002).

Zusammenfassend besitzen Termini aus linguistischer Sicht eine kommunikative, eine sprachliche und eine kognitive Dimension (FABER 2009). Das heißt, dass die sprachliche Form, ihr Gebrauch in Kommunikationssituationen und ihre Entstehung und Nutzung als Ergebnis menschlichen Denkens in den Fokus des Interesses treten.

An dieser Stelle möchte ich anknüpfen. Zu diesem Zweck führe ich eine meta-kommunikative Dimension ein. Als Teil eines Frame-Modells (bzw. als Wert in einem entsprechenden *slot*) bezieht sich diese Funktion auf den Nutzen eines Terminus in einer aus professionellem Wissen potenziell eintretenden Kommunikationssituation. Die in Abschnitt 3 unten vorgestellte Vergleichsstudie zielt darauf, das Vorhandensein einer meta-kommunikativen Funktion zu überprüfen, indem die Ergebnisse intellektueller Termextraktion mit denen maschineller Termextraktion verglichen werden.

Zuvor werden in zwei getrennten Unterabschnitten zunächst Details bezüglich der meta-kommunikativen Funktion sowie der computergestützten Terminologiearbeit erörtert.

2.3 Terminologiearbeit als Teil eines Arbeitsprozesses

In den vorangegangenen Unterabschnitten wurden Begriff und Zielsetzung von Terminologiearbeit beschrieben. Ebenso fand eine kritische Auseinandersetzung mit dem Terminus-Begriff statt, die auf der Literatur basiert. In diesem Abschnitt geht es um eine weitere Ergänzung des Terminus-Begriffs um die bereits erwähnte meta-kognitive Dimension.

In der theoretischen Auseinandersetzung zeigte sich, dass die systematische Terminologiearbeit zwar strategisch angelegt und als solche in einen Arbeitsprozess eingebunden ist. Es zeigte sich aber auch, dass

die Einbindung in einen Arbeitsprozess sich in den einschlägigen Definitionen bestenfalls indirekt widerspiegelt. Somit bleibt zu hinterfragen, inwiefern eine Inhaltsdefinition von *Terminus* der tatsächlichen Nutzbarkeit von Terminologie in Arbeitsprozessen gerecht wird.

Diese Frage ist sowohl aus terminologischer als auch aus metalexikographischer Perspektive überlegenswert. Metalexikographie (fr. *méta-lexikographie*) verstehe ich dabei in einem weiten Sinn mit HAUSMANN als

toute activité qui fait du dictionnaire un objet de réflexion et de recherche mais qui, elle-même, ne vise pas à la production de dictionnaire. (HAUSMANN 1989, 216)

Im selben Sinne grenzt WIEGAND (1998, 73) die Metalexikographie von der Lexikographie ab. HARTMANN (2001, 29) ist zuzustimmen, wenn er eine Erweiterung des Gegenstandsbereichs im Sinne einer „reference science“ vorausieht.

Die entsprechende Forschung steht allerdings laut HARTMANN (2001, 24) vor einer entscheidenden Schwierigkeit:

Dictionary compilers do not [...] communicate directly with their target usership, but via the product of their work, the dictionary text[.]

Da Terminologiearbeit aus der Sicht des vorliegenden Artikels Teil eines Arbeitsprozesses ist, trifft diese strikte Trennung auf sie nicht zu. Wenn Übersetzerinnen und Übersetzer selbst Terminologiearbeit betreiben, kann man sogar davon sprechen, dass die von HARTMANN (2001, 25) benannten Handlungsrollen des „compilers“ und „users“ zusammenfallen (s. a. Abbildung 2). Zumindest jedoch sind alle beteiligten Personen in einen gemeinsamen Arbeitsprozess eingebunden und antizipieren dadurch den Bedarf der jeweils anderen.

Aus diesem Zusammenhang erwächst ein besonderes empirisches Interesse der Metalexikographie an empirischen Daten über Terminologiearbeit. Denn durch ihre zentrale Stellung in den Arbeitsprozessen von technischer Redaktion und Übersetzen überbrückt sie in der Praxis die Kluft zwischen Erstellung und Nutzung von Terminologie (s. a. Abbildung 2).

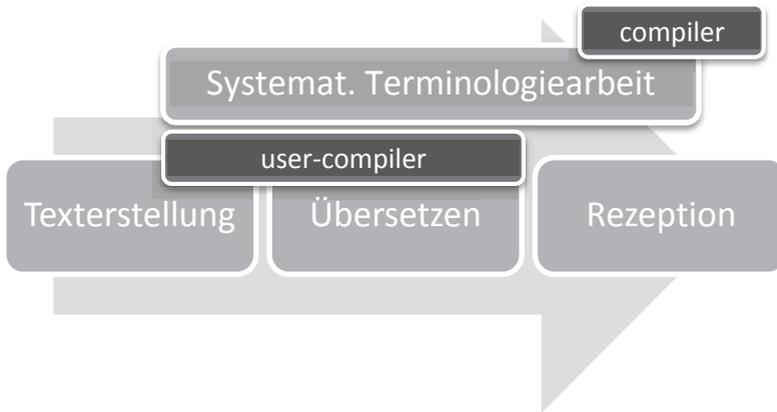


Abb. 2: Enge Verknüpfung von „user“- und „compiler“-Rolle bei der Terminologearbeit als Teil eines Arbeitsprozesses

Im vorliegenden Artikel wird die beschriebene Zwischenstellung als meta-linguistische, genauer: arbeitsprozessbezogene Dimension in die Analyse des Terminus-Begriffs einbezogen. Der hierzu gewählte Ansatz ist empirisch und vergleichend.

Die meta-kommunikative Dimension soll hier definiert sein als

Die Nützlichkeit von Termini für Textproduktionsaufträge im Rahmen eines Arbeitsprozesses.

Der Definition liegen zwei wesentliche Erweiterungen des Terminus_b-Begriffs zugrunde: erstens die Einbettung von Termini in einen textproduktiven Arbeitszusammenhang und zweitens die Fokusverschiebung auf die handelnde Person in ihrer Doppelrolle als antizipierender „compiler“ oder bewertender „user“.

Präziser dürfte eine arbeitsprozessbezogene Sicht auf Termini in der Praxis die Schnittmenge zweier Aspekte sein. Der erste Aspekt sind die explizit festgelegten (und immer noch gelehrten) Merkmale von Termini gemäß den Definitionen aus dem Abschnitt „Klassische Terminologielehre“. Den zweiten Aspekt bilden – in der Regel implizite – Annahmen über die Verwendung der Terminologie im Arbeitsprozess.

Daraus ergeben sich drei Kriterien für Terminus-Begriff aus Prozess-Sicht:

1. Der Termkandidat ist charakteristisch für das Thema/Fachgebiet des Textes.
2. Der Termkandidat ist schwierig zu übersetzen.
3. Es ist für die künftige Textproduktion oder -rezeption von erheblichem Nutzen, den Term(kandidaten) im Datenbestand zu finden.

Die Bezeichnung *Termkandidat* wird hier verwendet, um die Konsistenz über alle nachfolgenden Ausführungen sicherzustellen. Denn in der Praxis werden sowohl intellektuell als vor allem auch maschinell generierte Vorschläge für Termini in einem weiteren Arbeitsschritt überprüft, bevor sie endgültig in den Bestand der Termini aufgenommen werden. Besonders für die maschinelle Extraktion hat sich für die Ergebnisse der reinen Extraktion die Bezeichnung *Termkandidaten* (Singular: *Termkandidat*) etabliert (vgl. CABRÉ/VIVALDI PALATRESI 2013, 1496). Dies wird der Konsistenz wegen für den intellektuell-manuellen Extraktionsprozess übernommen.

Zwischen den Kriterien für den Terminus aus prozessorientierter Sicht ist eine gewisse Interdependenz anzunehmen, die das Dreieck des prozessorientierten Terminus-Begriffs in Abbildung 3 herausstellt:

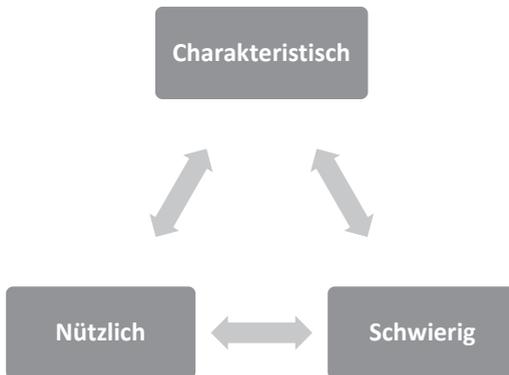


Abb. 3: Kriterien des prozessorientierten Terminus-Begriffs

Wie genau sich die Prozessdimension in den Termini niederschlägt, wird sich durch die ausgewählten Termini der Vergleichsstudie zeigen. Die intellektuell extrahierten Termkandidaten werden dabei mit einer Liste computerlinguistisch gewonnener verglichen. Dem voraus geht im nun folgenden Unterabschnitt die Auseinandersetzung mit dem Terminus-Begriff der Computerlinguistik, deren in diesem Artikel behandelte Anwendung in der Terminologearbeit als maschinelle Termextraktion bezeichnet wird (CABRÉ/VIVALDI PALATRESI 2013, 1489).

2.4 Computerlinguistisches Terminus-Modell

In den vorangegangenen Unterabschnitten stand stets ein Terminus-Begriff im Mittelpunkt. Dieser wurde kritisch erörtert und mit Dimensionen versehen. Im Hinblick auf die maschinelle Termextraktion stellt das Terminus-Modell das Gegenstück zum Terminus-Begriff dar. In der Modellierung spiegeln sich die Eigenschaften von Termini bzw. potenziellen Termini, sogenannten *Termkandidaten* wider. Diese werden in der Regel in einem abschließenden Schritt durch menschliche Prüfer als Termini angenommen oder auch abgelehnt.

Unter „computerlinguistischen Modellen“ soll hier Folgendes verstanden werden:

die Gesamtheit aller (software-)technisch implementierten Kriterien für die Auswahl von Termkandidaten.

Bei dieser Definition bleibt eine mögliche Diskrepanz zwischen Modell und Implementierung außen vor. Anknüpfend an die vorangehende Diskussion des Terminus-Begriffs und seiner Dimensionen muss aber mit CABRÉ/VIVALDI PALATRESI (2013, 1488) grundsätzlich festgestellt werden, dass

[t]he implementation of information technologies as tools for terminological work has [...] turned the terminographic methodology into a semasiological procedure[.]

Das heißt also, dass hier lediglich auf die Ausdrucksseite des sprachlichen Zeichens abgehoben wird. Dort, und nirgendwo anders, müssen die maschinell auswertbaren Eigenschaften zu finden sein, aufgrund derer die Entscheidung: „Termkandidat – ja oder nein“ getroffen wird.

Die computerlinguistische Umsetzung der Termextraktion wird auf breiter Basis erforscht und umfasst zahlreiche Ansätze, die sich stark voneinander unterscheiden (s. hierzu CABRÉ/VIVALDI PALATRESI 2013, 1489ff.). Eine grundsätzliche Auseinandersetzung mit Fragen der computerlinguistischen Modellierung von Termini findet sich bei KIT (2002). Mit der *unithood* diskutiert der Autor die zentrale Verbindung zwischen intelligiblem Konzept und dessen Umsetzung in maschinengerechte Regeln.

Unithood meint dabei

the degree of strength or stability of syntagmatic combinations or collocations. (KIT 2002)

Termhood – verstanden als Merkmal des Terminus-Begriffs – ist

the degree that a linguistic unit is related (or more straightforwardly, represents) domain-specific concepts. (ebd.)

Im Hinblick auf das computerlinguistische Terminus-Modell ist die entscheidende Frage, inwiefern sich *termhood* auf *unithood* abbildet.

Von den zahlreichen Ansätzen, mit denen dies versucht wird, soll hier nur kurz derjenige vorgestellt werden, der in der nachfolgenden Vergleichsstudie für die maschinelle Termextraktion zum Einsatz kam. Dieser ist ausführlich in GOJUN u. a. (2012) dokumentiert. Es handelt sich um eine so genannte „slim linguistic analysis“ (sla). Das heißt, dass für die Termextraktion keine sprachspezifischen Grammatikmodelle zum Einsatz kommen („generic term extraction tool“). Stattdessen beruht das Verfahren auf zwei Schritten:

Lemmatisierung und POS-Tagging mit dem Tool „TreeTagger“ (SCHMID 1994, s. GOJUN u. a. 2012, 651) und Aufstellung einer Rangfolge der gefundenen Termkandidaten nach Domänenspezifität (ds) (ebd.). Bei Letzterem handelt es sich um die relative Häufigkeit eines

Termkandidaten in einem fachspezifischen gegenüber einem allge-meinsprachlichen Korpus (AHMAD u. a. 1992).

Der Vollständigkeit halber soll nicht unerwähnt bleiben, dass der soeben skizzierte sla-Ansatz für das Deutsche um eine tiefere morphologische Analyse erweitert wurde. Anders war laut GOJUN u. a. (2012, 652) die Gewinnung der Zitierformen nicht realisierbar.

Der *Terminus*-Begriff, der sich aus der vorgestellten Modellierung ableiten lässt, besitzt demnach folgende Charakteristika: Termini sind sprachliche Zeichen, die in Texten über ein Fachgebiet überproportional häufig vorkommen (Domänenspezifitätswert ds). Weiterhin weisen Termini dem Modell zufolge spezifische morpho-syntaktische Muster der Benennungsbildung auf (POS-Tagging). Die Modellierung der *unithood* wird damit verfeinert und eine korrekte Identifizierung von fachspezifischen Mehrwortbenennungen ermöglicht (s. WELLER/BLAN-CAFORT u. a. 2011; WELLER/HEID 2012; GOJUN u. a. 2012).

Die vorangehenden Ausführungen machen jedoch vor allem eines deutlich: alle computerbasierte Termextraktion setzt immer an der Form des sprachlichen Zeichen (also Bühlers Symbolfunktion) an. Dies gilt – entgegen dem Anschein der Benennung – auch für so genannte „knowledge-rich approaches“. Sie berücksichtigen zwar weitergehende Modellierungsinformationen. Diese beziehen sich jedoch ihrerseits lediglich auf die Ausdrucksseite des sprachlichen Zeichens.

Zuvor war immer wieder von Termkandidaten die Rede. Dabei handelt es sich um die korrekte Bezeichnung von möglichen Termini. Ob intellektuell oder maschinell gewonnen, werden sie einem weiteren Prüfschritt durch menschliche Prüfer unterzogen. In der Praxis werden die maschinell erstellten Termlisten ggf. um die Funde menschlicher Terminologien ergänzt (CABRÉ/VIVALDI PALATRESI 2013, 1496). In der wissenschaftlichen Evaluation von Extraktionsverfahren erfolgt ein Abgleich mit menschlich erstellten Listen. Auf dieser Basis wird dann versucht, etwaige Lücken der softwaregenerierten Terminologie systematisch zu beschreiben. Ziel ist wiederum die Verbesserung der Term-suche bzw. Treffergenauigkeit der Software anhand von Merkmalen der sprachlichen Form (GOJUN u. a. 2012, 562f.).

Aus den Ausführungen des vorangehenden Absatzes geht hervor, dass die Charakteristika, die *termhood* ausmachen, implizit bleiben. Zwar gibt es in der Literatur Hinweise darauf, dass Terminologen beispielsweise „add (generic) terms to the collection in order to make it logically consistent“ (GOJUN u. a. 2012, 563). Die weiteren, nicht auf die Konsistenz des unterliegenden Wissensgebiets bezogenen Kriterien im Sinne des Arbeitsprozesses bleiben jedoch unhinterfragt. Um sie besser zu verstehen, legt der vorliegende Beitrag ein prozessorientiertes Modell von *termhood* an.

Im folgenden Abschnitt werden die Durchführung und Ergebnisse des Vergleichs zwischen maschinell und intellektuell extrahierten Termkandidaten dargestellt.

3. Vergleichende Studie zur Termkandidatenauswahl

Der vorangegangene zweite Abschnitt beinhaltete eine intensive Auseinandersetzung mit dem Terminus-Begriff. Dabei wurde vor allem auf den Unterschied zwischen dem Terminus-Begriff der Terminologielehre und dem Terminus-Modell computerlinguistischer Anwendungen unterschieden. Während der Terminus-Begriff für die Ziele des vorliegenden Aufsatzes eine Erweiterung um eine meta-kommunikative Dimension erfährt, sind die Modelle der Computerlinguistik zwangsläufig auf die Ausdrucksseite des sprachlichen Zeichens beschränkt – und zwar unabhängig von ihrem Detaillierungsgrad.

Dieser Abschnitt hat das Ziel, die Abweichungen des bei intellektueller und maschineller Termextraktion unterliegenden Begriffs bzw. Modells zu untersuchen. Dies geschieht, indem Unterschiede zwischen den Listen extrahierter Termini identifiziert und systematisiert werden. So wird die metakommunikative Dimension von Termini im Sinne des Aspekts „nützlich für den Arbeitsprozess“ beleuchtet.

3.1 Durchführung der Studie

In diesem Unterabschnitt werden die Ergebnisse der empirischen Studie vorgestellt. Für die Studie wurden die extrahierten Termkandidaten eines intellektuellen und eines maschinellen Extraktionsvorgangs verglichen. Die intellektuelle Extraktion wurde von Studierenden in 14 2er- bzw. 3er-Gruppen durchgeführt. Sowohl der intellektuellen als auch der maschinellen Extraktion lagen 13 verschiedene deutschsprachige Texte zugrunde, in denen unterschiedliche technologische Themen behandelt werden. Die Texte haben eine durchschnittliche Länge von 783 Wörtern (tokens) und entstammen der Zeitschrift „Technology Review“ von März 2006 bis Juni 2013. Die Textauswahl erfolgte aufgrund der Eignung für die Zielgruppe, der technologieorientierten Themen sowie von thematischen Überschneidungen zwischen den einzelnen Texten.

Aus denselben Texten wurden dann auch maschinell die Termkandidaten extrahiert. Das verwendete Verfahren wurde bereits weiter oben in Abschnitt 2.4 beschrieben (s. a. GOJUN u. a. 2012). Berücksichtigt wurden für die vorliegende Studie jedoch ausschließlich nominale Einwortbenennungen. Kriterien für den Vergleich sind die aus den beiden Verfahren hervorgegangenen Termkandidatenlisten sowie bei der intellektuell erstellten Kandidatenliste die Begründungen der Studierenden für deren Auswahl. Die Begründungen sollten während der Extraktionsaufgabe in ein dafür extra angelegtes Feld der Eingabemaske eingetragen werden. Die Eingrenzung der berücksichtigten Termini der maschinellen Liste stützt sich zudem auf die Werte der Domänenspezifizität (ds).¹ Eine nähere Betrachtung geschieht anhand der Benennungs- und Begriffsbildungsmuster sowie der Bezugnahmen auf diese in den Begründungen der Studienteilnehmer.

¹ Ich danke an dieser Stelle Nicole Buller, die im Rahmen einer Seminararbeit die Ergebnisse detailliert aufgearbeitet hat, für ihre Zuarbeit.

3.2 Intellektuell extrahierte Terminikandidaten

Die Anzahl der von den Studierenden manuell extrahierten Termkandidaten (nachfolgend in eckigen Klammern „<...>“ notiert) beträgt 61. Die Wahl der Bezeichnung *Termkandidat* geschieht hier aus Konsistenzgründen – es soll die Parallelität mit dem maschinell durchgeführten Arbeitsschritt gekennzeichnet werden. Aus der Liste der intellektuell extrahierten Termkandidaten wurden 5 (CCS-Technik, Einlagerung, Energieversorger, Emission(en) und Supraleitung) von zwei Gruppen extrahiert – die ersten vier aus demselben, doppelt vergebenen Text. Entsprechend verbleiben 56 manuell extrahierte Termkandidaten. Davon entfallen sechs auf Adjektive und Mehrwortbenennungen (Adjektiv-Nomen- und Nomen-Nomen-Kombinationen). Diese wurden im Vergleich intellektuell-manuell aufgrund der gewählten Einschränkung des maschinellen Extraktionsverfahrens nicht berücksichtigt. Es verbleiben also 49 Einwort-Termkandidaten (s. a. Tabelle 1), die mit den Ergebnissen der maschinellen Termextraktion verglichen werden können.

Zunächst sollen jedoch die Charakteristika der manuell extrahierten Termkandidaten an sich herausgearbeitet werden. Die Annahme eines unterliegenden Kriteriums für den Terminus-Begriff unterstreicht auch die Tatsache, dass nicht nur aus demselben Text, sondern auch aus unterschiedlichen Texten dieselben Termkandidaten extrahiert wurden. So wurde der Termkandidat <Supraleiter> aus zwei verschiedenen Texten als „Fachvokabular“ identifiziert. Beide Texte behandeln die Anwendung von Supraleitung, haben jedoch unterschiedliche Technologien und unterschiedliche Anwendungen zum Thema. Unabhängig davon hat sich die Bezeichnung <Supraleitung> für die Studierenden als thematisch einschlägig und als fachlich dargestellt (mehr zu Begründungen s. Abschnitt 4). Dass aus demselben Text von zwei Gruppen eine Schnittmenge von Termkandidaten, nämlich <CCS-Technik>, <Einlagerung>, <Energieversorger>, <Emission(en)> extrahiert wurden, unterstreicht die Annahme, dass es offenbar gemeinsame Kriterien gibt, aufgrund derer Studierende Termkandidaten auswählen.

Die soeben gemachten Ausführungen zeigen, dass es durchaus sinnvoll ist, manuell gewonnene Termkandidaten an und für sich zu

betrachten, statt ausschließlich eine maschinelle Vorauswahl manuell nachzubewerten. Um diesen möglichen Beitrag manueller Extraktion weiter zu untermauern, werden im nun folgenden Textabschnitt die Begründungen diskutiert, die Studierende für die Auswahl des jeweiligen Termkandidaten angegeben haben. Die Ergebnisse dieses Schritts sind in der nachstehenden Tabelle 1 aufgelistet:

Termkandidat	Begründung der Auswahl (wörtlich übernommen)
Auflösung	unterschiedliche Verwendung
Hologramm	unterschiedliche Verwendung
Interferenzmuster	Fachbegriff
Clusterauge	Fachbegriff
Bildchip	Fachbegriff
Lichtfeldkamera	eine spezielle Art von Kamera
Digitale Halluzination	unbekannte Wortkombination
Mikrolinsengitter	Fachbegriff
Tiefeninformation	unbekannte Wortkombination
Brennstoff	weil is so.
Brennstoffzelle	Oberbegriff+unterscheidendes Merkmal Kompositum
Petroleumlampen	siehe Brennstoffzelle
Autismus	Bei dem Text handelt es sich um einen Fachtext über Autismus.
Diagnose	Ist ein essentielles Verfahren in der Medizin.
Entwicklungsstörung	Fachausdruck, der im Text noch öfter auftreten könnte, da es um Autismus geht.
Genaktivität	Fachbegriff, der oft im text vorkommt und dem Übersetzer seine Arbeit erleichtert.
Elektrolyt	chemische Verbindung
Lithium-Polysulfide	chemische Elemente
Vanadium	chemisches Element
Elektrode	Elektronenleiter
Flüssigstickstoff	Fachspezifisch
Kernspintomografen	Medizinisches Gerät
Niob	Metallart
Plexiglas	bestimmtes Material
Supraleiter	Supraleiter
Windturbinen	technische Maschine
Yttrium-Barium-Keramik	Bestimmte Art von Keramik
Flüssighelium	bestimmtes Material

Termkandidat	Begründung der Auswahl (wörtlich übernommen)
CCS-Technik	Verfahren
Einlagerung	technischer Vorgang
Emissionen	Umweltverschmutzung, Bezug wichtig
Energieversorger	zusammengesetztes Wort
Carbon	chemisches Element
dezellularisierte	Bedeutung nicht klar
extrazellulären Matrix	Spezifizierung des Nomens durch das Adjektiv
Immunsystem	taucht nur in fachspezifischen Zusammenhängen auf
Tissue Engineering	Eigenname der zu bestimmten Fachbereich gehört
Transplantation	Dieser Begriff ist einem speziellen Fachgebiet zugeordnet (Medizin)
Tuberkulose	spezieller Eigenname
Alphagalaktose	Fremdwort im Bereich der Chemie
toter Winkel	Toter Winkel ist ein Begriff im Fachbereich der Autowelt, welcher es dem Fahrer unmöglich macht einen bestimmten Bereich einzusehen.
Wassermolekül	Es gehört zum Fachgebiet der Chemie.
Kühlfüssigkeit	Der Ausdruck ist nicht sehr geläufig.
Bundesstaat	andere Sprachen haben andere Bezeichnungen, politisches System nicht überall dasselbe
Einlagerung	kontextabhängig
Emission	schwer zu erklären, nicht allgemein bekannt
Energieversorger	wortwörtliche Übersetzung problematisch
Kohlekraft	Themenbereich nicht jedem bekannt
Kohlekraftwerk	schwer zu übersetzen, nicht wörtwörtlich (sic!) übersetzbar
Megawatt	Themenbereich nicht allen bekannt
Ölförderung	kontextabhängig
Stromkunde	kein gängiger Begriff
CCS-Technik	Fachbegriff und nicht allgemein bekannt
connection chemistry	Schlüsselwort für neue Medikamententechnik
Verbindungschemie	Schlüsselwort für neue Medikamententechnik
Catabasis	Eigenname, der in der Überschrift steht. Der Artikel handelt von dieser Firma.
Supraleiter	Fachvokabular
Supraleitung	termini aus einem fachgebiet
Widerstand	widerstand trägt mehrere bedeutungen
Draht	es gibt viele drähte

Tab. 1: Intellektuell ausgewählte Termkandidaten als Ergebnisse der studentischen Gruppenarbeit

An dieser Stelle werden die von den Studienteilnehmern gegebenen Begründungen für die Terminus-Eigenschaft des ausgewählten Termkandidaten klassifiziert und diskutiert. Formbezogene Argumente, also solche mit Bezug zur Ausdrucksseite des sprachlichen Zeichens, stehen dabei neben inhalts- und prozessbezogenen.

Mehrwortbenennungen scheinen als Termini für die Probanden eine gewisse Rolle zu spielen. Dies deutet auf eine Bewusstheit der *unithood*, also der Bezugnahme der Bezeichnung auf ein einzigartiges Konzept, hin. Dies gilt für alle vier Beispiele des Korpus, die nachfolgend mit der jeweiligen Begründung wiedergegeben sind:

- <extrazelluläre Matrix>: „Spezifizierung des Nomens durch das Adjektiv“
- <Tissue Engineering>: „Eigenname der zu bestimmten Fachbereich gehört“
- <toter Winkel>: „Toter Winkel ist ein Begriff im Fachbereich der Autowelt, welcher es dem Fahrer unmöglich macht einen bestimmten Bereich einzusehen“
- <digitale Halluzination>: „unbekannte Wortkombination“

Bei allen inhalts- und terminologiefachlichen Fehlern wird deutlich, dass die Spezifik des jeweils repräsentierten Begriffs ein Auswahlkriterium war. Im ersten Beispiel wird hier gar nur auf den Benennungsbildungsprozess abgehoben, in den beiden folgenden wird unmittelbar auf den dahinterstehenden Begriff verwiesen, es zeigt sich also das Kriterium der *unithood* als Denkeinheit.

Beispiele sind die Mehrwort-Termkandidaten wie <digitale Halluzination> oder <toter Winkel>. Auch wenn der Fachlichkeitsgrad in beiden Fällen mehr als diskussionswürdig erscheint, zeigen die Studierenden hier, dass sie zweierlei erkannt haben: Erstens, dass sich Termini durch den Bezug zu Begriffen auszeichnen, die für den jeweiligen fachlichen Kontext spezifisch und als Denkeinheiten unzerlegbar sind. Somit ist „Halluzination“ selbstverständlich kein domänenspezifisches Wort. <Digitale Halluzination> kann aber sehr wohl auf ein Konzept referenzieren, dem diese Qualität in bestimmten kommunikativen Kontexten zu eigen ist. Zweitens müssen Termini nicht dem Wortkörper

nach als solche zu identifizieren sein. So wurde <toter Winkel> gefunden, das terminologisierte Alltagssprache und zudem eine nicht aus ihren Bestandteilen erschließbare Mehrwortbenennung darstellt.

In weiteren Fällen dominiert offenbar die Thematik des Textes als Auswahlkriterium:

- <Supraleitung>: „termini (sic!) aus dem Fachgebiet“
- <Transplantation>: „Dieser Begriff ist einem speziellen Fachgebiet zugeordnet (Medizin)“
- <Autismus>: „Bei dem Text handelt es sich um einen Fachtext über Autismus.“

Der thematische Bezug, gepaart mit der Unbekanntheit, kommt bei einem intellektuell extrahierten Kandidaten zum Tragen:

- <Emission>: „Umweltverschmutzung, Bezug wichtig“ (eine Gruppe); „schwer zu erklären, nicht allgemein bekannt“ (weitere Gruppe)

Der Termkandidat <Elektrolyt> wird als „chemische Verbindung“ erklärt, was inhaltlich sicherlich an der Grenze des Akzeptablen und Nützlichen angesiedelt ist.

Probleme mit Mehrdeutigkeit bzw. die Notwendigkeit, bestimmte Übersetzungsstrategien anzuwenden, stehen bei der Auswahl der folgenden Beispiele im Mittelpunkt. Für <Energieversorger> wird angenommen: „wortwörtliche Übersetzung problematisch“, bezüglich <Bundesstaat> wird richtiggehend angeführt: „politisches System nicht überall dasselbe“, wobei natürlich gerade im letztgenannten Fall die Termeigenschaft als solche wegen sehr geringer Fachlichkeit angezweifelt werden kann. In einer Begründung heißt es: „[W]iderstand trägt mehrere [B]edeutungen“, für <Einlagerung> „kontextabhängig“, was auf eine Mehrdeutigkeit verweist, die hier von den Probanden erkannt bzw. für eine mögliche Übersetzung antizipiert wurde. Somit können also die spätere Übersetzung und die Einbettung in einen bestimmten Text und Kontext eine Rolle spielen.

Unabhängig von den Begründungen weisen die intellektuell extrahierten Termkandidaten gemeinsame Merkmale auf. Aus Gründen der Nachvollziehbarkeit und um Wiederholungen zu vermeiden, werden

diese in Abschnitt 4 beim Vergleich der Termkandidaten diskutiert. Zunächst folgt jedoch die Charakterisierung der maschinell extrahierten Termini.

3.3 Maschinell extrahierte Termkandidaten

Wie auch bei den intellektuell extrahierten Termkandidaten geschehen, geht es in diesem Abschnitt darum, unabhängig vom Vergleich die Ergebnisse der Extraktion (in ihrer maschinellen Variante) zu erörtern. Gemäß dem in Abschnitt 2.4 beschriebenen Verfahren wurden aus den Texten sämtliche Nomina extrahiert und auf ihre Termeigenschaft hin untersucht. Letzteres beruht auf dem Teil-Modell der Domänenspezifität (ds-Wert nach AHMAD u. a. 1992).

Insgesamt wurden mit dem maschinellen Verfahren 282 Termkandidaten mittlerer oder hervorragender Güte extrahiert, wenn man ds-Schwellenwerte von rund 14.000 bzw. rund 35.000 anlegt. Unter den 48 gefundenen Kandidaten der Top-Kategorie befinden sich jedoch 14 Eigennamen, unter den nachfolgenden 234 weitere 22. Hinzu kommen generische Funktionsbezeichnungen wie „Wissenschafts-Chef“ oder „Stanford-Forscher“. Letztere typisch wissenschaftsjournalistischen Bildungen (s. SCHNEIDER 1999, 74) werden interessant im Vergleich zu analogen Bildungen, bei denen der Firmen- oder Markenname aufgrund von Neuheit tatsächlich für eine terminologische Einheit steht. Die hohe Anzahl der Namen, aber auch das Außerachtlassen der Mehrwortbenennungen deuten darauf hin, dass auch die maschinelle Extraktion Fehler im Terminus-Modell aufweist.

In diesem Abschnitt werden die Charakteristika der maschinell extrahierten Termkandidaten näher beleuchtet werden. Aus Platzgründen findet dabei eine Konzentration auf diejenigen Wörter statt, die aufgrund eines hohen ds-Wertes als Top-Termkandidaten eingestuft wurden. Hier lässt sich mit hoher Wahrscheinlichkeit davon ausgehen, dass es sich am Ende tatsächlich um einen *Terminus* handelt. Die nachfolgende Tabelle 2 listet diese Top-Kandidaten auf.

n	ds	Termkandidat
5	117.100	Verbindungschemie
5	117.100	Flussbatterien
4	93.700	Start-up
4	93.700	Biomarker
3	70.300	Lithium-Schwefel-Batterien
3	70.300	Lichtfeldkamera
3	70.300	Lichtfeldfotografie
3	70.300	Keramikkabel
3	70.300	Half-Sweep
3	70.300	Genaktivität
3	70.300	Clusterauge
3	70.300	Bildsensor
2	47.000	Vanadium-haltige
2	47.000	Tiefeninformation
2	47.000	SynapDX-Test
2	47.000	Strombegrenzer
2	47.000	Mikrolinsengitter
2	47.000	Megapixel
2	47.000	Lithium-Polysulfide
2	47.000	Kochöfen
2	47.000	Hybrid-Batterie
2	47.000	Gigapixel
2	47.000	Gigapixel-Auflösung
2	47.000	Flüssigstickstoff
2	47.000	Filament
2	47.000	Depth
2	47.000	Defocus
2	47.000	CCS-Technik
2	47.000	Bildchip
2	47.000	Alphagalaktose
2	47.000	A4-Scanner
3	35.100	Kohlekraft
3	35.100	DoE

Tab. 2: Maschinell extrahierte Termkandidaten mit hoher Domänenspezifität (ds). n steht für die Anzahl der Vorkommen (tokens) im Korpus. Eigennamen wurden manuell aus der Liste entfernt.

Bei einem Blick auf die Daten fällt vor allem ein Charakteristikum ins Auge: der hohe Anteil von Komposita, die als spezifisch für die Fachwortbildung des Deutschen gelten. Im Gegensatz zu den Mehrwortbezeichnungen sind sie nicht vom Extraktionsverfahren ausgeschlossen.

Beispiele für Komposita unter den Termkandidaten der Top-Kategorie sind <Bildsensor> und <Genaktivität>. Legt man die genutzten Texte zugrunde, wird schnell deutlich, weshalb diese Termini extrahiert wurden. Der Text, aus dem <Bildsensor> stammt, ist vergleichsweise kurz. Entsprechend führt bereits einmaliges Auftreten in diesem Text zu einem hohen ds-Wert, sofern der Termkandidat im Vergleichskorpus wenig vertreten ist. Der Text, aus dem der nicht als Top-Kandidat eingeordnete <Biomarker> extrahiert wurde, ist fast doppelt so lang.

Wenn einige der maschinell extrahierten Kandidaten letztlich doch nicht als Termini angesehen werden können, dann aufgrund der Tatsache, dass ihre ausschließliche Nutzung zum Bezug auf eine Denkeinheit nicht gegeben – und somit deren *unithood* unterspezifiziert – ist. Dies ist in zwei der soeben behandelten Fälle aus dem Textumfeld leicht zu erkennen. So ist das einzige Auftreten von <Bildsensor> gekennzeichnet durch eine attributive Ergänzung („digitaler“), die allein typologisch auf eine Mehrwortbezeichnung hindeuten würde. Hinzu kommt, und auch dies ist typisch für den hier diskutierten Typ Termkandidat, dass eine mögliche Nutzung eines Technologietyps beschrieben wird: „Digitale Bildsensoren (CCDs) benötigen dazu Auflösungen im Gigapixelbereich, wie sie reguläre Kameras nicht erreichen.“² Dasselbe Schema liegt auch folgendem Exztrakt aus dem <Biomarker>-Text zugrunde: „Mindestens vier verschiedene Gruppen haben potenzielle Biomarker identifiziert[.]“³ „Biomarker“ (Plural) ist durch „potenzielle“ attribuiert. Daraus lässt sich klar erkennen, dass es sich nicht um einen spezifischen Sachverhalt „Biomarker“ handelt, sondern um einen Oberbegriff.

² <http://www.heise.de/tr/artikel/A4-Scanner-nimmt-Hologramme-auf-1883450.html> [2014-04-08]

³ <http://www.heise.de/tr/artikel/Autismus-Diagnose-per-Bluttest-1729089.html> [2014-04-08]

Hinzu kommt, wie bereits beim vorangegangenen Beispiel, die hypothetische Beschreibung technischer Möglichkeiten.

Weiterhin kritisiert NICOLE BULLER in ihrer in Fußnote 1 erwähnten Hausarbeit vor terminologiewissenschaftlichem Hintergrund für <Genaktivität>: „Als Kompositum hat [es] keine neue Bedeutung, es ist nicht spezifischer und beschreibt keinen Sachverhalt genauer [als die Phrase ‚Aktivität von Genen‘].“ Dies sieht sie richtiggehend als Argument gegen den ds-Faktor als Teil des Terminus-Modells an. Nicht unerwähnt bleiben soll auch, dass dieselbe Implementierung des Terminus-Modells zusätzlich Augenblicksbildungen zutage fördert. Teilweise handelt es sich dabei um wissenschaftsjournalistische, wie <Auflösungstestreihen> in der Phrase „die Auflösungstestreihen der US-Luftwaffe“. Hier schlägt sich ein ganzer Sachverhalt in einem Kompositum nieder (zu diesem Phänomen s. WEISSGERBER 2010, 176f.) ohne dass der Sachverhalt selbst hinreichend spezifisch wäre.

Ein positives Beispiel aus der Gruppe der Bindestrich-Komposita ist <Inline-Verfahren>. Bei diesem zeigen sich keine Hinweise in der Textstelle, dass der Termkandidat einen unterspezifizierten, hypothetischen Charakter haben könnte. Der Termkandidat steht in der Phrase „das sogenannte Inline-Verfahren“. Im unmittelbar darauffolgenden Satz wird das Verfahren genau beschrieben. Ein weiteres Beispiel für dieselbe Fallklasse ist der Termkandidat <Lithium-Schwefel-Batterien>. Im Textzusammenhang handelt es sich dabei um einen Begriff mit hoher *unithood*, d. h. es wird klar, dass eine bestimmte, fest definierte und mit diesem Wort beschriebene Technologie gemeint ist: „... eine Konstruktion [...], die die Technologien von Flussbatterien und von Lithium-Schwefel-Batterien kombiniert.“⁴ Hier zeigt sich ein Vorteil des maschinellen Verfahrens, das alle Termkandidaten eines bestimmten Bildungsmusters (POS-Tagging) und einer bestimmten relativen Häufigkeit vorschlägt. Betriebsblindheit, d. h. das Ausschließen von Termkandidaten aufgrund zuvor als wenig erfolgversprechend kennengelernter Muster (z. B. Komposita aus Alltagswörtern bei Wort-

⁴ <http://www.heise.de/tr/artikel/Die-Kraft-der-Polysulfide-1851263.html>
[2014-04-10]

bildungsart Terminologisierung) tritt nicht auf. Weshalb der Termkandidat manuell nicht extrahiert wurde, wird die weitere Diskussion im nächsten Unterabschnitt zeigen.

In diesem Unterabschnitt wurden die wesentlichen Charakteristika der maschinell extrahierten Termkandidaten herausgearbeitet. Im nun folgenden Unterabschnitt werden die maschinell extrahierten Termkandidaten mit den intellektuell extrahierten verglichen.

4. Vergleich der intellektuell und maschinell extrahierten Termkandidaten

In diesem Abschnitt erfolgt der Vergleich der maschinell mit den intellektuell identifizierten Termkandidaten. In einem ersten Schritt werden dazu die jeweils entstandenen Listen abgeglichen. Anschließend werden Termkandidaten, die nur in einer der beiden Listen auftreten, einer näheren Betrachtung unterzogen. Diese greift bei den intellektuell extrahierten Termkandidaten auf die Begründungen der Studierenden für die Auswahl zurück, für die maschinellen auf das Terminus-Modell. Weiterhin stützt sich der Vergleich auf Analysen unterliegender Begriffs- und Benennungsbildungsmuster. Ziel ist es, den Terminus-Begriff durch Vergleich stärker zu konturieren, um so die Anforderungen aus Sicht eines terminologischen bzw. lexikografischen Prozesses herauszuarbeiten.

Tabelle 3 stellt die intellektuell und maschinell gewonnenen Termkandidaten-Korpora gegenüber. Der Wert „Gesamt“ stellt die Vereinigungsmenge der gefundenen Termkandidaten dar, der Wert „Überschneidung“ die Schnittmenge. Bezüglich der Kriterien werden als Top-Kandidaten solche mit einem ds-Wert von über 23.000 eingestuft, als Mittel-Kandidaten solche mit einem ds-Wert zwischen 23.000 und 14.000. Die Kategorien „Komposita ≥ 3 “ und „Mehrwortbenennungen“ erklären sich aus den weiteren Ausführungen.

Kategorie/Kriterium	Kandidaten intell.	Kandidaten masch.	Gesamt	Überschneidung
Gesamt	56	212	268	15
Top-Kandidaten	13	34	34	13
Top- und mittel K.	15	212	212	15
Komposita > 3	5	28	28	5
Mehrwortbenn.	7	0	7	0

Tab. 3: Überschneidungen der Termkandidaten-Mengen aus unterschiedlichen Vergleichsperspektiven (s. nachfolgende Textabschnitte)

Da der vorliegende Artikel sich insbesondere für die prozessbezogenen Charakteristika von Termkandidaten interessiert, konzentriert sich dieser Abschnitt zunächst auf einen Teil der obigen Termkandidaten-Menge: diejenigen, die von der maschinellen Extraktion nicht berücksichtigt wurden, bei der intellektuellen aber sehr wohl auftauchten. Diese werden typisiert und ihr Ausgewähltwerden anhand der Begründungen sowie unter Rückgriff auf den in Abschnitt 2 eingeführten Terminus₆-Begriff erklärt.

Bei den maschinell nicht erfassten Termkandidaten handelt es sich neben Mehrwortbenennungen (die von der maschinellen Extraktion ausgeschlossen waren, s. Abschnitt 2.4) um das Determinandum bzw. die weiter hinten stehenden Komponenten von Komposita mit mehr als drei Elementen. D. h. es wurde maschinell ein mindestens dreiteiliges Kompositum identifiziert, von dem intellektuell nur das Determinandum in die Auswahl aufgenommen wurde.

Auf der anderen Seite wurden die 5 Komposita mit ≥ 3 Elementen, die intellektuell extrahiert wurden, auch von der computerlinguistischen Anwendung gefunden:

- <Lichtfeldkamera>
- <Mikrolinsengitter>
- <Lithium-Polysulfide>
- <Kernspintomograph>
- <Yttrium-Barium-Keramik>

Für die Fälle, in denen intellektuell lediglich Teil-Komposita gefunden wurden, listet Tabelle 4 die ds-Werte und absoluten Häufigkeiten für die Komposita-Teile sowie für die vollständigen Komposita auf:

Kompositums-Teil	ds	Anz.	Kompositum	ds	Anzahl
Brennstoffzelle	11.700	7	Kompaktbrennstoffzelle	23.400	1
			Festoxidbrennstoffzelle	23.400	1
			Brennstoffzellen-Ladegerät		1
Supraleiter	18.000	10	Hochtemperatur-Supraleiter	17.600	3
			Supraleiterstück	23.400	1
			Supraleiterschicht	23.400	1
Supraleitung	9.200	9	Supraleitungskabel	23.400	1
			Hochtemperatur-Supraleitung	23.400	1
Auflösung	33	9	2D-Auflösung	23.400	1
			Auflösungstestreihen	23.400	1
			Gigapixel-Auflösung	46.900	2
Transplantation	130	2	Transplantationsorgan	11.700	1
			Transplantationsmedizin	276	1
Autismus	2.700	8	Autismusspektrum	23.400	1
			Autismus-Diagnose	23.400	1
Diagnose	102	6	Gendiagnoseverfahren	23.400	1
			Autismus-Diagnose	23.400	1
			Diagnoseverfahren	2.900	2
Elektrolyt	11.700	8	Elektrolyt-Materialien	23.400	1
			Elektrolyt-Flüssigkeit	23.400	1
Emission	62	3	Emissionsstandards	23.400	1
			Emissionshandelssystem	23.400	1

Tab. 4: Komposita-Teile und zugehörige Komposita mit ≥ 3 Elementen

Der Grund für die Nicht-Beachtung der intellektuell extrahierten Termerkandidaten (linke Seite der Tabelle) durch das maschinelle Verfahren ist mit den sehr niedrigen Spezifitätswerten (ds) sofort ersichtlich.

Bezüglich der Teil-Komposita zeigen sich auf der Formseite zwei unterschiedliche Fälle. Zunächst einmal Teile von Komposita mit ≥ 3 Elementen, die ihrerseits Komposita sind (Anzahl im Korpus in Klammern): <Brennstoffzelle> (7), <Supraleiter> (10) und <Supraleitung> (9). Die Begründungen der Studienteilnehmer zeigen, dass teilweise bewusst die Oberbegriffe, also Begriffe mit weniger unterscheidenden Merkmalen, als Termkandidaten ausgewählt wurden:

- <Brennstoffzelle>: „Oberbegriff+unterscheidendes Merkmal“

Das gilt ebenso für die Fälle, in denen die extrahierten Termkandidaten selbst keine Komposita sind:

- <Auflösung>: „unterschiedliche Verwendung“
- <Diagnose>: „Ist ein essentielles Verfahren in der Medizin.“

Es zeigt sich, dass sowohl begriffs- als auch verwendungsbezogene Begründungen eine Rolle spielen.

Beide Begründungstypen zeigen sich auch bei weiteren Typen der intellektuell, aber nicht maschinell extrahierten Termkandidaten. So erscheint die Fachspezifik als Haupt-Auswahlkriterium bei Komposita mit zwei Elementen wie <Flüssighelium> (absolute Häufigkeit: 1; ds 23.400), und <Interferenzmuster> (2; ds 11.700). Die Begründungen lauten hier „bestimmtes Material“ bzw. „Fachbegriff“.

Interessant sind auch Termkandidaten, die oberflächlich Komposita-Charakteristika aufweisen, aber kaum noch als solche empfunden werden: <Plexiglas> (Häufigkeit 1, ds 254), <Megawatt> (3, ds 39) und <Kohlekraft> (3, ds 35.000). Hier erscheint die aristotelische Definition „Oberbegriff plus einschränkendes Merkmal“ bzw. eine entsprechende Motivierung der Benennung einen Einfluss zu entwickeln, wie die Begründung „bestimmtes Material“ für <Plexiglas> es nahelegt. Ein Vergleich der ds-Werte in der linken und rechten Tabellenhälfte zeigt auch, dass der ds-Wert des intellektuell extrahierten Teil-Kompositums fast immer geringer ist als der der zugehörigen Komposita mit ≥ 3 Elementen. Die Ausnahme bildet <Supraleiter>, wobei diese Anomalie auf die unzureichende Lemmatisierung im maschinellen Verfahren (Pluralformen nicht als solche identifiziert) zurückzuführen ist.

Nicht unerwähnt bleiben soll eine Paradoxie des Beispiels <Kohlekraft>. Diese Benennung als solche kann allein stehen und ist dann ein Abstraktum mit der Bedeutung „Nutzung von Kohle zur Energiegewinnung“ (vgl. DEUTSCHER WORTSCHATZ: *Kohlekraft*). Gleichzeitig steht sie als Definiens in längeren Komposita und bezeichnet mit diesen spezifische Einrichtungen der Energiegewinnung aus Kohle, z. B. <Kohlekraftwerk>. Dabei kommt der vermeintliche Unterbegriff mit der Häufigkeitsklasse 14 (vgl. DEUTSCHER WORTSCHATZ: *Kohlekraftwerk*) 16-mal häufiger vor als der vermeintliche Oberbegriffe (Häufigkeitsklasse 18, vgl. DEUTSCHER WORTSCHATZ: *Kohlekraft*).

Den vorangegangenen Ausführungen über intellektuell, aber nicht maschinell extrahierte Kandidaten soll nun der umgekehrte Vergleich gegenübergestellt werden. Dies ist die Betrachtung von Termkandidaten, deren Termeigenschaft maschinell als „hoch“ eingestuft wurde, die bei der intellektuellen Extraktion jedoch nicht ausgewählt wurden. Hier zeigen sich zwei einander ergänzende Tendenzen, nämlich die Aufnahme von Komposita (in Form von Zusammenschreibung) und von Bindestrich-Wörtern, wobei auf jeder Seite des Bindestrichs wiederum ein Kompositum steht

In die erstgenannte Kategorie fallen als Beispiele mit sehr hoher Spezifität (ds-Wert) <Bildsensor> und <Biomarker>. Da die Probanden naturgemäß nicht nach den von ihnen nicht gefundenen Termkandidaten befragt werden konnten, lässt sich über die Begründung nur mutmaßen: sie liegt möglicherweise in der Kürze (gemessen an Silbenzahl und Schriftbild), möglicherweise verbunden mit der allgemeinsprachlichen Anmutung der beiden Kandidaten. Bei <Bildsensor> kann zudem noch die alltagssprachlich-deutsche Herkunft des Attribut-Teils angeführt werden.

In die zweite Kategorie, Bindestrich-Bildungen, fällt beispielsweise <Lithium-Schwefel-Batterien>. Für die intellektuelle Extraktion ist der Ausdruck möglicherweise zu lang bzw. wird in einem Umfeld unterschiedlicher Komposita mit „Lithium-Schwefel-“ und anderer chemischer Elemente nicht in seiner *unithood* wahrgenommen. Dies steht im Übrigen im Widerspruch zu der Hypothese, menschliche Extraktoren

würden Bezeichnungen aus eigenem Hintergrundwissen ergänzen. Dies mag für bekannte Wissensgebiete zutreffen, für sich neu entwickelnde, die einer Erschließung durch Termextraktion bedürfen, sind Fälle wie der soeben erläuterte Beispiele für das Gegenteil.

Somit zeigt der Vergleich der Ergebnisse von intellektueller und maschineller Extraktion Folgendes: Zahlenmäßig betrachtet, werden maschinell mehr Termkandidaten gefunden. Dabei weisen die intellektuell gefundenen Termkandidaten auch maschinell einen hohen ds -Wert und somit eine hohe Wahrscheinlichkeit auf, auch tatsächlich spezifisch für das jeweils abgedeckte Fachgebiet zu sein. Auf der anderen Seite werden jedoch von menschlichen Probanden 41 (von insgesamt 56) Termkandidaten ausgewählt, die sich nicht im maschinellen Korpus widerspiegeln. Neben den Mehrwortbenennungen, die aufgrund der Auswahl des maschinellen Verfahrens ausgeschlossen waren, verbleiben 34 Termkandidaten in der Gruppe „intellektuell-nicht-maschinell“.

Für deren Auswahl zeichnet sich aufgrund der Begründungen und des Vergleichs mit Ergebnissen der maschinellen Extraktion ein Muster ab: Die intellektuelle Extraktion zielt teilweise absichtsvoll auf Termkandidaten mit der Funktion „übergeordneter Begriff“. Dieses Charakteristikum wird für die betreffenden Termkandidaten auch durch die niedrigen ds -Werte, die auf einen höheren Allgemeinheitsgrad hindeuten, unterstrichen. Ebenfalls eine Rolle könnte die absolute Häufigkeit des Vorkommens spielen, die die Termkandidaten als thematisch einschlägig für den jeweiligen Text erscheinen lässt. Hier scheint das maschinelle Verfahren den Eindruck menschlicher Probanden sehr gut zu reflektieren: Infolge der Anwendung des POS-Taggers werden Wörter gezählt, auch wenn sie nur als Teile von Komposita auftreten. Auf dieselbe Weise dürfte die absolute Häufigkeit im Text auch von den Studienteilnehmern wahrgenommen werden. Insgesamt zeigt sich eine stärkere Berücksichtigung themen- sowie verwendungs-/übersetzungsbezogener Aspekte bei den intellektuell extrahierten Termini.

5. Schluss: Prozessaspekte bei der Eintragsauswahl in Terminologie und Meta-Lexikographie

Ziel des vorliegenden Aufsatzes war es, den Kriterien auf den Grund zu gehen, die Probanden bei der tatsächlichen Arbeit mit Termini zugrunde legen. Zu diesem Zweck wurde der Terminus-Begriff der klassischen Terminologielehre in Fortführung der Literatur um eine prozessbezogene, meta-kommunikative Dimension ergänzt. Diese umfasst laut Definition die „Nützlichkeit von Termini für Textproduktionsaufträge im Rahmen eines Arbeitsprozesses“. Fragen der Nützlichkeit und des Anwendungsbezugs sind über die Terminologiearbeit hinaus auch im weiter gefassten Bereich der Meta-Lexikographie von großem Interesse.

Dem Artikel liegt eine explorative empirische Studie zugrunde, die die Ergebnisse einer intellektuellen Termextraktion durch studentische Probanden mit dem Output eines maschinellen Extraktionsverfahrens vergleicht. Die Analyse der Daten erfolgte dabei nur im ersten Schritt quantitativ, indem sie nach zweiseitigem Vergleich Anzahl und Schnittmengen der extrahierten Termkandidaten benannte. Darüber hinaus stützte sie sich auf eine typologisierende Analyse, die sich auf die anzugebenden Auswahl-Begründungen der Probanden sowie auf linguistische (morphologische und semantische) Merkmale der Termkandidaten stützte.

Diese Analyse erwies sich als sehr aufschlussreich. Nicht nur konnte sie zeigen, dass Verständlichkeit von Begriffen und die spätere (Wieder-)Verwendung von Termini häufige Gründe für deren Auswahl sind. Es zeigte sich auch, dass diese Gründe recht häufig über das Kriterium der hohen fachlichen Spezifität obsiegen. Die Ursache hierfür kann einerseits auf der Ausdrucksseite des sprachlichen Zeichens gefunden werden. Oberbegriffe kommen in ihrer Funktion als Teilkomposita vergleichsweise häufiger in einem Text vor als die jeweils aus ihnen gebildeten Unterbegriffe. Damit erscheinen sie Probanden möglicherweise bedeutender als ihre jeweiligen Unterbegriffe. Andererseits gibt es auch bedeutungsbezogene Begründungen, die auf der (bei transparenten Komposita möglichen) Erschließung mehrerer Begriffe mit Hilfe des Oberbegriffs beruhen. Dies deutet auch in Richtung einer

meta-kommunikativen bzw. prozessbezogenen Dimension, die weniger auf das Sammeln denn auf den Gebrauch von Datenbank- bzw. Wörterbucheinträgen fokussiert ist.

Die vorliegende Studie trägt somit inhaltlich zur Terminologielehre bei, indem sie in Fortführung sozio- und kognitionslinguistischer Ansätze den Terminus-Begriff über seine Bezüge zu Referenz und Ausdruck hinaus konturieren hilft. Der Meta-Lexikographie eröffnet sie einen Einblick in die expliziten und impliziten (durch linguistische Analyse rekonstruierten) Kriterien der *user-compiler* (Personen in der Doppelrolle von Nutzern und Verfassern) von Nachschlagewerken.

Die in diesem Artikel vorgestellte Studie hat gezeigt, wie viel ein Vergleich intellektuell und maschinell gewonnener Termkandidaten leisten kann. Eine Fortführung mit größeren und thematisch stärker begrenzten Korpora würde dazu noch einmal erheblich beitragen. Dabei sollten – ewiges Desideratum der Fachkommunikationsforschung – auch fertig ausgebildete Fachleute (Terminologen und Übersetzer) als Probanden einbezogen werden. Unabhängig von Aufbau und Datenmaterial künftiger Studien eröffnet die hier eingeführte meta-kommunikative Dimension die Möglichkeit, den (späteren) Nutzen der Einträge in Nachschlagewerken systematisch zu beachten – und entsprechende Auswahlkriterien für die terminologische Ausbildung in den Terminus-Begriff zu integrieren. Ebenso sollten Ergebnisse maschineller Extraktion besonders auf Abweichungen hinsichtlich der Spezifität und der (fehlenden) Prozessbezogenheit überprüft werden, um „das Beste aus zwei Welten“ letztlich verbinden zu können.

Literatur

- AHMAD, KHURSID/DAVIES, ANDREA/FULFORD, HEATHER/ROGERS, MARGARET (1992): What is a term? The semi-automatic extraction of terms from text. In: SNELL-HORNBY, MARY/PÖCHHACKER, FRANZ/KAINDL, KLAUS (Hrsg.): Translation Studies: An Interdiscipline. Selected Papers from the Translation Studies Congress, Vienna. Amsterdam/Philadelphia: John Benjamins, 267–278.

- ARNTZ, REINER/PICHT, HERIBERT/MAYER, FELIX (2004): Einführung in die Terminologearbeit. 5., verbesserte Auflage. Hildesheim/Zürich/New York: Olms (Studien zu Sprache und Technik. 2).
- CABRÉ, MARIA TERESA/VIVALDI PALATRESI, JORGE (2013): Acquisition of terminological data from text: Approaches. In: GOUWS, RUFUS H./HEID, ULRICH/SCHWEICKARD, WOLFGANG/WIEGAND, HERBERT ERNST (Hrsg.): Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/Boston: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.4), 1486–1497.
- CALZOLARI, NICOLETTA/CHOUKRI, KHALID/DECLERCK, THIERRY/DOĞAN, MEHMET UĞUR/MAEGAARD, BENTE/MARIANI, JOSEPH/ODIJK, JAN/PIPERIDIS, STELIOS (Hrsg.): Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012). Istanbul: ELRA.
- DEUTSCHER WORTSCHATZ. Portal. Online: <http://wortschatz.uni-leipzig.de/> [Abruf: 2014-03-25].
- DIN 2342:1992-10. Begriffe der Terminologielehre.
- DIN 2342:2011-08. Begriffe der Terminologielehre.
- DREWER, PETRA/ZIEGLER, WOLFGANG (2011): Technische Dokumentation. Eine Einführung in die übersetzungsgerechte Texterstellung und in das Content-Management. Würzburg: Vogel-Buchverlag.
- FABER, PAMELA (2009): The cognitive shift in terminology and specialized translation. In: MonTI – Monografías de Traducción y Interpretación 1, 107–134.
- GOJUN, ANITA/HEID, ULRICH/WEISSBACH, BERND/LOTH, CAROLA/MINGERS, INGA (2012): Adapting and evaluating a generic term extraction tool. In: CALZOLARI u. a. (Hrsg.), 651–656. Online: <http://www.lrec-conf.org/proceedings/lrec2012/summaries/746.html> [Abruf: 2014-03-25].
- HARTMANN, R. R. K. (2001): Teaching and researching lexicography. Harlow/New York: Longman (Applied linguistics in action).
- KIT, CHUNYU (2002): Corpus Tools for Retrieving and Deriving Termhood Evidence. In: Proceedings of the 5th East Asia Forum of Terminology, 69–80.
- SCHNEIDER, WOLF (1999): Deutsch für Profis. Wege zu gutem Stil. Vollst. Taschenbuchausgabe. München: Goldmann.
- WEISSGERBER, MONIKA (2010): Schreiben in technischen Berufen. Der Ratgeber für Ingenieure und Techniker: Berichte, Dokumentationen, Präsentationen, Fachartikel, Schulungsunterlagen. Erlangen: Publicis.
- WELLER, MARION/BLANCAFORT, HELENA/GOJUN, ANITA/HEID, ULRICH (2011): Terminology extraction and term variation patterns: a study of French and German data. In: HEDELAND, HANNA/SCHMIDT, THOMAS/

- WÖRNER, KAI (Hrsg.): Multilingual Resources and Multilingual Applications. Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011. Hamburg: Hamburger Zentrum für Sprachkorpora (Arbeiten zur Mehrsprachigkeit. Folge B. 96), 117–122. Online: http://www.ims.uni-stuttgart.de/institut/mitarbeiter/gojunaa/publ/variation_gscl_2011.pdf [Abruf: 2014-03-25].
- WELLER, MARION/GOJUN, ANITA/HEID, ULRICH/DAILLE, BÉATRICE/HARAS-TANI, RIMA (2011): Simple methods for dealing with term variation and term alignment. In: KAGEURA, KYO/ZWEIGENBAUM, PIERRE (Hrsg.): Proceedings of the 9th International Conference on Terminology and Artificial Intelligence. Paris: INALCO, 87–93. Online: <http://tia2011.crim.fr/tia2011/Proceedings/pdf/TIA15.pdf> [Abruf: 2014-05-12].
- WELLER, MARION/HEID, ULRICH (2012): Analyzing and Aligning German Compound Nouns. In: CALZOLARI u. a. (Hrsg.), 2395–2400. Online: <http://www.lrec-conf.org/proceedings/lrec2012/summaries/817.html> [Abruf: 2014-03-25].

SANDRA DENZER / FRANZISKA HORN

Die Arbeitsumgebung des Digitalen Familiennamenwörterbuch Deutschlands

Ein XML-basiertes Redaktionssystem

1. Einleitung

Familiennamen als Untergruppe der Eigennamen weisen besondere Eigenschaften auf, die sie von anderen Substantiven stark unterscheiden. Aus diesem Grund werden sie meist in speziellen Nachschlagewerken verzeichnet, aber auch in allgemeinsprachlichen Wörterbüchern sind sie oft erfasst. Jedoch weisen beide Typen von Nachschlagewerken eine ungenügende Beschreibungspraxis auf, die eine lexikographische Erfassung und Bearbeitung des Familiennamenbestandes in Deutschland nötig macht. Eine Aufarbeitung der Etymologie und wissenschaftliche Erklärung von Familiennamen ist eine Forschungsaufgabe, von der viele Disziplinen profitieren, da sie eine hervorragende Quelle für linguistische, aber auch kulturanthropologische, soziologische, kirchen- und siedlungsgeschichtliche Fragestellungen darstellen. Aus diesen Gründen wurde das von der Akademie der Wissenschaften und Literatur Mainz betreute „Digitale Familiennamenwörterbuch Deutschlands (DFD)“ ins Leben gerufen.

In diesem Beitrag werden Einblicke in die Wörterbuchwerkstatt eines Namenwörterbuchs gegeben, indem das Redaktionssystem des DFD mit seinen einzelnen Komponenten und Funktionalitäten näher vorgestellt wird. Nach einer kurzen Projektvorstellung wird im ersten Teil erläutert, welche spezifischen Anforderungen sich aufgrund der Besonderheiten und der bisherigen lexikographischen Erschließung von Familiennamen für das DFD stellen. Im zweiten Teil des Beitrags wird die Arbeitsumgebung in Hinblick auf vorhandene Kriterienkataloge zu

Redaktionssystemen (z. B. ATKINS/RUNDELL 2008; ABEL/KLOSA 2012) und die projektinternen Anforderungen, die sich aus der Konzeption des Wörterbuchs ergeben, vorgestellt.

2. Projektvorstellung

Ziel des DFD ist es, in einer Projektzeit von 24 Jahren (2012–2036) den rezenten Familiennamenbestand Deutschlands – auch unter Einbezug der fremdsprachigen Namen – lexikographisch zu erfassen und zu etymologisieren. Das Wörterbuch wird in enger Kooperation zwischen der Akademie der Wissenschaften und der Literatur Mainz, der Johannes Gutenberg-Universität Mainz und der Technischen Universität Darmstadt erstellt. Hierbei erfolgt die onomastische Erarbeitung der Wörterbuchartikel bei der Arbeitsstelle in der Mainzer Akademie, während die EDV-philologischen und informatischen Komponenten an der Technischen Universität Darmstadt in Zusammenarbeit mit der Digitalen Akademie Mainz entwickelt werden.

Dem DFD-Projekt liegen Daten zu Familiennamen zugrunde, die im Zuge des DFG-Projekts „Deutscher Familiennamenatlas (DFA)“ zusammengestellt wurden. Die für das DFA-Projekt entwickelte Software zur Kartierung¹ der Familiennamen enthält als Korpus die bei der Telekom registrierten Telefonanschlüsse (Stand 2005) mit den Daten: Familienname, Postleitzahl und Häufigkeit. Die Daten der Telekom wurden um Vornamen und Adressen, Verschreibungen, Firmen-, Vereinsanschlüsse u. Ä. und heiratsbedingte Doppelnamen bereinigt.² Durch die neue Möglichkeit, Familiennamen mit ihrer Häufigkeit systematisch geographisch zu verorten, bietet sich eine wichtige Methode, die Bedeutung des Namens zu erarbeiten und wissenschaftlich abzusichern. Durch diesen für die Namenforschung relativ neuen methodischen Zugang lassen sich beispielsweise Fehldeutungen älterer Forschung korrigieren, ergänzen und revidieren. Aber vor allem ist heraus-

¹ Entwickelt von Richard Kunze (tivano Software GmbH).

² Eine detaillierte Projektvorstellung findet sich bei FAHLBUSCH/HEUSER (2014).

zustellen, dass durch diese Methode viele Familiennamen erstmals ge-
deutet werden können. So ist beispielsweise der Familienname *Fixemer*
in keinem deutschen oder einschlägigen fremdsprachigen Nachschla-
gewerk aufgeführt. Durch seine geographische Verbreitung in der
Südpfalz und im Saarland lässt sich nun jedoch feststellen, dass es sich
um einen Herkunftsnamen handelt, der vom Namen der Siedlung Fi-
xem (Departement Moselle, nordöstlich von Thionville) abgeleitet ist.

3. Relevanz des Projekts

Eigennamen nehmen im Sprachsystem eine Sonderstellung ein, was
eine besondere lexikographische Behandlung erfordert (STRAUB 1989,
788). Deshalb sind sie meist in eigenen Nachschlagewerken verzeich-
net, aber auch in allgemeinsprachlichen Wörterbüchern sind Eigenna-
men als Einträge zu finden. Die lexikographische Umsetzung des Ge-
genstands ‚Eigenname‘ – und im Speziellen ‚Familienname‘ – zeigt in
beiden Arten von Wörterbuchtypen Defizite auf, die im Folgenden
weiter dargestellt und erläutert werden.

3.1 Sonderstellung der Eigennamen

Im Sprachsystem werden Eigennamen mit Appellativen unter die Klas-
se der Substantive subsumiert. Vor allem durch ihre spezielle Referenz-
funktion, aber auch durch ihr grammatisches Verhalten, nehmen Eigen-
namen im Sprachsystem jedoch eine Sonderstellung ein, die sie von den
anderen Substantiven abgrenzt (vgl. NÜBLING u. a. 2012, 64). So sind
Eigennamen monoreferent, d. h. sie referieren idealerweise auf nur ein
Denotat (vgl. NÜBLING u. a. 2012, 32) und haben somit die Funktion,
ein Individuum zu benennen bzw. zu identifizieren, während Appellati-
ve über ihre Semantik Klassen von Objekten charakterisieren (vgl.
FLEISCHER 1964, 377). Diese spezielle Referenzfunktion hat auch
Auswirkungen auf die grammatische Funktion. So erfasst der Plural
von Eigennamen nicht eine Gruppe ähnlicher Objekte – wie dies für
pluralische Appellative gilt –, vielmehr zeigt der Plural nur namentliche

Homonymie an (vgl. NÜBLING u. a. 2012, 73). Der onymische Genitiv besitzt fast ausschließlich die reine Possessivfunktion (vgl. NÜBLING u. a. 2012, 73). Zudem ergeben sich formale Unterschiede zwischen Eigennamen und Appellativen, die alle Ebenen der Sprache betreffen (NÜBLING u. a. 2012). So unterscheiden sich beispielsweise die onymischen Kasus- und Pluralsuffixe von den appellativischen, indem Eigennamen fast keine Allomorphie aufweisen: Zur Pluralbildung bei Namen kann nur das Flexiv *-s* verwendet werden, die anderen Allomorphe entfallen (vgl. NÜBLING u. a. 2012, 71–72). Die Kasusflexion ist fast vollständig abgebaut und beschränkt sich nur noch auf den Genitiv, wenn dieser nicht schon an einem vorangehenden Determinierer markiert wird (vgl. NÜBLING/SCHMUCK 2010, 154).

Ein weiteres wesentliches Alleinstellungsmerkmal von Eigennamen besteht darin, dass ihnen in der Regel kein lexikalisch-semantischer Inhalt zukommt (vgl. KUNZE 1998, 11), sondern nur eine etymologische Bedeutung. Eigennamen können noch lexikalische Strukturen enthalten, da sie sich oft aus Appellativen entwickelt haben, doch ist deren Semantik erloschen (vgl. HARNISCH/NÜBLING 2004, 1901). Wie MARYNISSEN/NÜBLING (2010, 311) herausstellen, konservieren vor allem Familiennamen, da sie relativ spät entstanden sind, morphologische und appellativische Reststrukturen. Die etymologische Deutung von Familiennamen rekonstruiert diejenige Bedeutung, die das einstige Appellativ (oder Adjektiv) zum Zeitpunkt der Festwerdung, d. h. der Erstarrung zum Namen (ca. 1200–1500) hatte. Hierdurch wird deutlich, dass Eigennamen andere Anforderungen an die lexikographische Beschreibung der Etymologie, aber eben auch der Grammatik, stellen als Appellative.³

³ Für eine detailliertere Darstellung der Unterschiede zwischen *Nomina propria* und *Nomina appellativa* ist auf die einschlägige Forschungsliteratur zu verweisen (hier in Auswahl): BLANÁR (2001), DEBUS (1977; 1980; 2005), FLEISCHER (1964), GERHARDT (1949/50), HARWEG (1997, 89–121), KALVERKÄMPER (1978; 1994), KOLDE (1995), LÖTSCHER (1995), NÜBLING (2005), VATER (1965).

3.2 Eigennamen in allgemeinsprachlichen Wörterbüchern

Aus dieser im vorigen Abschnitt beschriebenen Sonderstellung der Eigennamen heraus erscheint es angemessen, dass Eigennamen in eigenen speziellen Nachschlagewerken verzeichnet werden (vgl. MÖLLER 1995, 324). Jedoch finden sich in zahlreichen allgemeinsprachlichen Wörterbüchern Angaben zu Eigennamen, da ein Nachschlagebedürfnis beispielsweise zu deren Schreibung und Silbentrennung besteht (vgl. KLOSA/SCHOOAERT 2011, 193). Meist sind es geographische Namen – wie Länder- und Städtenamen – und Personennamen, die in allgemeinsprachlichen Wörterbüchern erfasst sind (vgl. KLOSA u. a. 2010, 656). An den Einträgen zu Eigennamen fällt zumeist die Inkonsistenz der metasprachlichen Beschreibung auf, da manche Einträge eine nicht-metasprachliche Form und andere eine metasprachliche Form aufweisen (vgl. KLOSA/SCHOOAERT 2001, 198). Anhand von eigenen Stichproben zu Personennamen im Rechtschreibduden (2009) zeigt sich eine unbeständige Handhabung, ob ein Eintrag sprachliches Wissen – „Edelbert (m. Vorname)“ – oder objektbezogenes Wissen – „Mozart (österreichischer Komponist)“ – beschreibt.⁴

3.3 Nachschlagewerke zu Familiennamen

Ein Blick auf die Namenbuchlandschaft in Bezug auf Familiennamen macht deutlich, dass bereits einige entsprechende Nachschlagewerke existieren. Bei genauerer Betrachtung zeigt sich jedoch, dass diese existierenden Nachschlagewerke erhebliche Defizite aufweisen. Ein Gesamtbild des Nameninventars in Deutschland ist nur eingeschränkt möglich, da nur ein Bruchteil des Namenbestands lexikographisch erfasst ist, wobei vor allem seltene und fremdsprachige Namen kaum in den Nachschlagewerken verzeichnet sind. Eine Stichprobe zeigt, dass mindestens jeder fünfte der heute vorkommenden Familiennamen in keinem der einschlägigen Nachschlagewerke erfasst ist. In Anbetracht

⁴ Eine detaillierte Aufarbeitung und Beschreibung dieses Aspekts bietet HARWEG (1999).

des Gesamtbestands von etwa 850.000 Familiennamen, die sich durch die Telekom-Daten ermitteln ließen, erfassen der Duden (2005; Datenbasis von 1997) mit etwa 20.000 Einträgen ca. 2,4 % und GOTTSCHALD (2006; letzte Neubearbeitung von 1932), als bisher umfangreichstes Nachschlagewerk zu Familiennamen, mit etwa 70.000 Einträgen ca. 8,2 % des Familiennameninventars. Abgesehen davon beschränken sich Namenbücher zum Teil auf bestimmte Regionen, wodurch Zusammenhänge der Bedeutung im deutschen Raum nicht zugänglich oder nur durch Hinzuziehen mehrerer Namenbücher offensichtlich werden. Durch Untersuchungen im Vorfeld des DFD-Projekts konnte zudem festgestellt werden, dass die vorhandenen Nachschlagewerke, zum Teil auch durch ihr Alter bedingt, eine Vielzahl an Familiennamen behandeln, die heute nicht mehr existieren. Neben diesen problematischen Aspekten des behandelten Nameninventars zeigen sich auch Unzulänglichkeiten in der Nutzbarkeit der existierenden Nachschlagewerke. Die Zugriffsstrukturen sind häufig schwierig zu handhaben, da Namen oft in Sammelartikeln vereint sind und Sublemmata nicht leicht ersichtlich dargestellt werden. Ein weiteres Defizit in der Zugriffsstruktur ergibt sich dadurch, dass beispielsweise bei GOTTSCHALD (2006) der alphabetischen Ordnung das Etymon zugrunde liegt. So ist zum Beispiel der Familienname *Robert* unter dem Eintrag *HROD* verzeichnet. Unter diesem althochdeutschen Namenglied finden sich 758 einzelne Namen, darunter so verschiedene wie *Gruhl*, *Krupp*, *Rhode* oder *Riedel*. Um eine vollständige Deutung zu *Robert* zu erlangen, muss zudem unter *BRECHT* nachgeschlagen werden, worauf beim Nestlemma *Robert* nicht explizit hingewiesen wird. Durch dieses Beispiel tritt deutlich hervor, dass linguistisches – vor allem sprachhistorisches – Vorwissen von Nöten ist, um einen Eintrag zu finden. Aber auch um dessen Inhalt vollständig zu verstehen, ist Vorwissen meist unabdingbar. Vorkenntnisse sind auch in Bezug auf die lexikographische Praxis der Textverdichtung durch Abkürzungen und der typographischen Strukturmarker nötig. Inhaltlich betrachtet lassen sich in den Namenbüchern zudem Lücken in der Namendeutung nachweisen. So sind meist nicht alle Bedeutungskonkurrenzen umfänglich beschrieben, die bei einer Vielzahl

von Familiennamen festzustellen sind. Ein Beispiel ist der Familienname *Wiedemann*, der aus *wīde* ‘Weide’ oder *widem* ‘kirchliches Grundstück’ (vgl. KUNZE 1998, 155) abgeleitet werden kann. Darüber hinaus sind auch fehlerhafte Deutungen in den Namenbüchern enthalten, die dem veralteten Wissensstand und der veralteten Methodik geschuldet sind. Durch die nun anwendbare moderne Methode der Familiennamengeographie lassen sich solche Schwächen der Namendeutung beheben. GOTTSCHALD (2006, 433) verzeichnet beispielsweise *Schillo* zu sorbisch *šylo* ‘Pfriem’, aber das Vorkommen im Saarland und in der Südpfalz legt einen Zusammenhang mit dem französischen Patronym *Gillot* (zum Rufnamen *Āgidius* mit Diminutivendung *-ot*) nahe.

Aus diesen benannten Defiziten in Bezug auf das Nameninventar und der Nutzbarkeit ergibt sich, dass für eine adäquate und umfassende Namendeutung immer mehrere Nachschlagewerke herangezogen werden müssen.⁵ Dies ist insbesondere für den interessierten Laien kaum zu bewerkstelligen und nicht hinnehmbar, da auch die Zugänglichkeit etwa durch Anschaffungskosten oder der Zugang zu Fachbibliotheken nicht immer gewährleistet ist.

3.4 Konzeptionelle Anforderungen an das DFD

Aus den Erläuterungen der vorangegangenen Kapitel wird deutlich, dass ein neues Nachschlagewerk für Familiennamen notwendig ist. Dies sollte vor allem in Bezug auf das Nameninventar umfassend und umfangreich, aber auch aktuell sein. Wissenschaftlich-methodisch und technologisch soll es auf dem neuesten Stand erarbeitet und entwickelt werden. Aus Sicht der NutzerInnen sollte es für jeden leicht verfügbar sein. Auch auf der strukturellen Ebene des Wörterbuchs muss ein einfacher, verständlicher Zugang gewährt sein. Die metasprachliche Be-

⁵ Für eine umfassende Namendeutung müssen aktuell mindestens sechs Nachschlagewerke konsultiert werden: BAHLOW (2005), BRECHENMACHER (1960–63), GOTTSCHALD (2006), KOHLHEIM/KOHLHEIM (2005), NAUMANN (2007), ZODER (1968).

schreibung muss dem Beschreibungsgegenstand ‚Familiennamen‘ angemessen sein und dabei für Laien verständlich bleiben.

Aus diesen Anforderungen haben sich die MitarbeiterInnen des DFD einen umfangreichen Arbeitsauftrag gesetzt. Zum einen ist geplant, dass ca. 200.000 Familiennamen⁶ lexikographisch erfasst und etymologisiert werden. Damit werden, selbst nach Abzug der niedrigfrequenten Namen, ca. 95 % aller Telefonanschlussbesitzer ihren Namen im DFD finden. Darüber hinaus ist es vorstellbar, dass alle ca. 850.000 Familiennamen des Datenbestands online mit den jeweiligen Angaben zur Häufigkeit und zum Rang, den der Name im gesamten Familiennamenbestand des Korpus einnimmt, zur Verfügung gestellt werden.

Jeder Name wird einen eigenen Namenartikel erhalten, d. h. jede Namensvariante ist ein Eintrag im Wörterbuch. Die Informationspositionen zur Häufigkeit und zum Rang des jeweiligen Namens sind durch die Daten in der Kartierungssoftware bereits vorhanden und lassen sich in die Datenbank des Wörterbuchs computergestützt übertragen. Die Häufigkeit des Familiennamens stellt ein Kriterium dar, um den Namenbestand umfangreich und systematisch für das Wörterbuch zu erschließen. Das Vorgehen ist, bei 10 Tokens⁷ eine pragmatische, arbeitsökonomische Grenze zu ziehen, wann ein Familienname etymologisiert wird. Die entsprechenden Types mit 10 Tokens lassen sich durch die Kartierungssoftware identifizieren. Die Herangehensweise ist jedoch nicht rein frequenzabhängig, damit ein breites Spektrum der Familiennamen zugänglich wird. Es werden auch Namen unter diesem Schwellenwert bearbeitet, wenn sie eine Variante eines Namens darstellen (zum Beispiel Komposita zu *Müller*, *Bauer* oder *Meyer*). Neben diesem Kriterium, das hauptsächlich durch die äußere Form des Namens bestimmt ist, wird der Namenbestand zudem thematisch aufgear-

⁶ Diese Zahl ergibt sich aus dem unten beschriebenen Kriterienkatalog aus Frequenzschwelle, Varianz und thematischer Zusammengehörigkeit.

⁷ In diesem Zusammenhang meint Tokens die Anzahl der Telefonanschlüsse eines Familiennamens in der vorliegenden Datenbasis. Der Familienname entspricht einem Type.

beitet. Es können beispielsweise solche Gruppen gebildet werden wie „alle Handwerksberufe, die sich mit der Brotherstellung beschäftigen“. Dieser modulare Ansatz aus Frequenz, Variante und Thema sorgt für ein vielfältiges Spektrum aufgenommener Namen. Darüber hinaus ist dadurch auch für die AutorInnen ein vereinfachtes systematisches und kontinuierliches Bearbeiten des Lemmabestandes gewährleistet.

Der vollständige Namenartikel umfasst als inhaltliche Positionen neben dem Lemma, der Token- und Rangangabe und der detaillierten etymologischen und linguistischen Beschreibung zudem Informationen zu Herkunftssprache, -land und -region. Häufige und besondere Schreibvarianten des Lemmas werden auch als Artikelposition aufgenommen. Neben der Darstellung der geographischen Verbreitung in Deutschland anhand einer dynamischen Karte⁸ werden Hinweise zur historischen Verbreitung und zum Vorkommen des Namens außerhalb der Bundesrepublik im Artikel erscheinen. Der Familienname wird, wenn möglich, auch anhand von historischen Belegen nachgewiesen. Um Informationen darüber hinaus direkt erreichbar zu machen, wird jedem Artikel weiterführende Literatur mitgegeben.

Neben dem eigentlichen Namenartikel werden im DFD zudem ergänzende Informationen erarbeitet, um dem/der NutzerIn einen breiten Einblick in das Themenfeld der Onomastik zu ermöglichen. Insbesondere inner- und außersprachliche Faktoren zur Entstehung des deutschen Namenssystems, kulturhistorische Hintergründe, Namenssysteme aus anderen Kulturen und bestimmte Namentypen (z. B. Lallnamen wie etwa „*Nonno*“ oder „*Poppo*“, die SEIBICKE (2004, 3540) als Beispiele nennt) werden in sogenannten Informationskästen bereitgestellt. Während der Projektlaufzeit entsteht weiterhin ein Glossar für Fachtermini als weitere Informationsquelle. Darüber hinaus werden den Familiennamen Kategorien (beispielsweise aus Rufnamen entstandene Familiennamen in Verbindung mit der Herkunftssprache) zugeordnet, wodurch ein weiterer direkter Zugriff zum Nameninventar über diese Großgruppen möglich ist.

⁸ Aus datenschutzrechtlichen Gründen wird eine Karte erst ab einem Vorkommen von über 5 Tokens zur Verfügung gestellt.

Auf das Zusammenwirken und die Rolle dieser vielschichtigen Informationspositionen des Artikels und der sekundären Informationen wird im nachfolgenden Kapitel eingegangen, da sie Kernaspekte der internen Anforderungen an die entwickelte Arbeitsumgebung darstellen.

4. Vorstellung der Arbeitsumgebung

Von ersten Überlegungen zur Konzeption einer Arbeitsumgebung in der Vorbereitungsphase eines Wörterbuchs über die Realisierung bis hin zum konkreten Einsatz zur Artikelerstellung und Publikation in den Phasen der Datenanalyse und Vorbereitung des Online-Releases eines Wörterbuchs – spielt das Redaktionssystem eine zentrale Rolle im gesamten lexikographischen Prozess (vgl. ABEL 2012, 104). Diesem relevanten Aspekt der Wörterbuchproduktion widmen sich einzelne Veröffentlichungen, die grundlegende Merkmale und Funktionalitäten eines Redaktionssystems aufzeigen und diese an konkreten Softwarelösungen verdeutlichen. Beispiele sind die Ausführungen von ATKINS/RUNDELL (2008) sowie RIDINGS (2003) unter Rückgriff auf die Software Onoma und ABEL/KLOSA (2012) für eine Arbeitsumgebung für *elexiko*.⁹

Im folgenden Abschnitt soll die Arbeitsumgebung für das DFD mit ihren Komponenten und Merkmalen näher vorgestellt werden. Dazu werden zunächst die projektinternen Anforderungen zusammengefasst, um anschließend die technische Realisierung darzulegen. Dabei werden die Anforderungen und Kriterien an Arbeitsumgebungen, die in der Forschungsliteratur aufkommen, herangezogen und es wird versucht, Parallelen zu ziehen, wenn Anknüpfungspunkte zu Funktionalitäten bei anderen Softwarelösungen ersichtlich werden.

⁹ Komponenten und Merkmale von Redaktionssystemen werden z. B. auch bei SVENSÉN (2009), DE SCHRYVER (2011) und ABEL (2012) benannt, wobei ABEL auch einen detaillierteren Überblick zu bestehenden, unterschiedlich ausgerichteten Softwarelösungen gibt.

4.1 Projektinterne Anforderungen an die Arbeitsumgebung

Innerhalb der Arbeitsumgebung muss es möglich sein, verschiedene Arten von Einträgen zu erarbeiten, da neben Wörterbuchartikeln auch sekundäre Informationen wie Glossareinträge und Informationskästen geplant sind. Das Editor-Interface hat einen Artikelaufbau mit vielfältigen Informationspositionen abzubilden. Zu berücksichtigen sind Angaben, die nicht nur sprachliches, sondern auch außersprachliches Wissen berühren. Dabei ist es insbesondere in Anbetracht der langen Projektlaufzeit erforderlich, dass Ergebnisse von Recherchen und Arbeitsschritte zur Erstellung der lexikographischen Inhalte direkt mit in den Einträgen verzeichnet werden. Für die Erarbeitung des DFD ist weiterhin zum einen ein Kartierungstool zu integrieren, um Aussagen über die Verbreitung der Namen treffen zu können. Zum anderen stellt sich die Integration einer Literaturdatenbank als Anforderung, die die Sichtung und Bewertung passender Veröffentlichungen unterstützt und organisiert. In Anbetracht der enormen Anzahl an Namenartikeln muss die Handhabbarkeit von großen Datenmengen innerhalb des Redaktionssystems gewährleistet sein. Da es im Kooperationsprojekt verschiedene Arbeitsstellen und extern beschäftigte MitarbeiterInnen gibt, ist es notwendig, dezentrales Arbeiten auf einem gemeinsamen Datenbestand zu organisieren. Dabei ist eine Datenmodellierung wichtig, bei der die lexikographischen Inhalte modernen Standards entsprechend medienneutral verwaltet werden. Dies ermöglicht es, verschiedene Ausgabeformate zu generieren. Das DFD ist primär auf eine Online-Publikation ausgerichtet, zu einem späteren Zeitpunkt kann somit jedoch u. a. auch eine Teilpublikation im Printmedium erfolgen. Eine medienneutrale Datenhaltung bildet weiterhin die Grundlage für Langzeitarchivierung und -sicherung sowie für Möglichkeiten der Auswertung und des Austausches hinsichtlich des Datenbestands. Die Anforderungen spiegeln die Konzeption des DFD als Online-Wörterbuch wider, das sich als Langzeitprojekt dem Beschreibungsgegenstand ‚Familiennamen‘ zuwendet.

4.2 Die technische Umsetzung der Arbeitsumgebung

Für das DFD wird ein hauseigenes lexikographisches Redaktionssystem verwendet. Zu den Kernbestandteilen gehört der XML-Editor oXygen, der für die Erstellung und Überarbeitung der lexikographischen Inhalte eingesetzt wird. Eine eXist-Datenbank wird zur Bereitstellung und Verwaltung des Namenbestandes genutzt und das Content-Management-System TYPO3 wird zur Präsentation des Wörterbuchs sowie zur Recherche der Daten verwendet. Somit wird den drei bei ABEL/KLOSA (2012, 414–415) identifizierten Komponenten eines lexikographischen Redaktionssystems Rechnung getragen – dem Editor zur Erarbeitung der Wörterbuchartikel, einer Datenbank für die Datenspeicherung und einer Softwarelösung für das Projektmanagement und die Veröffentlichung des Produkts.

Von Vorteil sind die Kenntnisse der EDV-MitarbeiterInnen in den Komponenten XML-Editor oXygen und dem CMS TYPO3, die beide weit verbreitete Standardlösungen sind. Somit verkürzt sich die Einarbeitungszeit der EntwicklerInnen entscheidend und auf Schwierigkeiten und Probleme kann effizient reagiert werden. Zudem lassen sich die beiden Produkte über eine Schnittstelle (API) verbinden. Da TYPO3 im Back-End für die Online-RedakteurInnen eine Eingabemaske zur leicht handhabbaren Erarbeitung des Website-Contents anbietet, erleichtert dies auch die Betreuung und Pflege der weiteren Inhalte der DFD-Website und des Portals „namenforschung.net“. So kann die Publikation des Wörterbuchs und des gesamten weiteren Inhalts, der über die Projektseite bereitgestellt wird, rein über TYPO3 vorgenommen werden. Bestätigt wurde das Vorgehen durch die positiven Erfahrungen der TELOTA-Initiative der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW), deren MitarbeiterInnen eine ähnliche Arbeitsumgebung für einzelne Projekte geschaffen haben. Ein weiterer Grund für das hauseigene Redaktionssystem ergibt sich aus den Integrationsmöglichkeiten der bereits im Rahmen des DFA-Projekts erstellten Komponenten wie dem Kartierungstool oder der Literaturdatenbank, die nicht im Hinblick auf ein lexikographisches Redaktionssystem erarbeitet wurden. Da das Projektmanagement per Microsoft SharePoint erfolgt,

brauchen bestimmte Funktionalitäten nicht im Redaktionssystem des DFD berücksichtigt sein. Diese Kernkomponente von vorgefertigten Produkten (sogenannte „Off-the-Shelf-Produkte“ wie beispielsweise TschwaneLex [TLex]¹⁰) wäre für das DFD nicht in dem Maße von Nöten.

Weiterhin ist für das lexikographische Redaktionssystem zu bedenken, dass im DFD sekundäre Informationen erarbeitet werden, die nicht wie ein Wörterbuchartikel aufgebaut sind und deshalb im Redaktionssystem eine angepasste Vorlage zur Erfassung (s. Abschnitt 4.2.1) erfordern.

Die Erarbeitung der Arbeitsumgebung erfolgt in enger Zusammenarbeit mit den BearbeiterInnen. Dies betrifft alle Aspekte des lexikographischen Redaktionssystems von der Eingabemaske, den Möglichkeiten der XML-Kodierung¹¹ bis hin zur Modellierung der Datenstruktur und Datenbankabfrage. Für die EntwicklerInnen ergibt sich daraus der Vorteil, dass vorab auf Merkmale des Beschreibungsgegenstands ‚Familiennamen‘ eingegangen werden konnte, und für die BearbeiterInnen steht ein System bereit, an dem sie selbst mitgewirkt haben. Die Arbeitsumgebung wurde den BearbeiterInnen während einer Schulung nähergebracht. Für Nachfragen zu einzelnen Funktionen oder für Anregungen zu Erweiterungen ist ein Forum zum gegenseitigen Austausch eingerichtet.

4.2.1 Editor

Der kostenpflichtige XML-Editor oXygen¹² wird wie bei elexiko (vgl. ABEL/KLOSA 2012, 416) und dem „Digitalen Wörterbuch der deutschen Sprache (DWDS)“ (vgl. GEYKEN 2011, 11) zur Erstellung und Überarbeitung der lexikographischen Inhalte eingesetzt. Die Datenmodellierung im Format XML entsprechend den Kodierungsrichtlinien der

¹⁰ Eine Beschreibung der Software findet sich bei JOFFE u. a. (2003).

¹¹ Näheres zu XML (Extensible Markup Language) über das W3C (World Wide Web Consortium) unter: <http://www.w3.org/TR/xml/>.

¹² Es wird derzeit die oXygen-Version 15.1. verwendet.

Text Encoding Initiative¹³, die auch bei Projekten wie dem DWDS (vgl. GEYKEN 2011, 12), dem Trierer „Wörterbuchnetz“ (vgl. HILDENBRANDT 2011) und dem „Digitalen Ortsnamenbuch Online (DONBO)“ (vgl. WINNER/BUCHNER 2011, 186–187) verfolgt wird, ist als Orientierung an De-facto-Standards (vgl. JANNIDIS 2009, 258) in den digitalen Geisteswissenschaften zu sehen. Eine tiefe Annotation relevanter Informationen¹⁴ bildet die Grundlage für allgemeine und erweiterte Suchfunktionen bei einer Online-Präsentation des Wörterbuchs, was auch MÜLLER-SPITZER (2011) bei der Beschreibung der Datenmodellierung für *lexiko* hervorhebt.

Bei der TEI-konformen Datenmodellierung für das DFD werden vorwiegend die Vorgaben des TEI-Moduls für Wörterbücher¹⁵ verwendet. Bei der Nutzung des Tagsets für die Kodierung eines Namenwörterbuches kommen viele Attribute zur Spezifikation von Informationen zum Einsatz. Wird zudem der Fokus von XML-Kodierung, Bedeutung zu explizieren, in den Blick genommen, sind einige Elemente für bestimmte Artikelangaben im Vergleich teilweise indirekt benannt. Dies kann die menschliche Lesbarkeit der XML-Dateien erschweren. Hier ist es von Vorteil, dass die Erstellung der Artikel nicht im XML-Code erfolgt. Das Editor-Interface ist entsprechend den Bedürfnissen und Kenntnissen der BearbeiterInnen über eine benutzerfreundliche Oberfläche mit einer WYSIWYG-Ansicht¹⁶ zur Visualisierung der Annotationen realisiert. Genutzt werden hierbei Möglichkeiten des Editors oXygen, über eine sogenannte Dokumententypen-Zuordnung ein eigenes Framework für das DFD-Projekt zu erstellen. Dieses beinhaltet

¹³ Nähere Informationen zur Text Encoding Initiative (TEI) und deren Empfehlungen zur XML-basierten Auszeichnung von geisteswissenschaftlichen Texten sowie deren Austausch und Langzeitarchivierung lassen sich über die Internetseite der TEI einsehen, unter: <http://www.tei-c.org/index.xml>.

¹⁴ Näher beschrieben wird die Datenmodellierung für das DFD als Aspekt des lexikographischen Prozesses bei DENZER/HORN (erscheint).

¹⁵ Eine kurze Einführung zum TEI-Modul für die Kodierung von Wörterbüchern gibt HILDENBRANDT (2011, 22–24).

¹⁶ *WYSIWYG* steht für *What You See Is What You Get*.

Vorgaben zur Darstellung der lexikographischen Inhalte mittels CSS¹⁷, XML-Schemata zur Validierung, Vorlagen für Einträge sowie projektspezifische Einstellungen wie angepasste Menüleisten und Buttons. Ein solch konfiguriertes Editor-Interface findet auch beim DWDS und dem Editionsvorhaben „Schleiermacher in Berlin 1808–1834“¹⁸ Verwendung. Zu den verschiedenen Visualisierungsmöglichkeiten, die ein Editor bieten sollte (vgl. ABEL/KLOSA 2012, 414; auch ATKINS/RUNDELL, 114–115), gehört bei oXygen neben der WYSIWYG-Ansicht u. a. auch eine Baumstrukturansicht. Diese verdeutlicht übersichtlich die Gliederung der Artikelinhalte.

Weiterhin bietet der Editor wie bei ABEL und KLOSA beschrieben (2012, 414) die Möglichkeit, Teile der lexikographischen Inhalte zu kopieren, einzufügen und zu verschieben. Die BearbeiterInnen dokumentieren ihre einzelnen Arbeitsschritte innerhalb eines integrierten Redaktionsformulars. Das schließt ein, dem Dokument einen bestimmten Status zuzuweisen, um kenntlich zu machen, ob der Artikel zur Überprüfung durch eine/n weitere/n RedakteurIn und nachfolgend zur Publikation bereitsteht. So wird Transparenz bei der Erstellung des Wörterbuchs gewahrt und lexikographische Entscheidungen können auch im späteren Projektverlauf, z. B. bei der Überarbeitung oder Aktualisierung eines Artikels nachvollzogen und berücksichtigt werden.

¹⁷ Näheres zu CSS (Cascading Style Sheets) über das W3C unter: <http://www.w3.org/TR/CSS2/>.

¹⁸ Einblicke in die Arbeitsumgebung bieten DUMONT/FECHNER (2012), Hinweise zur Einrichtung und Nutzung eines oXygen-Frameworks finden sich bei DUMONT (2013).

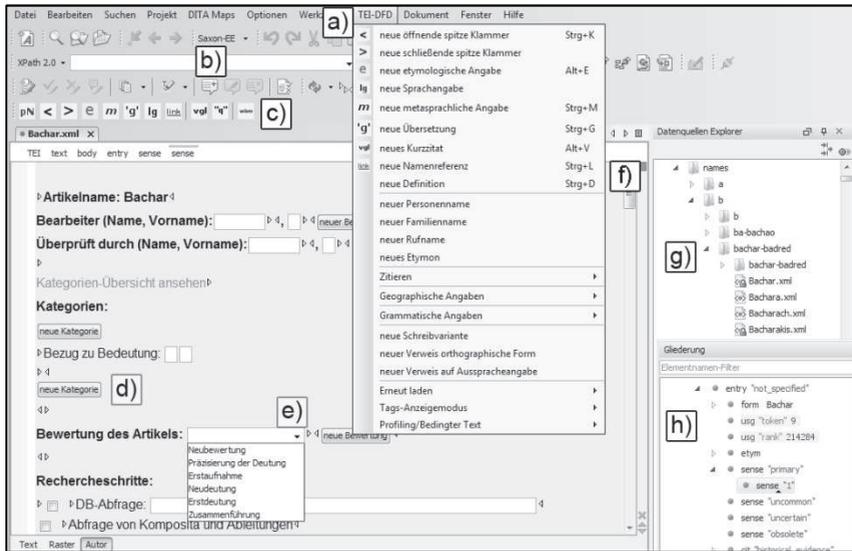


Abb. 1: XML-Editor oXygen mit angepasster *Author*-Ansicht. Die Beschriftungen markieren a) Menü-Reiter für DFD-Auszeichnungen, b) Kommentarfunktion, c) Button-Leiste für die DFD-Annotation, d) Buttons (hier Kategorie) zum Hinzufügen eines Angabebereichs, e) Dropdown-Menü, f) Validierungsanzeige, g) Ordner-Ansicht der Datenbank, h) Baumstruktur-Ansicht der XML-Kodierung

Die Struktur der Einträge wird entsprechend den Anforderungen an einen Editor bei ABEL/KLOSA (2012, 414) validiert, indem während der Dateneingabe ein Abgleich mit dem XML-Schema, der zugrunde liegenden Dokumentengrammatik, erfolgt. Der Vorteil einer permanenten Überprüfung der Dateien im Gegensatz zu einer Validierung, die erst vor dem Speichern aktiv wird, ist darin zu sehen, dass inkonsistente Inhalte schneller auffallen. Bei festgestellten Verstößen wird eine entsprechende Fehlermeldung ausgegeben. Um zu gewährleisten, dass die lexikographischen Inhalte konsistent und somit auch Schema-bewusst erarbeitet werden (vgl. ABEL/KLOSA 2012, 414), sind verschiedene Funktionen mithilfe von oXygen realisierbar. Beispiele hierfür sind:

- Drop-down-Menüs, z. B. zur Auswahl von Sprachen oder Ländern,
- Datumswähler, z. B. zur Dokumentation des letzten Zugriffs auf URLs,
- Checkboxes, z. B. zur Bestätigung von Arbeitsschritten,
- Texteingabefelder, z. B. zur Eingabe des/r zuständigen Redakteurs/in.

Für weitere Konsistenz sorgt zudem das Glossar der Fachbegriffe, welches nicht nur den NutzerInnen als Hilfe dient, sondern auch den BearbeiterInnen im Schreibprozess als Referenz zur Verfügung steht, welche Termini innerhalb des Wörterbuchs verwendet werden.

Eine automatische Rechtschreibkontrolle (vgl. ABEL/KLOSA 2012, 414; auch ATKINS/RUNDELL 2008, 116) kann bei oXygen im *Author*-Modus durch den Artikelschreibenden aktiviert werden, um Tippfehler zu vermeiden. Sonderzeichen im Unicode-Standard können über eine Zeichentabelle benutzerfreundlich ausgewählt und eingefügt werden. ABEL/KLOSA (2012, 414) erachten weiterhin ein im Editor integriertes Redaktionshandbuch als hilfreich, um bei Bedarf schnell und direkt Tipps z. B. zu einzelnen Funktionalitäten des Programms oder zu Richtlinien der Artikelgestaltung zu erhalten. Das Handbuch für die Arbeitsumgebung des DFD soll zukünftig direkt über die Editor-Oberfläche für die BearbeiterInnen verfügbar gemacht werden.

Vorlagen für häufig wiederkehrende Artikel(-Bestandteile) anzubieten, wird als Anforderung an ein lexikographisches Redaktionssystem gesehen (vgl. ABEL/KLOSA 2012, 414; auch ATKINS/RUNDELL 2008, 115–116). Zur Unterstützung der lexikographischen Arbeit lassen sich zahlreiche Funktionen innerhalb des Editor-Interfaces einfügen. Beispiele sind Buttons, über die sich eine neue Bedeutung oder ein zusätzlicher historischer Beleg mit entsprechender XML-Kodierung ergänzen lassen, oder Menüpunkte, die die Inline-Auszeichnung von etymologischen und metasprachlichen Passagen ermöglichen. Bei der Inline-Auszeichnung von Literaturverweisen gibt es Auszeichnungsvorlagen, bei denen automatisch Kommata und Klammern eingefügt werden. Auf das automatische Erzeugen und Einfügen von Zeichen wie Kommata und Klammern in der Funktion von nichttypographischen

Strukturanzeigern (vgl. ABEL/KLOSA 2012, 414) wird jedoch verzichtet. Das DFD mit dem Fokus auf der Online-Publikation verzichtet auf solche textverdichtenden Mittel. Mit der Ausrichtung auf das Internet und prinzipiell „unbegrenzte“ Platzmöglichkeiten lässt sich ebenso begründen, warum eine Funktion zum Zählen der Wörter eines Artikels nicht berücksichtigt ist. Eine solche Option bietet zum Beispiel die Software Onoma, die für das „Woordeboek van die Afrikaanse Taal“ eingesetzt wird, das gedruckt, auf CD-ROM und online erscheint (vgl. RIDINGS 2003, 206).

Für die Erarbeitung des DFD gibt es neben Vorlagen für einzelne Artikelbestandteile auch Templates für die lexikographischen Inhalte. So stehen die Namenartikel mit einem XML-Grundgerüst zur weiteren Bearbeitung zur Verfügung. Dieses enthält bereits automatisch aus der DFA-Datenbank generierbare Inhalte wie den Artikelnamen, die Häufigkeit und den Rang innerhalb der Familiennamen in Deutschland. Ein anderes Wörterbuchprojekt, das Angaben teilweise automatisch kompiliert, ist *elexiko*, bei dem Korpusbelege für Stichwörter per Mechanismus ermittelt werden (vgl. MÜLLER-SPITZER 2011b, 37). Computergestützt ist zudem jeder Namendatei beim DFD eine eindeutige ID (*identifier*) zugewiesen, um die einzelnen Artikel identifizieren zu können. Für die Glossareinträge und Informationskästen gibt es weiterhin Templates, sodass neu erstellte Dateien bereits ein Grundgerüst an XML-Kodierung aufweisen und automatisch eine eindeutige ID erhalten.

Der Editor für das lexikographische Redaktionssystem befindet sich in der Entwicklung. Geplant, aber derzeit noch nicht implementiert, ist neben der Integration der Literaturdatenbank die automatische Neunummerierung, nachdem eine neue Bedeutung innerhalb eines Artikels hinzugefügt wird. Ähnlich verhält es sich mit einer automatisierten Aktualisierung von Querverweisen, was bei *elexiko* über die eigens entwickelte Software „*vernetziko*“ gelöst wird (vgl. MEYER 2011). Die genannten Funktionen berühren den Bereich der Unterstützung bei unterschiedlichen fehleranfälligen Aufgaben der BearbeiterInnen (vgl. ABEL/KLOSA 2012, 414). Als Erweiterung des Editors ist es ferner

geplant, über oXygen auf das Kartierungstool zugreifen zu können, mit dem die Verbreitungskarten für die Namen erstellt werden.

4.2.2 Datenbank

Als zentraler Bestandteil des lexikographischen Redaktionssystems dient die Datenbank der Speicherung der Daten. Bei der Vorstellung dieser Komponente steht im Vordergrund, wie die Inhalte dezentral bearbeitet, importiert, exportiert und durchsucht werden können.

Mit der Open-Source-Software eXist¹⁹ wird eine native XML-Datenbank eingesetzt. Auf diese Datenbank können die BearbeiterInnen über eine WebDAV-Verbindung innerhalb ihres oXygen-Editors zugreifen. Ortsunabhängiges Arbeiten an einem gemeinsamen Datenbestand stellt eine wichtige Anforderung an diese Komponente des Redaktionssystems (vgl. ABEL/KLOSA 2012, 415) dar und erscheint insbesondere für die MitarbeiterInnen des DFD erforderlich, die einzelne Lemmastrecken zu fremdsprachigen Namen als externe Fachkräfte bearbeiten.

Eine weitere wichtige Funktionalität stellt das Exportieren und Importieren von Dateien unterschiedlicher Formate dar (vgl. ABEL/KLOSA 2012, 415). Während oXygen die Möglichkeit bietet, die XML-Dateien als PDF-Dateien zu speichern, lassen sich Namenartikel, Glossareinträge und Informationskästen, die im XML-Format in der Datenbank vorliegen, über XSLT²⁰ nach HTML²¹ exportieren. Solche Transformationen sind für die Vorbereitung und Durchführung der Online-Präsentation der lexikographischen Inhalte im HTML-Format notwendig. Neben der Weiterverarbeitung ist ein Einspielen von Dateien in verschiedenen Dateiformaten in die eXist-Datenbank realisierbar. So können beispielsweise XQuery-Dateien mit Suchanfragen gespeichert und die Verbreitungskarten derzeit im PDF-Format und perspektivisch als

¹⁹ Verwendet wird eXist in der Version 2.0.

²⁰ Näheres zu XSLT (Extensible Stylesheet Language Transformations) über das W3C unter: <http://www.w3.org/TR/xslt20/>.

²¹ Weiterführende Informationen zu HTML (HyperText Markup Language) über das W3C unter: <http://www.w3.org/TR/html401/>.

SVG-Dateien abgelegt werden. Um sicherzustellen, dass sämtliche Zeichen erkannt werden können, wird in der Datenbank entsprechend ABEL/KLOSA (2012, 415) Unicode verwendet. Dies erscheint insbesondere angemessen, da bei der Bearbeitung von fremdsprachigen Namen häufig das Zeicheninventar für nichteuropäische Schriftsysteme benötigt wird.

Für die als wichtig erachteten Recherchen innerhalb der Datenbank (vgl. ABEL/KLOSA 2012, 415; auch ATKINS/RUNDELL 2008, 117) kommt die Abfragesprache XQuery²² zum Einsatz. Für Suchanfragen sind folglich weniger Kenntnisse zu einer speziellen Abfragesprache einer Datenbanklösung erforderlich, sondern vielmehr Vertrautheit mit den sogenannten X-Technologien. Beispiele für Auswertungen des Datenbestandes sind Recherchen, in denen Einträge bestimmter RedakteurInnen gesucht werden oder Artikel mit bestimmten Informationspositionen samt deren Häufigkeiten ermittelt werden.

Mit der eXist-Datenbank steht eine geeignete Lösung für die Speicherung der großen Anzahl an Einträgen zur Verfügung, deren voller Funktionsumfang, was Recherche- und Exportfunktionalitäten betrifft, bei Auswertungen und der Vorbereitung der Online-Publikation im weiteren Projektverlauf noch eingehender genutzt wird.

4.2.3 Verwaltungstools für das Projektmanagement und die Veröffentlichung

Die im Folgenden beschriebene Komponente der Arbeitsumgebung befasst sich mit der Verwaltung von Arbeitsabläufen, die bei der Integration neuer Inhalte und dem Zugriff auf Daten beginnen und vom Durchsuchen, der Überarbeitung sowie der Publikation bis hin zur (langfristigen) Sicherung der erarbeiteten Einträge reichen.

ABEL/KLOSA (2012, 415) stellen fest, dass eine „einfache und schnelle Integration großer Datenmengen in die Datenbank“ realisierbar sein sollte. Beim DFD hat das Einspielen der etwa 850.000 Namenarti-

²² Über das W3C stehen nähere Informationen zu XQuery (XML query language) bereit: <http://www.w3.org/TR/xquery-30/>.

kel als XML-Grundgerüst mit bereits automatisch generierten Inhalten angesichts des enormen Datenumfangs etwa zwölf Stunden erfordert. Die Herausforderung im Umgang mit einer derart großen Zahl von Einträgen zeigt sich weiterhin, wird bedacht, dass eine feingliedrige Ordnerstruktur zur Ablage der XML-Dateien gewählt ist, um Performanzprobleme beim Aufrufen über den Editor zu vermeiden. Der Bestand an Namen in der Datenbank ist so zu bewerten, dass es keine zusätzlich eingespielten, neuen Lemmata geben wird. Ergänzt werden zukünftig lexikographische Inhalte, die vor der Bereitstellung der Datenbank während einer ersten Testphase lokal in oXygen erstellt wurden.

Ein zentraler Aspekt der Verwaltungstools ist das Rechtemanagement (vgl. ABEL/KLOSA 2012, 415) innerhalb der eXist-Datenbank. Lese- und Schreibrechte sind entsprechend den jeweiligen Aufgabenbereichen im Team verteilt. Da das zu bearbeitende Nameninventar anhand der DFA-Datenbank als Datenbasis festgelegt ist, sind die Schreibrechte zum Beispiel derart begrenzt, dass in der Datenbank kein neuer Namenartikel erstellt werden kann. Innerhalb der DFD-Arbeitsumgebung gibt es keinen automatisierten Mechanismus (vgl. ABEL/KLOSA 2012, 415; auch ATKINS/RUNDELL 2008, 116), der den MitarbeiterInnen von der Redaktion ausgewählte bestimmte Datensets zur Bearbeitung zuweist. Die Artikel werden selbstständig in Absprache mit der Redaktion und entsprechend den Auswahlkriterien erstellt.

Um ein zeitgleiches Bearbeiten von Einträgen durch mehrere MitarbeiterInnen zu verhindern, sollten Dateien gesperrt werden können (vgl. ABEL/KLOSA 2012, 412; auch ATKINS/RUNDELL 2008, 117). Der Schreibschutz wird in der DFD-Arbeitsumgebung über die WebDAV-Verbindung gewährleistet und über ein Schloss-Symbol an den jeweiligen Dateien angezeigt. Auf Artikelenebene gibt es Vorschläge (vgl. ABEL/KLOSA 2012, 412; auch ATKINS/RUNDELL 2008, 117), Informationspositionen eines Eintrags wie zum Beispiel Ausspracheangaben zur Bearbeitung zu sperren, um diese SpezialistInnen vorzubehalten. Dies kommt beim DFD nicht zum Tragen, da ein/e BearbeiterIn jeweils für einen gesamten Eintrag verantwortlich ist. Außerdem werden die

geplanten Ausspracheangaben für die Namen erst zu einem späteren Zeitpunkt bearbeitet und sind daher derzeit in der Visualisierungsansicht des Artikels ausgespart.

Zur Unterstützung der Projektverwaltung erscheinen Suchmöglichkeiten hilfreich, die Auskunft über das Fortschreiten des gesamten Projektes und über einzelne Teilstrecken geben (vgl. ABEL/KLOSA 2012, 412; auch ATKINS/RUNDELL 2008, 116). Beispiele für Anfragen sind gezielte Recherchen zu Namenstrecken, die von einem/r bestimmten MitarbeiterIn bearbeitet oder überprüft werden, und Abfragen zu sämtlichen Artikeln mit einer Freigabe zur Publikation. Innerhalb der DFD-Arbeitsumgebung sind solche Suchen über XQuery durchführbar.

Werden einzelne Einträge überprüft, können die KorrektorInnen im Editor-Interface Kommentare hinterlassen. Diese bei ABEL/KLOSA (2012, 412) erwähnte Möglichkeit, gezielt Rückmeldungen zu geben, erleichtert den Austausch. Gleichzeitig können die BearbeiterInnen ihre Entscheidungen mit Kommentaren belegen, wie es auch bei der Software Onoma möglich ist (vgl. RIDINGS 2003, 208).

Werden Anpassungen und Aktualisierungen der lexikographischen Inhalte notwendig, zum Beispiel aufgrund von Änderungen der inhaltlichen Richtlinien, nach denen Einträge erarbeitet werden, bietet es sich an, diese automatisiert auf dem Bestand der Daten durchzuführen (vgl. ABEL/KLOSA 2012, 415).²³ Bei den XML-basierten Daten innerhalb der DFD-Datenbank ist dies über XQuery bzw. XSLT realisierbar, wobei Korrekturfunktionen zunächst an Einträgen auf einem Testserver ausprobiert werden können.

Änderungen an dem Datenbestand, die automatisiert durchgeführt werden, sind neben der regulären lexikographischen Arbeit des Redaktionsteams ein Grund dafür, zukünftig eine technische Lösung zur Versionierung zu implementieren. In der Arbeitsumgebung für das norwegische Wörterbuch „Norsk Ordbok 2014“ werden zum Beispiel sämtliche Änderungen in der Datenbank mit Angaben wie Datum und ver-

²³ Erfahrungswerte bei der Überarbeitung von Einträgen gibt MÜLLER-SPITZER (2011a) wieder, wobei auch auf den Aufwand manueller inhaltlicher Nachbearbeitung hingewiesen wird.

antwortlicher Person gespeichert (vgl. GRØNVIK/SMITH ORE 2013, 255). Versionenkontrolle wird auch bei ABEL/KLOSA (2012, 415) als relevantes Merkmal eines Redaktionssystems angesehen. Perspektivisch ist es weiterhin geplant, für die Langzeitarchivierung der lexikographischen Inhalte Repositorien wie zum Beispiel TextGrid²⁴ zu nutzen. Zur Sicherung der Daten, wie es auch ABEL/KLOSA (2012, 415) als erforderlich ansehen, werden bereits täglich automatisch gesteuert Back-ups durchgeführt.

Zur Veröffentlichung der Einträge ist eine Schnittstelle mit dem Content-Management-System TYPO3²⁵ eingerichtet, das derzeit bereits für den Webauftritt des Projekts verwendet wird. Es handelt sich um eine Open-Source-Software, für deren Verwendung die langjährige und positive Erfahrung der Digitalen Akademie spricht.²⁶ Hier zeigt sich, welchen Einfluss es auf die Zusammensetzung der Arbeitsumgebung hat, wenn bestimmte Softwarelösungen bereits bei Institutionen etabliert sind und dort intensiv genutzt werden.

Die Verwaltungstools für das Projektmanagement und die Veröffentlichung werden folglich über das Zusammenspiel verschiedener Softwarelösungen realisiert: der Editor oXygen, die Datenbank eXist sowie das Content-Management-System TYPO3.

5. Resümee

Für das DFD-Vorhaben ist die Erarbeitung einer projektspezifischen Arbeitsumgebung erforderlich, die sich an den selbstgesteckten Aufgaben und Zielen des Projekts orientiert, die sich wiederum aus den Fak-

²⁴ Nähere Informationen zur virtuellen Forschungsumgebung finden sich u. a. bei RAPP (2007) und auf der Internetseite von TextGrid unter: <http://www.textgrid.de/>.

²⁵ Es wird derzeit die TYPO3-Version 4.7.17 verwendet.

²⁶ Parallelen lassen sich hier zu OWID ziehen. Bei der Begründung für das Oracle-Datenbankmanagementsystem wird mit erwähnt, dass die Software bereits langjährig am IDS verwendet wird (vgl. MÜLLER-SPITZER/SCHNEIDER 2009, 200).

toren des besonderen Status der Eigennamen und den Defiziten in den allgemeinsprachlichen und den Spezialwörterbüchern im Hinblick auf Familiennamen ergeben.

Das lexikographische Redaktionssystem mit einer eXist-Datenbank, die über WebDAV aus oXygen bearbeitbar ist, ermöglicht ortsunabhängig performantes Arbeiten an einem großen gemeinsamen Datenbestand. Flexibilität und Erweiterbarkeit sind Merkmale der Arbeitsumgebung, um dem heterogenen Datenbestand mit Artikeln und sekundären Informationen in Form von Glossareinträgen und Informationskästen gerecht zu werden und die vielen und unterschiedlichen Informationspositionen des Artikels abbilden zu können. Weiterhin stehen die zu integrierenden Komponenten Kartierungstool und Literaturdatenbank für projektspezifisch nötige Anpassungen der Arbeitsumgebung.

Der Artikelerarbeitung innerhalb des lexikographischen Redaktionssystems liegt zudem eine Datenmodellierung zugrunde, die bei strikten Richtlinien die Konsistenz der Einträge ermöglicht und gleichzeitig flexibel ist. Diese Flexibilität ist nötig, um der Beschreibung von unterschiedlichen Familiennamen gerecht zu werden, wobei neben sprachlichem Wissen auch außersprachliches Wissen im Wörterbuch kodiert wird sowie die Recherche- und Arbeitsschritte des Langzeitprojekts dokumentiert werden. Unterstützt wird die Erarbeitung der unterschiedlichen lexikographischen Inhalte durch eine Editor-Oberfläche bei oXygen, die angepasste Eingabehilfen und diverse Vorlagen bietet. Mit der tiefen TEI-konformen Auszeichnung entsprechend modernen Standards wird die Grundlage für spätere Suchfunktionen gelegt und Langzeitarchivierung und -sicherung sowie Auswertungs- und Austauschmöglichkeit sichergestellt.

Eine wichtige Einflussgröße bei der Erarbeitung der Datenmodellierung sowie darauf aufbauend der gesamten Entwicklung der Arbeitsumgebung ist die stetige und enge Zusammenarbeit zwischen den BearbeiterInnen und den EDV-MitarbeiterInnen. Regelmäßiger und intensiver Austausch fördert das gegenseitige Verständnis für unterschiedliche Arbeitsbereiche im Wörterbuchvorhaben und ist somit un-

erlässlich, um ein lexikographisches Redaktionssystem entstehen zu lassen, mit dem ein modernes und umfassendes Wörterbuch zu den Familiennamen Deutschlands erarbeitet werden kann. Im Ergebnis ist es bereits in der Startphase des Langzeitprojektes gelungen, eine XML-basierte Arbeitsumgebung zu erstellen, die aktuelle technologische sowie wissenschaftlich-methodische Standards berücksichtigt und dabei flexibel an neue Anforderungen und Technologieentwicklungen anpassbar ist.

Literatur

- ABEL, ANDREA (2012): Dictionary writing systems and beyond. In: GRANGER, SYLVIANE/PAQUOT, MAGALI (Hrsg.): *Electronic Lexicography*. Oxford, 83–106.
- ABEL, ANDREA/KLOSA, ANNETTE (2012): Der lexikographische Arbeitsplatz – Theorie und Praxis. In: VATVEDT FJELD, RUTH/TORJUSEN, JULIE MATILDE (Hrsg.): *Proceedings of the 15th EURALEX International Congress*, 7–11 August, 2012, Oslo. Oslo, 413–421.
- ATKINS, SUE/RUNDELL, MICHAEL (2008): *The Oxford Guide to Practical Lexicography*. Oxford.
- BAHLOW, HANS (2005): *Deutsches Namenlexikon. Familien- und Vornamen nach Ursprung und Sinn erklärt*. 16. Auflage. Frankfurt.
- BLANÁR, VINCENT (2001): *Theorie des Eigennamens*. Hildesheim.
- BRECHENMACHER, JOSEF KARLMANN (1960–63): *Etymologisches Wörterbuch der deutschen Familiennamen*. 2 Bände. Limburg.
- DEBUS, FRIEDHELM (1977): Aspekte zum Verhältnis Namen – Wort. In: STEGER, HUGO (Hrsg.): *Probleme der Namenforschung im deutschsprachigen Raum*. Darmstadt, 3–25.
- DEBUS, FRIEDHELM (1980): Onomastik. In: ALTHAUS, HANS PETER/HENNE, HELMUT/WIEGAND, HERBERT ERNST (Hrsg.): *Lexikon der Germanistischen Linguistik*. Band 1. Tübingen, 187–198.
- DEBUS, FRIEDHELM (2005): Lexikologie und Onomastik. In: CRUSE, DAVID ALAN/HUNDSNURSCHER, FRANZ/JOB, MICHAEL/LUTZEIER, PETER ROLF (Hrsg.): *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. 2. Teilband. Berlin/New York (*Handbücher zur Sprach- und Kommunikationswissenschaft*. 21.2), 1838–1846.
- DENZER, SANDRA/HORN, FRANZISKA (erscheint): Die ersten Schritte eines Langzeitprojekts. *Das Digitale Familiennamenwörterbuch Deutschlands*.

- In: HILDENBRANDT, VERA (Hrsg.): Der lexikografische Prozess bei Internetwörterbüchern. 4. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim (OPAL – Online publizierte Arbeiten zur Linguistik).
- DE SCHRYVER, GILLES-MAURICE (2011): Why Opting for a Dedicated, Professional, Off-the-shelf Dictionary Writing System Matters. In: AKASU, KAORU/UCHIDA, SATORU (Hrsg.): ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto Terra, Kyoto, Japan, August 22–24, 2011. Kyoto, 647–656.
- DUDEN – Die deutsche Rechtschreibung (2009). 25., völlig neu bearbeitete und erweiterte Auflage. Mannheim.
- DUMONT, STEFAN/FECHNER, MARTIN (2012): Digitale Arbeitsumgebung für das Editionsprojekt »Schleiermacher in Berlin 1808–1834«. In: digiversity – Webmagazin für Informationstechnologie in den Geisteswissenschaften. Online: <http://digiversity.net/2012/digitale-arbeitsumgebung-fur-das-editionsvorhaben-schleiermacher-in-berlin-1808-1834/> [Abruf: 2014-04-30].
- DUMONT, STEFAN (2013): Tutorial: Wie baue ich ein eigenes Framework für Oxygen XML? In: digiversity – Webmagazin für Informationstechnologie in den Geisteswissenschaften. Online: <http://digiversity.net/2013/tutorial-wie-baue-ich-ein-oxygen-xml-framework/> [Abruf: 2014-04-30].
- EICHLER, ERNST/HILTY, GEROLD/LÖFFLER, HEINRICH/STEGE, HUGO/ZGUSTA, LADISLAV (Hrsg.) (1995): Ein internationales Handbuch zur Onomastik. Berlin/New York (Handbücher zur Sprach- und Kommunikationswissenschaft. 11).
- FAHLBUSCH, FABIAN/HEUSER, RITA (2014): Das „Digitale Familiennamenwörterbuch Deutschlands“. Möglichkeiten und Ziele am Beispiel regionaler Namen. In: GILLES, PETER/KOLLMANN, CRISTIAN/MULLER, CLAIRE (Hrsg.): Familiennamen zwischen Maas und Rhein. Frankfurt am Main [u. a.], 209–226.
- FLEISCHER, WOLFGANG (1964): Zum Verhältnis von Name und Appellativum im Deutschen. In: Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig, Heft 13. Leipzig, 369–378.
- GERHARDT, DIETRICH (1949/50): Über die Stellung der Namen im lexikalischen System. In: Beiträge zur Namenforschung 1, 1–24.
- GEYKEN, ALEXANDER (2011): Die dynamische Verknüpfung von Kollokationen mit Korpusbelegen und deren Repräsentation im DWDS-Wörterbuch. In: KLOSA/MÜLLER-SPITZER (Hrsg.), 9–20.
- GOTTSCHALD, MAX (2006): Deutsche Namenkunde. 6., durchgesehene und bibliographisch aktualisierte Auflage. Berlin/New York.

- GRØNVIK, ODDRUN/SMITH ORE, CHRISTIAN-EMIL (2013): What should the electronic dictionary do for you – and how? In: KOSEM, IZTOK/KALLAS, JELENA/GANTAR, POLONA/KREK, SIMON/LANGEMETS, MARGIT/TUULIK, MARIA (Hrsg.): Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn. Estonia. Ljubljana/Tallinn, 243–260.
- HARWEG, ROLAND (1997): Namen und Wörter. 1. Halbband. Bochum.
- HARWEG, ROLAND (1999): Eigennamen als Einträge in Wörterbüchern und Lexika. In: HARWEG, ROLAND (Hrsg.): Studien zu Eigennamen. Aachen, 61–123.
- HILDENBRANDT, VERA (2011): TEI-basierte Modellierung von Retrodigitalisaten (am Beispiel des Trierer Wörterbuchnetzes). In: KLOSA/MÜLLER-SPITZER (Hrsg.), 21–35.
- JANNIDIS, FORTIS (2009): TEI in a crystal ball. In: Literary and Linguistic Computing 24/3, 253–65.
- JOFFE, DAVID/DE SCHRYVER, GILLES-MAURICE/PRINSLOO, DANIEL JACOBUS (2003): Introducing TschwaneLex – A New Computer Program for the Compilation of Dictionaries. In: DE SCHRYVER, GILLES-MAURICE (Hrsg.): TAMA 2003 South Africa: Conference Proceedings. Pretoria, 97–104.
- KALVERKÄMPER, HARTWIG (1978): Textlinguistik der Eigennamen. Stuttgart.
- KALVERKÄMPER, HARTWIG (1994): Eigennamen in Texten. In: CANISIUS, PETER/HERBERMANN, CLEMENS-PETER/TSCHAUDER, GERHARD (Hrsg.): Text und Grammatik. Festschrift für Roland Harweg zum 60. Geburtstag. Bochum, 205–238.
- KLOSA, ANNETTE/SCHNÖRCH, ULRICH/SCHOOOLAERT, SABINE (2010): Stichwort, Stichwortliste und Eigennamen in *lexiko*: Einflüsse der Korpusbasiertheit und Hypermedialität auf die lexikografische Konzeption. In: DYKSTRA, ANNE/SCHOONHEIM, TANNEKE (Hrsg.): Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010). Ljouwert, 653–663.
- KLOSA, ANNETTE/MÜLLER-SPITZER, CAROLIN (Hrsg.) (2011): Datenmodellierung für Internetwörterbücher. 1. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim (OPAL – Online publizierte Arbeiten zur Linguistik 2/2011).
- KLOSA, ANNETTE/SCHOOOLAERT, SABINE (2011): Die lexikografische Behandlung von Eigennamen in *lexiko*. In: KLOSA, ANNETTE (Hrsg.): *lexiko*. Erfahrungsberichte aus der lexikografischen Praxis eines Internetwörterbuchs. Tübingen, 193–211.
- KOHLHEIM, ROSA/KOHLHEIM, VOLKER (2005): Duden Lexikon der Familiennamen. Herkunft und Bedeutung von 20 000 Familiennamen. Mannheim.

- KOLDE, GOTTFRIED (1995): Grammatik der Eigennamen. In: EICHLER u. a. (Hrsg.), 400–408.
- KUNZE, KONRAD (1998): dtv-Atlas Namenkunde. Vor- und Familiennamen im deutschen Sprachgebiet. München.
- LÖTSCHER, ANDREAS (1995): Der Name als lexikalische Einheit: Denotation und Konnotation. In: EICHLER u. a. (Hrsg.), 448–457.
- MARYNISSEN, ANN/NÜBLING, DAMARIS (2010): Familiennamen in Flandern, den Niederlanden und Deutschland. Ein diachroner und synchroner Vergleich. In: DAMMEL, ANTJE/KÜRSCHNER, SEBASTIAN/NÜBLING, DAMARIS (Hrsg.): Kontrastive germanistische Linguistik. Hildesheim/Zürich/New York (Germanistische Linguistik. 206–209/2010), 311–364.
- MEYER, PETER (2011): *vernetziko*: A Cross-Reference Management Tool for the Lexicographer's Workbench. In: KOSEM, IZTOK/KOSEM, KARMEN (Hrsg.): Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011, Bled, 10–12 November 2011. Ljubljana, 191–198. Online: <http://www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-25.pdf> [Abruf: 2014-06-01].
- MÖLLER, LUCIE A. (1995): Methods and problems in proper name lexicography. In: EICHLER u. a. (Hrsg.), 324–328.
- MÜLLER-SPITZER, CAROLIN (2011a): Der Einsatz einer maßgeschneiderten, feingranularen XML-Modellierung im lexikografischen Prozess. In: KLOSA, ANNETTE (Hrsg.): *ellexiko*. Erfahrungsberichte aus der lexikografischen Praxis eines Internetwörterbuchs. Tübingen, 173–191.
- MÜLLER-SPITZER, CAROLIN (2011b): Der Aufbau einer maßgeschneiderten XML-basierten Modellierung für ein Wörterbuchnetz. In: KLOSA/MÜLLER-SPITZER (Hrsg.), 37–51.
- MÜLLER-SPITZER, CAROLIN/SCHNEIDER, ROMAN (2009): Ein XML-basiertes Datenbanksystem für digitale Wörterbücher – Ein Werkstattbericht aus dem Institut für Deutsche Sprache. In: *it – Information Technology* 51/4, 197–206.
- NAUMANN, HORST (2007): Das große Buch der Familiennamen. Augsburg/München.
- NÜBLING, DAMARIS (2005): Zwischen Syntagmatik und Pragmatik: Grammatische Eigennamenmarker und ihre Typologie. In: *Zeitschrift für Germanistische Linguistik* 33, 25–56.
- NÜBLING, DAMARIS/FAHLBUSCH, FABIAN/HEUSER, RITA (2012): Namen. Eine Einführung in die Onomastik. Tübingen.
- NÜBLING, DAMARIS/SCHMUCK, MIRJAM (2010): Die Entstehung des s-Plurals bei Eigennamen als Reanalyse vom Kasus- zum Numerusmarker. In: *Zeitschrift für Dialektologie und Linguistik* 77/2, 145–182.

- RAPP, ANDREA (2007): „TextGrid. Modulare Plattform für verteilte und kooperative wissenschaftliche Textdatenverarbeitung – ein Community-Grid für die Geisteswissenschaften“. Chancen und Perspektiven für eine neue Wissenschaftskultur in den Geisteswissenschaften. In: ARBEITSGEMEINSCHAFT HISTORISCHER FORSCHUNGSEINRICHTUNGEN IN DER BUNDESREPUBLIK DEUTSCHLAND (Hrsg.): Jahrbuch der historischen Forschung in der Bundesrepublik Deutschland. Berichtsjahr 2006. München, 61–68.
- RIDINGS, DANIEL (2003): Lexicographic Workbench: A Case History. In: VAN STERKENBURG, PIET (Hrsg.): A Practical Guide to Lexicography. Amsterdam (Terminology and Lexicography Research and Practice. 6), 204–214.
- SEIBICKE, WILFRIED (2004): Über die Geschichte und Typen der deutschen Personennamen. In: BESCH, WERNER/BETTEN, ANNE/REICHMANN, OSKAR/SONDEREGGER STEFAN (Hrsg.): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 2., völlig neu bearbeitete und erweiterte Auflage. 4. Teilband. Berlin/New York (Handbücher zur Sprach- und Kommunikationswissenschaft. ²2.4), 3535–3552.
- STRAUß, GERHARD (1989): Angabe traditioneller Wortarten oder Beschreibung nach funktionalen Wortklassen im allgemeinen einsprachigen Wörterbuch? In: HAUSMANN, FRANZ JOSEF/REICHMANN, OSKAR/WIEGAND, HERBERT ERNST/ZGUSTA, LADISLAV (Hrsg.): Wörterbücher: Ein internationales Handbuch zur Lexikographie. 1. Teilband. Berlin/New York (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.1), 788–796.
- SVENSÉN, BO (2009): A Handbook of Lexicography: The Theory and Practice of Dictionary-Making. Cambridge.
- TEXTGRID. Virtuelle Forschungsumgebung für die Geisteswissenschaften. Online: <https://www.textgrid.de/> [Abruf: 2014-05-07].
- TEXT ENCODING INITIATIVE. Online: <http://www.tei-c.org/index.xml> [Abruf: 2014-05-07].
- VATER, HEINZ (1965): Eigennamen und Gattungsbezeichnungen. In: Muttersprache 75, 207–213.
- WINNER, SABINA/BUCHNER, MARTINA (2011): Digitales Ortsnamenbuch Online (DONBO): Neue Perspektiven der Namenforschung. In: ZIEGLER, ARNE/WINDBERGER-HEIDENKUMMER, ERIKA (Hrsg.): Methoden der Namenforschung. Methodologie, Methodik und Praxis. Berlin, 183–198.
- W3C (1999): HTML 4.01 Specification. W3C Recommendation 24 December 1999. Online: <http://www.w3.org/TR/html401/> [Abruf: 2014-05-07].
- W3C (2007): XSL Transformations (XSLT) Version 2.0. W3C Recommendation 23 January 2007. Online: <http://www.w3.org/TR/xslt20/> [Abruf: 2014-05-07].

- W3C (2008): Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation 26 November 2008. Online: <http://www.w3.org/TR/xml/> [Abruf: 2014-05-07].
- W3C (2011): Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. W3C Recommendation 07 June 2011. Online: <http://www.w3.org/TR/CSS2/> [Abruf: 2014-05-07].
- W3C (2014): XQuery 3.0: An XML Query Language. W3C Recommendation 08 April 2014. Online: <http://www.w3.org/TR/xquery-30/> [Abruf: 2014-05-07].
- ZODER, RUDOLF (1968): Familiennamen in Ostfalen. 2 Bände. Hildesheim.

PETER MEYER

Entlehnungsketten in einem Internetportal für Lehnwörterbücher

IT-Infrastruktur und computerlexikographischer Prozess in einem
Projekt zu polnisch vermittelten Germanismen im Ostslavischen

1. Einleitung

Im Rahmen eines auf drei Jahre angelegten und von der Deutschen Forschungsgemeinschaft geförderten Kooperationsprojekts entstehen zur Zeit drei Wörterbücher zu deutschen Lehnwörtern in den ostslavischen Sprachen Russisch, Weißrussisch und Ukrainisch. Erfasst werden in diesen Wörterbüchern ausschließlich solche schriftsprachlich belegten Lehnwörter, für die eine Vermittlung über das Polnische wahrscheinlich oder doch wenigstens möglich ist. Hintergrund ist eine historisch komplexe, langanhaltende Sprachkontaktsituation, zu der es eine größere Anzahl an Einzelstudien gibt (vgl. beispielweise zur Frage der polnischen Vermittlung ukrainischer Germanismen BESTERS-DILGER 2002), aber bislang noch keine umfassende, den gesamten ostslavischen Bereich abdeckende lexikographische Untersuchung. Die drei Wörterbücher sollen ab 2016 im *Lehnwortportal Deutsch*, einer am Institut für Deutsche Sprache (IDS) betreuten, unter <http://lwp.ids-mannheim.de> frei zugänglichen Online-Ressource zu Wörterbüchern deutscher Lehnwörter in anderen Sprachen, der Öffentlichkeit zur Verfügung stehen. Die lexikographischen Arbeiten werden am sprachwissenschaftlichen Lehrstuhl des Instituts für Slavistik der Universität Oldenburg durchgeführt bzw. koordiniert, während die computerlexikographische Infrastruktur in der Abteilung Lexik am IDS betreut wird.

Der vorliegende Beitrag beschreibt den lexikographischen Prozess der Erstellung der drei genannten Lehnwörterbücher aus computerlexi-

kographischer Perspektive. Wesentliche Besonderheiten dieses Prozesses sind

- die große Zahl an konsultierten und neu zu erzeugenden Wörterbüchern,
- der angestrebte hohe Vernetzungsgrad der neuen Wörterbücher untereinander und mit ihrer gemeinsamen lexikographischen Ausgangsbasis, einem Wörterbuch polnischer Germanismen,
- der verteilte und kooperative Charakter des lexikographischen Prozesses, der hohe Anforderungen an die IT-Infrastruktur stellt,
- das Fehlen einer permanenten Netzanbindung in einigen Arbeitsstellen, das die Verwendung einer vollständig server- und damit internetbasierten Lösung unmöglich macht
- sowie die möglichst zeitnahe Einbettung der Projektergebnisse in die komplexe graphenbasierte Datenmodellierung des Lehnwortportals.

Die genannten Punkte seien hier kurz näher ausgeführt. Operationaler lexikographischer Ausgangs- und Bezugspunkt für alle drei Wörterbücher sind die polnischen Germanismen im *Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache* (DE VIN-CENZ/HENTSCHEL 2010; im Weiteren WDLP) mitsamt ihren ausdrucksseitigen Varianten und den davon im Polnischen gebildeten, im WDLP gebuchten abgeleiteten Formen (Derivaten und Komposita). Insbesondere dient auch das im WDLP für die einzelnen Lemmata angesetzte Lesartenspektrum als ein Ausgangspunkt für die bedeutungsseitige Beschreibung der ostslavischen Lehnwörter. Daraus ergibt sich eine starke und von Anfang an auch für die Konzeption der Redaktionsumgebung zu berücksichtigende Vernetzung der Wörterbücher mit dem WDLP.

Die drei zu erstellenden Wörterbücher (im Folgenden als *Zielwörterbücher* bezeichnet) werden in Oldenburg von mehreren Lexikographinnen bearbeitet; dabei sollen ostslavische Lemmata, die Entsprechungen zu einem und demselben WDLP-Lemma sind (und im Folgenden als *korrespondierende Lemmata* bezeichnet werden), möglichst parallel redigiert werden. Dies erfordert eine kollaborative Redaktions-

umgebung mit zentraler Datenhaltung, die im vorliegenden Fall auf Servern des IDS in Mannheim realisiert wird.

Zu korrespondierenden Lemmata in den drei Zielwörterbüchern soll es einen *sprachübergreifenden Kommentar* geben, der die aufgrund des lexikographisch-philologischen Belegmaterials möglichen Entlehnungswege unter Berücksichtigung chronologischer und semantischer Faktoren diskutiert und durch eine interaktive Visualisierung (Pfeilgraphik) der diskutierten Entlehnungswege ergänzt wird. Zielwörterbücher und sprachübergreifende Kommentare werden im Weiteren als *Zielressourcen* bezeichnet.

Das lexikographische Material wird nicht neu auf der Grundlage eigener Textkorpora erhoben, vielmehr werden für jedes der drei Wörterbücher die bereits in größerer Zahl und in überwiegend sehr guter Qualität zur Verfügung stehenden historischen und zeitgenössischen wissenschaftlichen Wörterbücher der betreffenden Sprachen – hier als *Quellwörterbücher* bezeichnet – ausgewertet. Ein Teil des Materials ist noch unveröffentlicht und muss direkt in den russischen, weißrussischen bzw. ukrainischen Karteien der betreffenden großen Wörterbuchprojekte von Kooperationspartnern ausgewertet werden. Die Ergebnisse sollen möglichst direkt in der Oldenburger Redaktionsumgebung zur Verfügung stehen. Es kann jedoch für die Karteien keine verlässliche permanente Internet-Anbindung an die zentrale Serverstruktur vorausgesetzt werden. Die Karteien müssen daher mit der Redaktionssoftware in einem dezentralen, lokalen Modus arbeiten, die Ergebnisse aber dennoch regelmäßig in die zentrale Datenhaltung einspeisen können.

Die Zielwörterbücher sind von vornherein als Onlinepublikationen, genauer als Module für das *Lehnwortportal Deutsch*, konzipiert. Das Lehnwortportal bietet durch eine granulare Vernetzungsstruktur und durch eine wörterbuchübergreifende Metalemmaliste der deutschen Etyma komplexe Suchmöglichkeiten und soll es Benutzern langfristig insbesondere ermöglichen, ausgehend von deutschen Etyma deren vielfältige Entlehnungswege in eine größere Zahl von Nehmersprachen

nachzuvollziehen. Dabei soll auch die Darstellung von und Suche nach Entlehnungsketten möglich sein.

Der Aufbau des Beitrags sieht wie folgt aus. Abschnitt 2 gibt einen kurzen Überblick über die für das Projekt relevanten lexikalischen Ressourcen (WDLP, Quell- und Zielwörterbücher, Lehnwortportal). Abschnitt 3 skizziert die verwendeten Repräsentationsformate für die lexikographischen Daten. Im Abschnitt 4 werden die einzelnen Etappen des lexikographischen Prozesses aus computerlexikographischer Perspektive vorgestellt, mit einem Schwerpunkt auf der im Projekt entwickelten technischen Infrastruktur (Redaktions- und Serversoftware). Im abschließenden Abschnitt 5 werden Ergebnisse, Probleme und Perspektiven des Projekts, im Besonderen auch das Verhältnis von technologischem Aufwand und lexikographischem Ertrag des Projekts, diskutiert.

2. Lexikographische Ressourcen des Projekts

2.1 WDLP und Quellwörterbücher

Die beiden wichtigsten lexikographischen Vorentscheidungen für das vorzustellende Projekt sind folgende: (i) Berücksichtigt werden ausschließlich diejenigen polnischen Germanismen, die im WDLP verzeichnet sind; (ii) die Zielwörterbücher werden auf der Grundlage der Auswertung einer größeren Anzahl von Quellwörterbüchern erstellt, das heißt, es wird kein lexikographisches Primärmaterial neu erhoben. Durch diese beiden operationalen Beschränkungen wird der erforderliche Arbeitsaufwand kontrollierbar gemacht, andererseits aber auch der Skopus der Untersuchung eingeschränkt.

Die Arbeiten am WDLP (DE VINCENZ/HENTSCHEL 2010) begannen Anfang der 80er Jahre des zwanzigsten Jahrhunderts und wurden 2010 mit einer Online-Publikation abgeschlossen. Ende 2012 wurde das WDLP zusätzlich mit wesentlich umfangreicheren wörterbuchspezifischen sowie portalweiten Suchmöglichkeiten im *Lehnwortportal Deutsch* veröffentlicht. Beide Veröffentlichungen sind frei zugänglich.

Das WDLP hat mit seiner seinerzeit neuartigen technischen Umsetzung in einer relationalen Datenbankarchitektur Pionierarbeit geleistet (vgl. HENTSCHEL 1989) und ist mit seinen über 2400 Lemmata mit hochgranularer Mikrostruktur eine einzigartige Informationsquelle in der slavistischen Lexikographietradition. Verzeichnet werden direkte Entlehnungen aus dem Deutschen, deren Etyma germanischen Ursprungs sind, im schrift- und standardsprachlichen Polnisch von den Anfängen des polnischen Schrifttums im 13./14. Jahrhundert bis in die Mitte des 20. Jahrhunderts. Die Artikel des WDLP weisen eine komplexe Neststruktur auf, die hier etwas ausführlicher skizziert werden soll, da sie auch die Grundlage für den Aufbau der neu zu erstellenden Zielwörterbucheinträge darstellt. Verzeichnet werden im WDLP insbesondere alle im ausgewerteten Quellenmaterial belegten ausdrucksseitigen Varianten, sofern diese nicht rein orthographischer Natur sind; die attestierten Lesarten jeweils mit polnischer und deutscher Bedeutungsparaphrase; innerpolnisch vom Lehnwort abgeleitete Komposita und Derivate; mit dem Lehnwort im Polnischen konkurrierende Synonyme; die deutschen Etyma; und ein diskursiver Kommentar mit philologischen und sprachhistorischen Anmerkungen. Dem oft sehr granularen Lesartenspektrum – im Falle von poln. *waga* ‘Waage’ werden 24 Lesarten unterschieden – kommt bereits im WDLP eine über bloße Bedeutungsangaben hinausreichende strukturierende Funktion zu: Zum einen wird bei den einzelnen deutschen Etyma vermerkt, für welche Lesarten des Lehnwortes es als „Vorlage“ gedient haben kann. Zum anderen wird bei Komposita, Derivaten und Konkurrenten notiert, welche Lesarten des Simplex für die Semantik des Wortes „einschlägig“ sind, falls hier, wie häufig der Fall, Einschränkungen vorliegen.

Für das WDLP wurde in großem Umfang vorliegendes, z. T. unveröffentlichtes lexikographisches Material ausgewertet, in beschränktem Maße jedoch auch zusätzlich eigenes Textmaterial erhoben. Auf dieser Grundlage verzeichnet das WDLP insbesondere für jede Ausdrucksvariante und jede Lesart Belege mit Zitaten sowie eine Buchungsgeschichte der für den betreffenden Artikel ausgewerteten Wör-

terbücher. Besonderes Gewicht wird auf Aussagen zu Erst- und Letztbelegen gelegt.

Für jede der drei Sprachen Russisch, Weißrussisch und Ukrainisch wird im Rahmen des hier vorzustellenden Projekts eine größere Zahl von *Quellwörterbüchern* aus der sehr reichen und langen lexikographischen Tradition dieser Sprachen ausgewertet, darunter aktuelle und historische einsprachige Gegenwartswörterbücher sowie historische einsprachige Wörterbücher. Dabei handelt es sich ausdrücklich i. A. nicht um etymologische oder Lehnwörterbücher, so dass in der Regel in den exzerpierten Werken keine Informationen zur sprachlichen Herkunft zu finden sind. Die Wörterbücher werden systematisch auf Entsprechungen zu den im WDLP gebuchten Lemmata samt Varianten und Derivaten durchsucht. Ob es sich dabei tatsächlich um Entlehnungen von polnischen Germanismen handelt oder aber z. B. um direkte Entlehnungen aus dem Deutschen, spielt beim Exzerpierungsprozess zunächst keine Rolle; dies wird im Rahmen der Endredaktion mit philologisch-sprachhistorischen Methoden im einzelnen geprüft. Einige Wörterbücher haben für das Projekt „kanonischen“ Status, d. h. sie werden in allen Fällen konsultiert, während eine offene Anzahl weiterer lexikographischer Quellen nur nach Bedarf hinzugezogen wird. Für das Ukrainische und Russische haben je sechs Wörterbücher kanonischen Status, im Weißrussischen sind es vier. Für das Russische reicht die Spanne der obligatorisch zu prüfenden Wörterbücher vom historischen *Wörterbuch der russischen Sprache des 11. bis 17. Jahrhunderts* (derzeit 29 Bände; BARCHUDAROV 1975ff.) bis zum abgeschlossenen 17-bändigen *Wörterbuch der modernen russischen Literatursprache* (ČERNYŠEV 1948–1965) sowie einem jüngeren, viel genutzten einbändigen Wörterbuch.

Ein zentrales Problem, das in ähnlicher Form bereits für die Erstellung des WDLP relevant war, besteht darin, dass viele der zu konsultierenden lexikographischen Werke noch nicht vollständig publiziert sind und daher unveröffentlichtes Material aus den Karteien dieser langfristig angelegten Großprojekte gesucht und exzerpiert werden muss. Konkret handelt es sich um die Moskauer Kartei des *Wörterbuchs der russi-*

schen Sprache des 11. bis 17. Jahrhunderts (BARCHUDAROV 1975ff.), die Sankt Petersburger Kartei des *Wörterbuchs der russischen Sprache des 18. Jahrhunderts* (SOROKIN 1984ff.), die Minsker Kartei des *Historischen Wörterbuchs der weißrussischen Sprache* (ŽURAŪSKI 1982ff.) sowie die Lemberger Kartei des *Wörterbuchs der ukrainischen Sprache des 16. und der ersten Hälfte des 17. Jahrhunderts* (HRYNČYŠYN u. a. 1994ff.).

2.2 Zielwörterbücher

Die drei zu erstellenden Zielwörterbücher zu ukrainischen, weißrussischen bzw. russischen Entlehnungen polnischer Germanismen weisen eine identische Artikelmikrostruktur auf, die eng an die oben skizzierte Mikrostruktur der WDLP-Artikel angelehnt bzw. in einigen Fällen sogar umfangreicher ist. Das Lesartenspektrum des WDLP spielt auch für die Exzerpierung der Quellwörterbücher sowie für die Erstellung der Zielwörterbücher eine wesentliche strukturierende Rolle: Für jede einzelne exzerpierte bzw. im Rahmen der Redaktion der Zielwörterbücher neu in deutscher Metasprache beschriebene Lesart eines Wortes wird von der Lexikographin in der Redaktionsumgebung eingetragen, mit welchen Lesarten des entsprechenden WDLP-Lemmas sie ganz und mit welchen sie zumindest teilweise korrespondiert; vgl. den Screenshot in Abb. 1, der dies für die zu polnisch *waga* ‘Waage’ notierten Lesarten veranschaulicht. Die Liste der vollständigen und der teilweisen Lesart-Entsprechungen wird im Folgenden als *WDLP-Profil* einer ostslavischen Lesartenangabe bezeichnet. Die WDLP-Profile der Bedeutungsangaben, die aus den Quellwörterbüchern exzerpiert werden, spielen eine wesentliche Rolle bei der in Abschnitt 4.2 näher beschriebenen Erstellung der Zielwörterbuchartikel.

Nur am Rande sei erwähnt, dass die Zuordnung beispielsweise von exzerpierten Ausdrucksvarianten zu einer der drei ostslavischen Sprachen bei älteren Belegen durchaus problematisch ist, da den drei modernen kodifizierten ostslavischen Standardsprachen in älteren Zeiten ein komplexes Dialektkontinuum gegenübersteht. So sind auch die

Sprachenzuordnungen in älteren Wörterbüchern mit großer Vorsicht zu beurteilen; eine operationalisierbare Zuordnung zu modernen Sprachabgrenzungen kann allenfalls über die geographische Lokalisierung von Belegen erfolgen.

Korrespondierende, also demselben WDLP-Lemma zuzuordnende, Artikel der drei Zielwörterbücher werden mit einem wörterbuchübergreifenden, diskursiven Kommentar versehen, der insbesondere die Plausibilität verschiedener Entlehnungswege der betreffenden Lexeme zum Gegenstand hat. Die aufgrund der lexikographisch-philologischen Evidenz möglichen Entlehnungswege sollen überdies später in einer schematisch-kartographischen Visualisierung online veranschaulicht werden.

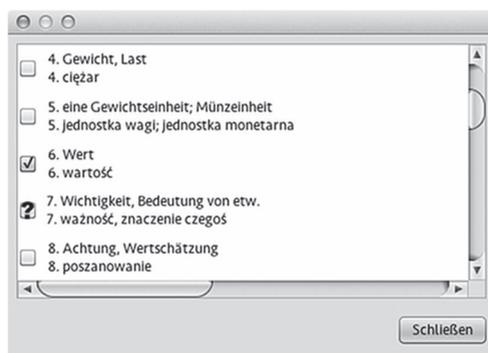


Abb. 1: Dialog zur Auswahl von korrespondierenden WDLP-Lesarten für Bedeutungsangaben in den ostslavischen Exzerpten und Zielwörterbüchern

Einen Überblick über die Rolle der verschiedenen lexikographischen Ressourcen des Projekts gibt Abb. 2.: Für einen gegebenen WDLP-Artikel *xyz* wird in jeder der drei Zielsprachen in einer Anzahl von Wörterbüchern nach Entsprechungen gesucht. Die gefundenen Quellwörterbuchartikel werden nach einem einheitlichen Schema, das eng an die Zielwörterbuch-Mikrostruktur angelehnt ist, exzerpiert; dabei werden an verschiedenen Stellen Informationen aus dem WDLP referenziert, insbesondere in Form der eben diskutierten WDLP-Lesarten-

profile. Bei der anschließenden Erstellung der Artikel der drei Zielwörterbücher wird die Information aus den Exzerpten der betreffenden Sprache in einem komplexen, in Abschnitt 4.2 näher beschriebenen Verfahren in das Format der neu zu erstellenden Artikel überführt. Die Lemmatisierung dieser mit WDLP-artiger Neststruktur versehenen Artikel ist prinzipiell von den ohnehin divergierenden Lemmatisierungskonventionen der Quellwörterbücher unabhängig. In einem letzten Schritt werden die wörterbuch- und artikelübergreifenden Kommentare erstellt, die mit den einzelnen Zielwörterbuchartikeln sowie deren Inhalten vernetzt sind.

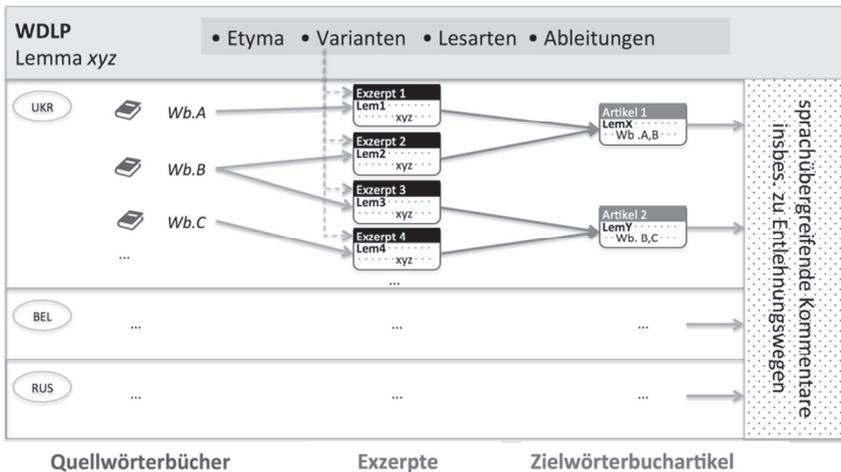


Abb. 2: Schematische Darstellung des lexikographischen Prozesses

Die Quellwörterbuchexzerpte und die Artikel der Zielwörterbücher weisen, wie bereits angedeutet, letztlich dieselbe Mikrostruktur auf, auch wenn einige lexikographische Angabetypen in den Exzerpten fehlen, also nicht spezifiziert werden. Diese grundsätzliche Modellierungsentscheidung vereinfacht viele technische Abläufe beträchtlich und erlaubt es im Weiteren, Dokumente der beiden genannten Typen zusammenfassend und abstrahierend als (*ostslavische*) *Einträge* zu

bezeichnen. Es sei abschließend betont, dass die Zuordnung von Einträgen zu einer ostslavisches Einzelsprache, auch aus den oben erwähnten sprachhistorischen Gründen, eine untergeordnete Rolle für den lexikographischen Prozess sowie für die computerlexikographischen Abläufe spielt.

3. Datenmodellierung und verteilte Infrastruktur des Projekts

Zum besseren Verständnis der computerlexikographischen Hintergründe des Projekts wird in diesem Abschnitt ein Blick auf die informationstechnische Repräsentation der lexikographischen Daten der ostslavisches Einträge geworfen. Um auch Lesern ohne entsprechende fachliche Vorkenntnisse einen Einblick zu ermöglichen, wird hier der etwas unübliche Weg einer eher didaktisch ausgerichteten, bewusst stark vereinfachenden Darstellung gewählt, die die verwendeten Konzepte skizzenhaft und möglichst voraussetzungsarm zu illustrieren versucht. Der nachfolgende Abschnitt zum lexikographischen Prozess kann jedoch weitestgehend unabhängig von diesen Ausführungen gelesen werden.

Für die Betrachtung der Datenrepräsentation erweist es sich als zweckmäßig, zwischen der Ebene der Bearbeitung der Einträge und ihrer persistenten Speicherung (Datenhaltung) in einer Datenbank zu unterscheiden. Diese beiden Ebenen werden nun nacheinander betrachtet.

3.1 Objektorientierte Modellierung in der Redaktionsumgebung

Die Erzeugung, Bearbeitung und Prüfung der ostslavisches Einträge findet, wie in Abschnitt 4 detailliert beschrieben, überwiegend in einer am IDS für das Projekt entwickelten Software (Redaktionsumgebung) statt, die im Weiteren kurz als *Editor* bezeichnet wird. Im Zentrum der computerlexikographischen Architektur steht ein *objektorientiertes logisches Datenschema*, bei dem die zu beschreibenden Entitäten – in diesem Fall lexikographische Angaben – und ihre Beziehungen zueinander durch ein Geflecht von als *Objekten* bezeichneten Datenstrukturi-

ren abgebildet und beschrieben werden. Die Datenmodellierung der Einträge – also die Festlegung der lexikographischen Angaben mitsamt ihrer logischen Eigenschaften und ihrer Verhältnisse zueinander – erfolgt im Rahmen der insgesamt objektorientierten Programmierung des Editors. An dieser Stelle kann keine Einführung in das sehr umfangreiche informatische Themengebiet der objektorientierten Programmierung gegeben werden, das auf elementarem Niveau in allen Lehrbüchern moderner objektorientierter Programmiersprachen wie Java, Python oder C# abgehandelt wird. Es soll aber nachstehend dennoch versucht werden, interessierten Lesern einen Einblick in die Denkweise dieses Paradigmas zu geben, soweit dies für die vorliegende Übersicht von Belang ist. Aufgabe des Editors ist es aus informatischer Sicht, Benutzern über eine graphische Benutzeroberfläche die Anzeige und ggf. auch Erzeugung und Manipulation von *Objekten* zu ermöglichen. Ein Objekt ist zunächst nichts anderes als eine Ansammlung von Daten (Informationen), die sich, vereinfacht gesagt, in einem bestimmten Bereich im Arbeitsspeicher des Computers befinden. So ist beispielsweise jede Lesart eines WDLP-Artikels im Editor als ein Objekt vom Typ WDLPBedeutung repräsentiert. Objekte dieser *Klasse* – wie solche Objekttypen üblicherweise genannt werden – haben bestimmte *Attribute* oder Eigenschaften, deren Werte (Inhalte) i. A. selbst wieder Objekte sind. In der objektorientierten Sprache Groovy, die neben Java zur Implementierung des Editors verwendet wird, könnte der technisch auf das hier Wesentliche reduzierte Code zur Definition der Klasse WDLPBedeutung vereinfacht und zurechtgestellt so aussehen:

```
class WDLPBedeutung {
    String    deutscheParaphrase
    String    polnischeParaphrase
    Integer   wdlpID
    Integer   laufendeNummer
}
```

Dieser Code bringt zum Ausdruck, dass Objekte der Klasse WDLPBedeutung vier Attribute haben, nämlich eine Zeichenkette (String), die die deutsche Bedeutungsparaphrase der Lesart ist, eine weitere Zei-

chenkette für die polnische Bedeutungsparaphrase, eine Ganzzahl (Integer), die der internen Identifikationsnummer des WDLP-Lemmas entspricht, zu dem diese Lesart gehört, sowie eine weitere Ganzzahl, die der laufenden Nummer der Lesart im WDLP-Artikel entspricht. Die im Code vergebenen ‚Namen‘ der Attribute sind frei wählbar. Objekte der Klasse `WDLPBedeutung` können als Werte von Attributen anderer Objekte auftreten, zum Beispiel in folgender wiederum etwas fiktiven Definition der Klasse `EintragBedeutung`, deren Objekte Lesarten von ostslavischen Einträgen repräsentieren:

```
class EintragBedeutung {
    String                paraphrase
    List<WDLPBedeutung>  wdlpProfil
    List<Beleg>          belege
}
```

Jedes `EintragBedeutung`-Objekt verfügt demnach neben einem Zeichenkettenattribut für die deutschsprachige Bedeutungsparaphrase über zwei weitere Attribute, deren Werte *Listen* von Objekten sind: Das Attribut `wdlpProfil` repräsentiert eine für die vorliegende Darstellung vereinfachte Form eines WDLP-Profiles, ist also eine Liste von `WDLPBedeutung`-Objekten und gibt an, welche Lesarten des zugehörigen WDLP-Lemmas semantisch unter die betreffende Lesart des ostslavischen Lemmas subsumierbar sind. Die Wörterbuchbelege für das `EintragBedeutung`-Objekt sind eine Liste von Objekten der ebenfalls benutzerdefinierten Klasse `Beleg`. Objekte haben eine referentielle Identität. Auf ein und dasselbe Objekt (etwa eine konkrete `WDLPBedeutung`) kann also in verschiedenen Objekten (beispielsweise in mehreren `EintragBedeutung`-Objekten) verwiesen werden.

Eine weitere zentrale Idee der objektorientierten Programmierung ist die Möglichkeit, Klassen mit *Methoden* auszustatten. Methoden sind Blöcke von Programmcode, der jeweils speziell zum Auslesen und Verändern von Objekten dieser einen Klasse konzipiert sind. So könnte etwa die Klasse `WDLPBedeutung` eine Methode enthalten, die eine an

menschliche Benutzer adressierte „Selbstbeschreibung“ von Objekten dieser Klasse als Zeichenkette erzeugt:

```
class WDLPBedeutung {
    String    deutscheParaphrase
    String    polnischeParaphrase
    Integer   wdlpID
    Integer   laufendeNummer

    String    erzeugeBeschreibung() {
        "Lesart Nr. $laufendeNummer " +
        "von WDLP-Lemma $wdlpID: " +
        "'$deutscheParaphrase'"
    }
}
```

Wenn nun im Code ein Verweis auf ein bestimmtes Objekt der Klasse `WDLPBedeutung` vorkommt, beispielsweise mit dem ‚Namen‘ `bed`, kann einfach mit `bed.erzeugeBeschreibung()` die Selbstbeschreibung abgerufen werden, die etwa so aussieht: „Lesart Nr. 2 von WDLP-Lemma 1234: 'Gerät zur Gewichtsmessung'“. Eine solche Selbstbeschreibung kann dann an passender Stelle im Editor angezeigt werden. Eine analoge Methode für die obige Klasse `EintragBedeutung` könnte von sämtlichen `WDLPBedeutung`-Objekten, auf die ein `EintragBedeutung`-Objekt `O` mit seinem Attribut `wdlpProfil` verweist, deren Selbstbeschreibungen anfordern und diese dann als Überblick über das WDLP-Profil in die Selbstbeschreibung von `O` ‚einbauen‘. Ganz entsprechend können Methoden einer Klasse statt bloßer Zeichenketten Objekte liefern, die eine inhaltliche Validierung der lexikographischen Information repräsentieren, und die dafür wiederum entsprechende Validierungsinformationen von ‚untergeordneten‘ Objekten in ihren Attributen abrufen.

Insgesamt steht mit der objektorientierten Programmierung und Datenmodellierung ein seit Langem ausgereiftes und sehr mächtiges Werkzeug zur Verfügung, um Verhältnisse in den zu bearbeitenden Daten elegant und transparent in Programmiercode abbilden zu können.

Bei der objektorientierten Beschreibung von ostslawischen Einträgen entstehen so komplexe Geflechte von Objekten: Ein Objekt, das einen kompletten Eintrag repräsentiert, enthält Verweise auf Objekte, die für Ausdrucksvarianten, Lesarten, Derivate usw. stehen; diese verweisen wiederum z. B. auf Objekte, deren Klasse lexikographische Belege modelliert, und so fort.

3.2 XML-basierte Datenhaltung in der serverseitigen Architektur

Die konkreten Objekte, die eine Benutzerin in der Redaktionssoftware über eine graphische Benutzeroberfläche erzeugt und bearbeitet, existieren, wie bereits betont, als maschinenlesbare Datenstrukturen im Arbeitsspeicher des Rechners und würden verloren gehen, sobald das Programm beendet wird. Zur dauerhaften Sicherung müssen diese Objekte auf der Festplatte des Rechners bzw. in einer Datenbank auf einer Festplatte eines Servers abgelegt werden. Im Rahmen des hier vorgestellten Projekts wurde ein Vorgehen gewählt, bei dem der Editor eine textuelle, für Menschen lesbare textuelle Repräsentation der verwalteten lexikographischen Objekte erzeugt. Dieser Prozess wird auch als *Serialisierung* bezeichnet. Die Serialisierung eines konkreten Objekts der obigen Klasse `EintragBedeutung`, dessen WDLP-Profil zwei WDLP-Lesarten umfasst, könnte im verwendeten XML-Format ungefähr wie folgt aussehen:

```
<EintragBedeutung>
  <paraphrase>Waage</paraphrase>
  <wdlpProfil>
    <list>
      <wdlpBedeutung>
        <deutscheParaphrase>
          Gerät zur Gewichtsmessung
        </deutscheParaphrase>
        <polnischeParaphrase>
          przyrząd służący do pomiaru ciężaru
        </polnischeParaphrase>
      <wdlpID>2418</wdlpID>
    </list>
  </wdlpProfil>
</EintragBedeutung>
```

```
        <laufendeNummer>1</laufendeNummer>
    </wdlpBedeutung>
    <wdlpBedeutung> ... </wdlpBedeutung>
</list>
</wdlpProfil>
<belege> ... </belege>
</EintragBedeutung>
```

Auf das Auszeichnungsformat XML kann hier nicht detailliert eingegangen werden; vgl. als gut zugängliche Einführung etwa HOLZNER (2009). Es ist aber leicht erkennbar, dass den Objekten und ihren Attributen jeweils hierarchisch ineinandergeschachtelte ‚Textblöcke‘ (*Elemente*) im XML-Dokument entsprechen, die mit einem entsprechend benannten, in Spitzklammern notierten *Starttag* (z. B. `<polnischeParaphrase>`) eingeleitet und mit einem korrespondierenden *Endtag* mit zusätzlichem Schrägstrich (`</polnischeParaphrase>`) abgeschlossen werden. XML wird als in der Computerlexikographie am weitesten verbreitetes Datenaustauschformat beispielsweise auch im *Lehnwortportal Deutsch* verwendet. Je nach Betriebsmodus des Editors (s. u., Abschnitt 4.1) werden die XML-Dateien in einem komprimierten Format auf der Festplatte des Rechners oder aber in der zentralen Datenbank abgelegt. Die zentrale Datenbank (Oracle) ist relational, d. h. als System von Tabellen, organisiert. In der Tabelle der ostslavischen Einträge korrespondiert jede Tabellenzeile einem Eintrag. Eine Zelle in einer solchen Zeile enthält den Text des XML-Dokuments, weitere Zellen enthalten zu Zwecken schneller Suchen Informationen, die aus dem XML-Dokument extrahiert werden, wie etwa das zugeordnete WDLP-Lemma, sowie Verwaltungs- und Statusinformationen zum Eintrag, insbesondere dazu, ob der Eintrag aktuell von einer bestimmten Benutzerin bearbeitet und daher für andere gesperrt ist.

Bei der oben beispielhaft gezeigten Form der XML-Serialisierung gehen die referentiellen Identitäten von Objekten verloren. So ist nicht zu erkennen, ob eine textuelle Gleichheit zweier XML-Elemente in einem Dokument zufällig ist, d. h. ob es sich um zwei verschiedene Objekte mit gleichen Attributwerten handelt oder ob wirklich an zwei Stellen dasselbe Objekt referenziert wird. Dieses Problem wird durch

Einführung von sogenannten ID-Attributen gelöst: Ein als `<wdlpBedeutung id="9876">...</wdlpBedeutung>` serialisiertes Objekt kann an anderer Stelle desselben Dokuments als `<wdlpBedeutung ref="9876"></wdlpBedeutung>` wiederholt werden – also ohne Textmaterial zwischen Start- und Endtag –, um Referenzidentität auszudrücken. Dieses Verfahren kommt auch mit sogenannten zyklischen Referenzen zurecht, bei denen zwei Objekte gegenseitig aufeinander verweisen. Ein gravierender praktischer Nachteil dieses Verfahrens ist, dass Elemente innerhalb eines XML-Dokuments, wenn sie ‚Wiederholungen‘ der gezeigten Art, also Elemente mit einem `ref`-Attribut, enthalten, nicht mehr unabhängig interpretiert werden können und somit auch bei Anpassung der verwendeten IDs nicht ohne weiteres in andere XML-Dokumente hineinkopiert werden dürfen.

Das Gegenstück zur Serialisierung ist die *Deserialisierung*: Ruft eine Lexikographin im Editor einen bestimmten Eintrag zur Ansicht oder Bearbeitung auf, so wird aus der Datenbank oder von der lokalen Festplatte das entsprechende XML-Dokument geholt und ein den Angaben im Dokument genau entsprechendes Objekt im Arbeitsspeicher erzeugt.

3.3 Graphenbasierte Datenrepräsentation im *Lehnwortportal Deutsch*

Die durch XML-Serialisierung von Editor-Eintragsobjekten gewonnenen Dokumente, die Artikeln in den Zielwörterbüchern entsprechen, dienen später im *Lehnwortportal Deutsch* als Basis für die Anzeige der Artikel. Dazu kann beispielsweise die im Portal auch sonst benutzte Standardtechnologie XSLT (vgl. HOLZNER 2009) verwendet werden, die auf sehr effektive Weise XML-Dokumente u. a. in HTML-Dokumente, also Webseiten, ‚übersetzen‘ kann. Für alle weiteren Funktionalitäten wird jedoch im Portal eine völlig andere, wörterbuchübergreifend eingesetzte Datenstruktur verwendet, nämlich ein *gerichteter azyklischer Graph* (vgl. MEYER/ENGELBERG 2011 sowie MEYER 2013b). Diese Datenstruktur ermöglicht insbesondere sehr effiziente und granulare portalweite Suchmöglichkeiten (vgl. MEYER 2013a). Graphen-

orientierte Technologien werden mittlerweile in verschiedenen Projekten im Umfeld der Lexikographie eingesetzt (siehe z. B. SPOHR 2012, POLGUÈRE 2009, TRIPPEL 2006). Ein Graph ist ganz allgemein eine Menge von Objekten (*Knoten*), zwischen denen Beziehungen (*Kanten*) bestehen können. Im speziellen Fall des Lehnwortportals sind die Knoten einfach die Wörter, die in den verschiedenen lexikographischen Ressourcen des Lehnwortportals gebucht sind, also deutsche Etyma ebenso wie Lehnwörter mitsamt deren ausdrucksseitigen Varianten, Derivaten, Komposita usw. Zwischen bestimmten gebuchten Wörtern besteht jeweils paarweise eine Beziehung; an einem konkreten, leicht vereinfachten Beispiel illustriert:

- Neuhochdeutsch *Mörder* **hat als diasystematische** (hier: diachrone) **Variante** das mittelhochdeutsche *mordaere* ‘Mörder’.
- Mittelhochdeutsch *mordaere* **ist Etymon zum** polnischen **Lehnwort** *morderz* ‘Mörder’.
- Polnisch *morderz* **bildet ein Derivat** *morderca* ‘dasselbe’.
- Polnisch *morderca* **ist Etymon des russischen Lehnwortes** *morderca*.

Die vorstehend genannten Informationen lassen sich auf naheliegende Weise graphisch veranschaulichen, indem man die gebuchten Wörter z. B. als kleine Kreise und die paarweisen Beziehungen durch Pfeile von einem Kreis zu einem anderen symbolisiert. Diese Darstellungsweise, die im Lehnwortportal für jedes deutsche Etymon abrufbar ist (vgl. Abb. 3), motiviert auch die Terminologie von Knoten und Kanten. Die in der vorstehenden Aufzählung fett gesetzten Typen von Beziehungen sind gerichtet (asymmetrisch), es ist also z. B. *mordaere* Etymon zu *morderz*, aber natürlich nicht umgekehrt. Daher müssen in der Visualisierung eines solchen gerichteten Graphen die Kanten sinnvollerweise wirklich als Pfeile, nicht einfach nur als verbindende Linien, dargestellt werden. Für Suchabfragen sind in der Regel nicht nur die direkten Verbindungen zwischen Wörtern relevant, sondern auch indirekte. So gelangt man von neuhochdeutsch *Mörder* zu russisch *morderca*, indem man entsprechend der obigen Aufzählung eine Abfolge von insgesamt vier aneinanderhängenden Pfeilverbindungen, einen *Pfad* im

In der relationalen Datenbank des Lehnwortportals wird der Graph in zwei Tabellen gespeichert: In der *Knotentabelle* repräsentiert jede Zeile einen Knoten (ein gebuchtes Wort) mitsamt dessen Eigenschaften, zu denen auch jeweils eine den Knoten eindeutig identifizierende ID-Nummer gehört. In der *Kantentabelle* entspricht jede Zeile einer Kante, also einer Beziehung, mitsamt Angaben zu ihrem Typ. Eine Spalte der Tabelle enthält dabei jeweils die ID des ‚Ausgangswortes‘ der gerichteten Beziehung, eine weitere die ID des ‚Zielwortes‘.

Der oben genannte Pfad erstreckt sich im Graphen des Portals über Wörter aus drei verschiedenen lexikographischen Ressourcen:

1. Die neuhochdeutsche Form *Mörder*, die im WDLP-Artikel selbst gar nicht als Etymon genannt wird, ist ein im Rahmen der Portalerstellung definiertes deutsches *Metalemma* (typischerweise ein neuhochdeutsches standardsprachliches Wort), dem etymologisch möglichst genau entsprechende diasystematische Varianten, so wie sie in den verschiedenen Wörterbüchern als Etyma gebucht sind, zugeordnet sind. Im vorliegenden Fall ist dem Metalemma *Mörder* auch die im slovenischen Lehnwörterbuch des Portals gebuchte mittelhochdeutsche Schreibvariante *mordære* zugeordnet, zu der im selben Lehnwörterbuch noch zwei diasystematische Varianten *mörder* und *mordere* aufgeführt sind. Die deutschen Metalemmata ermöglichen so eine wörterbuchübergreifende Vernetzung von etymologisch zusammengehörigen Wortformen und dienen als Lemmata des ‚umgekehrten Lehnwörterbuchs‘ (ENGELBERG 2010) des Portals, das es Benutzern ermöglicht, ausgehend von (nach Möglichkeit) aus dem heutigen Deutsch bekannten Wörtern nach zugehörigen Entlehnungen in den im Portal vertretenen Sprachen zu suchen.

2. Die Graphenkanten *mordaere* → *morderz* und *morderz* → *morderca* werden nach einem in MEYER (2014b) detailliert beschriebenen Verfahren aus dem XML-Dokument des WDLP-Artikels gewonnen.

3. Die Kante poln. *morderca* → russ. *morderca* wird aus den XML-Daten des hier präsentierten Projekts extrahiert werden. Entlehnungsketten über beliebig viele Sprachen hinweg lassen sich direkt und sehr anschaulich als Pfade in einem gerichteten Graphen modellieren und so

für den Benutzer auch auffindbar machen (vgl. Abschnitt 5). Dies gilt umso mehr, wenn es mehrere verschiedene aufgrund der Beleglage mögliche Entlehnungswege gibt, so wie dies bei polnischen Entlehnungen im Ostslavischen häufig der Fall ist.

4. Der lexikographische Prozess des Projekts aus computerlexikographischer Sicht

Dieser Abschnitt erörtert die einzelnen Phasen des in Abschnitt 2 überblicksartig skizzierten lexikographischen Prozesses und legt dabei einen Schwerpunkt auf die Darstellung der verteilten IT-Infrastruktur, die für diesen Prozess im Rahmen des Projekts entwickelt wurde und noch wird. Die in Abb. 2 dargestellten Phasen der Redaktionsarbeit werden für jeden WDLP-Eintrag und grundsätzlich sogar – bis auf die abschließende Erstellung der eintrags- und wörterbuchübergreifenden Kommentare – für jeden einzelsprachlichen Zielwörterbucheintrag separat durchlaufen.

4.1 Exzerpierungsphase

Für die Erstellung und Bearbeitung sowohl der Quellwörterbuchexzerpte als auch der Zielwörterbuchartikel wird im Projekt, wie bereits erläutert, eine eigens entwickelte kollaborative Softwareplattform verwendet, in deren Zentrum aus Benutzersicht die Redaktionsumgebung (Editor) steht.

Abb. 4 zeigt einen Screenshot des Hauptfensters der vollständig zweisprachig (deutsch und russisch) verfügbaren Editor-Benutzeroberfläche. Ausgangspunkt für jeden lexikographischen Arbeitsgang ist die Liste aller WDLP-Lemmata (1). Nach Auswählen (Anklicken) eines WDLP-Lemmas erscheint im Textbereich (2) eine Ansicht des zugehörigen WDLP-Artikels und in einer weiteren Listendarstellung (3) eine Übersicht über die bisher angelegten ostslavischen Einträge zu diesem Lemma. Vergrößert ist ein Beispiel für diese Übersicht in Abb. 5 zu sehen. Für jeden Eintrag werden in der Liste (3) unterhalb des Lemmas,

jeweils mit entsprechender Überschrift versehen, die bislang angelegten Varianten, Lesarten und Ableitungen (Derivate, Komposita usw.) angezeigt; erst in der unten besprochenen Endredaktion werden auch Konkurrenten zu den Lehnwörtern angelegt. Für jede Ableitung kann wiederum eine beliebige Anzahl von Ausdrucksvarianten und Lesarten im Editor angelegt werden. Alle genannten Einträge dieser hierarchisch-baumartig strukturierten Liste, einschließlich der Überschriften, werden im Folgenden als *Buchungsobjekte* bezeichnet. Sie korrespondieren aus lexikographischer Perspektive mit Abschnitten und Unterabschnitten (Angabebereichen) der Artikelmikrostruktur, aus Sicht der Implementierung in der Tat mit Objekten im Sinne der in Abschnitt 3 skizzierten objektorientierten Modellierung.

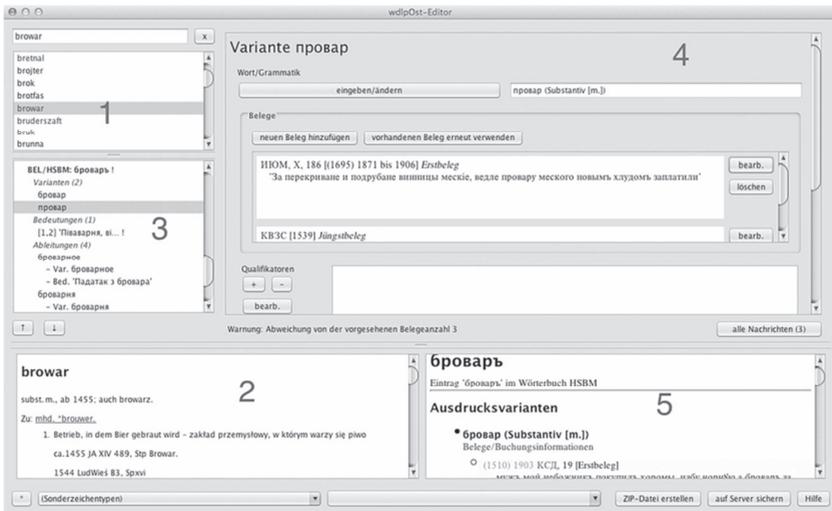


Abb. 4: Screenshot: Benutzeroberfläche des Editors

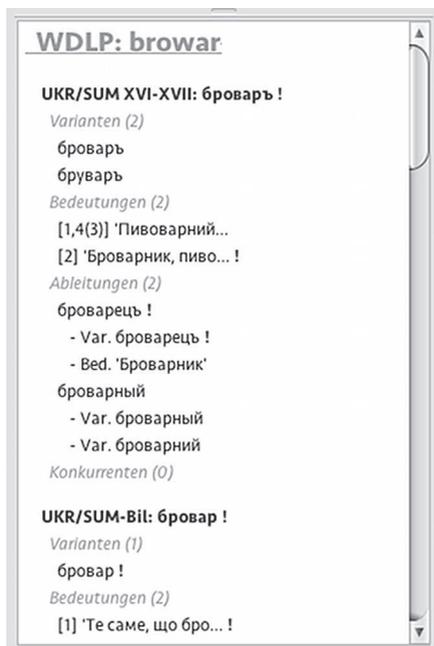


Abb. 5: Liste der ostslavischen Einträge und ihrer zugehörigen Buchungsobjekte (Abschnitte und Unterabschnitte der Mikrostruktur)

Wählt die Benutzerin nun ein Buchungsobjekt aus der Liste (3), beispielsweise eine bestimmte Lesart, durch Klicken aus, erscheint im Haupteingabebereich (4) des Editors ein im Folgenden als *Buchungs-panel* bezeichnetes Feld mit allen zugehörigen Eingabeoptionen, die je nach Betriebsmodus des Editors unterschiedlich umfangreich ausfallen können. So haben in der Exzerpierungsphase die Karteien etwas reduzierte Eingabemöglichkeiten als die Redakteure in Oldenburg; noch zahlreicher sind die Optionen in der Phase der Endredaktion. Im Textbereich (5) wird im Editor eine stets aktuelle Vorschau des gesamten Artikels gegeben. Der Editor verfügt über eine umfangreiche Validierungssystematik; separat in jedem Buchungspanel werden Fehler (insbesondere fehlende oder inkonsistente Informationen) und potentielle Probleme des zugehörigen Buchungsobjekts ausgewiesen. Der Validie-

rungsstatus eines Buchungsobjekts wird auch in der Buchungsliste (3) durch farbige Symbole neben dem zugehörigen Eintrag markiert. Erst dann, wenn das Validierungssystem keine Fehler mehr meldet, erlaubt es der Editor der Benutzerin, einen Eintrag als „abgeschlossen“ zu markieren. Dieser Status kann, außer von Administratoren, nicht mehr rückgängig gemacht werden und verhindert jegliche weitere Veränderung an den Eintragsdaten. Insbesondere dienen die „abgeschlossenen“ Exzerpte als Grundlage zur Erstellung von Zielwörterbuchartikeln; hier würde ein nachträgliches Ändern der Exzerptdaten unweigerlich zu Dateninkonsistenzen führen.

Jede im Editor vorgenommene Änderung wird automatisch durch eine Serialisierung des angezeigten Datenbestandes dauerhaft gespeichert, so wie in Abschnitt 3.2 kurz erläutert: Die lexikographischen Daten des Eintrags werden als für Menschen lesbare XML-Textdokumente kodiert. Der Editor kennt, wie schon erwähnt, zwei grundsätzlich verschiedene Betriebsmodi. Im kollaborativen *Netz-Betriebsmodus*, der am IDS und in der lexikographischen Redaktion in Oldenburg genutzt wird, werden die XML-Dokumente in der zentralen Datenbank gespeichert (vgl. ebenfalls Abschnitt 3.2 oben). Der Editor kommuniziert zu diesem Zweck über das Internet mit einer auf dem Mannheimer Server laufenden Anwendung, einem *Webservice*, der mit der Datenbank verbunden ist und die jeweils benötigten XML-Dokumente von dort abrufen bzw. von Editoren gelieferte geänderte Fassungen prüft und wieder in die Datenbank zurückschreibt. Bei diesen Vorgängen wird das betreffende Dokument jeweils durch den aktuellen Bearbeiter für die Bearbeitung durch andere gesperrt.

Vor allem für die Arbeitsplätze in den osteuropäischen Karteien ist der *lokale Betriebsmodus* des Editors konzipiert, der sämtliche XML-Dokumente in komprimierter Form auf der Festplatte des jeweiligen Rechners ablegt und daher keine Internetanbindung erfordert. In diesem Modus sieht der Benutzer nur die von ihm selbst bearbeiteten Einträge (Exzerpte) aus seiner Kartei. Es kann jederzeit per Knopfdruck eine Sicherungskopie der lokal abgelegten Daten in einer gesonderten Tabelle der Datenbank abgelegt werden, sofern der betreffende Rechner

zumindest zeitweise mit dem Internet verbunden werden kann. Andernfalls kann eine ZIP-Datei mit allen lokal gespeicherten XML-Dokumenten erzeugt werden, die von einem anderen Rechner aus als E-Mail-Anhang an die Redaktion gesendet und dort in die besagte gesonderte Tabelle der Datenbank eingefügt wird. – Einträge aus den Karteien, die von den dortigen Bearbeitern als „abgeschlossen“ markiert wurden und daher in den Karteien nicht mehr geändert werden können, werden automatisch aus der gesonderten Tabelle in die zentrale Datenbanktabelle übernommen und sind dann auf allen im Netz-Betriebsmodus laufenden Systemen sichtbar.

Abschließend soll noch auf einige speziellere Funktionen des Editors eingegangen werden. Ausdrucksvarianten und Lesarten können mit beliebig vielen *Belegen* verknüpft werden. Belege umfassen immer eine Datierung sowie optional und je nach lexikographischer Anforderung auch eine Quellen- und Belegstellenangabe und ggf. ein Zitat. Die Eingabe von Belegen erfolgt in einem eigenen Dialogfenster. Für Datierungen wird von dort aus wiederum ein separater Eingabedialog aufgerufen, der bei nur ungefähr bekannter Datierung statt exakter Jahresangaben auch Datumstypen wie „erstes Drittel des *x*.ten Jahrhunderts“ akzeptiert und zwischen Entstehungs- und Publikationsdatum von Quellen zu unterscheiden ermöglicht. Quellenbezeichnungen mit Quellenkürzel und zugehöriger Datierung werden automatisch, für jedes exzerpierte Wörterbuch getrennt, gesondert gespeichert, im regulären Netzbetrieb des Editors in der zentralen Datenbank; per Mausclick kann eine bereits früher eingegebene Quelle dann bequem für eine neue Belegeingabe übernommen werden.

In bestimmten Fällen muss derselbe Beleg mehrfach verwendet werden, so etwa bei der einzigen Ausdrucksvariante und der einzigen Lesart eines Hapaxlegomenons. Auch dies ist (unter Ausnutzung der oben in Abschnitt 3.1 erläuterten referentiellen Identität von Objekten) im Editor möglich; spätere Änderungen an einem solchen Beleg erscheinen dann automatisch bei allen Verwendungsinstanzen dieses Belegs.

Bearbeiter können in der Exzerpierungsphase nach Belieben neue Exzerpte im Editor anlegen und nach und nach ausfüllen. Zu Beginn des Projekts wurde während der Implementierung der IT-Infrastruktur jedoch sowohl in den Karteien als auch in Oldenburg eine „Vorexzerpierung“ mittels eines Tabellenkalkulationsprogramms durchgeführt, bei der für jedes WDLP-Lemma in einer einfachen Tabellenstruktur mögliche Entsprechungen (Ausdrucksvarianten, Derivate, zum Teil auch Lesartenangaben) in den verschiedenen Wörterbüchern notiert wurden. Die Vorexzerpierung diente auch der Abschätzung des zu erwartenden Arbeitsaufwandes insbesondere in den Karteien. Über mehrere Zwischenschritte (bereinigte und systematisierte Fassung in einer MS-Access-Datenbank, exportiert als XML) konnten die Ergebnisse der Vorexzerpierung in die zentrale Datenbank übernommen bzw. als „Erstversionen“ der karteienspezifischen lokalen Datenhaltung verwendet werden. In der eigentlichen, editorgestützten und sehr viel detaillierteren Exzerpierung konnte so von Anfang an mit bereits angelegten, aber eben nur zum Teil ausgefüllten Einträgen gearbeitet werden.

Abschließend sei betont, dass der Editor keinen bestimmten sequentiellen Arbeitsstil erzwingt: Alle lexikographischen Angaben der Einträge lassen sich in beliebiger Reihenfolge angeben und später jederzeit noch weiter ändern.

4.2 Endredaktionsphase

Aufgabe der in Oldenburg zu leistenden Endredaktion ist es, für jedes WDLP-Lemma und jede der drei ostslavischen Sprachen aus womöglich zahlreichen zugehörigen Wörterbuchexzerpten die entsprechenden Artikel der drei Zielwörterbücher zu erzeugen. Die im Editor verfügbaren Werkzeuge für diese Aufgabe befinden sich zwar derzeit noch in der Entwicklung, sind jedoch von der computerlexikographischen Gesamtkonzeption her von zentraler Bedeutung für das Projekt. Nachfolgend eine skizzenartige Aufzählung der Prozesse zur Erzeugung eines Zielwörterbuchartikels für eine der drei ostslavischen Sprachen, der einem bestimmten WDLP-Artikel zugeordnet ist:

- Die Lexikographin wählt den WDLP-Artikel aus, legt eine ostslavische Lemmaform für den neuen Artikel fest und wählt in einem Dialog aus den als „abgeschlossen“ markierten Exzerpten der betreffenden Sprache, die dem betreffenden WDLP-Artikel zugeordnet sind, die lexikographisch passenden aus. Die Lemmatisierung der Zielartikel ist, wie bereits betont, völlig unabhängig von den Lemmatisierungen der Quellwörterbücher definierbar.
- Das System erzeugt automatisch einen neuen Zielwörterbuchartikel, in dem zunächst sämtliche Varianten, Lesarten und Ableitungen der ausgewählten Exzerpte als voneinander verschiedene Varianten, Lesarten und Ableitungen aufgelistet sind. Bei diesem Prozess müssen bereits die allgemeinen lemmabezogenen lexikographischen Angaben (die *Artikelköpfe*) der Einzelexzerpte im Artikelkopf des Zielartikels zusammengeführt werden. Textuelle Angaben werden einfach – mit Verweisen auf den Herkunftsartikel – aneinandergesetzt und müssen später manuell überarbeitet und im Falle von exzerpierten slavischsprachigen Angaben ggf. auch ins Deutsche übersetzt werden. Bei anderen Angabetypen können sich die Informationen aus verschiedenen Wörterbüchern aber auch formal widersprechen. In solchen Fällen wird die Benutzerin um eine Auswahl aus den konfligierenden Möglichkeiten gebeten oder eine Defaultangabe eingefügt sowie gleichzeitig ein interner Kommentar erzeugt, der auf das Problem hinweist.
- Wesentliches Ziel der Endredaktion ist die Beseitigung von Redundanz bei den einzelnen Angabetypen. Viele Ausdrucksvarianten werden typischerweise in einer größeren Anzahl von Quellwörterbüchern belegt sein. Im Normalfall wird man homographie Varianten aus verschiedenen Exzerpten miteinander identifizieren wollen, d. h. sie sollen nicht als separate Varianten im Zielartikel erscheinen, sondern als eine einzige Variante mit Buchungen in mehreren Quellwörterbüchern. Für diesen Identifizierungsvorgang ist ein separates Bearbeitungsfenster zuständig, das sämtliche vorgefundenen, graphematisch unterschiedlichen Varianten auflistet und zu jeder Variante die vorgefundenen Buchungen präsentiert. Der Lexi-

kograph kann diese Listen manuell verändern. Insbesondere ist es möglich, die Buchungen von rein orthographischen Varianten zusammenzufassen oder auch einzelne Buchungen durch bloßes Verschieben mit der Maus einer anderen oder auch einer manuell neu angelegten Variante zuzuordnen.

- Analog verläuft der Prozess der Konsolidierung von exzerpierten Bedeutungsangaben. Auch hier kann das System zunächst eine automatische Zusammenfassung aufgrund der WDLP-Profile der exzerpierten Lesarten erzeugen: Exzerpierte Lesarten mit identischen Profilen werden zu einer einzigen Lesart im Zielartikel zusammengefasst. Unter bestimmten Bedingungen müssen exzerpierte Lesarten jedoch auch in mehrere unterschiedene ‚Ziellesarten‘ zergliedert werden, insbesondere häufig dann, wenn ihr WDLP-Profil mehrere WDLP-Lesarten umfasst. Da die Metasprache der Zielartikel das Deutsche ist, muss die lexikographische Bearbeiterin die so erzeugten und wiederum ggf. durch manuelle Umgruppierung modifizierten Lesarten selbst mit einer deutschen Übersetzung versehen.
- Ableitungen verfügen über je eigene Listen an Varianten und Lesartenangaben. Hier ist die Konsolidierung entsprechend komplexer, da außer der Identifizierung von ausdrucks gleichen Varianten und profildeichen Bedeutungen auch noch die Zusammenführung der Informationen aus den varianten- und lesartenübergreifenden Angaben, den *Ableitungsköpfen*, durchgeführt werden muss.
- In allen Fällen der Identifizierung von Varianten und Lesarten müssen auch die zugehörigen Beleglisten konsolidiert, also zusammengeführt und reduziert werden. Der Editor kann, insbesondere aufgrund von Standardvorgaben für Anzahl und zeitliche Eigenschaften von Belegen, Vorschläge für das Streichen von Belegen machen, die die Bearbeiterin annehmen oder eigenständig modifizieren kann.

Die fertig endredigierten Artikel können direkt als XML-Dokumente in die Datenbank des Lehnwortportals übernommen werden. Um jedoch die Struktur der zugehörigen Teilgraphen korrekt ermitteln zu können,

ist ein weiterer Arbeitsschritt erforderlich, da zwar die Beziehungen der in den Artikeln gebuchten Wörter untereinander, nicht aber die möglichen Entlehnungswege, also die Beziehungen zu Wörtern aus je anderen ostslawischen Sprachen, dem Polnischen bzw. dem Deutschen, aus den lexikographischen Angaben der Zielwörterbuchartikel ableitbar sind. Zu allen korrespondierenden Zielartikeln, also Artikeln, die demselben WDLP-Artikel zugeordnet sind, gehört jeweils ein artikelübergreifender sprachhistorischer Kommentar, der nach Abschluss der Endredaktion der Einzelartikel als diskursiver und formatierter Text mit Verweisen auf Buchungen in den zugehörigen Einzelartikeln verfasst wird und die möglichen Entlehnungswege von Wörtern diskutiert. Die lexikographischen Bearbeiter erhalten außerdem in einem eigenen Bearbeitungsfenster die Möglichkeit, direkt anhand einer visuellen Darstellung des entsprechenden Teilgraphen passend zum Kommentar die möglichen Entlehnungswege als Kanten mit Gewichtungen (Entlehnungsweg sicher, wahrscheinlich, möglich oder unplausibel) interaktiv in die Graphendarstellung einzutragen. Die so definierten Kanten werden in separaten XML-Dokumenten abgespeichert und dienen anschließend neben den Zielartikel-Dokumenten als weitere Datenquelle zur Erzeugung des Gesamtgraphen in der Datenbank (zu den Einzelheiten der Graphenerzeugung vgl. MEYER 2014b). Im Gesamtgraphen werden fortgeschrittene Nutzer des Lehnwortportals beliebig komplexe Suchen auch beispielsweise nach Lehnwörtern mit bestimmten Entlehnungswegen durchführen können (vgl. hierzu MEYER 2014a und den folgenden Abschnitt 5.1).

5. Abschließende Bemerkungen

5.1 Recherchemöglichkeiten zu Entlehnungsketten

Das *Lehnwortportal Deutsch* verfolgt das langfristige Ziel, mit seiner feinkörnigen und ressourcenübergreifenden graphenbasierten Datenhaltung und verschiedenen darauf abgestimmten Suchwerkzeugen der Sprachkontaktforschung ein leistungsfähiges Instrument zur Verfügung

zu stellen. Der vorliegende Beitrag hat eine computerlexikographische Infrastruktur zur Erstellung von spezialisierten Portalwörterbüchern zu indirekten ostslavischen Entlehnungen aus dem Deutschen vorgestellt. Um einen Eindruck von den zukünftigen Nutzungsmöglichkeiten des Portals hinsichtlich solcher Entlehnungsketten zu vermitteln, sei an dieser Stelle am konkreten, derzeit aber noch fiktiven, Beispiel vorgeführt, wie eine etwas komplexere Suchanfrage zu Entlehnungsketten im Portal aussehen wird. Für wissenschaftliche Zwecke stellt das Lehnwortportal bereits jetzt eine *domänenspezifische*, also speziell auf den konkreten fachlichen Anwendungsfall hin konzipierte, an das Deutsche angelehnte Abfragesprache bereit, in die sich auch informatisch nicht einschlägig vorgebildete Nutzer einarbeiten können (<http://lwp.ids-mannheim.de/search/prof>). In dieser Abfragesprache lassen sich konkrete Konfigurationen im Graphen des Portals beschreiben, nach denen die Nutzerin sucht. Angenommen nun, es werden in den Portalressourcen vorkommende Fälle gesucht, in denen

- ein auf *-aere* oder *-ære* endendes mittelhochdeutsches Substantiv **DE**
- ins Polnische entlehnt wurde, wobei dieser polnische Germanismus **POL**
- eine Ableitung **POL_DER** auf *-ka* oder *-ca* aufweist,
- die direkt oder indirekt (z. B. über das Ukrainische) ins Russische entlehnt wurde (Lehnwort **RUS**).

Diese Suchkriterien lassen sich *deklarativ* in der Abfragesprache ausdrücken; vereinfacht gesagt bedeutet dies, dass man dem System nur mitteilen muss, wonach es suchen soll, nicht aber, wie es diesen Suchvorgang durchführen muss. Die Abfrage besteht aus einzelnen, mit einem Punkt endenden Suchklauseln. Zunächst werden die zu suchenden Wörter, also Knoten im Graphen, mit frei wählbaren Namen versehen und gleichzeitig als Metalemma, Etymon oder Lehnwort klassifiziert, ähnlich, wie dies oben geschehen ist:

```
suche etymon DE.
suche lehnwort POL.
suche lehnwort POL_DER.
suche lehnwort RUS.
```

Hier ist zu bemerken, dass auch Varianten, Ableitungen usw. von Lehnwörtern als ‚lehnwort‘ klassifiziert werden. – Anschließend spezifiziert man Eigenschaften dieser Graphenknoten bzw. Wörter:

die sprache von **DE** ist 'mittelhochdeutsch'.

DE ist substantiv.

DE endet auf 'aere' oder **DE** endet auf 'ære'.

die sprache von **POL** ist 'polnisch'.

die sprache von **POL_DER** ist 'polnisch'.

POL_DER endet auf 'ka' oder **POL_DER** endet auf 'ca'.

die sprache von **RUS** ist 'russisch'.

Schließlich werden die gewünschten Beziehungen zwischen den Wörtern, also die Graphenkanten oder -pfade und deren Eigenschaften, spezifiziert:

POL ist lehnwort zu **DE**.

POL_DER ist derivat zu **POL**.

POL_DER ist vorgaenger zu **RUS**.

Die ersten beiden Klauseln spezifizieren Kanten, während die letzte einen beliebig langen Pfad im Graphen beschreibt, der von **POL_DER** zu **RUS** führt.

Suchresultate sind allgemein sämtliche geordneten Tupel von Graphknoten, für die alle Bedingungen der Suchanfrage erfüllt sind, im vorliegenden Fall also beispielsweise folgendes Wortquadrupel:

DE: Etymon *mordaere* im Artikel *morderz* im WDLP;

POL: Lehnwortvariante *morderz* im Artikel *morderz* im WDLP;

POL_DER: Ableitung *morderca* im Artikel *morderz* im WDLP;

RUS: Lehnwort *morderca* im russischen Zielwörterbuch des Projekts.

Es ist vorgesehen, künftig nicht nur die geordneten Tupel von Wörtern mit Links zu den zugehörigen Wörterbucheinträgen als Suchresultate anzubieten, sondern auch zu jedem Tupel den entsprechenden Ausschnitt aus dem Gesamtgraphen darzustellen, der im Falle von mehreren möglichen Entlehnungswegen sämtliche dieser Wege als Pfade im Teilgraphen repräsentiert.

Beschränkungen solch komplexer Suchoptionen ergeben sich in erster Linie aus Einschränkungen hinsichtlich der in den einzelnen Portalressourcen überhaupt verfügbaren lexikographischen Angaben. So können Suchen, die als Kriterium auch Angaben zum Entlehnungszeitraum enthalten, natürlich nur über Ressourcen durchgeführt werden, die solche Angaben auch enthalten.

5.2 Kosten und Nutzen der Neuimplementierung eines lexikographischen Redaktionssystems

Der hohe Aufwand, der erforderlich ist, um die hier beschriebene Infrastruktur von Grund auf und auf die Erfordernisse des Projekts zugeschnitten zu konzipieren und zu entwickeln, bedarf der Rechtfertigung. So wird in DE SCHRYVER (2011, 647) im Zusammenhang mit der projektspezifischen Entwicklung lexikographischer Software treffend von einem permanenten „reinventing the wheel“ gesprochen. In der Tat dürfte in sehr vielen Fällen die Anschaffung einer langjährig entwickelten und erprobten kommerziellen Lösung erheblich günstiger und auch in Hinblick auf den bei technischen Problemen erforderlichen Support sinnvoller sein als die Neuentwicklung eines Systems, bei der viele anderswo längst befriedigend gelöste Probleme noch einmal von vorne angegangen werden müssen. Das vorgestellte Projekt unterscheidet sich jedoch erheblich von der Entwicklung eines konventionellen ein- oder zweisprachigen Wörterbuchs, insbesondere durch den mehrstufigen Redaktionsprozess: Die im ersten Schritt erstellten ostslavischen Exzerpte sind selbst bereits neu erstellte Wörterbuchartikel, die unter komplexer Bezugnahme auf ein bereits existierendes digitales Wörterbuch (das WDLP) erstellt werden und dann in einem inhaltlich anspruchsvollen, softwareunterstützten Prozess in wiederum neue, von der Mikrostruktur her etwas komplexere Artikel von drei parallel erarbeiteten und miteinander auch noch durch übergreifende Kommentare vernetzten Wörterbüchern zusammengeführt werden müssen. Hinzu kommen Anforderungen wie die zentrale Quellenverwaltung und die Notwendigkeit, artikelintern Belege mehrfach verwenden zu können.

Leistungsfähige kommerzielle lexikographische Redaktionssysteme wie das von DE SCHRYVER näher besprochene *TLex* bieten zwar häufig flexible Anpassungsmöglichkeiten durch die Verwendung einer spezifischen Programmiersprache. Damit können zwar auf einfache Weise zusätzliche lexikographische Prozesse definiert und automatisiert werden, die Umsetzung grundsätzlich abweichender Anforderungen und fachlicher Abläufe ist jedoch problematisch bis unmöglich bzw. führt dazu, dass bestimmte im System vorgesehene Datenstrukturen und Bedienungsabläufe gewissermaßen missbräuchlich verwendet werden müssen, um das gewünschte Ziel zu erreichen. Häufig ist es schwierig, im Vorhinein hinreichend verlässlich zu evaluieren, ob ein solches Problem auftreten wird, da Detailprobleme in der Regel erst in der konkreten Implementierungsarbeit zu Tage treten.

Eine interessante und auch in einigen größeren und anspruchsvolleren Projekten genutzte Alternative ist die Verwendung einer sog. Wiki-Software, wie sie von den Projekten der Wikimedia Foundation (*Wikipedia*, *Wiktionary*) bekannt ist. Wikisysteme stehen in zahlreichen, oft quelloffenen Varianten zur Verfügung und bieten mittlerweile auch Zusatzmodule für spezifische lexikographische Anforderungen (vgl. etwa BON/NOWAK 2013). Beschränkungen ergeben sich hier jedoch aus den vom jeweiligen System unterstützten Datenmodellierungen sowie aus der Schwierigkeit, solche Systeme, auch wenn der Quellcode verfügbar ist, um komplexe eigene Werkzeuge zu erweitern. Das im hier vorgestellten Projekt erforderliche Nebeneinander von lokalem und netzgebundenem Betriebsmodus dürfte mit den üblichen Wikisystemen auch nur unter erheblichen Schwierigkeiten umsetzbar sein.

Eine gerade auch im akademischen Bereich viel genutzte alternative Vorgehensweise ist die direkte Bearbeitung von Wörterbucheinträgen in einem handelsüblichen XML-Editor. In diesem Fall ist XML also das primäre Datenformat zur Bearbeitung von lexikographischem Material. Ein großer Vorteil ist hierbei der Umstand, dass für die manuelle wie automatische Be- und Verarbeitung von XML-Dokumenten eine große Zahl von bewährten informatischen Standardwerkzeugen zur Verfügung steht.

Vor allem die oben in Abschnitt 4.2 beschriebenen Prozesse der Weiterverarbeitung und Zusammenführung von Exzerpten im Rahmen der Endredaktion würden beim direkten Operieren mit XML-Dokumenten jedoch komplexe und sehr unübersichtliche Arbeitsabläufe des Kopierens von Fragmenten aus verschiedenen XML-Dokumenten in jeweils ein und dasselbe Zielartikel-Dokument erforderlich machen. Der Aufwand, um solche Abläufe durch Entwicklung zusätzlicher Software (z. B. sogenannter Plugins für XML-Editoren) in einem derartigen Fall handhabbar zu machen und in einem kollaborativen, servergestützten Umfeld durch umfangreiche Konsistenzprüfungen gegen drohende Bedienungsfehler abzusichern, ist erheblich, insbesondere angesichts der praktischen Einschränkungen, die eine rein XML-basierte Datenmodellierung gegenüber der sehr viel abstrakteren und flexibleren objektorientierten Modellierung aufweist.

5.3 Perspektiven: Erweiterbarkeit und Anpassbarkeit des Redaktionssystems

Eine grundsätzliche Herausforderung für jedes größere lexikographische Projekt ist die Notwendigkeit, eine bereits definierte Datenmodellierung bzw. Artikelmikrostruktur vor dem Hintergrund gesammelter Erfahrungen oder neuer Anforderungen revidieren zu müssen. Alle in Abschnitt 5.2 diskutierten technischen Alternativen für Redaktionssysteme haben spezifische Beschränkungen hinsichtlich der Anpassbarkeit von Modellierungen im laufenden Projekt. In aller Regel sind weitreichende nachträgliche Umbauten der hierarchischen Mikrostruktur schwierig oder unmöglich. Die in Abschnitt 3.1 vorgestellte objektorientierte Modellierungsweise erlaubt es immerhin, ohne größere Schwierigkeiten eine vorhandene Modellierung um neue Elemente, also neue lexikographische Angabetypen oder neue Auswahlmöglichkeiten bei bestehenden Angabetypen, zu *erweitern*. Die bereits erstellten XML-Dokumente ohne die neuen Erweiterungen bleiben für das System ohne Änderungen lesbar (deserialisierbar), sofern im Zuge der Modellierungserweiterung im Programmcode sinnvolle Standardvorga-

ben für die bei fehlenden Angaben automatisch einzusetzende lexikographische Information spezifiziert werden.

Diese Flexibilität ist für das hier präsentierte Vorhaben von erheblicher Bedeutung. So ist beabsichtigt, die Lesarten der deutschen und entlehnten polnischen Etyma sowie der ostslavischen Lehnwörter in eine thematische Gliederung einzuordnen, um Bedeutungswandelprozesse in der Datenbasis systematisch erschließbar zu machen. Weil die thematische Gliederung jedoch erst im Rahmen des Projekts entwickelt werden kann, müssen die entsprechenden Details der Modellierung im laufenden Projekt nachgeführt werden. – Eine mittelfristige Perspektive des Lehnwortportals ist ganz allgemein die Anreicherung von Bedeutungsangaben zu Etyma und Lehnwörtern mit i. w. S. semantischen Annotationen wie etwa einer Klassifikation im Rahmen einer linguistischen Ontologie oder eines Thesaurus. Damit würde nicht nur ein onomasiologischer Zugriff jenseits der jetzt schon vorhandenen Suche nach Zeichenketten in den – stets deutschsprachigen – Bedeutungsangaben möglich, sondern auch die Nutzung des Portals und seiner Suchmöglichkeiten als Instrument zur Erforschung und Visualisierung von semantischen Prozessen bei Entlehnungsvorgängen.

Literatur

- BARCHUDAROV, STEPAN GRIGOR'EVIC (1975ff.): Slovar' russkogo jazyka XI–XVII vv. Moskva.
- BESTERS-DILGER, JULIANE (2002): Deutsche lexikalische Entlehnungen im Ukrainischen. Zur Frage der polnischen Vermittlung und heutigen Aktualität. In: POSPIŠIL, IVO (Hrsg.): Crossroads of Cultures: Central Europe. Brno (Litteraria Humanitas. XI), 25–51.
- BON, BRUNO/NOWAK, KRZYSZTOF (2013): Wiki Lexicographica. Linking Medieval Latin Dictionaries with Semantic Media-Wiki. In: KOSEM u. a. (Hrsg.), 407–420. Online: http://eki.ee/elex2013/proceedings/eLex2013_28_Bon+Nowak.pdf [Abruf: 2014-07-24].
- ČERNÝŠEV, VASILIJ IL'IC (1948–1965): Slovar' sovremennogo russkogo literaturnogo jazyka. Moskva/Leningrad.
- DE SCHRUYVER, GILLES-MAURICE (2011): Why Opting for a Dedicated, Professional, Off-the-shelf Dictionary Writing System Matters. In: AKASU, KAORU/

- UCHIDA, SATORU (Hrsg.): ASIALEX2011 Proceedings. Lexicography: Theoretical and Practical Perspectives. Papers Submitted to the Seventh ASIALEX Biennial International Conference, Kyoto Terrsa, Kyoto, Japan, August 22–24, 2011. Kyoto, 647–656.
- DE VINCENZ, ANDRZEJ/HENTSCHEL, GERD (2010): Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache. Von den Anfängen des polnischen Schrifttums bis in die Mitte des 20. Jahrhunderts. (*Studia slavica Oldenburgensia*. 20). Oldenburg. Online: <http://diglib.bis.uni-oldenburg.de/bis-verlag/wdlp/> [Abruf: 2014-07-24].
- ENGELBERG, STEFAN (2010): An inverted loanword dictionary of German loanwords in the languages of the South Pacific. In: DYKSTRA, ANNE/SCHOONHEIM, TANNEKE (Hrsg.): Proceedings of the XIV EURALEX International Congress, Ljouwert/Leeuwarden, 6–10 July 2010. Ljouwert/Leeuwarden, 639–647.
- HENTSCHEL, GERD (1989): A dictionary of language contact as a relational data base. In: MCCRANK, LAWRENCE J. (Hrsg.): Databases in the humanities and social sciences, 4: Proceedings of The International Conference on Databases in the Humanities and Social Sciences, Auburn University at Montgomery, July 1987. Medford, NJ, 293–302.
- HOLZNER, STEVEN (2009): XML: A Beginner's Guide. New York.
- HRYNČYŠYN, DMYTRO H. u. a. (1994ff.): Slovník ukraïns'koï movy XVI – peršoï polovyny XVII st. L'viv.
- KOSEM, IZTOK/KALLAS, JELENA/GANTAR, POLONA/KREK, SIMON/LANGEMETS, MARGIT/TUULIK, MARIA (Hrsg.) (2013): Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia. Ljubljana/Tallinn.
- MEYER, PETER/ENGELBERG, STEFAN (2011): Ein umgekehrtes Lehnwörterbuch als Internetportal und elektronische Ressource: Lexikographische und technische Grundlagen. In: HEDELAND, HANNA/SCHMIDT, THOMAS/WÖRNER, KAI (Hrsg.): Multilingual Resources and Multilingual Applications. Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011 (Arbeiten zur Mehrsprachigkeit/Working Papers in Multilingualism, Series B. 96). Hamburg, 169–174.
- MEYER, PETER (2013a): Advanced graph-based searches in an Internet dictionary portal. In: KOSEM u. a. (Hrsg.), 488–502. Online: http://eki.ee/elex2013/proceedings/eLex2013_34_Meyer.pdf [Abruf: 2014-07-24].
- MEYER, PETER (2013b): Ein Internetportal für deutsche Lehnwörter in slavischen Sprachen. Zugriffsstrukturen und Datenrepräsentation. In: KEMPGEN, SEBASTIAN/FRANZ, NORBERT/JAKIŠA, MIRANDA/WINGENDER, MONIKA (Hrsg.): Deutsche Beiträge zum 15. Internationalen Slavistenkon-

- gress, Minsk 2013. München (Die Welt der Slaven. Sammelbände. 50), 233–242.
- MEYER, PETER (2014a): Graph-Based Representation of Borrowing Chains in a Web Portal for Loanword Dictionaries. In: ABEL, ANDREA/VETTORI, CHIARA/RALLI, NATASCIA (Hrsg.): Proceedings of the XVI EURALEX International Congress: The User in Focus. 15–19 July 2014, Bolzano/Bozen. Bolzano/Bozen, 1135–1144.
- MEYER, PETER (2014b): Von XML zum DAG: Der lexikographische Prozess bei der Erstellung eines graphenbasierten Wörterbuchportals. In: DOMÍNGUEZ VÁZQUEZ, MARÍA JOSÉ/MOLLICA, FABIO/NIED CURCIO, MARTINA (Hrsg.): Zweisprachige Lexikographie im Spannungsfeld zwischen Translation und Didaktik. Berlin/Boston (Lexicographica, Series Maior. 147), 303–322.
- POLGUÈRE, ALAIN (2009): Lexical systems: graph models of natural language lexicons. In: Language Resources and Evaluation 43, 41–55.
- SOROKIN, JURIJ S. (1984ff.): Slovar' russkogo jazyka XVIII veka. Leningrad/St. Peterburg.
- SPOHR, DENNIS (2012): Towards a Multifunctional Lexical Resource. Design and Implementation of a Graph-based Lexicon Model. Berlin/Boston (Lexicographica, Series Maior. 141).
- TRIPPEL, THORSTEN (2006): The Lexicon Graph Model: A generic model for multimodal lexicon development. Saarbrücken.
- ŽURAŪSKI, ARKADZIJ I. (1982ff.): Histaryčny sloŭnik belaruskaj movy. Minsk.

MÓNICA MIRAZO BALSA

Überlegungen und Vorschläge zur Strukturierung und Darstellung lexikographischer Information in kontrastiven Onlinewörterbüchern

1. Einleitung¹

Die technologische Entwicklung hat in den letzten Jahrzehnten die Lexikographie grundlegend umgestaltet (LUQUE DURÁN 1996; MÜLLER-SPITZER 2007; DE SCHRYVER 2003; STORRER 2001), indem sie einen tiefgreifenden Wandel nicht nur des Prozesses zur Herstellung von Wörterbüchern und allen daran beteiligten Aspekten (Erstellung von Textcorpora, Datenbanken, Werkzeugen zur Datenausgabe, usw.), sondern auch der Art der Dateiabfrage von den Nutzern verursacht hat.

Bereits seit ungefähr zwei Jahrzehnten gerät die sogenannte elektronische Lexikographie ins Zentrum der lexikographischen Fachdiskussion. Zurzeit scheint es aber, dass das Internet sich als die dominante elektronische Plattform für Wörterbücher konsolidiert hat und es gibt keine sichtbaren Anzeichen dafür, dass diese Situation sich in der näheren Zukunft ändern könnte.

Bis in die 60er Jahre des 20. Jahrhunderts war das Papier das einzige Medium zur Verteilung von lexikographischen Werken, aber die Entwicklung der neuen Technologien und die Anwendung computerbasierter Techniken zur Herstellung von Wörterbüchern, vor allem die

¹ Diese Arbeit steht in Zusammenhang mit dem Forschungsprojekt „Spanisch-deutsche kontrastive Untersuchung der Nominalphrase. Erstellung eines computergestützten bilingualen Corpus und eines Online-Wörterbuchs“ und dem seit 2012 von María José Domínguez Vázquez koordinierten Forschungsprojekt *PORTLEX* „Portal Lexicográfico: Diccionario online modular multilingüe y corpus informatizado anotado de la frase nominal“ (gefördert vom spanischen Bildungsministerium, MINECO: FFI2012-32456).

Umsetzung von Datenbanken in die lexikographische Arbeit, haben eine Revolution dieser Disziplin mit sich gebracht, wobei der Herstellungsprozess von Nachschlagewerken umgewandelt worden ist und neue Wörterbuchtypen² entstanden sind.

Die vielfältigen Möglichkeiten der elektronischen Medien und ihre Anwendung im lexikographischen Prozess haben dazu geführt, dass das Papier nicht mehr als das einzige Medium zur Verteilung von lexikographischen Arbeiten gilt, so dass der Benutzer die lexikographischen Informationen auf dem Bildschirm erhält und jederzeit und überall auf sie zugreifen kann. Das elektronische Medium gilt aber nicht nur als ein neues Publikationsmedium, sondern ihm wird ein hohes Verbesserungspotenzial für die lexikographische Praxis zugeschrieben.³

Das neue Medium spielt in der Lexikografie nicht nur als Publikationsmedium eine Rolle. Ein wichtiger Bereich ist auch der Einsatz neuer Technologien in der Herstellung von Wörterbüchern, also im lexikographischen Prozess (MÜLLER-SPITZER 2007, 12).

In diesem Punkt setzt der informatische Begriff *Datenbank* bereits eine tiefe Veränderung der Konzeption der lexikographischen Aktivität voraus (ÁGUILA ESCOBAR 2006). Die Bedeutungen, die uns Wörterbücher wie z. B. *Duden*, *Wahrig*, *Langenscheidt* oder *Pons* anbieten, stellen bereits wichtige Veränderungen im Wörterbucherstellungsprozess sowie einige Vorteile von digitalen Wörterbüchern gegenüber Printwörterbüchern heraus:

DUDEN (2001): *Technische Anlage, in der große Bestände an Daten zentral gespeichert sind.*

WAHRIG (2006): *System, bestehend aus einer EDV-Anlage, einer Menge von Daten einschließlich der Menge der Beziehungen zw. diesen Daten u. speziellen Verwaltungsprogrammen, die mehreren Benutzern einen schnellen Zugriff auf die Daten ermöglichen.*

² Dazu siehe ENGELBERG/LEMNITZER (⁴2009).

³ Über die Möglichkeiten und Vorteile des elektronischen Mediums siehe auch DE SCHRYVER (2003) oder HAß/SCHMITZ (2010).

- LANGENSCHIEDT DAF (2008): *E-e große Sammlung von Daten (mst in e-m Computer), die nach verschiedenen Kriterien organisiert sind u. auch abgerufen werden können.*
- PONS (2011): *Eine große Menge von Daten, die in einem Computer nach bestimmten Kriterien organisiert sind und komplexe Abfragen zulassen.*

Diese Definitionen setzen zum einen voraus, dass der Einsatz von Datenbanken im lexikographischen Prozess die Überwindung der üblichen Praxis mit sich bringt, ein Formular für jedes Lemma bzw. jede Bedeutung im Handsatz herzustellen; zum anderen, dass die Informatik die strukturierte Speicherung von einer großen Menge von Daten in einem deutlich kleineren und einfach verarbeiteten Platz ermöglicht, sowie einen schnellen und simultanen Zugriff von mehreren Benutzern auf die Daten, was im Gegensatz zum Nachschlagen in gedruckten Werken steht, da diese von der begrenzten Anzahl der verfügbaren Kopien abhängig ist. Der Inhalt der Wörterbucheinträge wird dank des elektronischen Mediums in unterschiedlichen Feldern einer Datenbank strukturiert und kontrolliert gespeichert, so dass die eingegebenen Daten schnell überprüft und bearbeitet werden können und der Benutzer nicht auf die Veröffentlichung einer neuen Ausgabe warten muss.

Das World Wide Web (WWW), die Hypertext-Plattform im Internet, macht ein solches digitales Wörterbuch nicht nur weltweit zugänglich, sondern unterstützt auch die rasche und unkomplizierte Aktualisierung und Erweiterung der lexikographischen Daten (STORRER 2001, 54).

Im Gegensatz zu Printwörterbüchern erleichtert der Einsatz der neuen Technologien im lexikographischen Prozess aber auch u. a.

a) den vielfältigen Zugriff auf die Informationen: das gedruckte Medium bietet nur einen linearen Zugang zu den lexikographischen Daten, d. h. man kann auf die Information nur durch die Lemmata (normalerweise alphabetisch geordnet) und durch ein lineares der Mikrostruktur zugreifen. Nachschlagewerke werden aber nicht linear gelesen, sondern „in ihnen wird punktuell nachgeschlagen“ (MÜLLER-SPITZER 2007, 12).

b) die hypertextuellen Verweise innerhalb und außerhalb des Wörterbuchs: Das ermöglicht zum einen, dass der Benutzer nicht ständig hin- und herblättern muss; zum anderen, dass er durch Anklicken einen direkten Zugang zu anderen Wörterbüchern, zu den Textcorpora oder zu anderen nicht-lexikographischen Angeboten hat (vgl. TORRUELLA I CASAÑAS 2002 oder HAB/SCHMITZ 2010).

c) eine flexiblere Darstellung der lexikographischen Daten (MÜLLER-SPITZER/KOPLINIG 2014; DE SCHRYVER 2003; STORRER 2001).

Aber wie sollen die lexikographischen Daten in zwei- oder mehrsprachigen Onlinewörterbüchern strukturiert und dargestellt werden, um den Informationsbedarf der möglichen Benutzer zu erfüllen? Mit dieser Frage werden wir uns im Weiteren beschäftigen, indem wir unseren Vorschlag für ein kontrastives Onlinewörterbuch zur Substantivvalenz vorstellen werden.

2. Die Darstellung und Strukturierung der Information am Beispiel eines kontrastiven Onlinewörterbuchs zur Substantivvalenz

Das Projekt *CSVEA*⁴ ist im Jahr 2009 entstanden und sein Zweck ist die Erstellung eines annotierten spanisch-deutschen kontrastiven Corpus und eines bidirektionalen Onlinewörterbuchs zur Substantivvalenz. Die Frage nach den syntaktischen und semantischen Kombinationsmöglichkeiten der unterschiedlichen lexikalischen Einheiten ist der Ausgangspunkt für die lexikographische Entwicklung von valenzzentrierten Nachschlagewerken, die sich vor allem auf das Verb konzentriert haben (z. B. ENGEL/SCHUMACHER 1978, HELBIG/SCHENKEL 1973, HERBST u. a. 2004, SCHUMACHER u. a. 2004 als monolinguale Wörterbücher; oder z. B. BIANCO 1996, BUSSE/DUBOST 1977, CIRCO u. a. 1995, DJORDJEVIĆ/ENGEL 2009, ENGEL/SAVIN 1983, OCTAVIAN/ENGEL 2012 u. a. als zweisprachige Werke). In Vergleich zu anderen Ländern

⁴ *Das spanisch-deutsche Substantivvalenzwörterbuch (CSVEA)* war ein von María José Domínguez Vázquez und Stefan Schierholz koordiniertes Forschungsprojekt, das von der galicischen Landesregierung gefördert wurde (XUNTA: INCITE09 204 074 PR).

bleibt aber die Valenzlexikographie in Spanien weit zurück. Heutzutage verfügen wir über kein vollständiges einsprachiges Valenzwörterbuch.⁵ Was die kontrastive Valenzlexikographie anbelangt, verfügen wir auch über kein veröffentlichtes Nachschlagewerk mit dem Spanischen als Bezugssprache⁶, d. h. es gibt in der spanischen Lexikographie heutzutage weder ein vollständiges Valenzwörterbuch noch ein veröffentlichtes kontrastives Nachschlagewerk mit Bezug auf Valenz.

CSVEA entstand dann zur Füllung einer Lücke in der spanischen und kontrastiven Lexikographie, einem Bereich, in dem nur wenige Studien über das Substantiv und seine Valenz aus einer zweisprachigen bzw. mehrsprachigen Perspektive zu finden sind. Zu erwähnen sind u. a. die Werke von BASSOLA (2003; 2009: Deutsch–Ungarisch) als zweisprachiges Wörterbuch oder die lexikographischen Vorschläge von COSMA (2004: Deutsch–Rumänisch), DOMÍNGUEZ (2011: Deutsch–Spanisch), KUBCZAK/CONSTANTINO (1998: Deutsch–Französisch) oder SCHIERHOLZ (2000: Deutsch–Englisch–Portugiesisch). Diese Lücke in der spanischen Lexikographie erklärt auch die diesem Wörterbuchtyp wenig geschenkte Aufmerksamkeit seitens Deutsch lernender Hispanophoner (anvisierte Adressatengruppe unseres Wörterbuchs), wie aus der von DOMÍNGUEZ/MIRAZO/VIDAL (2013) durchgeführten empirischen Studie über Wörterbuchbenutzung hervorgeht. Die Ergebnisse der UDALPE⁷-Befragung zeigen, dass zweisprachige/kontrastive Wörterbücher am häufigsten von den Befragten benutzt werden, während sich

⁵ Das Werk mit der umfangreichsten syntaktischen Information ist CUERVO (1992–1994), aber es handelt sich dabei um kein Valenzwörterbuch im engeren Sinne.

⁶ Hier muss aber das *Spanisch-deutsche kontrastive Verbalenzwörterbuch (DCVVEA)* erwähnt werden, eine an der Universität Santiago de Compostela entwickelte Pilotstudie. Für weitere Information siehe <http://gramatica.usc.es/proyectos/valencia/?lang=es&id=descripcion> oder DOMÍNGUEZ u. a. (2008a; 2008b).

⁷ Online-Umfrage über die Benutzung deutscher und deutsch-spanischer Wörterbücher seitens spanischsprechender Deutschlernender, die vom 21. März 2011 bis zum 3. Juni 2011 im Internet (<http://www.e-encuesta.com/answer.do?testid=OdClwsKsv3w=>) freigeschaltet war.

Valenzwörterbücher (11,39 % Bekanntheitsgrad) als die am wenigsten konsultierten Nachschlagewerke herausstellen:

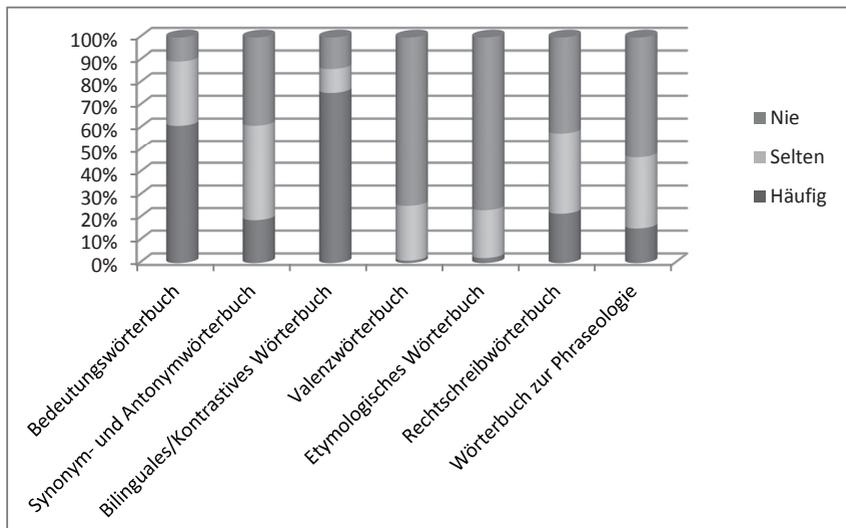


Abb. 1: UDALPE-Umfrage: Nutzungshäufigkeit von ausgewählten Wörterbuchtypen

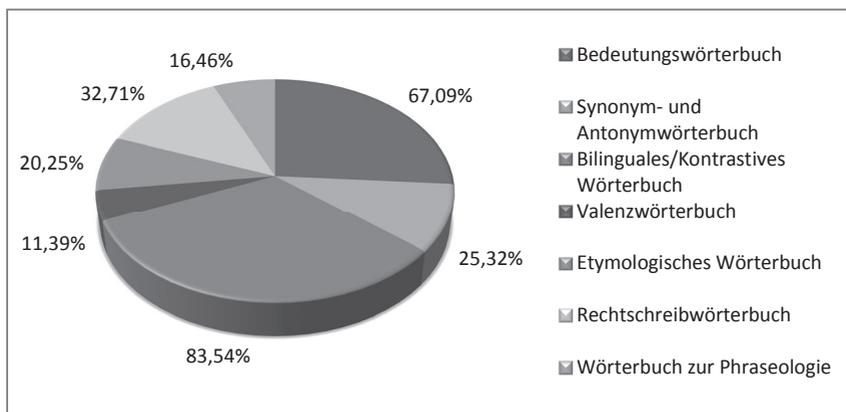


Abb. 2: UDALPE-Umfrage: Bekanntheitsgrad ausgewählter Wörterbuchtypen

Als Onlinewörterbuch positioniert sich CSVEA im Rahmen der aktuellsten Strömungen der lexikographischen Forschung und konkret im Rahmen der elektronischen Lexikographie. Das Netz und der Einsatz neuer technischer Fortschritte im lexikographischen Prozess bieten die Möglichkeit, neue Verfahren zu entwickeln, die den Lexikographen zur Optimierung der Nachschlagewerke Information über die konkreten Benutzerbedürfnisse, Benutzungssituationen (LEW 2012; MÜLLER-SPITZER u. a. 2012; SANMARCO 2008) und über die Vor- und Nachteile der unterschiedlichen Layouts und Interfaces (DOMÍNGUEZ/MIRAZO/VIDAL 2013; DOMÍNGUEZ/MOLLICA/NIED 2014 oder MÜLLER-SPITZER/KOPLÉNIG 2014) anbieten sollen. Aber trotz der zunehmenden Auseinandersetzungen mit der Wörterbuchbenutzungsforschung (RIPFEL/WIEGAND 1988; SCHARNHORST 2004; WELKER 2006; WIEGAND 1987; 1998; 2008 u. a.) ist dieses Forschungsgebiet nach wie vor wenig erforscht (WIEGAND 2008, 1) und viele Forscherinnen und Forscher fordern schon lange mehr empirische Untersuchungen, die den Lexikographen dabei helfen, die spezifischen Benutzungssituationen und -bedürfnisse verstehen zu können (HARTMANN 2000; TARP 2008). Nach diesen Benutzungssituationen und Benutzerinteressen können dann die lexikographischen Daten modelliert werden (vgl. STORRER 2001, 87), so dass sie nach individuellen Bedürfnissen und nach verschiedenen Suchkriterien abgerufen und auf dem Bildschirm dargestellt werden können.

Die Resultate solcher empirischer Untersuchungen über die Benutzerbedürfnisse, die Wörterbuchtypologie und die Häufigkeit und Art der Suche und des Umgangs mit dem Wörterbuch zeigen, dass Benutzer Online-Nachschlagewerke aber auch eine modulare Strukturierung der Information bevorzugen (DOMÍNGUEZ/MIRAZO/VIDAL 2013; DOMÍNGUEZ/MOLLICA/NIED 2014; DOMÍNGUEZ/VALCÁRCEL 2014 oder MÜLLER-SPITZER/KOPLÉNIG 2014).

Im Fall von CSVEA sollen die Ergebnisse dieser an den angesprochenen Benutzerkreis gerichteten UDALPE-Umfrage als Ausgangspunkt für die Herstellung der Benutzeroberfläche und die Datenmodellierung dienen.

2.1 Art und Architektur der Information in CSVEA

Ein elektronisches oder digitales Wörterbuch ist viel mehr als eine alphabetische Liste von Lemmata mit vollständigen lexikographischen Informationen. Es handelt sich um eine Datenbank, in der eine große Menge von Daten in unterschiedlichen Feldern gespeichert sind, die nachträglich von der Informatik je nach den Bedürfnissen, für die sie konzipiert wurden, abgerufen werden. Genauso wie TARP (2013, 304) hinweist, muss man aber deutlich zwischen der Datenbank selbst und den Daten, die auf dem Bildschirm gezeigt werden, unterscheiden:

La base de datos debe incorporar la mayor cantidad de datos posible para poder dar respuesta a todas las necesidades del grupo usuario previsto en las situaciones previstas y puede incluso alimentar y dar soporte a varios diccionarios simultáneamente. A su vez, el diccionario, es decir, los „artículos“ que aparecen en la pantalla al consultarlo, debe recoger la menor cantidad de datos posibles, o sea, sólo los que sean precisos para satisfacer las necesidades del usuario en cada tipo de situación o actividad y en cada consulta específica, puesto que los datos irrelevantes, en el mejor de los casos, son superfluos para la solución del problema concreto del usuario, y en el peor de los casos, pueden dificultar y, hasta obstaculizar, el acceso a los datos relevantes y la extracción, a partir de estos datos, de la información requerida en cada consulta [...].

Zum einen sollte dann die Datenbank so viele lexikographischen Daten wie möglich enthalten, um die Bedürfnisse aller anvisierten Benutzergruppen und alle Verwendungssituationen zu erfüllen; zum anderen sollen auf der Präsentationsebene nur diejenige Daten gezeigt werden, die dem Bedarf des Nutzers in jeder Situation und in jeder spezifischen Abfrage erforderlich sind. Dazu auch STORRER (2001, 60):

Die Art und Weise, wie Daten in einem Datenbank- oder Hypertextsystem strukturiert sind, muss nicht der Art und Weise entsprechen, wie diese Daten dem Benutzer am digitalen Lesegerät (dem Bildschirm, dem Handy, dem eBook) präsentiert werden. Die Zielsetzung bei der Informationsmodellierung besteht vielmehr gerade darin, Daten so zu strukturieren, dass aus ein und demselben Datenpool für verschiedene Anwendungs-

zwecke und Nutzungskontexte die jeweils relevanten Informationen herausgegriffen und in geeigneter Weise präsentiert werden können.

Die nachstehende Abbildung zeigt die Informationsblöcke, die jeder Artikel unseres Wörterbuchs enthält:

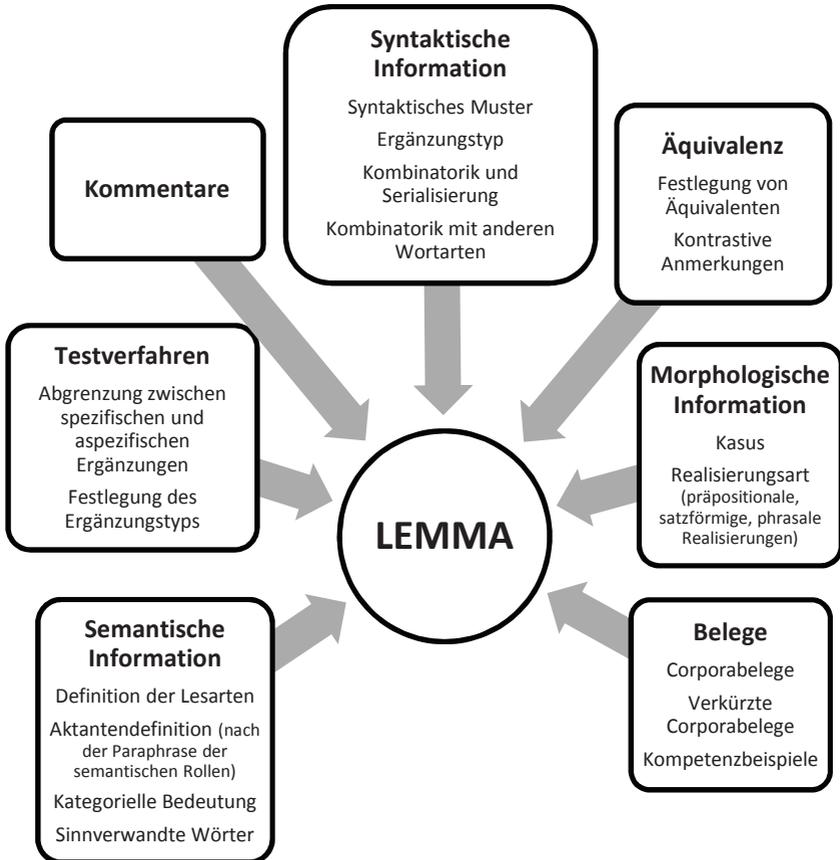


Abb. 3: Skizzierung der für jeden Wörterbucheintrag angegebenen Information

Methodologisch gesehen wird jedes spanische bzw. deutsche Substantiv mithilfe von unterschiedlichen Bedeutungswörterbüchern⁸ in verschiedene Lesarten eingeteilt, welche mit Beispielen aus unterschiedlichen Corpora⁹ belegt werden. Diese aus den Corpora erhaltenen authentischen Beispiele bilden dann die empirische Grundlage unseres Wörterbuchs, auf die sich sowohl die valenzielle Beschreibung als auch die Festlegung der Entsprechungen in der Zielsprache stützt. Nach der Einteilung jedes Lemmas in unterschiedlichen Bedeutungsvarianten werden diese syntaktisch und semantisch nach den Parametern der Valenztheorie beschrieben (siehe dazu DOMÍNGUEZ 2011).

Die Resultate dieser Analysen werden nachher elektronisch in unterschiedlichen Feldern einer Datenbank gespeichert, die nachträglich von dem Benutzer über eine Schnittstelle und nach unterschiedlichen Suchparametern abgerufen werden können.

Nach STORRER (2001, 64) sollte diese Benutzerschnittstelle „an Typen von Benutzungssituationen adaptierbar sein“. Die lexikographischen Daten können dann auf dem Bildschirm auf vielfältige Weise präsentiert werden.¹⁰ Obwohl an der Benutzeroberfläche dieses Projekts noch gearbeitet wird, ist CSVEA ein corpusgestützt bidirektional konzipiertes Onlinewörterbuch, in dem die Suche nach der Information einer modularen Strukturierung folgt. Das ermöglicht zum einen die Berücksichtigung von unterschiedlichen Benutzergruppen (Studenten, Übersetzer, Forscher

⁸ Für das Spanische werden Wörterbücher wie *DRAE* (Diccionario de la Lengua Española. Real Academia Española. Online: <http://www.rae.es/rae.html>), *CLAVE* (Online: <http://clave.smdiccionarios.com/app.php>), MOLINER (2008) oder SECO/ANDRÉS/RAMOS (1999) u. a. nachgeschlagen; für das Deutsche DUDEN (2010), *DWDS* (Digitales Wörterbuch der deutschen Sprache. Online: <http://www.dwds.de/>) u. a.

⁹ Für das Spanische stützen wir uns auf die aus *CREA* (Referenzcorpus für die spanische Sprache. Online: <http://corpus.rae.es/creanet.html>) erhaltenen Belege; für das Deutsche auf *DeReKo* („Das Deutsche Referenzkorpus“ des Instituts für Deutsche Sprache. Online: <https://cosmas2.ids-mannheim.de/cosmas2-web/>) und auf das Taz-Corpus der Universität Erlangen-Nürnberg (Online: <http://corpora.linguistik.uni-erlangen.de/>).

¹⁰ Einen Überblick über unterschiedliche Datenpräsentationen in elektronischen Wörterbüchern bieten z. B. MÜLLER-SPITZER/KOPLINIG (2014).

oder Deutsch- und Spanischlehrer) und ihren auch unterschiedlichen Benutzungsbedürfnissen und -interessen; zum anderen eine spätere mögliche Einbeziehung von weiteren Informationsmodulen (siehe DOMÍNGUEZ 2013).

Wie oben erwähnt, wird die Suche nach der Information in unserem Wörterbuch modular aufgebaut sein, so dass der Benutzer je nach der Vertiefung seiner Suche Zugriff auf verschiedene Inhaltsblöcke hat. Zur Erklärung unseres Vorschlags zeigen wir im Folgenden die unterschiedlichen Suchphasen:

i) Nach der Auswahl eines bestimmten Substantivs würde der Benutzer zuerst einen **Überblick über die unterschiedlichen Bedeutungsvarianten**¹¹ bekommen, indem sowohl ein Musterbeispiel als auch die Synonyme in der Ausgangssprache und die Äquivalente in der Zielsprache aufgeführt werden (vgl. Abb. 4). Der Einsatz von Links wird dem Nutzer auch ermöglichen, die Suche nach präziseren Informationen fortzusetzen.

¹¹ Die hier gezeigten Abbildungen stellen die Arbeitsschnittstelle des Lexikographen dar, d. h. es handelt sich nicht um die Präsentation der Daten auf der Benutzerschnittstelle.

sentido

Masculino

Acepción	Sinónimos	Equivalentes
<i>El sentido del olfato de los animales</i> Capacidad para percibir, mediante determinados órganos corporales, impresiones externas.	Percepción, Sensación	Der Sinn, Der -sinn
<i>¿Cuál es el sentido de mi vida?</i> Lógica, finalidad o razón de ser.	Objetivo, Finalidad, Intención	Das Ziel, Der Zweck, Die Absicht
<i>Mi sentido de la lealtad.</i> Modo particular de entender algo.	Idea	Der Sinn, Das Verständnis, Die Auffassung
<i>El origen y el sentido del texto se exponen claramente en varios tramos del mismo.</i> Significado de una palabra, de un conjunto de palabras o expresiones de cualquier tipo.	Significado	Die Bedeutung
<i>Las calles afectadas por los cortes son la avenida del General Perón, entre la calle Orense y la plaza de Lima en ambos sentidos.</i> Cada una de las dos orientaciones que tiene una misma dirección.	Dirección	Die Richtung
<i>El sentido musical de los negros africanos es muy agudo.</i> Capacidad que se tiene para realizar algo.	Capacidad	Das -gefühl

Abb. 4: Überblick über die unterschiedlichen Bedeutungsvarianten, Synonyme und Äquivalente des spanischen Substantivs *sentido*

ii) Nach der Auswahl der gewünschte Lesart erfolgt in einer zweiten Phase eine **Beschreibung der ausgewählten Bedeutungsvariante**, indem dem Benutzer ein Überblick über die Ergänzungen oder Aktanten angeboten wird, sowie ihre syntaktische (syntaktisches Muster und Ergänzungstypen) und semantische (Definition der Lesarten, Synonyme in der Ausgangssprache, Aktantendefinitionen nach einer Paraphrase der semantischen Rollen) Beschreibung. In dieser Suchphase bekommt er auch morphologische Informationen wie einen Überblick über die Realisierungsarten jedes Aktanten. Zur Veranschaulichung der in diesem Modul angebotenen Informationen dient in Abb. 5 das deutsche Substantiv *Weg*, das als Äquivalent von dem spanischen Substantiv *camino* in dem Sinne „Vía, trayecto, itinerario o dirección“ registriert wurde:

Weg
Masculino

Equivalentes	Sinónimos
Camino	Der Gang, Die Fahrt
Via, trayecto, itinerario o dirección. (Singular&Plural) Ej: Der Weg von lech nach Schröcken durch das Auenfeld.	
Utilización del artículo:	
CS →	Aquel/Aquello que realiza una acción der Weg + Gen
CS →	Aquel que realiza una acción der Weg von jemandem
Cadvorg →	Origen del desplazamiento der Weg aus + Dat der Weg von + Dat
Cadvpas →	Paso/travesía del desplazamiento der Weg durch + Acu der Weg über + Acu der Weg via + Acu
Cadvexp →	Expansión der Weg von + Dat
Cadvdird →	Dirección/destino del desplazamiento der Weg auf + Acu der Weg nach + Dat der Weg zu + Dat der Weg in + Acu der Weg bis + Dat

Abb. 5: Beschreibung einer Lesart am Beispiel des deutschen Substantivs *Weg*

iii) Ausführliche syntaktische, semantische und morphologische Informationen zu jeder Ergänzung bekommt man im nächsten Schritt. Dem Benutzer wird hier eine **detaillierte Beschreibung jedes Aktanten** angeboten, indem er auf die möglichen formalen Realisierungen, die

kategorielle Bedeutung sowie die aus den Corpora erhaltenen Belege¹² Zugriff hat. Ein weiteres Feld mit Kommentaren und den Testverfahren, auf die sich die ausgewählte Ergänzung stützt, steht auch zur Verfügung. Abb. 6 stellt die in dieser Suchphase angebotenen Informationen am Beispiel einer der möglichen Realisierungen der Subiectivergänzung bei *Diskussion* als Entsprechung des spanischen Substantivs *discusión* dar:

Aquel/Aquello que existe/es	
zwischen + Dat / [hum], [inst] / die Diskussion zwischen den Regierungsparteien; die Diskussion zwischen den Männern; die Diskussion zwischen konservativen und progressiven Juden	
<u>Ej corpus:</u> Das sogenannte Rettungsübernahmegesetz war am Freitag noch in der Diskussion zwischen den Regierungsparteien.	
Bei der Diskussion zwischen den Männern geht es munter zu, bisweilen wird es sogar laut.	
Die Diskussion zwischen konservativen und progressiven Juden kann beginnen.	
Comentario:	<i>Coord./Pl.</i>
Frecuencia:	
Tipo test:	<i>Test de transformación en oración, Test de derivación, Test de especificidad</i>
Observaciones:	

Abb. 6: Detaillierte Beschreibung jeder Realisierungsform der erfassten Ergänzungen

iv) Die **Kombinationsmöglichkeiten** der für jede Lesart erfassten Aktanten, d. h. die Interaktion der Ergänzungen untereinander, werden auch wie folgt in dem vierten Modul dargestellt:

¹² Auf dem Bildschirm wird aber nur eine Auswahl von Belegen präsentiert. Durch Anklicken erhält man weitere Belege sowie die Quellen.

Combinatoria

Ejemplo: die Liebe der Künstlerin für das Reisen
Comentario: Gen [anim] + für [-]

Ejemplo: die Liebe der Eltern zum Kind
Comentario: von [anim] + zu [-]

Ejemplo: die Liebe von meinem Vater für die Oper
Comentario: von [hum, zool] + für [-]

Ejemplo: die Liebe von Kindern zu ihren Eltern
Comentario: von [hum, zool] + zu [-]

Abb. 7: Kombinationsmöglichkeiten der unterschiedlichen Ergänzungen einer Bedeutungsvariante

v) **Kontrastive Anmerkungen** werden in einem weiteren Modul dargeboten, in dem diejenige Aspekte in Betracht kommen, die sich aus dem Vergleich zwischen der beiden zu analysierenden Sprachen ergeben, wie z. B. Anmerkungen zur quantitativen Valenz und der Kombinatorik, zur qualitativen Valenz des jeweiligen Aktanten oder Unterschiede zwischen dem nachgeschlagenen Substantiv und seinen Entsprechungen in der Zielsprache.

Die eigenen Eigenschaften dieses Projekts erfordern auch einen komplexeren Aufbau der Datenbank, da zur Erfüllung der Benutzermerkmale und -bedürfnisse sehr unterschiedliche Informationen benötigt werden. Die kontrastive und bidirektionale Ausrichtung unseres Wörterbuchs bedarf:

a) einer mehrsprachig konzipierten Datenbank, die die unterschiedlichen Benutzergruppen und ihre Verwendungsperspektiven (aktiv oder passiv) beim Abrufen der Information berücksichtigt. Darauf weist bereits STORRER (2001, 64) hin:

Ein Wörterbuch mit dem Sprachenpaar deutsch-englisch wird benutzt von englischen und deutschen Muttersprachlern, zur Hin- und zur Herüber-

setzung, zum Verstehen und zum Produzieren fremdsprachlicher Texte, zur Suche nach Übersetzungsäquivalenten, nach grammatischen Eigenschaften fremdsprachlicher Einheiten oder nach typischen Kollokationen.

Die Datenbank sollte die Anzeige der analysierten Daten in beliebigen Sprachen erlauben, wobei die Ausgangssprache immer die von dem Benutzer ausgewählte Sprache sein sollte, d. h. in ihrer Konzeption sollen sowohl die aktiven und passiven Verwendungsperspektiven als auch die Mutter- und Fremdsprachen des Benutzers berücksichtigt werden. Das bedeutet aber ein komplexeres Design der logischen und konzeptuellen Struktur der Datenbank, was auch eine intensive lexikographische Arbeit mit sich bringt.

b) einer komplexeren Vernetzung der Daten in der Datenbank, die bei einem späteren Abruf die Erfüllung des Benutzerbedürfnisses erlaubt. Im Gegensatz zu Printwörterbüchern bieten Online-Nachschlagewerke vielfältige und komplexere Recherchemöglichkeiten an. Suchanfragen wie z. B. „alle Substantive, bei denen eine Subiectivergänzung: Agens und eine Lokativergänzung vorhanden sind“ können in Online-Valenzwörterbüchern wie CSVEA verfügbar sein, was in Printwörterbüchern nicht möglich ist.

Sowohl das methodologische Verfahren als auch das Beschreibungsmodell von CSVEA dienen als Ausgangspunkt für die Entwicklung von PORTLEX¹³, dessen Hauptziel die Erstellung eines lexikographischen Portals ist, das in der nahen Zukunft unterschiedliche modulare und multidirektionale Onlinewörterbücher beinhalten wird. Die lexikographische Beschreibung von CSVEA wird dann auch auf die folgenden Sprachen angewendet: Spanisch, Deutsch, Italienisch, Französisch und Galicisch. Das würde die Anpassung des kontrastiven bidirektionalen Wörterbuchs an das kontrastive multidirektionale Nachschlagewerk bedeuten, was auch wichtige Veränderungen in der Konzeption der Datenbank mit sich bringen wird.

¹³ Siehe Fußnote 1. Für weitere Informationen zu PORTLEX siehe auch DOMÍNGUEZ/MIRAZO/VALCÁRCEL (2014).

3. Schlussfolgerung

Lexikographische Informationen können in Onlinewörterbüchern auf vielfältige Weise strukturiert sein. Am Beispiel unseres kontrastiven Substantivvalenzwörterbuchs schlagen wir in diesem Beitrag aber eine modulare Strukturierung der Information vor, die dem Benutzer die Suche nach der verlangten Information erleichtern wird.

Diese modulare Entwicklung und Darstellung der Information bringt nach unserer Meinung einige Vorteile mit sich:

a) Die nach linguistischen Kriterien konzipierten Module der Datenbank erlauben, dass die Suchanfrage seitens der möglichen Benutzer immer erfüllt wird, d. h. auf dem Bildschirm können entweder der Gesamtartikel oder nur die nach bestimmten Suchparametern verlangte Information präsentiert werden. Sucht der Nutzer nach semantischen Kriterien, bekommt er eine Einteilung der Substantive nach Bedeutungsvarianten, eine Beschreibung der semantischen Rollen und der den unterschiedlichen Aktanten zugeschriebenen kategoriiellen Bedeutungen und den Zugriff auf sinnverwandte Wörter; erfolgt die Suchanfrage nach syntaktischen Parametern, kann der Benutzer auf das syntaktische Muster der Bedeutungsvariante, aber auch auf die Beschreibung des Ergänzungstyps und die möglichen Kombinationsmöglichkeiten zugreifen. Dieser modulare Aufbau ermöglicht dann vielfache Recherchemöglichkeiten.

b) Die Berücksichtigung aller möglichen Benutzer: Die hier von uns vorgeschlagene Strukturierung der Daten berücksichtigt auch den breiten Benutzerkreis und seine konkreten Benutzungssituationen und –bedürfnisse. Es ist eindeutig, dass nicht alle Nutzergruppen z. B. die Informationen über die unterschiedlichen Testverfahren zur Festlegung des Ergänzungstyps verlangen. Hier wäre es vielleicht sinnvoller, einen Abruf der Daten nach individuellen Bedürfnissen bereitzustellen, indem die Daten für die verschiedenen Zielgruppen jeweils anders dargestellt werden.

Eine Optimierung der hier vorgeschlagenen Struktur und die Entwicklung einer benutzerfreundlichen Schnittstelle muss aber noch durchgeführt werden.

Literatur

- ÁGUILA ESCOBAR, GONZALO (2006): Las nuevas tecnologías al servicio de la lexicografía: los diccionarios electrónicos. In: VILLAYANDRE LLAMAZARES, MILKA (Hrsg.): Actas del XXXIV Simposio Internacional de la Sociedad Española de Lingüística. León: Universidad de León. Online: <http://fhyc.unileon.es/SEL/actas.htm> [Abruf: 2014-08-03].
- BASSOLA, PETER (Hrsg.) (2003): Német-magyar fonévi valenciaszótár. Deutsch-ungarisches Wörterbuch zur Substantivvalenz. Szeged: Grimm Kiadó.
- BIANCO, MARIA TERESA (1996): Valenzlexikon Deutsch–Italienisch. Dizionario della valenza verbale. Heidelberg: Groos.
- BUSSE, WINFRIED/DUBOST, JEAN-PIERRE (Hrsg.) (1977): Französisches Verblexikon. Die Konstruktion der Verben im Französischen. Stuttgart: Klett-Cotta.
- CAMPOS SOUTO, MAR/GONZÁLEZ GONZÁLEZ, EVA MARÍA/PÉREZ PASCUAL, JOSÉ IGNACIO (Hrsg.) (2008): La lexicografía bilingüe y didáctica: ayer y hoy. Anexos Revista de Lexicografía 8. A Coruña: Universidade da Coruña.
- CIRKO, LESLAW/MORCINIEC, NORBERT/ZIOBRO, RYSZARD (1995): Wörterbuch zur Valenz deutscher Verben. Wrocław: Wyd. Uniwers. Wrocławskiego.
- COSMA, RUXANDRA (2004): Über Einträge in zweisprachigen Lexika des Nomens. Wie sie sind und wie sie in einem rumänisch-deutschen Ansatz außerdem noch sein könnten. In: Zeitschrift der Germanisten Rumäniens, 243–254. Online: <http://www.e-scoala.ro/germana/ggr4.html> [Abruf: 2010-12-12].
- CUERVO, RUFINO J. (1992–1994): Diccionario de construcción y régimen de la lengua castellana. Santafé de Bogotá, vol. 1 (1886) von A–B erweitert von Instituto Caro y Cuervo.
- DE SCHRYVER, GILLES-MAURICE (2003): Lexicographers’ Dreams in the Electronic-Dictionary Age. In: International Journal of Lexicography 16/2, 143–199.
- DJORDJEVIĆ, MILOJE/ENGEL, ULRICH (2009): Wörterbuch zur Verbvalenz, Deutsch–Bosnisch/Kroatisch/Serbisch. München: Iudicium.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ (2011): Kontrastive Grammatik und Lexikographie, spanisch-deutsches Wörterbuch zur Valenz des Nomens. München: Iudicium.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ (2013a): Kontrastive Valenzwörterbücher im spanischen Sprachraum: eine neue Wende? In: DOMÍNGUEZ (Hrsg.), 19–39.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ (Hrsg.) (2013b): Trends in der deutsch-spanischen Lexikographie. Frankfurt am Main [u. a.]: Peter Lang.

- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/GÓMEZ GUINOVART, XAVIER/VALCÁRCEL RIVEIRO, CARLOS (Hrsg.) (im Druck): Lexicografía de las lenguas románicas II. Aproximaciones a la lexicografía contemporánea y contrastiva. Wissenschaftliche Beratung von M^a DOLORES SÁNCHEZ PALOMINO und M^a JOSÉ DOMÍNGUEZ VÁZQUEZ. Berlin: De Gruyter.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/LÜBKE, BARBARA/MELISS, MEIKE/PAREDES, GEMMA/POULSEN, PIA/VÁZQUEZ ROZAS, VICTORIA (2008a): Presentación del Diccionario de valencias verbales: español-alemán. I. Características generales del proyecto. In: CAMPOS/GONZÁLEZ/PÉREZ (Hrsg.), 51–61.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/LÜBKE, BARBARA/MELISS, MEIKE/PAREDES, GEMMA/POULSEN, PIA/VÁZQUEZ ROZAS, VICTORIA (2008b): Presentación del Diccionario de valencias verbales: español-alemán. II. Problemas descriptivos particulares. In: CAMPOS/GONZÁLEZ/PÉREZ (Hrsg.), 63–74.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/MIRAZO BALSA, MÓNICA (2014): Lexicografía contrastiva online: Tipología de usuarios y arquitectura de la información en una base de datos para la frase nominal. In: Actas del V. Congreso Internacional de Lexicografía Hispánica. Getafe (im Druck).
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/MIRAZO BALSA, MÓNICA/VALCÁRCEL RIVEIRO, CARLOS (2014): Evolución del diccionario bilingüe al multilingüe: de CSVEA a PORTLEX. In: MELISS, MEIKE/SÁNCHEZ PALOMINO, M^a DOLORES/SANMARCO BANDE, M^a TERESA (Hrsg.): A lexicografía das linguas románicas: Estado da cuestión. München: Iudicium (im Druck).
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/MIRAZO BALSA, MÓNICA/VIDAL PÉREZ, VANESSA (2013): Wörterbuchbenutzung: Erwartungen und Bedürfnisse. Ergebnisse einer Umfrage bei Deutsch lernenden Hispanophonen. In: DOMÍNGUEZ (Hrsg.), 135–172.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/MOLLIKA, FABIO/NIED CURCIO, MARTINA (2014): Simplex-Verben im Italienischen und Spanischen vs. Präfix- und Partikelverben im Deutschen. Eine Untersuchung zum Gebrauch von Online-Wörterbüchern bei der Übersetzung. In: DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/MOLLIKA, FABIO/NIED CURCIO, MARTINA (Hrsg.): Zweisprachige Lexikographie im Spannungsfeld zwischen Translation und Didaktik. Berlin/Boston: De Gruyter (Lexicographica, Series Maior. 147), 179–220.
- DOMÍNGUEZ VÁZQUEZ, M^a JOSÉ/VALCÁRCEL RIVEIRO, CARLOS (2014): Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias? In: DOMÍNGUEZ/GÓMEZ/VALCÁRCEL (Hrsg.) (im Druck).
- DUDEN (2001): Duden. Deutsches Universalwörterbuch. 4., neu bearbeitete und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- DUDEN (2010): Duden. Das Bedeutungswörterbuch. 4., aktualisierte und erweiterte Auflage. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

- ENGEL, ULRICH/SAVIN, EMILIA (1983): Valenzlexikon deutsch-rumänisch. Dictionar de valenta german-român. Heidelberg: Groos.
- ENGEL, ULRICH/SCHUMACHER, HELMUT (1978): Kleines Valenzlexikon deutscher Verben. 2. Auflage. Tübingen: Narr.
- GRANGER, SYLVIANE/PAQUOT, MAGALI (Hrsg.): Electronic lexicography. Oxford: Oxford University Press.
- HARTMANN, REINHARD (2000): European Dictionary Culture. The Exeter Case Study of Dictionary Use among University Students, against the Wider Context of the Reports and Recommendations of the Thematic Network Project in the Area of Languages (1996–1999). In: HEID, ULRICH/EVERT, STEFAN/LEHMANN, EGBERT/ROHRER, CHRISTIAN (Hrsg.): Proceedings of the Ninth EURALEX International Congress, EURALEX 2000. Stuttgart, Germany, August 8th–12th, 2000. Stuttgart, 385–391.
- HAB, ULRIKE/SCHMITZ, ULRICH (2010): Lexikographie im Internet 2010 – Einleitung. In: Lexicographica 26, 1–18.
- HELBIG, GERHARD/SCHENKEL, WOLFGANG (1973): Wörterbuch zur Valenz und Distribution deutscher Verben. 2. Auflage. Leipzig: Verl. Enzyklopädie.
- HERBST, THOMAS/HEATH, DAVID/ROE, IAN F./GÖTZ, DIETER (2004): A Valency Dictionary of English. A corpus based analysis of the complementation patterns of English verbs, nouns and adjectives. Berlin: De Gruyter.
- KUBCZAK, JACQUELINE/CONSTANTINO, SYLVIE (1998): Exemplarische Untersuchungen für ein syntagmatisches Wörterbuch Deutsch–Französisch/Französisch–Deutsch. In: BRESSON, DANIEL/KUBCZAK, JACQUELINE (Hrsg.): Abstrakte Nomina, Vorarbeiten zu ihrer Erfassung in einem zweisprachigen syntagmatischen Wörterbuch. Tübingen: Narr, 11–119.
- LANGENSCHIEDT DAF (2008): GÖTZ, DIETER/HAENSCH, GÜNTHER/WELLMANN, HANS (Hrsg.): Langenscheidt Großwörterbuch Deutsch als Fremdsprache. Berlin: Langenscheidt.
- LEW, ROBERT (2012): How can we make electronic dictionaries more effective? In: GRANGER/PAQUOT (Hrsg.), 343–361.
- LUQUE DURÁN, JUAN DE DIOS (1996): Tipos de diccionarios y el diccionario del futuro (sobre las aplicaciones de la informática y de la lingüística computacional a la lexicografía). In: LUQUE DURÁN, JUAN DE DIOS/PAMIES BERTRÁN, ANTONIO (Hrsg.): Segundas jornadas sobre estudio y enseñanza del léxico. Granada, 93–102.
- MOLINER, MARÍA (2008): Diccionario de uso del español. 3^a edición. Madrid.
- MÜLLER-SPITZER, CAROLIN (2007): Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis. Tübingen: Narr (Studien zur Deutschen Sprache. 42).

- MÜLLER-SPITZER, CAROLIN (2014) (Hrsg.): *Using Online Dictionaries*. Berlin/Boston: De Gruyter (Lexicographica, Series Maior. 145).
- MÜLLER-SPITZER, CAROLIN/KOPLÉNIG, ALEXANDER (2014): Requisitos y expectativas de un buen diccionario online. Resultados de estudios empíricos en la investigación sobre el uso de diccionarios con especial atención a los traductores. In: DOMÍNGUEZ/GÓMEZ/VALCÁRCEL (Hrsg.) (im Druck).
- MÜLLER-SPITZER, CAROLIN/KOPLÉNIG, ALEXANDER/TÖPEL, ANTJE (2012): Online dictionary use: Key findings from an empirical research project. In: GRANGER/PAQUOT (Hrsg.), 425–457.
- OCTAVIAN, NICOLAE/ENGEL, ULRICH (2012): *Dictionar de verbe germane. Iasi: Polirom*.
- PONS (2011): *Pons Großwörterbuch Deutsch als Fremdsprache*. Stuttgart: Klett Sprachen.
- RIPPEL, MARTHA/WIEGAND, HERBERT ERNST (1988): Empirische Wörterbuchbenutzungsforschung. In: WIEGAND, HERBERT ERNST (Hrsg.): *Studien zur neuhochdeutschen Lexikographie VI. 2. Teilband*. Hildesheim/Zürich/New York: Olms (Germanistische Linguistik. 87–90/1986), 491–520.
- SANMARCO BANDE, M^a TERESA (2008): La lexicografía del siglo XXI. Hacia el diccionario ilimitado. In: CAMPOS/GONZÁLEZ/PÉREZ (Hrsg.), 133–140.
- SCHARNHORST, JÜRGEN (2004) (Hrsg.): *Sprachkultur und Lexikographie. Von der Forschung zur Nutzung von Wörterbüchern*. Frankfurt am Main [u. a.]: Peter Lang.
- SCHIERHOLZ, STEFAN (2000): Governed Prepositions. A Database for German, English and Portuguese Nouns. In: MOGENSEN, JENS E./PEDERSEN, VIGGO/ZETTERSTEN, ARNE (Hrsg.): *Symposium on Lexicography IX. Proceedings of the Ninth International Symposium on Lexicography April 23–25, 1998 at the University of Copenhagen*. Tübingen: Niemeyer (Lexicographica, Series Maior. 103), 143–164.
- SCHUMACHER, HELMUT/KUBCZAK, JACQUELINE/SCHMIDT, RENATE/DE RUITER, VERA (2004): *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Narr.
- SECO, MANUEL/ANDRÉS, OLIMPIA/RAMOS, GABINO (1999): *Diccionario del español actual*. Madrid: Aguilar.
- SOMMERFELDT, KARL-ERNST/SCHREIBER, HERBERT (1983a): *Wörterbuch zur Valenz und Distribution deutscher Adjektive*. 3. Auflage. Leipzig: Bibliographisches Institut.
- SOMMERFELDT, KARL-ERNST/SCHREIBER, HERBERT (1983b): *Wörterbuch zur Valenz und Distribution der Substantive*. 3. Auflage. Leipzig: Bibliographisches Institut.
- STORRER, ANGELIKA (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: LEMBERG, INGRID/

- SCHRÖDER, BERNHARD/STORRER, ANGELIKA (Hrsg.): Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen: Niemeyer (Lexicographica, Series Maior. 107), 53–69.
- TARP, SVEN (2008): Lexicography in the borderland between knowledge and non-knowledge: General lexicographical theory with particular focus on learner's lexicography. Tübingen: Niemeyer (Lexicographica, Series Maior. 134).
- TARP, SVEN (2013): El diccionario del futuro. In: RUÍZ MIYARES, LEONEL/ÁLVAREZ SILVA, MARÍA ROSA/MUÑOZ ALVARADO, ÁLEX (Hrsg.): Actualizaciones en Comunicación Social. Santiago de Cuba, 304–308.
- TORRUELLA I CASAÑAS, JOAN (2002): Los diccionarios en la Red. Barcelona. Online: http://seneca.uab.es/sfi/CURSOWEB/SESIONES/Diccionarios_en_la_Red.pdf [Abruf: 2014-03-24].
- WAHRIG (2006): WAHRIG-BURFEIND, RENATE (Hrsg.): Wahrig Deutsches Wörterbuch. 8., vollständig neu bearbeitete und aktualisierte Auflage. München: Wissen Media.
- WELKER, HERBERT ANDREAS (1998): Dicionário de verbos alemão-português. Online: <http://www.unb.br/il/let/welker/dici/index> [Abruf: 2010-03-08].
- WELKER, HERBERT ANDREAS (2006): O Uso de Dicionários. Panorama geral das pesquisas empíricas. Brasília: Thesaurus.
- WIEGAND, HERBERT ERNST (1987): Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung. In: Lexicographica 3, 178–227.
- WIEGAND, HERBERT ERNST (1998): Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Berlin/New York: De Gruyter.
- WIEGAND, HERBERT ERNST (2008): Wörterbuchbenutzung bei der Übersetzung. Möglichkeiten ihrer Erforschung. In: JESENŠEK, VIDA/LIPAVIC OŠTIR, ALJA (Hrsg.): Wörterbuch und Übersetzung. 4. Internationales Kolloquium zur Lexikographie und Wörterbuchforschung. Universität Maribor. 20. bis 22. Oktober 2006. Hildesheim/Zürich/New York: Olms (Germanistische Linguistik. 195–196), 1–43.
- ZIFONUN, GISELA/HOFFMANN, LUDGER/STRECKER, BRUNO u. a. (1997): Grammatik der deutschen Sprache. 3 Bände. Berlin/New York: De Gruyter (Schriften des Instituts für Deutsche Sprache. 7).

AXEL HEROLD

Das Wörterbuch als zentrale Komponente eines digitalen lexikalischen Systems

1. Was ist ein digitales lexikalisches System?

Ein System ist im allgemeinsten Sinne eine Menge von Einzelelementen, sowie der Beziehungen, in denen diese Einzelelemente zueinander stehen. Die Konzeption eines digitalen lexikalischen Systems (DLS) orientiert sich eng an der Idee des mentalen Lexikons, einer Standardannahme der Linguistik zur Beschreibung der Organisation des nicht-regelhaften Anteils menschlichen Sprachwissens (vgl. KLEIN 2004, sowie darauf aufbauend KLEIN/GEYKEN 2010). Die zugrundeliegenden Einzelelemente eines solchen Systems werden im Folgenden als lexikalische Einheiten bezeichnet, um die prinzipielle Unabhängigkeit der Idee eines DLS von konkreten linguistischen Analyseeinheiten zu verdeutlichen. Eine lexikalische Einheit kann sowohl ein (beispielsweise orthographisches oder phonologisches) Wort repräsentieren als auch eine lexikalisierte Phrase oder ein Morphem. Die zulässigen Arten lexikalischer Einheiten sind abhängig von der konkret zugrunde gelegten linguistischen Theorie. Für die folgende allgemeine Betrachtung ist eine genauere Spezifizierung nicht notwendig.

KLEIN (2004) folgend wird eine lexikalische Einheit durch Bündel von Merkmalen gebildet. Die wichtigsten Merkmalsbündel sind die Zeichengestalt (beziehungsweise die Menge der jeweils möglichen Gestaltformen), die semantischen sowie die morphosyntaktischen Eigenschaften. Sowohl ihre semantischen als auch ihre morphosyntaktischen Eigenschaften bestimmen letztlich die Arten der möglichen Beziehungen einer lexikalischen Einheit zu anderen Einheiten. In diesem Sinne ist ein DLS tatsächlich ein Modell des mentalen Lexikons und die einzelnen Elemente eines solchen Systems sind wiederum Modelle

lexikalischer Einheiten. Die konkrete Form der Beschreibung beziehungsweise Modellierung einer lexikalischen Einheit sowie deren Beziehungen zu anderen Einheiten kann von System zu System stark variieren. So soll die Beschreibung einerseits möglichst konsistent einer vereinbarten linguistischen Theorie folgen, andererseits muss sie typischerweise aus (zeit-)ökonomischen Gründen auf ausgewählte Aspekte dieser Theorie beschränkt bleiben.

Die bis hierher skizzierte formale Beschreibung eines DLS lässt sich allerdings ohne Weiteres ebenso auf „klassische“ gedruckte Wörterbücher beziehen. Auch diese zielen auf die Darstellung eines Ausschnitts des mentalen Lexikons im Rahmen eines vereinbarten Modells beziehungsweise einer linguistischen Theorie ab. Sie verzeichnen lexikalische Einheiten in der Form von Merkmalsbündeln sowie vielfach ebenfalls Relationen zwischen den individuellen Einheiten. Der eigentliche qualitative Unterschied zwischen einem traditionellen gedruckten Wörterbuch und einem DLS lässt sich folglich nur auf der Basis des Attributs „digital“ bestimmen, genauer: auf der Basis der Möglichkeiten, die die digitale Repräsentation der linguistischen Informationen impliziert oder zumindest ermöglicht.

Im Folgenden wird unter einer solchen digitalen Repräsentation linguistischer und speziell lexikografischer Informationen eine Darstellungsform verstanden, die zum einen maschinenlesbar ist, also rechnergestützt be- und verarbeitet werden kann, und in der zum anderen die grundlegenden Informationseinheiten explizit identifiziert und adressierbar sind.

Impliziert wird durch diese beiden Vorbedingungen der prinzipiell direkte und verzögerungsfreie Zugriff auf jede lexikalische Einheit. Für lexikografisch aufbereitete Ressourcen erstreckt sich dieser Zugriff ebenso auf die auftretenden Mikrostrukturelemente. Es können dadurch gezielt beispielsweise Lesarten innerhalb von Artikeln semasiologischer Wörterbücher ermittelt, morphosyntaktische Eigenschaften oder etymologische Beschreibungen von Wörtern lokalisiert oder generell die gesamte Ressource nach beliebigen Kriterien gefiltert und durchsucht werden. Diese Implikation unter den oben geforderten Vorausset-

zungen entspricht damit letztlich einer radikalen Erweiterung der Zugriffsstruktur eines Wörterbuchs.

Zu den nicht notwendigerweise folgenden, jedoch oft geforderten Möglichkeiten, die sich auf der Basis der beiden Vorbedingungen aus der digitalen Repräsentation ergeben können, gehören vor allem:

- Aspekte der Informationspräsentation: Es treten keine medienspezifischen Platzbeschränkungen auf, die Präsentation kann dynamisch an aktuelle Nutzerwünsche und -bedürfnisse angepasst werden, die erweiterten Zugriffsstrukturen können umfassend ausgewertet und genutzt werden, weitere Kommunikationsmodi wie Ton und Bewegtbild können zur Repräsentation verwendet werden und vieles mehr.
- Aspekte der Informationsgenerierung und Ressourcenvernetzung: Die simultane beziehungsweise synoptische Darstellung unterschiedlicher Ressourcen wird möglich, ebenso wie die Synthese komplexer lexikalischer Beschreibungen aus den Einzelinformationen verschiedener Ressourcen. So könnte beispielsweise ein Bedeutungswörterbuch durch wortgeschichtliche Angaben eines etymologischen Wörterbuchs ergänzt werden; in einem allgemeinsprachlichen Wörterbuch ursprünglich nicht angesetzte Fachtermini könnten mit Hilfe eines entsprechenden fachsprachlichen Wörterbuchs ergänzt werden.
- Aspekte der Beschleunigung und Verbesserung des lexikografischen Prozesses selbst: Die Bearbeitung der Daten kann modular erfolgen, der lexikografische Arbeitsplatz kann vollkommen rechnergestützt konzipiert werden. Vgl. GEYKEN (im Druck) für eine detaillierte Darstellung der entsprechenden Möglichkeiten.

Aus der Fülle aktueller Forschungsliteratur, die sich mit der Darstellung dieser Möglichkeiten befasst, sei hier nur exemplarisch auf den Jahrgangsband 2010 (Band 26) der *Lexicographica* verwiesen; eine etwas neuere, ergänzende Darstellung mit stärkerem Fokus auf den lexikografischen Prozess ist der Sammelband von GRANGER/PAQUOT (2012).

2. Das Wörterbuch im komplexen digitalen lexikalischen System

Die Frage nach dem Status eines Wörterbuchs innerhalb eines lexikalischen Systems ist trivial zu beantworten, solange der oben angesprochene Aspekt der Informationsgenerierung und Ressourcenvernetzung keine Rolle spielt. In diesem Fall sind das in Frage stehende Wörterbuch und das DLS praktisch identisch. Das ist typischerweise bei wichtigen retrodigitalisierten Wörterbüchern der Fall, die als einzelstehende Ressource für die elektronische Nutzung erschlossen wurden (beispielsweise Adelungs „Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart“¹ von 1811).

Häufiger treten digitale lexikalische Systeme auf, die eine Reihe unterschiedlicher lexikalischer Ressourcen für die gemeinsame Darstellung integrieren (beispielsweise das Trierer „Wörterbuchnetz“², aber auch vielfach verlagslexikografische Angebote, in denen verlagseigene Wörterbücher simultan konsultiert werden können). Bei solchen Systemen, die selbst aus prinzipiell eigenständig nutzbaren (Sub-)Systemen bestehen, ergibt sich dadurch eine Form rekursiver Einbettung lexikalischer Systeme: Ein digitales lexikalisches (Gesamt-)System dieser Art umfasst dann Wörterbücher, die ihrerseits wiederum der Definition eines DLS entsprechen. Derartige DLS werden im Folgenden als komplexe DLS bezeichnet.

In einem komplexen DLS aus unabhängig voneinander entwickelten lexikalischen Subsystemen wird es zwangsläufig zu Widersprüchen und Inkonsistenzen in der Beschreibung einzelner lexikalischer Einheiten kommen. Die möglichen Gründe hierfür sind vielfältig; einige der wichtigsten sind:

- Unterschiede im Wörterbuchgegenstand (beziehungsweise bezogen auf den Ausgangspunkt dieser Darstellung: verschiedene Ausschnitte des mentalen Lexikons) können dazu führen, dass einzelne linguistische Merkmale unterschiedlich stark gewichtet oder teil-

¹ <http://lexika.digitale-sammlungen.de/adelung/online/angebot>

² <http://woerterbuchnetz.de/>

weise gar nicht dargestellt werden. Beispielsweise könnte ein Wörterbuch, das einen Fachwortschatz darstellt, für einige Massennomen einen Sortenplural beschreiben, der in einem allgemeinsprachlichen Wörterbuch nicht aufgeführt wird.

- Sprachwandelprozesse bewirken, dass sich die semantische Struktur, aber auch alle anderen linguistischen Merkmale lexikalischer Einheiten (wie Zeichengestalt oder morphosyntaktische Eigenschaften) mit der Zeit ändern. Das führt dazu, dass auch Wörterbücher, die einen vergleichbaren Gegenstand – allerdings bezogen auf unterschiedliche Zeiträume – beschreiben (beispielsweise die städtische Umgangssprache des 19. im Gegensatz zum 21. Jahrhundert), inkongruente Beschreibungen ihrer einander eigentlich entsprechenden lexikalischen Einheiten aufweisen können.
- Unterschiede in den verwendeten Wörterbuchquellen (die ja den Gegenstandsbereich in ihrer Gesamtheit repräsentieren sollen) und den darauf aufsetzenden Analyseverfahren führen möglicherweise dazu, dass der semantische Raum, der einem lexikalischen Zeichen zugeordnet ist, unterschiedlich gegliedert wird. Das kann sich in fehlenden (weil seltenen) Lesarten niederschlagen oder in inkongruenten Beschreibungen einzelner Lesarten oder Verwendungsweisen. So gibt das „Wörterbuch der deutschen Gegenwartssprache“ (WDG) im Artikel zum Verb PFEIFEN eine übertragene Verwendung in der Phrase „jmd. pfeift auf dem letzten Loch“ (mit der Bedeutungsparaphrase „jmdm. geht es sehr schlecht“) an; in den Korpora zur deutschen Gegenwartssprache lässt sich jedoch vergleichbar häufig eine Verwendung nachweisen, in der anstelle der Präposition „auf“ die Präposition „aus“ auftritt. Bei einer vollständigen kollokationsstatistischen Auswertung der Wörterbuchbasis (an die zum Entstehungszeitpunkt des WDG allerdings noch nicht zu denken war) träte diese Verwendungsweise zu Tage.
- Unterschiede in der Identifizierung einzelner lexikalischer Einheiten bewirken, dass sich diese Einheiten zwischen unterschiedlichen Ressourcen im Allgemeinen nicht stets uneindeutig (bijektiv) zueinander in Relation setzen lassen. Ein augenscheinliches Beispiel für

diesen Phänomenbereich ist die inkompatible Ansetzung von Homonymen. Auch eine nur in einer Richtung eindeutige Zuordnung ist zwischen zwei Ressourcen bereits nicht mehr gegeben, wenn in der Ausgangsressource lexikalische Einheiten auftreten, die in der Zielressource nicht verzeichnet sind (weil sie zum Beispiel zu selten in der Wörterbuchbasis auftreten). HEROLD u. a. (2012, 37ff.) diskutieren diese Problematik im Detail.

Nicht nur die Bündelung von Wörterbüchern in einem gemeinsamen Repräsentationssystem erzeugt ein komplexes DLS. Auch die Integration der Wörterbuchbasis (in Form von Korpora oder Belegsammlungen) oder relevanter Analyseverfahren des lexikografischen Prozesses, die die Genese der lexikografischen Information transparent und nachvollziehbar machen, führen zu komplexen Systemen. Die Nachvollziehbarkeit ist dabei durchaus nicht nur auf die „Begreifbarkeit“ des Prozesses auf der Grundlage dokumentierter Arbeitsschritte und Einzelergebnisse zu beziehen, sondern ebenso sehr auf den Umstand, dass der lexikografische Prozess durch die Nutzerinnen zumindest prinzipiell repliziert werden kann. Der Zugriff auf die Wörterbuchbasis und die zu ihrer Analyse verwendeten statistischen Werkzeuge ermöglicht es damit in einem solchen DLS praktisch jedermann, sich selbst als Lexikograf zu betätigen.

Die Informationen, die in einem DLS verfügbar sind, das neben den Resultaten lexikografischer Arbeit auch die zugrundeliegenden Daten und Verfahren für die Nutzer öffnet, sind notwendigerweise von unterschiedlicher Aussagekraft und Explizitheit in Bezug auf den Gegenstand ihrer Beschreibung. Während bloße Korpuskonkordanzen zwar alle Vorkommen eines Wortes in diesem Korpus lückenlos nachweisen, kann dabei jedoch keine direkt lexikografisch relevante Klassifizierung geleistet werden. Statistische Verfahren ermöglichen dagegen solche Klassifizierungen, weisen jedoch in der Regel systematische Unsicherheiten und Fehler auf. Durch menschliche Bearbeiter, die auf der Grundlage vereinbarter lexikografischer Richtlinien arbeiten, kann

schließlich eine intersubjektiv verlässliche und akzeptierte lexikografische Klassifikation und Beschreibung erzielt werden.

Sowohl durch die Gegenüberstellung verschiedener lexikalischer Ressourcen als auch durch die Offenlegung und Erschließung der Wörterbuchbasis und der lexikografischen Werkzeuge wird zudem eine kritische Auseinandersetzung mit dem Wörterbuch als Teil des DLS gefördert. Es verliert dadurch gegenüber seinen Nutzern teilweise den Status einer für sich stehenden – und nicht selten als normativ empfundenen – Ressource; sein sprachdokumentarischer Charakter tritt dagegen stärker hervor. Das Wörterbuch in Beziehung zu seiner untersuchbaren Basis macht deutlich, dass es eine mögliche, aber nicht die einzig mögliche Interpretation des Quellenmaterials ist. Es lädt dazu ein, als Ausgangspunkt für darüber hinaus gehende Untersuchungen zu dienen und von der Konsultation des Wörterbuchs zur Exploration des gesamten DLS überzugehen.

3. Das DWDS als digitales lexikalisches System

Das „Digitale Wörterbuch der deutschen Sprache“³ (DWDS, vgl. KLEIN/GEYKEN 2010) ist ein komplexes DLS in Hinsicht auf beide in Abschnitt 2 diskutierten Ausprägungen. Es ist zunächst ein Wörterbuchverbund aus den folgenden Ressourcen:

- DWDS-Wörterbuchkomponente (DWDSWB) auf der Grundlage der Retrodigitalisierung des WDG, einem allgemeinen Bedeutungswörterbuch, das von 1968 bis 1977 in sechs Bänden erschien und einen Umfang von ca. 120.000 Artikeln unterschiedlichen Darstellungsumfangs aufweist (HEROLD 2011). Das WDG wird in einer druckgetreuen Repräsentation ebenfalls vorgehalten.
- „Etymologisches Wörterbuch des Deutschen“ von Wolfgang Pfeifer (EtymWB),
- Erstausgabe des „Deutschen Wörterbuchs“ von Jacob und Wilhelm Grimm (¹DWB),

³ <http://www.dwds.de/>

- GermaNet⁴, ein lexikalisch-semantisches Wortnetz des Deutschen, sowie
- OpenThesaurus⁵ in Form eines Wortnetzes.

Alle lexikalischen Ressourcen des DWDS, mit Ausnahme von ¹DWB und WDG, werden aktiv bearbeitet und fortlaufend erweitert.

Ergänzt werden die lexikalischen Ressourcen des DWDS außerdem durch eine breite Auswahl von Korpora (Referenzkorpora, Zeitungskorpora sowie eine Reihe von thematisch spezialisierten kleineren Korpora), statistischen Analysewerkzeugen (graphische und numerische Häufigkeitsauswertungen, Kollokationsstatistiken, Wortprofile) und weitere lexikografische Hilfsmittel wie beispielsweise einem Extraktor für „gute Beispiele“, der Korpusbelege anhand konfigurierbarer linguistischer Bedingungen filtert und bewertet, um Lexikografinnen eine qualitativ hochwertige Auswahl von Belegen zu bieten, die möglichst das gesamte Verwendungsspektrum abdecken und dabei gleichzeitig das typische Verhalten eines Wortes illustrieren (DIDAKOWSKI u. a. 2012). Von den in Abschnitt 1 dargestellten Bereichen, in denen ein DLS potentiell über die Möglichkeiten von traditionellen Wörterbüchern hinausgeht, liegt der Fokus des DWDS demzufolge momentan vor allem auf dem Aspekt der Informationsgenerierung und Ressourcenintegration sowie auf dem Aspekt der Beschleunigung des lexikografischen Prozesses.

Als digitales lexikalisches System ist das DWDS nicht nur das Produkt lexikografischer Arbeit, sondern es dient gleichzeitig auch als primäres Recherche- und Analysemittel der Lexikografen des Vorhabens. Damit weist es die in Abschnitt 2 skizzierte Rekursion auf: Das DLS dient auch als Werkzeug zur Bearbeitung eines seiner Module und ist dadurch nicht nur Produkt, sondern gleichzeitig ein integrales Hilfsmittel des lexikografischen Prozesses. Anders gesagt: Der lexiko-

⁴ <http://www.sfs.uni-tuebingen.de/GermaNet/>

⁵ <http://www.openthesaurus.de/>

grafische Arbeitsplatz und das DLS gehen hier – jedenfalls zum Teil – ineinander über.

Diese Situation macht deutlich, dass eine explizite und möglichst genaue Kennzeichnung des jeweiligen Status der innerhalb des DLS gebotenen Informationen notwendig ist. Bei manchen Informationen handelt es sich um nicht lexikografisch klassifizierte und validierte „Rohdaten“. Die Daten des Wortprofils (DIDAKOWSKI/GEYKEN 2014) und die „guten Beispiele“ beruhen auf linguistisch motivierten, aber automatisch ablaufenden Verfahren. Die Daten der meisten angebotenen Wörterbücher sind das Resultat manueller lexikografischer Bearbeitung, allerdings in der Regel nicht auf der Basis der Korpora und Statistiken des DWDS.

Die Artikel des DWDSWB als zentraler Wörterbuchkomponente des DWDS schließlich vereinen teilweise Daten mit verschiedenen Status und unterschiedlicher Herkunft. Sie können daher in begrenztem Umfang bereits als synthetische Wörterbuchartikel angesehen werden: Im Zuge der Retrodigitalisierung des WDG konnten beispielsweise die Angaben zur Wortbildung von Komposita (und teilweise von Ableitungen) weitgehend automatisch aus der Neststruktur der Artikel gewonnen werden. Abb. 1 zeigt den Artikel HANDBALL in der momentan aktuellen Panel-Darstellung des DWDSWB. Die morphologische Zerlegung ist unterhalb der Lemmazeichengestaltangabe skizziert: Es handelt sich um eine Komposition aus HAND (als erstem Bestandteil) und BALL¹ (als zweitem Bestandteil). Auf der Grundlage dieser morphologischen Daten können nun Wortfelder dynamisch erzeugt werden, die sich in dieser Form im WDG nicht finden. Das Wortfeld, das dem Artikel HANDBALL zugeordnet ist, wird dadurch erzeugt, dass das gesamte Wörterbuch automatisch nach Artikeln durchsucht wird, die eine morphologische Zerlegung aufweisen, in der „Handball“ als zweiter Bestandteil auftritt (FELDHANDBALL, HALLENHANDBALL), als erster Bestandteil (HANDBALLFAN, HANDBALLSPIELER, ...) oder als Basis einer Ableitung (HANDBALLER).

Es werden stets nur die unmittelbaren morphologischen Konstituenten in einem Artikel kodiert. Eine vollständige morphologische

Analyse – etwa bei mehrgliedrigen Komposita – ist in dieser Repräsentationsform nicht notwendig, da sich die etwaige Zerlegung der Konstituenten wiederum in den entsprechenden Artikeln findet. Die wörterbuchinterne Konsistenz und Vollständigkeit der Wortfelder ist sichergestellt, solange für jeden Artikel korrekte morphologische Analysen vorliegen. Neu hinzugenommene Artikel führen bei der skizzierten Abfrage der Wörterbuchsubstanz sofort zu einer Erweiterung der entsprechenden Wortfelder. Grundlegende Voraussetzung für diese Wortfeldbildung ist die in Abschnitt 1 beschriebene Identifikation und Adressierung der Mikrostruktur eines DLS.



Abb. 1: Artikel HANDBALL (DWDSWB)

Die Mikrostrukturadressierung in einem DLS lässt sich auch für weitere phänomenbasierte und lokal beschränkte Artikelbearbeitungen ausnutzen. So wurden die Lemmazeichengestaltungsangaben für alle Artikel des DWDSWB gemäß der aktuell gültigen Rechtschreibnorm manuell

durchgesehen und gegebenenfalls überarbeitet sowie mit den jeweils einschlägigen Regeln annotiert. In einem parallel und unabhängig dazu durchgeführten Arbeitsgang wurden für einen Großteil der Artikel durch eine professionelle Sprecherin akustische Ausspracheangaben zu den Artikeln hinzugefügt. Weitere jeweils autonome Bearbeitungsschritte betrafen unter anderem die Expansion der Stichwortvorkommen in Belegen und Kompetenzbeispielen, sofern diese im WDG ursprünglich in abgekürzter Form auftraten; in Bedeutungsparaphrasen und Kompetenzbelegen wurde die Orthographie der aktuell gültigen Norm gemäß überarbeitet.

Der in Abb. 2 dargestellte Artikel SCHNITTSTELLE repräsentiert einen Artikeltyp aus der Lemmastrecke der Neubearbeitung des DWDSWB. Sogenannte Basisartikel wie SCHNITTSTELLE werden in einem halbautomatischen Verfahren arbeitsteilig aus einzelnen autonom erstellten Artikelteilen erzeugt. Die morphologische Zerlegung und das Flexionsparadigma werden mit computerlinguistischen Methoden ermittelt und anschließend doppelt manuell geprüft. Historische Rechtschreibvarianten sowie die entsprechenden Regelangaben werden manuell eingepflegt. Mit Hilfe des Extraktors für „gute Beispiele“ wird aus den Korpora des DWDS eine Menge von Belegen gewonnen und jeweils von einem Lexikografen qualitativ gefiltert, gegebenenfalls geschnitten und semantisch vorklassifiziert (vergleiche den Abschnitt „Textbelege sortiert nach vorläufiger Bedeutungsunterscheidung“ in der Abbildung). Das DWDS-Wortprofil wird schließlich genutzt, um typische Kollokationen zu ermitteln, die zukünftig den identifizierten Lesarten vorangestellt werden (in der Abbildung noch nicht dargestellt). Alle Arbeitsgänge zur initialen Datengewinnung und (teilweisen) Vorklassifizierung können wiederum parallel und unabhängig voneinander durchgeführt werden. Die aus der Synthese der Angaben resultierenden Basisartikel sind bereits auf der DWDS-Webseite verfügbar. Die abschließende lexikografische Bearbeitung, insbesondere die Sicherstellung artikelinterner und möglichst DLS-weiter Konsistenz, die diasystematische Annotation der Angaben, die Formulierung von Bedeutungsparaphrasen und das Vermerken wörterbuchinterner

sowie möglicherweise auch DLS-weiter (lexikalisch-semantischer) Verweise erfolgt zeitlich entkoppelt von den genannten Vorarbeiten zu einem späteren Zeitpunkt. Durch diese nachgelagerten Arbeitsschritte werden die Basisartikel zu umfassenden Vollartikeln ausgebaut.

Sowohl für die aus dem WDG übernommen Bestandsartikel als auch für die neu bearbeiteten Artikel werden die Quelle der Daten und der Zeitpunkt ihrer Hinzufügung durch das Artikelredaktionssystem bei der Bearbeitung automatisch vermerkt beziehungsweise aktualisiert. Dadurch können Nutzerinnen zukünftig zusätzlich zu der bereits jetzt gezeigten Versionsangabe des Wörterbuchs (vgl. die Fußzeilen in den Abbildungen: „Version: 0.4.23“) spezifisch für Mikrostrukturbereiche einzelner Artikel das Veröffentlichungsdatum sowie die Quelle der jeweiligen Angaben bei Bedarf feststellen.

The screenshot shows a window titled "DWDS-Wörterbuch" with a search bar and navigation icons. The main content area displays the entry for "Schnittstelle" (fem., -, -n) with the composition "Zusammensetzung mit: Stelle". Below this, it indicates "Artikeltyp: Basisartikel" and "Textbelege sortiert nach vorläufiger Bedeutungsunterscheidung:". A list of four text examples follows, each with a diamond icon and a plus sign for expansion. The examples are:

- Die Zentrale der Bundesbank in Frankfurt ist die einzige deutsche Schnittstelle mit der Europäischen Zentralbank. – Frankfurter Allgemeine Zeitung, 06.11.2001
- Was der digitalen Musik der Zukunft noch fehlt, ist eine Schnittstelle zum menschlichen Körper. – Frankfurter Allgemeine Zeitung, 09.09.2000
- Man hegt keine Vorurteile gegenüber anderen Bereichen und bemüht sich an Schnittstellen darum, den anderen die Arbeit nicht unnötig zu erschweren. – Kellner, Hedwig, Das geheime Wissen der Personalchefs, Frankfurt a.M.: Eichborn 1998, S. 168
- Auf den Schnittstellen sitzen Rosetten oder Palmettblüten. – o.A., D, in: Harald Olbrich (Hg.), Lexikon der Kunst Band 1: Cin - Gree, Leipzig: Seemann 1989, S. 7042

 The footer of the window shows "Version: 0.4.23 | Artikeltyp: Basisartikel" and "Kompakt | Details".

Abb. 2: Artikel SCHNITTSTELLE (DWDSWB)

4. Zusammenfassung und Ausblick

In einem komplexen DLS wie dem DWDS lassen sich Tendenzen zur Konvergenz von Wörterbuchbasis (in Form von Korpora), lexikografischer Rechercheumgebung und lexikografischem Produkt – bedingt durch die gemeinsame Präsentationsform – erkennen. Insbesondere durch die Transparenz der Wörterbuchbasis und die Verfügbarkeit von statistischen Analysefunktionen für die Nutzerinnen des DWDS geraten Fragen der systeminternen Konsistenz und Widerspruchsfreiheit in den Blick. Die Möglichkeit, einzelne Artikelteile autonom und gruppenweise zu bearbeiten, macht es notwendig, den Zeitpunkt der jeweiligen Änderungen explizit anzugeben, damit Nutzer sich einen Eindruck von der Aktualität der Angaben eines Wörterbuchartikels machen können.

Beim Blick auf komplexe DLS, die verschiedene Wörterbücher integrieren und bei denen versucht wird, interne Widersprüche und Inkonsistenzen zu minimieren, drängt sich die Frage auf, ob solche Systeme in Richtung eines umfassenden Wörterbuchtyps konvergieren können. HARM/SCHIEDER (2012) versuchen nachzuweisen, dass das zumindest für diachrone Wörterbücher, die Sprachwandelprozesse dokumentieren, nicht gilt: Diese lassen sich nicht aus einzelnen Sprachstadienwörterbüchern ableiten. Betrachtet man jedoch einen synchronen Schnitt, gibt es mit dem kollaborativ erstellten *Wiktionary* bereits den praktischen Versuch eines umfassenden gegenwartssprachlichen Wörterbuchs – multilingual in Bezug auf Objekt- und Metasprache (vgl. MEYER/GUREVYCH 2012).

Viele lexikografische Aufgaben müssen weiterhin manuell gelöst werden, weil sie momentan nicht zufriedenstellend automatisierbar sind. Dazu gehören insbesondere die Formulierung von Bedeutungsparaphrasen, die systematische Strukturierung des semantischen Raums und die lesartenspezifische Zuordnung von Belegen (RUNDELL/KILGARRIFF 2011).

Wo auch immer man die Anfänge der Lexikografie historisch verortet, umfasst die Periode der rechnergestützten Lexikografie im Vergleich bisher eine sehr kurze Zeitspanne. Es steht zu vermuten, dass dieser Zweig der Disziplin sich noch am Anfang seiner Entwicklung befindet und mit ihm seine Produkte, die digitalen lexikalischen Systeme.

Literatur

- DIDAKOWSKI, JÖRG/GEYKEN, ALEXANDER (2014): From DWDS corpora to a German Word Profile – methodological problems and solutions. In: ABEL, ANDREA/LEMNITZER, LOTHAR (Hrsg.): Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern. Mannheim (OPAL – Online publizierte Arbeiten zur Linguistik 2/2014), 39–47.
- DIDAKOWSKI, JÖRG/LEMNITZER, LOTHAR/GEYKEN, ALEXANDER (2012): Automatic example sentence extraction for a contemporary German dictionary. In: FJELD, RUTH VATVEDT/TORJUSEN, JULIE MATILDE (Hrsg.): Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo. Oslo, 343–349.
- GEYKEN, ALEXANDER (im Druck): Methoden bei der Wörterbuchplanung in Zeiten der Internetlexikographie. In: *Lexicographica* 30.2014 (im Druck).
- GRANGER, SYLVIANE/PAQUOT, MAGALI (Hrsg.) (2012): *Electronic Lexicography*. Oxford.
- HARM, VOLKER/SCHIEDER, MARCO (2011): Modul statt Monument? Zur Perspektive der historischen Lexikographie nach dem Abschluß der DWB-Neubearbeitung. In: *Jahrbuch für germanistische Sprachgeschichte* 2, 179–192.
- HEROLD, AXEL (2011): Retrodigitalisierung und Modellierung des Wörterbuchs der deutschen Gegenwartssprache. In: KRAFFT, ANDREAS/SPIEGEL, CARMEN (Hrsg.): *Sprachliche Förderung und Weiterbildung – transdisziplinär*. Frankfurt am Main (Forum angewandte Linguistik. 51), 197–213.
- HEROLD, AXEL/LEMNITZER, LOTHAR/GEYKEN, ALEXANDER (2012): Integrating Lexical Resources through an Aligned Lemma List. In: CHIARCOS, CHRISTIAN/NORDHOFF, SEBASTIAN/HELLMANN SEBASTIAN (Hrsg.): *Linked data in Linguistics. Representing and Connecting Language Data and Language Metadata*. Heidelberg [u. a.], 35–44.
- KLEIN, WOLFGANG (2004): Vom Wörterbuch zum Digitalen Lexikalischen System. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 136, 10–55.
- KLEIN, WOLFGANG/GEYKEN, ALEXANDER (2010): Das digitale Wörterbuch der deutschen Sprache (DWDS). In: *Lexicographica* 26, 79–93.
- MEYER, CHRISTIAN M./GUREVYCH, IRYNA (2012): Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In: GRANGER/PAQUOT (Hrsg.), 259–291.
- RUNDELL, MICHAEL/KILGARRIFF, ADAM (2011): Automatic creation of dictionaries: Where will it all end? In: MEUNIER, FANNY/DE COCK, SYLVIE/GILQUIN, GAETANELLE/PAQUOT, MAGALI (Hrsg.): *A Taste for Corpora*. Amsterdam/Philadelphia (Studies in Corpus Linguistics. 45), 257–281.

ANTJE TÖPEL

Die Beständigkeit von Wortartikeln im Onlinewörterbuch am Beispiel von *elexiko*

1. Das Projekt *elexiko* und seine Ansprüche zu Beginn

Seit mehr als zehn Jahren wird am Institut für Deutsche Sprache in Mannheim das Onlinewörterbuch *elexiko* (kurz für: elektronisches, lexikalisch-lexikologisches und korpusbasiertes Informationssystem) erstellt.¹ Dieses einsprachige Wörterbuch zum Gegenwartsdeutschen gehört zum Wörterbuchportal OWID.² Die Wörterbuchartikel für *elexiko* werden mithilfe der korpuslinguistischen und statistischen Methoden erarbeitet, die COSMAS II und die Kookkurrenzdatenbank CCDB zur Verfügung stellen.³ Die Textgrundlage bildet dabei das bis Ende 2013 regelmäßig aktualisierte *elexiko*-Korpus, das mit seinen rund 2,7 Milli-

¹ Die Planungen für das Projekt, das zuerst *LEXXIS* und anschließend *Wissen über Wörter* hieß, reichen bis in die Neunzigerjahre des vergangenen Jahrhunderts zurück (vgl. HAB-ZUMKEHR 2001a und 2004). Die ursprüngliche theoretische Konzeption von *elexiko* wird in HAB (Hrsg.) vorgestellt. Zur praktischen Umsetzung nach einigen Jahren Artikelarbeit vgl. KLOSA (Hrsg.). Einen kurzen Einblick in das Projekt bieten auch die Internetseiten unter <http://www.owid.de/wb/elexiko/start.html>. Diese drei Quellen behandeln auch einzelne Aspekte wie die Stichwortliste oder das *elexiko*-Korpus ausführlicher.

² Einzelheiten dazu, wie OWID konzipiert wurde und wie es sich entwickelt hat, werden in MÜLLER-SPITZER (2007) und (2010), in ENGELBERG u. a. sowie auf den Internetseiten des Projekts unter <http://www.owid.de> dargelegt.

³ Details zu COSMAS II und zur Kookkurrenzdatenbank CCDB finden sich auf den jeweiligen Internetseiten der Projekte unter <https://cosmas2.ids-mannheim.de/cosmas2-web/> bzw. <http://corpora.ids-mannheim.de/ccdb/>.

arden Textwörtern aus deutschen, österreichischen sowie Schweizer Zeitungs- und Zeitschriftentexten zusammengestellt ist.⁴

Die Stichwortliste von *lexiko* umfasst rund 300.000 Einträge, die in unterschiedlicher Tiefe bearbeitet werden. Der Ausbau von *lexiko* erfolgt dabei nicht von A bis Z, sondern modular.⁵ Alle Lemmata werden zunächst mit breit angelegten, einfachen Informationen (z. B. zur Orthografie oder Worttrennung) sowie drei automatisch ausgewählten Belegen ausgestattet. Begrenzte Wortschatzbereiche, aktuell das *Lexikon zum öffentlichen Sprachgebrauch*, erhalten umfangreiche Informationen in einer sehr komplexen Artikelstruktur. Solche redaktionell bearbeiteten Wortartikel schließen Angaben zur Bedeutung, Verwendung, Grammatik, Rechtschreibung und Wortbildung ein. Außerdem bieten gesprochensprachliche Belege die Möglichkeit, die Aussprache und Betonung des Stichwortes im Satzzusammenhang zu hören. Wo es sinnvoll ist, begleiten ein bis drei Illustrationen in einer oder mehreren Lesarten eines Stichwortes die Bedeutungserläuterung.

Das im Aufbau befindliche Onlinewörterbuch *lexiko* wird nicht nur modular erarbeitet, auch seine Veröffentlichung im Internet erfolgt Schritt für Schritt. Seit der Publikation der Stichwortliste und des Moduls *Demonstrationswortschatz* im Jahr 2004 werden neu erarbeitete Wortartikel zu Elementen des Moduls *Lexikon zum öffentlichen Sprachgebrauch* kontinuierlich veröffentlicht, sodass fast täglich neue Inhalte zum Wörterbuch hinzukommen.⁶

Der Start des Projektes *lexiko* fiel in eine Zeit, in der nicht auf Erfahrungen aus vergleichbaren Projekten zurückgegriffen werden konnte. Genuin neu erarbeitete, für das noch junge Medium Internet konzi-

⁴ Näheres zum dynamischen *lexiko*-Korpus findet sich im Abschnitt 3.1.

⁵ KLOSA (2011a, 17f.) beschreibt die modulare Erarbeitung von *lexiko* und einzelne Module wie das *Lexikon zum öffentlichen Sprachgebrauch* genauer.

⁶ Für diese Art von Wörterbüchern wird zum einen der auf STORRER/FREISE (1996, 117f.) zurückgehende Begriff *Aufbauwörterbuch* verwendet, zum anderen der Terminus *Ausbauwörterbuch*, der von SCHRÖDER (1997, 60) stammt und von STORRER (2001, 65) aufgegriffen wird. Der lexikografische Prozess bei Onlinewörterbüchern wird im Abschnitt 2.1 näher dargestellt.

pierte Wörterbücher des Gegenwartsdeutschen gab es noch nicht. Diese Umstände erklären zumindest teilweise die hohen Ansprüche, die *lexiko* sich stellte:

Wissenschaftliches Ziel des Projekts ist, auf dem aktuellen Stand der lexikografischen, lexikologischen und korpuslinguistischen Forschung und Methodik ein langfristig angelegtes Nachschlageinstrument zu Fragen des deutschen Gegenwartswortschatzes zu entwickeln, es auszubauen und immer wieder zu aktualisieren. (HAß 2005a, 1)

Zur permanenten Überarbeitung der Wörterbuchinhalte heißt es:

Bei *lexiko* handelt es sich prinzipiell um ein dynamisches System, das fortlaufend Änderungen, Verbesserungen, Verfeinerungen etc. unterworfen ist [...] – anders als im gedruckten Wörterbuch, in dem Neuerungen gebündelt erst in einer Folgeauflage berücksichtigt werden können. (SCHNÖRCH 2005, 107)

Die Möglichkeit, ein Internetwörterbuch kontinuierlich aktualisieren zu können, wird in der metalexikografischen Forschung grundsätzlich als ein Vorteil gegenüber Printwörterbüchern beschrieben (vgl. u. a. NESI 2000, 840f. und DEBUS-GREGOR/HEID 2013, 1004). LEMBERG äußert sich dazu folgendermaßen:

Das Publikationsmedium Internet bietet die Chance, von der Statik des Wörterbuchs auf eine Dynamik zu wechseln. Dynamik meint in diesem Fall fortgesetzte Korrektur- und Ergänzungsmöglichkeiten des Wörterbuchs auch in seinen bereits geschriebenen Teilen. (LEMBERG 2001, 82)

Im konkreten Arbeitsalltag sollte die Dynamik von *lexiko* ursprünglich unter anderem so umgesetzt werden, dass die Wörterbuchautoren nach Ablauf einer bestimmten Frist von einigen Jahren über eine Funktion der Datenbank an eine anstehende Aktualisierung ihrer alten Wörterbuchartikel erinnert werden sollten. Für diese Art der Erinnerung bürgerte sich unter den *lexiko*-Lexikografen die Bezeichnung *Wecker* ein:

Das dynamisch ausgebaute WiW-Korpus [Wissen-über-Wörter-Korpus, vorheriger Name von *lexiko*; A. T.] ermöglicht die Nachzeichnung des lexikalischen Wandels. [...] Aber nicht nur die beschriebene Sprache

wandelt sich, auch die Beschreibung, ihre Methode und die lexikografischen Erkenntnisse sind ja dem Wandel unterworfen. [...] In vorhersehbaren Fällen ist es möglich, eine Art Wecker in die Datenbank einzubauen, der die Lexikografen auffordert, zu überprüfen, ob z. B. *Bundeskanzlerin* immer noch auf die Inhaberin eines schweizerischen politischen Amtes, auf eine bundesdeutsche Utopie oder eines schönen Tages auf eine existente Person referiert. (HAB 2004, 325f.)

Diese Pläne wurden jedoch nie realisiert, da aus Mangel an Personal der Schwerpunkt auf der Neuerarbeitung von Stichwörtern lag. Zudem war angedacht, auch die Stichwortliste in einer besonderen Weise dynamisch zu gestalten: Die Entscheidung darüber, welche Lexeme als Stichwort und welche als morphologische Variante dazu verzeichnet werden sollten, sollte von der jeweiligen Frequenz der beiden möglichen Stichwörter abhängig gemacht werden. Eine variable Textgrundlage wie das *ellexiko*-Korpus (vgl. Abschnitt 3.1) hätte jedoch dazu geführt, dass bei jeder Änderung im Korpus auch überprüft hätte werden müssen, ob sich das Verhältnis von Stichwort und Variante inzwischen nicht umgedreht hat. Aus diesem Grund wurde die Entscheidung über Stichwort und morphologische Variante schließlich alphabetorientiert getroffen. Nichtsdestoweniger ist die Stichwortliste von *ellexiko* dynamisch gestaltet: So werden nicht verzeichnete Lexeme, die beispielsweise als paradigmatische Relationspartner eines bearbeiteten Wortes auftreten und im *ellexiko*-Korpus inzwischen ausreichend häufig vorkommen, nachlemmatisiert. Im umgekehrten Fall werden beispielsweise in der Stichwortliste verzeichnete Partizipien aus dieser entfernt, wenn sich bei der lexikografischen Bearbeitung des zugrunde liegenden Verbs zeigt, dass die Partizipien keine eigenständige, über die Semantik des Basisverbs hinausgehende Bedeutung und Verwendung aufweisen (vgl. ERB 2005).

Allein diese beiden Ideen für das Projekt *ellexiko* verdeutlichen, dass die ständige Aktualisierbarkeit eines Onlinewörterbuchs nicht nur Vorteile bringt. Im lexikografischen Alltag kann sie auch dazu verleiten, bereits publizierte Inhalte immer wieder zu überarbeiten, sie zu verbessern und auf den neuesten Stand zu bringen. Dies bindet personelle Ressourcen, die dann bei der Erarbeitung neuer Inhalte fehlen,

sodass letztendlich weniger zusätzliche Wortartikel entstehen und der eigentliche Ausbau des Wörterbuchs langsamer vorangeht (vgl. HAHN u. a. 2008, 82). Um dies zu verhindern, muss zu manchen Zeitpunkten der Projektarbeit die Entscheidung getroffen werden, Inhalte auf einem bestimmten Stand zu lassen und nicht ununterbrochen „nach hinten“ zu arbeiten. Dieses Phänomen ist eng mit dem lexikografischen Prozess bei Onlinewörterbüchern verknüpft, der im nächsten Abschnitt detaillierter beleuchtet wird.

2. Problemfelder im Kontext der Veränderbarkeit von Wortartikeln

2.1 Der lexikografische Prozess bei Onlinewörterbüchern

Für die Herstellung traditioneller Printwörterbücher ist der lexikografische Prozess gut erforscht und beschrieben (vgl. WIEGAND u. a. 2010, 8ff.). Er kann in die folgenden fünf Phasen eingeteilt werden: Vorbereitung, Datenbeschaffung, Datenaufbereitung, Datenauswertung sowie Satz- und Druckvorbereitung (vgl. WIEGAND u. a. 2010, 10ff.).

Auch bei einem computerunterstützten lexikografischen Prozess ändert sich an diesem Ablauf nichts Grundlegendes, außer dass an vielen Stellen Computer eingesetzt werden.⁷ Allerdings verändern sich die Inhalte der einzelnen Phasen zum Teil gravierend (vgl. WIEGAND u. a. 2010, 17ff.). So muss in die Vorbereitungsphase auch die Auswahl von Hard- und Software einbezogen werden. Bei der Datenbeschaffung und -aufbereitung kann auf digitale Korpora zurückgegriffen werden, die computerlinguistisch aufbereitet sind und z. B. über Annotationen verfügen. In der Phase der Datenauswertung unterstützen beispielsweise Kollokationsanalysen den lexikografischen Prozess. Je nach angestreb-

⁷ Ein lexikografischer Prozess, zielt – unabhängig davon, ob er computerunterstützt ist oder nicht – auf die Produktion von Wörterbüchern für Menschen. Im Gegensatz dazu werden durch computerlexikografische Prozesse sogenannte Computerlexika für die maschinelle Sprachverarbeitung hergestellt (vgl. WIEGAND u. a. 2010, 9).

tem Produkttyp wird in der Phase der Satz- und Druckvorbereitung die Druckvorlage erstellt (Printwörterbuch) oder die Präsentation der Daten in einer Benutzerschnittstelle ausgearbeitet (elektronisches Wörterbuch).

Bei Printwörterbüchern oder bei digitalen Wörterbüchern, die erst nach der Fertigstellung veröffentlicht werden, laufen die beschriebenen Phasen des lexikografischen Prozesses überwiegend nacheinander ab.⁸ Ganz anders sieht es bei Onlinewörterbüchern aus, die kontinuierlich publiziert werden (vgl. HAHN u. a. 2008, 81f.; WIEGAND u. a. 2010, 17 und KLOSA 2013, 518f.). In Anlehnung an KLOSA (2013, 519ff.) besteht der lexikografische Prozess bei im Aufbau befindlichen Onlinewörterbüchern aus diesen Phasen:

1. **Vorbereitung:** Entwurf, Planungen der Organisation (Finanzierung, Arbeitsablauf, Zeitplan – auch für die Publikation, Mitarbeiter) und der Inhalte (lexikografisches Handbuch zum Korpus, zu den Wörterbuchinhalten, zum Verfassen der Artikel, zur Auszeichnung der Daten; Erstellen von Musterartikeln; Planung der technischen Unterstützung; Entwurf des Onlinedesigns; Fragen der Benutzungsforschung), Pilotstudie,
2. **Datensammlung:** Auswahl der Primärquellen (Korpus), der Sekundärquellen (andere Wörterbücher) und der weiteren Quellen (z. B. Grammatiken), Beschaffung von Audio- und Videodateien, von Illustrationen, Diagrammen etc.,
3. **Computerisierung:** Annotierung und Lemmatisierung des Korpus, Programmierung eines Werkzeugs für die Korpussuche, Installation eines Redaktionssystems, Klärung von Datenabruf und -speicherung,
4. **Datenverarbeitung:** Erstellung einer Stichwortkandidatenliste, Ermittlung der Frequenzen der Stichwörter und Definition von Fre-

⁸ Werden aus derselben Datenbasis unterschiedliche lexikografische Produkte hergestellt (beispielsweise für verschiedene elektronische Geräte oder auch gedruckte und elektronische Wörterbücher), spricht man von einem medienneutralen lexikografischen Prozess bzw. von Cross-media-Publishing (vgl. WIEGAND u. a. 2010, 17 und DEBUS-GREGOR/HEID 2013, 1004).

- quenzschichten, Analyse der Kookkurrenzen, Spezifikation der Datenbankstruktur, Auszeichnung der Audio- und Videodateien,
5. **Datenanalyse:** automatisch erstellte Daten (z. B. Frequenz, Kollokationen, Belege), Verfassen der Wortartikel, Hyperlinks, Einbindung von Illustrationen, Hörbelegen usw.,
 6. **Vorbereitung für die Onlineveröffentlichung:** Korrekturlesen (inhaltlich und formal) sowie Testen (der Links, der multimedialen Elemente, der Onlinepräsentation, der Recherchemöglichkeiten), Verfassen und Verlinken der Wörterbuchaußentexte, Implementierung einer Führung.

Bei einem im Aufbau befindlichen Onlinewörterbuch gibt es kein konsequentes Nacheinander dieser Schritte, vielmehr gehen sie ineinander über. Wie KLOSA (2013, 522) ausführlich beschreibt, können sich verschiedene Module, aber auch einzelne Lemmata oder Angabebereiche ein und desselben Wörterbuchs zu einem bestimmten Zeitpunkt in ganz unterschiedlichen Phasen des lexikografischen Prozesses befinden. Außerdem verläuft der lexikografische Prozess bei digitalen Aufbauwörterbüchern nicht zwingend in nur eine Richtung: Durch veränderte lexikografische Ansprüche oder durch veränderte Ansprüche von außen (beispielsweise durch Erkenntnisse aus der Benutzerforschung) kann jederzeit in die Phase der Vorbereitung zurückgekehrt werden, wenn beispielsweise einzelne Angabebereiche konzeptionell überarbeitet werden. Dann beginnt der Kreislauf von Neuem: „While other lexicographic processes lead to an end (i. e. the publication of the dictionary), theoretically, working on an online dictionary under construction could go on forever“ (KLOSA 2013, 522).

Die mögliche Zirkularität des lexikografischen Prozesses bei Onlinewörterbüchern, die kontinuierlich publiziert werden, bildet eine notwendige Voraussetzung für einige Fragestellungen, die im Kontext der Veränderbarkeit von Wortartikeln auftreten können. Auf diese wird im Abschnitt 3 näher eingegangen.

2.2 Archivierung und Versionierung

Wenn Printwörterbücher inhaltlich oder äußerlich verändert werden sollen, muss eine neue Auflage erscheinen. Jeder Nutzer kann dadurch nachvollziehen, auf welchem Stand die enthaltenen Daten sind und sie im wissenschaftlichen Umfeld problemlos zitieren. Dies gilt auch noch für einen Teil der elektronischen Wörterbücher, die beispielsweise auf CD-ROM statisch gespeichert sind.

Onlinewörterbücher, die sich im Aufbau befinden, funktionieren vielfach anders: Zu jeder Zeit können Elemente in die Stichwortliste aufgenommen oder aus ihr entfernt werden, bereits verfasste Wortartikel können geändert oder gelöscht werden (vgl. u. a. KLOSA 2011a, 20).⁹ Diese Änderungen werden in vielen Fällen nicht kenntlich gemacht, wie auch HAB-ZUMKEHR (2001b, 378) kritisiert: „Welche Veränderungen wann vorgenommen wurden, welche Elemente herausgenommen und welche eingefügt wurden – alles dies ist durch Nutzer nicht mehr nachvollziehbar.“ In diesem Kontext sieht sie den Handlungsbedarf bei

den Autoren und Verwaltern, die nicht nur jede Bildschirmseite mit Datum und Verfasserangabe kennzeichnen könnten, sondern auch in der Lage wären, die jeweilige Version eines Nachschlagewerks zu jedem beliebigem [sic!] Zeitpunkt zu speichern und auch für Nutzer zugänglich zu archivieren. (HAB-ZUMKEHR 2001b, 378)

STORRER argumentiert in ähnlicher Weise und fordert:

Es muss auch in einem Ausbauwörterbuch gesichert sein, dass die abrufbaren Informationen verbindlich, verlässlich und zitierbar sind. Ein digitales Wörterbuch, das seine Wortartikel in vom Benutzer nicht nachvollziehbarer Weise verändert oder löscht, verliert für die Zwecke an Wert, bei denen es auf Verlässlichkeit und Zitierbarkeit ankommt. Werkzeuge und Verfahren der Versionenverwaltung können hier Abhilfe schaffen,

⁹ Die Veränderbarkeit von Onlinewörterbüchern thematisieren auch STORRER/FREESE (1996, 117f.), LEMBERG (2001, 82), HAHN u. a. (2008, 82) und KLOSA (2013, 523).

wenn die Probleme frühzeitig erkannt und berücksichtigt werden. (STORRER 2001, 65)

Das Bewusstsein dafür, auch bei Onlinewörterbüchern die unterschiedlichen Versionen zu kennzeichnen sowie frühere Entwicklungsstufen in einem Archiv zu speichern und den Nutzern zur Verfügung zu stellen, ist in den vergangenen Jahren innerhalb der Wörterbuchprojekte gestiegen. So beschäftigte sich das *Netzwerk Internetlexikografie* in seinem vierten Arbeitstreffen, das dem lexikografischen Prozess gewidmet war, auch mit Fragen der Aktualisierung und der Bestandssicherung bei Onlinewörterbüchern.¹⁰

Am dringendsten stellt sich die Frage der Versionierung bei kollaborativ erstellten Onlinewörterbüchern wie dem *Wiktionary*. Hier kann man sich die gesamte Versionsgeschichte eines Wortartikels anzeigen lassen und die verschiedenen Fassungen auch miteinander vergleichen lassen. Da beim *Wiktionary* jeder Interessierte an Wortartikeln mit-schreiben kann, gibt es vor allem bei brisanten Wörtern zum Teil un-entwegte Änderungen durch verschiedene Benutzer. Bei einem Teil des Wortschatzes findet sogar ein regelrechter Vandalismus statt, der zu Einschränkungen des Wiki-Prinzips führt.¹¹

Als Beispiele für redaktionelle Internetwörterbücher, die ihre Wortartikel bereits mit Versionsbezeichnungen versehen, können das *DWDS* oder das *OED* genannt werden. Im *DWDS* werden derzeit für drei Panels – das *DWDS*-Wörterbuch, das etymologische Wörterbuch und das Wortprofil 3.0 – die Fassungen in Versionsnummern wie „0.4.22“, „1.0.62“ und „3.0“ angegeben (vgl. Abb. 1).¹² Zwar ermöglichen diese Ziffern eine bessere Zitierbarkeit der Wörterbuchinhalte, leider bleiben

¹⁰ Das Programm des Treffens sowie Details zu dem durch die Deutsche Forschungsgemeinschaft geförderten Netzwerk, das seine Ergebnisse in Form eines Handbuchs veröffentlichen wird, finden sich auf den Projektseiten unter <http://www.internetlexikografie.de>.

¹¹ Solche Wortartikel wie *Nazi*, die regelmäßigem Vandalismus zum Opfer fallen, sind in ihrer Bearbeitung für nicht angemeldete oder neue Benutzer gesperrt. Diese Benutzer können den Wortartikel nur lesen und darüber diskutieren bzw. eine Entsperrung beantragen.

¹² Die Versionsnummer „3.0“ des Wortprofils 3.0 ist in Abb. 1 nicht sichtbar.

sie für den Benutzer ähnlich kryptisch wie die Versionsnummer einer Software. Auf welchem Stand die Daten sind, wann also Version „0.4.22“ genau aktualisiert wurde, wird nicht angegeben. In dieser Hinsicht ist die Angabe im Panel des *OpenThesaurus* hilfreicher – „2013-07-16“ weist klar auf das Datum der letzten Aktualisierung hin (vgl. Abb. 1).

The screenshot shows the DWDS interface for the word "Palme". The interface is divided into several panels. The top left panel shows the word "Palme" and its grammatical information. The middle left panel contains a description of the word and a list of related terms, with a version number "0.4.22" circled in red. The bottom left panel shows the "OpenThesaurus" section with a version number "2013-07-16" circled in red. The right panel shows the "Etymologisches Wörterbuch" section with a version number "1.0.62" circled in red. The bottom right panel shows the "Wortprofil 3.0" section with a search bar and a list of related words.

Abb. 1: Ausschnitt aus dem *DWDS*-Wortartikel *Palme* mit Versionsangaben [Markierungen durch mich; A. T.]

Das *OED* gibt für jedes Lemma Details zu seiner Aktualisierung an, im Beispiel *tree* (vgl. Abb. 2) ist in warnendem Rot zu lesen: „This entry has not yet been fully updated (first published 1914).“ Wortartikel neueren Datums wie *eco* erhalten in Blau den Hinweis „This is a new entry (OED Third Edition, March 2008).“, oder wie *liveware* einen Hinweis auf die letzte Aktualisierung „This entry has been updated (OED Third Edition, September 2009).“ Weiterhin kann der Nutzer auf die Vorgän-

gerversion des Artikels sowie auf eine Publikationsgeschichte zugreifen (vgl. Abb. 2).

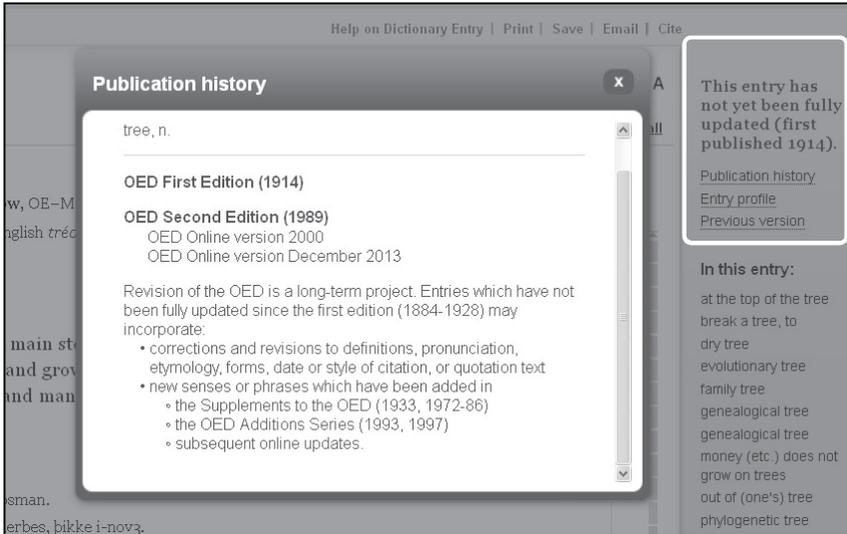


Abb. 2: Ausschnitt aus dem *OED*-Wortartikel *tree* mit Angaben zur Aktualisierung und zur Publikationsgeschichte [Markierung durch mich; A. T.]

STORRER weist allerdings zu Recht darauf hin, dass sich die Zitierbarkeit eines Onlinewörterbuchs durch eine Versionierung nur dann verbessert, „wenn die Probleme frühzeitig erkannt und berücksichtigt werden“ (STORRER 2001, 65). Denn um eine Kennzeichnung unterschiedlicher Stadien oder Fassungen von Wortartikeln vornehmen zu können, muss diese auch in den Daten erfasst sein. Eine Angabe von Versionen ist vor allem bei Wörterbuchprojekten sinnvoll, die über einen längeren Zeitraum Wortartikel erarbeiten und die währenddessen vielleicht sogar ihre Methodik verändern.

Für das Onlinewörterbuch *lexiko* wurde der dringende Bedarf einer Versionierung der Wörterbuchartikel und der Archivierung von älteren Fassungen zwar bereits vor einigen Jahren festgestellt (vgl.

HAHN u. a. 2008, 83), allerdings wurden diese Probleme für das aktuell bearbeitete Modul *Lexikon zum öffentlichen Sprachgebrauch* noch immer nicht gelöst. Da diese Frage jedoch alle in OWID enthaltenen Wörterbücher betrifft, wird hierzu zukünftig eine wörterbuchübergreifende Portallösung angestrebt.

Im Zusammenhang mit der Frage der Versionierung muss für die Onlinelexikografie noch festgelegt werden, was genau als neue Version eines Wörterbuchartikels gilt. Denn von echten Änderungen in den Wörterbuchinhalten abgesehen kann die Darstellung eines Artikels bereits allein durch den verwendeten Browser oder das benutzte Gerät stark variieren.¹³ Bei einer entsprechenden Datengrundlage, wie sie beispielsweise *ellexiko* besitzt, ist es zudem technisch sehr leicht möglich, bestimmte Angaben online freizuschalten, nicht anzuzeigen oder in einem veränderten Design zu präsentieren (vgl. MÜLLER-SPITZER 2011).

3. Ursachen für die Veränderbarkeit von Wortartikeln in *ellexiko*

3.1 Monitorkorpus

Die Grundlage für das korpusbasierte Erstellen von Wortartikeln in *ellexiko* bildet das digitale *ellexiko*-Korpus, das systematisch aus dem am Institut für Deutsche Sprache erstellten *Deutschen Referenzkorpus (DeReKo)* aufgebaut wurde.¹⁴ Fast über den gesamten Zeitraum, in dem

¹³ Einige Onlinewörterbücher, beispielsweise das *DWDS* oder das *Wiktionary*, bieten bereits „mobile“ Ansichten ihrer Wörterbuchinhalte an. Solche reduzierte Ansichten nutzen oft explorerartige Darstellungen und sind besonders gut für vergleichsweise kleine Displays (wie von Handys) geeignet.

¹⁴ Der Aufbau und die Zusammensetzung des *ellexiko*-Korpus in der Anfangsphase des Projekts 2004 wird in STORJOHANN (2005a) beschrieben. Die aktuelle Zusammensetzung wird unter dem Stichwort *ellexiko-Korpus* im Glossar auf den Projektseiten vorgestellt. Einzelheiten zu *DeReKo* lassen sich auf den Internetseiten des Projekts unter <http://www.ids-mannheim.de/kl/projekte/korpora/> nachlesen.

Wortartikel verfasst wurden, war das *lexiko*-Korpus dynamisch, also ein Monitorkorpus.

Der Begriff des Monitorkorpus wurde von SINCLAIR geprägt (vgl. SINCLAIR 1982, 4). Es zeichnet sich vor allem dadurch aus, dass es in seinem Umfang und seiner Zusammensetzung veränderlich ist, dass somit neue Texte hinzukommen, aber auch alte Texte entfernt werden. SINCLAIR schreibt mit Bezug auf das Englische: „the corpus will have a large and up-to-date selection of current English available; it will have a historical dimension, and it will have a comprehensive word list because of its elaborate record-keeping“ (SINCLAIR 1991, 25). Damit sind Monitorkorpora als lexikografische Arbeitsgrundlage besonders gut geeignet (vgl. MCENERY/WILSON 2001, 30), denn auf diese Weise können

u. a. schnell neue Lexeme entdeckt, Bedeutungsveränderungen nachgewiesen, kontextuelle Neuerungen erkannt oder Veränderungen hinsichtlich des Vorkommens aufgespürt werden. Wandelt sich die Bedeutung eines Stichwortes oder sein Gebrauch, kann vor allem der Zeitpunkt und die Art einer solchen Veränderung in einem Monitorkorpus gut und schnell erkannt werden. (STORJOHANN 2005a, 56f.)

Die Modifikationen in einem dynamischen Korpus wirken sich demzufolge einerseits auf die Inhalte der einzelnen Wortartikel aus, wenn sich die Bedeutung und Verwendung eines Wortes ändern (vgl. Abschnitt 3.3). Andererseits ist davon auch die Stichwortliste betroffen, wenn durch die Aufnahme zusätzlicher Texte neue Lexeme belegt werden oder aber durch das Entfernen anderer Texte bestimmte Lexeme aus dem Korpus verschwinden.

Das *lexiko*-Korpus wurde im Jahr 2013 zum letzten Mal ergänzt. Seitdem wird es nicht mehr verändert, um bestimmte lexikografische Abschlussarbeiten für die Beendigung des Moduls *Lexikon zum öffentlichen Sprachgebrauch* Ende 2014 zu ermöglichen. Leider ist das *lexiko*-Korpus wegen urheberrechtlicher Gründe bis heute nicht als Ganzes, sondern nur in Teilen öffentlich zugänglich.

Vergleicht man die ersten Korpora von 2003 und 2004, die für *lexiko* zusammengestellt wurden und auf deren Grundlage das Modul

„*ellexiko*-Demonstrationswortschatz“ erarbeitet wurde, mit dem aktuellen und letzten Stand des *ellexiko*-Korpus von 2013, fallen einige Unterschiede auf.¹⁵ Die ältesten enthaltenen Texte stammen heute aus dem Jahr 1946, die jüngsten von 2012. Damit hat sich die Zeitspanne der einbezogenen Texte von ursprünglich 52 bzw. 54 Jahren (2003: Texte von 1949 bis 2001; 2004: Texte von 1949 bis 2003) auf nun 66 Jahre etwas vergrößert, der Fokus liegt allerdings immer noch auf aktuellen Texten. Für das *ellexiko*-Pilotkorpus war ein Umfang von einer Milliarde laufender Wörter angestrebt, es umfasste dann 1,19 Milliarden Textwörter im Jahr 2003 bzw. 1,27 Milliarden Textwörter im Jahr darauf (vgl. STORJOHANN 2005a, 65). Im Laufe der Zeit ist das Korpus schließlich auf mehr als 2,7 Milliarden laufende Wörter im Jahr 2013 angewachsen. Der Umfang hat sich also mehr als verdoppelt. Das Ausgangskorpus von 2003 bestand zu 59,25 Prozent aus bundesdeutschen, zu 27,21 Prozent aus österreichischen und zu 13,54 Prozent aus Schweizer Textwörtern (bzw. 2004: 61,83 Prozent bundesdeutsche, 25,48 Prozent österreichische und 12,68 Prozent Schweizer Textwörter). In Orientierung an die Zahlen der Deutschsprecher in den einzelnen Ländern sollte auf eine Verteilung von 70 Prozent bundesdeutschen, 20 Prozent österreichischen und 10 Prozent Schweizer Textwörtern hingearbeitet werden (vgl. STORJOHANN 2005a, 61). Dieses Ziel wurde erreicht, denn die endgültigen Anteile liegen 2013 bei 71,83 Prozent bundesdeutschen, 16,7 Prozent österreichischen und 11,47 Prozent Schweizer Textwörtern. Die Zahl der im *ellexiko*-Korpus enthaltenen Quellen hat sich von 19 in den Jahren 2003 und 2004 auf 28 im Jahr 2013 erhöht.¹⁶ Ein weiteres Ziel war es, die enthaltenen Zeitungen regional besser zu verteilen und mehr ost- und norddeutsche Texte

¹⁵ Das erste *ellexiko*-Korpus wurde 2003 zusammengestellt. Das *ellexiko*-Korpus von 2004, das in STORJOHANN (2005a) näher beschrieben wird, wurde im Vergleich zu diesem Korpus von 2003 bereits verändert, vor allem durch die Neuaufnahme neuerer Jahrgänge der Zeitungen und Zeitschriften.

¹⁶ Dabei gilt eine Zeitung oder Zeitschrift, beispielsweise *Die Zeit*, als eine Quelle.

aufzunehmen, um die Dominanz süd- und mitteldeutscher Presseergebnisse aufzubrechen (vgl. STORJOHANN 2005a, 68). Hier konnte durch die Akquise neuer Quellen für das *Deutsche Referenzkorpus* Abhilfe geschaffen werden. Im Vergleich zum *lexiko*-Ausgangskorpus von 2003 (und 2004) sind im endgültigen *lexiko*-Korpus elf neue Quellen enthalten, dafür fehlen das *Bonner Zeitungskorpus*¹⁷ und die österreichische *Neue Kronen-Zeitung*. Insgesamt besteht das endgültige *lexiko*-Korpus aus folgenden Quellen: *Berliner Morgenpost*, *Berliner Zeitung*, *Braunschweiger Zeitung*, *Der Spiegel*, *die tageszeitung*, *Die Welt*, *Die Zeit*, *dpa-Meldungen*, *Frankfurter Allgemeine Zeitung*, *Frankfurter Rundschau*, *Hamburger Morgenpost*, *Handbuchkorpus 1985–1988*¹⁸, *Hannoversche Allgemeine Zeitung*, *Mannheimer Morgen*, *Neues Deutschland*, *Rhein-Zeitung*, *stern*, *Burgenländische Volkszeitung*, *Die Presse*, *Kleine Zeitung*, *Niederösterreichische Nachrichten*, *Oberösterreichische Nachrichten*, *Salzburger Nachrichten*, *Tiroler Tageszeitung*, *Vorarlberger Nachrichten*, *Die Südostschweiz*, *St. Galler Tagblatt* und *Zürcher Tagesanzeiger*.

Ein dynamisches Korpus wie das *lexiko*-Korpus kann durch die darin vorgenommenen Veränderungen eine Ursache dafür sein, dass nach einer gewissen Zeit die Inhalte eines Wortartikels von den Inhalten des Korpus abweichen. In welchem Maß das bei den *lexiko*-Artikeln der Fall ist, wird in Abschnitt 4 gezeigt.

¹⁷ Das *Bonner Zeitungskorpus* umfasst ca. 3,1 Millionen Textwörter aus den Zeitungen *Neues Deutschland* und *Die Welt* aus verschiedenen Jahrgangsquerschnitten von 1949 bis 1974. Näheres dazu findet sich unter <http://www.ids-mannheim.de/kl/projekte/korpora/archiv/bzk.html> auf den Projektseiten von *DeReKo*.

¹⁸ Im *Handbuchkorpus 1985–1988* sind rund elf Millionen Textwörter aus den Jahren 1985 bis 1988 enthalten, überwiegend aus den Zeitungen *Mannheimer Morgen* und *Die Zeit* sowie aus der Zeitschrift *stern*. Weitere Informationen dazu enthalten die Projektseiten von *DeReKo* unter <http://www.ids-mannheim.de/kl/projekte/korpora/archiv/hbk.html>.

3.2 Lexikografische Erfahrungen

Neben den Veränderungen in einem Korpus können vor allem bei mittel- und langfristigen lexikografischen Projekten auch die Erfahrungen der beteiligten Lexikografen dazu führen, dass sich die Art und Weise, wie Wortartikel verfasst werden, über die Zeit verändert. So kann das anfängliche Bedürfnis des Lexikografen, als interessant wahrgenommene sprachliche Phänomene eines zu beschreibenden Stichwortes im Detail zu analysieren sowie diese im Wortartikel zu kommentieren, nach und nach nachlassen oder sich schließlich ganz verlieren. Durch das tägliche Umgehen mit einem zu Beginn neuen Werkzeug wie der bei *elexiko* genutzten Kookkurrenzanalyse von COSMAS II sowie mit einer großen Menge an Belegen kann bei den Lexikografen über die Zeit eine Art Gewöhnungseffekt eintreten, was ehemals Bemerkenswertes betrifft. Werden bestimmte Erscheinungen wie Gebrauchsbesonderheiten beim ersten Mal vielleicht noch beschrieben oder kommentiert, kann das gleiche Phänomen beim wiederholten Auftreten als ganz normal und nicht erwähnenswert empfunden werden.

Typische Stellen, die je nach lexikografischer Erfahrung eines Wörterbuchautors besonders unterschiedlich ausfallen können, sind deshalb z. B. die lexikografischen Kommentare, die Angaben zu pragmatischen Besonderheiten und die Bedeutungsgliederung. Bei der Bedeutungsgliederung kann durch die längere Arbeit an einem Wörterbuch die Tendenz zunehmen, Lesarten stärker zusammenzufassen und das Bedeutungsspektrum weniger stark auszudifferenzieren. Bei den lexikografischen Kommentaren und den Angaben zu pragmatischen Besonderheiten kann die Gewöhnung des Lexikografen an die Wörterbucharbeit zur Folge haben, dass bestimmte Beobachtungen im Wortartikel weniger ausführlich oder gar nicht mehr dargestellt werden.

LEMBERG sieht die längere Arbeit an einem Wörterbuch positiv, da „sich bei den Lexikographinnen und Lexikographen im Laufe der Bearbeitungszeit auch das Wissen um den Beschreibungsgegenstand und damit die lexikographische Beschreibungscompetenz“ (LEMBERG 2001, 81) erhöhen. Jedoch kann nicht nur das Maß an lexikografischer Kompetenz und Erfahrung, sondern auch eine gewisse „Tagesform“ des

Lexikografen dazu führen, dass man mit einem gewissen zeitlichen Abstand einen selbst verfassten Artikel am liebsten umschreiben würde. REICHMANN beschreibt diesen Fall wie folgt:

Die Stunde der Ernüchterung schlägt in der Regel dann, wenn man bei einem einige Monate nach der Artikelformulierung vollzogenen Korrekturgang oder bei späterem Nachschlagen im eigenen Wörterbuch seinen eigenen Text nicht mehr versteht oder jedenfalls in Teilen nicht mehr nachvollziehen kann. Die verzweifelte Frage „Was habe ich da bloß gemacht?“ wird keinem Lexikographen fremd sein. [...] [Es] bleibt immer mal wieder das Gefühl, bestimmte Artikel lieber nicht selbst geschrieben zu haben. (REICHMANN 2012, 218)

Neben diesen intraindividuellen Differenzen bestehen zusätzlich noch interindividuelle Unterschiede zwischen den an einem lexikografischen Projekt beteiligten Mitarbeitern. Kein Redaktionshandbuch, sei es auch noch so ausführlich, kann durch redaktionelle Vorgaben vollständig verhindern, dass Wörterbuchartikel unterschiedliche Handschriften tragen, die von der Erfahrung, der „Tagesform“ und der Persönlichkeit des Lexikografen abhängen.

An welchen Stellen sich die Erfahrung eines Lexikografen beim Schreiben eines Wortartikels wie stark niederschlägt, ist im Einzelfall schwer nachzuweisen. Die Veränderungen, die sich nach fünf Jahren lexikografischer Praxis grundsätzlich für *lexiko*-Artikel ergeben haben, sind in KLOSA (Hrsg.) ausführlich beschrieben.

3.3 Mikrodiachronie

Werden Wörterbücher über einen längeren Zeitraum hinweg verfasst, können ihre Inhalte durch Prozesse des Sprachwandels eingeholt werden. Wenn bei einer Überarbeitung solcher Wortartikel die synchron nicht mehr gültigen Angaben in einen Angabebereich zur Bedeutungsgeschichte überführt werden, erhält das primär synchron orientierte Wörterbuch diachrone Züge.

Solche Angaben zur jüngeren sprachgeschichtlichen Entwicklung eines Wortes werden in der Mikrodiachronie erfasst. Sie müssen aller-

dings nicht durch die Bearbeitung eines veralteten Wortartikels entstehen, sondern können sich bei der Neubearbeitung eines Wortes auch allein durch die Analyse eines chronologisch gestaffelten Korpus bzw. eines Monitorkorpus ergeben (vgl. Abschnitt 3.1).

In *elexiko* wird die Mikrodiachronie lesartenübergreifend im Bereich *Herkunft und Wandel* unter der Überschrift *Wandel seit 1945* dargestellt (vgl. HAB 2005a, 11 und STORJOHANN 2005b, 319ff.).¹⁹ Hierfür wird das *elexiko*-Korpus daraufhin überprüft, ob sich die Befunde der Kookkurrenzanalyse von COSMAS II diachron interpretieren lassen. Falls sich jüngere sprachliche Veränderungen zeigen, werden diese narrativ beschrieben, die Schwerpunkte liegen dabei auf dem Bedeutungswandel, der besonders die Veränderungen im Wortgebrauch dokumentiert, und der Diskursgeschichte, die Unterschiede in der diskursiv-thematischen Verwendung des Wortes dokumentiert. Aber auch Unterschiede zum synchronen Zustand beim Zusammenhang der Lesarten, beim Referenzbereich und bei den Kollokationen werden erfasst.

Ein eindrückliches Beispiel aus dem *Lexikon zum öffentlichen Sprachgebrauch*, das über mikrodiachrone Angaben verfügt, ist das Lemma *Aids*. Hier werden die diskursiven Besonderheiten dargestellt, die das Stichwort in den 1980er-Jahren im Vergleich zu Korpus-texten jüngeren Datums aufweist. Die Analyse der Kollokationen offenbart für diesen Zeitraum Sets wie „Wie wird Aids charakterisiert? – ansteckend, todbringend, tödlich, übertragbar, unheilbar“, „Wie wird Aids politisch bekämpft? – Maßnahmenkatalog, Meldepflicht, Zwangsmaßnahme, Zwangsmeldung, Zwangstest“ und „Womit wird Aids verglichen? – Epidemie, Pandemie, Pest, Seuche, Volksseuche, Weltseuche“. In der Diskursgeschichte werden die Ergebnisse in erzählender Form präsentiert: Ausführlich wird der Bogen von der Ausweitung der Gruppe der Betroffenen über die anfängliche Stigmatisierung der Erkrankten in den Medien bis hin zur Diskussion politischer Maßnahmen zur Bekämp-

¹⁹ Weitere Details zur Mikrodiachronie finden sich auch im *elexiko*-Glossar unter dem Stichwort *Wandel seit 1945* sowie den untergeordneten Stichwörtern auf den Internetseiten unter <http://www.owid.de/wb/elexiko/start.html>.

fung der Krankheit gespannt und jeweils mit geeigneten Korpusbelegen untermauert.

Inwieweit bei einer Aktualisierung von Wortartikeln die bisherige Beschreibung ergänzt werden kann oder wann Daten aus der synchronen Beschreibung in die Mikrodiachronie verschoben werden müssen, wird im folgenden Abschnitt diskutiert.

4. Einige Beispiele aus *lexiko* – Wortartikel gestern und heute

Die ersten in *lexiko* verfassten Wortartikel gehören zum Modul *Demonstrationswortschatz* und stammen aus dem Jahr 2003. Sie sind inzwischen also mehr als zehn Jahre alt. Wie wirken sich die oben beschriebenen Faktoren wie das Monitorkorpus (vgl. Abschnitt 3.1) und die Mikrodiachronie (vgl. Abschnitt 3.3) nach einem Zeitraum von rund zehn Jahren auf die Inhalte der älteren *lexiko*-Artikel aus?

Das *lexiko*-Korpus ist in seiner Schlussversion von 2013 mehr als doppelt so groß wie das Ausgangskorpus aus dem Jahr 2003. Deshalb ist es nicht verwunderlich, dass die Stichwörter 2013 viel häufiger belegt sind als 2003 oder 2004.²⁰ In welchem Maß die Anzahl der Belege in diesem Zeitraum zugenommen hat, ist jedoch vom einzelnen Wort abhängig. Das Lemma *kreativ* ist beispielsweise 2004 24.373-mal und 2013 86.445-mal belegt, also 3,5-mal so häufig. Auch die Belegmenge von *urkomisch* (720 Belege 2004, 2.184 Belege 2013) wächst mit der Verdreifachung der Belege überdurchschnittlich stark. Im Gegensatz dazu erhöht sich die Anzahl der Belege bei Stichwörtern wie *Gewähr* (7.835 Belege 2004, 10.569 Belege 2013, Steigerung um 35 Prozent) oder *Innenrand* (22 Belege 2004, 28 Belege 2013, Steigerung um 27 Prozent) im gleichen Zeitraum weniger stark.

²⁰ Die Wörter aus dem Modul *Demonstrationswortschatz* wurden 2003 und 2004 lexikografisch bearbeitet. In dieser Zeit hat sich das *lexiko*-Korpus bereits verändert (vgl. Abschnitt 3.1). Da die gewählten Beispiele alle 2004 erarbeitet wurden, erfolgt der Vergleich im Folgenden zwischen den *lexiko*-Korpora von 2004 und 2013.

In das *lexiko*-Korpus sind in der Zwischenzeit nicht nur neue Texte aufgenommen worden, einige Texte wurden auch daraus entfernt. Das Korpus von 2004 ist also nicht vollständig im Korpus von 2013 enthalten. Das kann dazu führen, dass Belege aus dem Ausgangskorpus heute nicht mehr zum *lexiko*-Korpus gehören. Für einige niedrigfrequente Stichwörter wurde überprüft, inwieweit sich die Beleglage im *lexiko*-Korpus zu den beiden Zeitpunkten deckt. Beim Lemma *Gartengott* (11 Belege 2004, 13 Belege 2013) gehören alle elf Belege von 2004 auch zum Korpus von 2013, lediglich zwei neue Belege sind dort hinzugekommen. Von den 22 Belegen des Stichwortes *Innenrand* im Jahr 2004 sind 15 auch im 2013er Korpus enthalten, in dem *Innenrand* insgesamt 28-mal belegt ist. Das Stichwort *zulässigerweise* hat im Ausgangskorpus 20 Belege, von denen 13 auch im Korpus von 2013 vorkommen (2013 ist *zulässigerweise* insgesamt 54-mal belegt). Das Lemma *hohnvoll* ist 2004 22-mal belegt, 13 dieser Belege finden sich auch im 2013-er Korpus (dort gibt es für *hohnvoll* insgesamt 29 Belege). Der Fall *hohnvoll* ist auch ein gutes Beispiel dafür, dass Belege, die im Wortartikel angeführt werden und die im *lexiko*-Ausgangskorpus enthalten waren, nicht mehr Bestandteil des *lexiko*-Korpus in seiner Schlussversion von 2013 sind. Von den sechs Belegen (drei bei der Bedeutungserläuterung, einer bei den Gebrauchsbesonderheiten und zwei im Angabebereich *Grammatik*), die zum Wortartikel *hohnvoll* gehören, sind drei nicht mehr im 2013er Korpus enthalten (der Beleg aus den Gebrauchsbesonderheiten und die beiden Belege aus dem Angabebereich *Grammatik*). Dies spielt für den Nutzer jedoch keine Rolle, da das *lexiko*-Korpus nicht öffentlich zugänglich ist. Die Lexikografen – und auch die Nutzer – könnten die inzwischen fehlenden Belege bei Bedarf noch im *Deutschen Referenzkorpus (DeReKo)* finden, aus dem das *lexiko*-Korpus zusammengestellt wurde.

Durch das wachsende Monitorkorpus, in das kontinuierlich neue Jahrgänge von Zeitungen und Zeitschriften aufgenommen werden, können Prozesse des Sprachwandels besonders gut dokumentiert werden (vgl. Abschnitt 3.1 und 3.3). So kann es auch nach einem relativ kurzen Zeitraum – wie hier die knapp zehn Jahre von 2004 bis 2013 –

dazu kommen, dass Angaben in einem Wortartikel ergänzt werden müssten. Für das Lemma *hochmobil* aus dem Demonstrationswortschatz war die Beleglage für eine adverbiale Verwendung im damaligen *ellexiko*-Korpus nicht ausreichend. Im Korpus von 2013 sieht das anders aus – hier finden sich einige Belege, die eine adverbiale Verwendung von *hochmobil* im Satz zeigen. Sie stammen, bis auf eine Ausnahme, alle aus der Zeit nach 2004:

- (1) Lesart ‚ortsungebunden‘: Das Märchen vom Hasen und Igel lehre, so Geißler, dass „jene rasch zu Tode kommen, die **hochmobil** zwischen den Zielen hin und her rasen, und es denen gutgeht, die zu zweit sind und die sitzen bleiben können“. (Die Zeit, 22.12.1999, Luxusartikel Zeit)
- (2) Lesart ‚ortsungebunden‘: Gestern, am Pfingstsamstag, sind auf der Malanseralp im Calfeisental ein drittes Mal Bartgeier ausgesetzt worden. [...] **Hochmobil** streifen sie in elegantem Flug mal hierhin, mal dorthin, auf der Suche nach Nahrung, nach Altersgenossen oder Thermik. (Die Südschweiz, 27.05.2012, Gemeinsam zum Ziel)
- (3) Lesart ‚flexibel‘: Die Schweiz braucht keinen Mentalitätswandel, den Architekten und Planer predigen. Viel mehr braucht sie Vordenker, die sehen: Die Schweizer verhalten sich wie Städter – sie wollen nur nicht so genannt werden. Sie leben **hochmobil** und vernetzt. Sie träumen nicht vom Leben in der Pampa. (Die Zeit, 31.05.2012, Die Zufriedenen)
- (4) Lesart ‚flexibel‘: Carmen Losmann hat einen zutiefst beunruhigenden Film über moderne Arbeitswelten gedreht. [...] Der Film heftet sich an die Fersen einer High-Tech-Arbeiterschaft, die **hochmobil** und leidenschaftlich ihre Arbeit zum Leben machen soll. (Mannheimer Morgen, 17.11.2012, Arbeit zum Leben machen?)

Bei einer Überarbeitung des Wortartikels müsste in beiden Lesarten im Angabebereich *Grammatik* die adverbiale Verwendung von *hochmobil* ergänzt und belegt werden. Bei einem Ausbau der Angaben in der Mikrodiachronie und einer Einbeziehung von Angaben zum grammatischen Wandel wäre zu überlegen, ob die allein attributive und prädikative Verwendung unter der Überschrift *Wandel seit 1945* beschrieben

werden sollte, da die adverbiale Verwendung erst seit wenigen Jahren im *elexiko*-Korpus belegt ist.²¹

Das dynamische *elexiko*-Korpus mit seinen wechselnden Inhalten wirkt sich auch auf die Ergebnisse der durch COSMAS II bereitgestellten Kookkurrenzanalyse aus. Exemplarisch werden hier die Kookkurrenzen des Lemmas *Bevölkerung* von 2004 und 2014 (*elexiko*-Korpus auf dem Stand von 2013) miteinander verglichen. 2004 ist *Bevölkerung* im *elexiko*-Korpus 167.517-mal belegt. Durch COSMAS II werden insgesamt 324 Kookkurrenzen präsentiert. Zehn Jahre später gibt es im *elexiko*-Korpus für *Bevölkerung* mehr als doppelt so viele Belege (338.803) und COSMAS II stellt 1.141 Kookkurrenzen zur Verfügung.²²

Tabelle 1 stellt jeweils die fünfzig Kookkurrenzen zu *Bevölkerung* aus den Jahren 2004 und 2014 gegenüber, die statistisch am signifikantesten sind. Nicht hervorgehoben sind die Kookkurrenzen, die in beiden Jahren unter den ersten fünfzig Kookkurrenzen zu finden sind. Grau unterlegt sind jeweils die Kookkurrenzen, die im Vergleichsjahr nicht zu den ersten fünfzig Kookkurrenzen gehören oder aber gar nicht in der Kookkurrenzliste auftauchen. Der entsprechende Platz aus dem Vergleichsjahr wird in Klammern aufgeführt. Insgesamt finden sich 28 der 50 ersten Kookkurrenzen in beiden Jahren, die übrigen 22 unterscheiden sich voneinander. Auffällig ist, dass die Kookkurrenzen im vorderen Teil der Tabelle (bis zur Nummer 18) 2004 und 2014 in der Regel noch übereinstimmen, erst im hinteren Teil der Tabelle (ab Nummer 19) häufen sich die Unterschiede. Eine Gruppe von Kookkurrenzen, die 2004 zu den ersten fünfzig Kookkurrenzen gehören und 2014 nicht mehr, sind Adjektive der Herkunftsbezeichnung wie *österreichischen*, *deutschen* oder *Tiroler*. 2014 tauchen hingegen Kookkurrenzen wie *Alterung* oder *schrumpft* neu auf.

²¹ Gleiches gilt auch für das *Deutsche Referenzkorpus*, wo, von zwei Ausnahmen aus dem Jahren 1999 und 2000 abgesehen, *hochmobil* erst seit 2008 in adverbialer Verwendung belegt ist.

²² 2004 wurde für die Analyse die Version 3.3.0.0 von COSMAS II benutzt, 2014 die Version 4.5.4.

	2004	2014
1	Prozent	Mehrheit
2	Mehrheit	Akzeptanz
3	Teil (2014: –)	Mithilfe
4	Teile	Versorgung
5	Versorgung	Rückhalt
6	Drittel	einheimischen
7	Hinweise	Prozent
8	Teilen	Hinweise
9	Hälfte (2014: 76)	bittet
10	gesamte	Anteilnahme (2004: 56)
11	eingeladen	breite
12	Großteil	einheimische
13	Akzeptanz	eingeladen
14	österreichischen (2014: –)	Hinweisen
15	Vertrauen	Drittel
16	Viertel	Teile
17	Mithilfe	Vertrauen
18	Hinweisen	Teilen
19	breite	Unmut (2004: 135)
20	ist (2014: –)	Sicherheitsgefühl (2004: 56)
21	weiten (2014: 55)	Ängste
22	Kopf (2014: –)	breiten
23	ganze (2014: –)	Alterung (2004: –)
24	bittet	Schutz
25	jüdischen (2014: 159)	arbeitenden
26	Unterstützung	Großteil
27	einheimische	notleidende (2004: 56)
28	deutschen (2014: –)	herzlich
29	herzlich	Armutsgrenze (2004: 71)
30	Stimmung (2014: 389)	Überalterung (2004: 145)
31	Tiroler (2014: –)	Fünftel (2004: 137)
32	Schutz	Bewusstsein (2004: 155)

	2004	2014
33	Rückhalt	sensibilisieren (2004: 84)
34	Ängste	Schichten
35	einheimischen	Verunsicherung (2004: 112)
36	großer (2014: 254)	Viertel
37	90 (2014: –)	Wohle (2004: 197)
38	Schweizer (2014: –)	albanische (2004: 54)
39	betroffenen (2014: 88)	Sensibilisierung (2004: 92)
40	Anteil (2014: 71)	schrumpft (2004: –)
41	ausmachen (2014: 65)	albanischen
42	rund (2014: –)	aufgerufen (2004: 55)
43	Bedürfnisse (2014: 77)	gesamte
44	albanischen	erwerbstätigen (2004: 68)
45	serbische (2014: 59)	Grossteil (2004: –)
46	Interesse (2014: 100)	ersucht (2004: 66)
47	arbeitenden	arbeitende (2004: –)
48	Schichten	Unzufriedenheit (2004: 142)
49	Zustimmung (2014: 98)	Unterstützung
50	breiten	hungernde (2004: 77)

Tab. 1: Die ersten 50 Kookkurrenzen zu *Bevölkerung* 2004 und 2014

Diese Kollokationen müssten bei einer Überarbeitung des Wörterbuchartikels ergänzt werden. Wenn die Thematisierung der alternden oder schrumpfenden Bevölkerung noch an weiteren Stellen des *lexiko*-Korpus auffällt, sollte dies als themengebundene Verwendung des Lemmas *Bevölkerung* auch bei den *Besonderheiten des Gebrauchs* beschrieben werden.

Weitere Beispiele dafür, wie innerhalb weniger Jahre durch veränderte Korpora neue Kollokationen zu Wörtern belegt werden können, sind die Lemma *beruflich* und *Behinderter*, die in *lexiko* ebenfalls innerhalb des Moduls *Demonstrationswortschatz* bearbeitet wurden. Zu *beruflich* treten rund zehn Jahre nach der Erstellung des Wortartikels unter anderem folgende neue Kollokationen auf: *Bildung, Perspektiven,*

Qualifizierung, Vorsorge, Wiedereinstieg, Fortkommen, Verpflichtungen, Orientierung, Fortbildung und Eingliederung. Damit zeigt sich deutlich, dass Fragen im Kontext von (*Aus-*)*Bildung (Bildung, Qualifizierung, Fortbildung), Karriere (Perspektiven, Fortkommen, Wiedereinstieg, Eingliederung, Orientierung)* oder *Aufgaben (Verpflichtungen)* eine zunehmend wichtige Rolle spielen. Bei *Behinderter* spiegelt sich die Änderung in der gesellschaftlichen Diskussion, anstelle der Integration auf die Inklusion von Behinderten zu zielen, beispielsweise in der Kollokation *Inklusion* wider. Inwieweit dieser Einstellungswandel bei den Besonderheiten des Gebrauchs oder sogar in der Mikrodiachronie dargestellt werden sollte, müsste eine umfangreichere Untersuchung zeigen.

Die zwei Versionen von COSMAS II liefern auf Basis der beiden Korpora nicht nur zum Teil unterschiedliche Kookkurrenzen, auch die Präsentation der Daten unterscheidet sich durch technische Umstrukturierungen stark. 2004 werden nur für die ersten Kookkurrenzen sogenannte Cluster angegeben, die die weitere Umgebung des Kollokators zeigen. 2014 sind diese Cluster viel umfangreicher und tiefer, außerdem werden syntagmatische Muster aufgeführt, die der Lexikograf als Grundlage für die Formulierung der Konstruktionen verwenden kann (vgl. MÖHRS 2011). 2004 mussten die Konstruktionen noch wesentlich arbeitsaufwendiger aus dem Korpus ermittelt werden. Für den Lexikografen hat sich die Arbeit an dieser Stelle demzufolge vereinfacht. Die Konstruktionen gehören zu den Angabebereichen, in denen durch das veränderte Korpus, aber auch durch verfeinerte Werkzeuge Ergänzungen in den älteren Artikelbeständen vorzunehmen wären. In den Wortartikel *Bevölkerung* wurden auf Basis der Kookkurrenzanalyse von 2004 folgende Konstruktionen aufgenommen: *fast die Hälfte der Bevölkerung, die Mehrheit der Bevölkerung, ein großer Teil der Bevölkerung, große Teile der Bevölkerung, in weiten Teilen der Bevölkerung, pro Kopf der Bevölkerung, rund [Zahl] Prozent der Bevölkerung, zum Wohle der Bevölkerung* und *Hinweise aus der Bevölkerung*. Aus der Analyse der ersten 50 Kookkurrenzen ergeben sich 2014 folgende weitere Konstruktionen, durch die dieser Angabebereich bei einer Überar-

beitung ergänzt werden könnte: die [z. B. arbeitende, notleidende] Bevölkerung, ein [Zahl, z. B. Drittel, Fünftel] der Bevölkerung, [z. B. die Mithilfe, die Überalterung] der Bevölkerung, die Versorgung der Bevölkerung mit [z. B. Lebensmitteln], aus allen Schichten der Bevölkerung, zum Schutz der Bevölkerung, [z. B. die Akzeptanz, der Unmut] in der Bevölkerung, unter großer Anteilnahme der Bevölkerung, die Bevölkerung ist [z. B. aufgerufen, eingeladen] [...], die Bevölkerung für [z. B. das Problem] sensibilisieren und [z. B. die Polizei] bittet die Bevölkerung um [z. B. Hilfe].

5. Zusammenfassung

Der Beitrag hat vorgestellt, mit welchen ambitionierten Zielen das Projekt *lexiko* gestartet ist – ein wissenschaftliches Onlinewörterbuch zum Gegenwartsdeutschen nicht nur von Grund auf neu zu erarbeiten, sondern seine Inhalte auch fortlaufend zu aktualisieren. Eine solche Arbeit ist zwar unter den gegebenen Bedingungen nicht zu leisten, sie entspricht allerdings in besonderem Maße dem zirkulär verlaufenden lexikografischen Prozess bei Onlinewörterbüchern und erfüllt zudem metalexikografische Forderungen aus dem Bereich der Internetwörterbücher. Die potenziell ständige Veränderbarkeit von Wortartikeln in Onlinewörterbüchern wirft auch Fragen der Archivierung und Versionierung neu auf, die auch im Kontext der technischen Möglichkeiten bei Internetwörterbüchern wie *lexiko* zu sehen sind.

Für *lexiko* wurden das dynamische Monitorkorpus, die wachsende Erfahrung der Lexikografen sowie Fragen der Mikrodiachronie als mögliche Ursachen für die Veränderbarkeit von Wortartikeln ausgemacht und beschrieben. Die gewählten Beispiele aus dem Modul *Demonstrationswortschatz* haben bestätigt, dass sich auch bereits nach wenigen Jahren in nahezu allen Angabebereichen Ansatzpunkte für eine Überarbeitung der Wortartikel ergeben: Belege sind nicht mehr Teil des aktuellen *lexiko*-Korpus, es entwickeln sich neue Kollokationen und daraus neue Konstruktionen, neue Satzfunktionen können entstehen und andere Verwendungen können zu Gebrauchsbesonderheiten

des Lemmas führen. Ob die Veränderungen noch innerhalb der lesartenbezogenen Angaben dargestellt werden oder ob sie bereits auf eine entstehende Mikrodiachronie hinweisen und dementsprechend dort beschrieben werden sollten, hängt vom Einzelfall ab. Auch die Gestaltung der dynamischen Stichwortliste steht mit den genannten Faktoren, hauptsächlich mit dem Monitorkorpus, in Zusammenhang.

Die Wortartikel aus *elexiko*, für die Möglichkeiten der Aktualisierung aufgezeigt wurden, haben auch die entscheidende Rolle des dynamischen Korpus als Grund für potenzielle Veränderungen verdeutlicht. Durch eine Analyse der sprachlichen Daten, die das aktualisierte *elexiko*-Korpus zur Verfügung stellt, können am verlässlichsten die lexikografischen Angaben ermittelt werden, die bei einer Überarbeitung des Wortartikels ergänzt werden müssten.

Es wurde deutlich, dass es sich bei *elexiko* um ein typisches Aufbauwörterbuch handelt, dessen Ausbau jedoch nicht allein durch die Neuerarbeitung von Wortartikeln erfolgt, sondern auch durch Veränderungen im Stichwortbestand und mögliche Überarbeitungen bereits ausgearbeiteter Wortartikel. Daraus ergeben sich für die Nutzer einerseits Nachteile, wie die nicht vorhandene Versionierung. Andererseits sind damit aber auch Vorteile für die Wörterbuchbenutzer verbunden. Hier ist die Aktualität der Wörterbuchinhalte zu nennen. Diese spielt für Nutzer eine große Rolle, wie eine Umfrage im Rahmen des Projekts „BZV*elexiko* – Benutzeradaptive Zugänge und Vernetzungen in *elexiko*“ mit knapp 700 Teilnehmern zeigte: Bei einer Bewertung von zehn möglichen Merkmalen guter Onlinewörterbücher landete die regelmäßige Aktualisierung der Wörterbuchinhalte hinter den Merkmalen *inhaltliche Verlässlichkeit* und *Übersichtlichkeit* auf Rang 3. Innerhalb des Aspekts der Aktualisierung war es den Befragten am wichtigsten, dass aktuelle sprachliche Entwicklungen (wie z. B. die neue Rechtschreibung im Deutschen oder neue typische Kontexte) schnell in das Onlinewörterbuch einfließen und dass neue Wörter zeitnah beschrieben werden (vgl. MÜLLER-SPITZER/KOPLINIG 2014).

Insgesamt hat der Beitrag illustriert, dass es neben der beständigen Basis der Wortartikel zahlreiche Stellen gibt, die im Prinzip permanent

überarbeitet und auf den neuesten Stand gebracht werden könnten, sobald sich im aktualisierten Korpus eine sprachliche Veränderung widerspiegelt. Dass sich diese Entwicklungen in einem so rasanten Tempo zu vollziehen scheinen, liegt an der kaum noch zeitversetzten Reflexion sprachlicher Debatten in den Korpora, die durch ihre Aktualität auch kurzlebige sprachliche Erscheinungen einbeziehen.

Literatur

- CCDB – Kookkurrenzdatenbank. V3.3. Online:
<http://corpora.ids-mannheim.de/ccdb/> [Abruf: 2014-04-01].
- COSMAS II. Online: <https://cosmas2.ids-mannheim.de/cosmas2-web/> [Abruf: 2014-04-01].
- DEBUS-GREGOR, ESTHER/HEID, ULRICH (2013): Design criteria and ‘added value’ of electronic dictionaries for human users. In: GOUWS u. a. (Hrsg.), 1001–1013.
- DEREKO. Ausbau und Pflege der Korpora geschriebener Gegenwartssprache. Das Deutsche Referenzkorpus – DeReKo. Online:
<http://www.ids-mannheim.de/kl/projekte/korpora/> [Abruf: 2014-04-01].
- DWDS. Digitales Wörterbuch der Deutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften. Online: <http://www.dwds.de/> [Abruf: 2014-04-01].
- ELEXIKO (2003ff.). In: Institut für Deutsche Sprache, Mannheim (Hrsg.): OWID – Online-Wortschatz-Informationssystem Deutsch. Online:
http://www.owid.de/elexiko_/index.html [Abruf: 2014-04-01].
- ENGELBERG, STEFAN/KLOSA, ANNETTE/MÜLLER-SPITZER, CAROLIN (2009): Internet lexicography as a challenge: The Internet dictionary portal at the Institute for German Language. In: Kernerman Dictionary News (Juli), 10–19.
- ERB, SABINE (2005): Stichwortansetzung von Partizipien. In: HAß (Hrsg.), 91–95.
- GOUWS, RUFUS H./HEID, ULRICH/SCHWEICKARD, WOLFGANG/WIEGAND, HERBERT ERNST (Hrsg.) (2013): Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/Boston: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.4).

- GRAMMIS 2.0 – das grammatische Informationssystem des Instituts für Deutsche Sprache. Online: <http://hypermedia.ids-mannheim.de/index.html> [Abruf: 2014-04-01].
- HAHN, MARION/KLOSA, ANNETTE/MÜLLER-SPITZER, CAROLIN/SCHNÖRCH, ULRICH/STORJOHANN, PETRA (2008): *elexiko* – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche. In: KLOSA, ANNETTE (Hrsg.): Lexikografische Portale im Internet. Mannheim: Institut für Deutsche Sprache (OPAL Sonderheft. 1/2008), 57–85.
- HAB, ULRIKE (2005a): *elexiko* – das Projekt. In: HAB (Hrsg.), 1–17.
- HAB, ULRIKE (Hrsg.) (2005b): Grundfragen der elektronischen Lexikografie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz. Berlin/New York: De Gruyter (Schriften des Instituts für Deutsche Sprache. 12).
- HAB-ZUMKEHR, ULRIKE (2001a): Die deutschsprachige Lexikografie der Gegenwart im Schnittpunkt gesellschaftlicher und wissenschaftlicher Interessen. In: Jahrbuch der ungarischen Germanistik, 157–178.
- HAB-ZUMKEHR, ULRIKE (2001b): Deutsche Wörterbücher. Brennpunkt von Sprach- und Kulturgeschichte. Berlin/New York: De Gruyter (De Gruyter-Studienbuch).
- HAB-ZUMKEHR, ULRIKE (2004): Das Projekt „elexiko – Wissen über Wörter“ des Instituts für Deutsche Sprache. In: SCHARNHORST, JÜRGEN (Hrsg.): Sprachkultur und Lexikographie. Von der Forschung zur Nutzung von Wörterbüchern. Frankfurt am Main [u. a.]: Peter Lang (Sprache, System und Tätigkeit. 50), 311–330.
- KLOSA, ANNETTE (2011a): Einleitung. In: KLOSA (Hrsg.), 9–26.
- KLOSA, ANNETTE (Hrsg.) (2011b): *elexiko*. Erfahrungsberichte aus der lexikografischen Praxis eines Internetwörterbuchs. Tübingen: Narr (Studien zur Deutschen Sprache. 55).
- KLOSA, ANNETTE (2013): The lexicographical process (with special focus on online dictionaries). In: GOUWS u. a. (Hrsg.), 517–524.
- LEMBERG, INGRID (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: LEMBERG/SCHRÖDER/STORRER (Hrsg.), 71–91.
- LEMBERG, INGRID/SCHRÖDER, BERNHARD/STORRER, ANGELIKA (Hrsg.) (2001): Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen: Max Niemeyer (Lexicographica, Series Maior. 107).

- MCENERY, TONY/WILSON, ANDREW (2001): *Corpus Linguistics. An Introduction*. Second Edition. Edinburgh: Edinburgh University Press (Edinburgh Textbooks in Empirical Linguistics).
- MÖHRS, CHRISTINE (2011): Die typischen Verwendungen in *lexiko*. In: KLOSA (Hrsg.), 81–98.
- MÜLLER-SPITZER, CAROLIN (2007): Das *lexiko*-Portal: Ein neuer Zugang zu lexikografischen Arbeiten am Institut für Deutsche Sprache. In: REHM, GEORG/WITT, ANDREAS/LEMNITZER, LOTHAR (Hrsg.) (2007): *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Proceedings of the Biennial GLDV Conference 2007 (April 11–13, 2007, Eberhard Karls Universität Tübingen)*. Tübingen: Narr, 179–188.
- MÜLLER-SPITZER, CAROLIN (2010): OWID – A dictionary net for corpus-based lexicography of contemporary German. In: DYKSTRA, ANNE/SCHOONHEIM, TANNEKE (Hrsg.): *Proceedings of the XIV Euralex International Congress. Leeuwarden, 6–10 July 2010*. Leeuwarden: Fryske Akademy, 445–452.
- MÜLLER-SPITZER, CAROLIN (2011): Der Einsatz einer maßgeschneiderten, feingranularen XML-Modellierung im lexikografischen Prozess. In: KLOSA (Hrsg.), 173–191.
- MÜLLER-SPITZER, CAROLIN/KOPLINIG, ALEXANDER (2014): Online dictionaries: expectations and demands. In: MÜLLER-SPITZER (Hrsg.) (2014): *Using Online Dictionaries*. Berlin/Boston: De Gruyter (Lexicographica, Series Maior. 145), 143–188.
- NESE, HILARY (2000): Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art. In: EVERT, STEFAN/HEID, ULRICH/LEHMANN, EGBERT/ROHRER, CHRISTIAN (Hrsg.) (2000): *Proceedings of the Ninth EURALEX International Congress, Stuttgart, Germany, 8–12 August 2000*. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, 839–847.
- NETZWERK INTERNETLEXIKOGRAFIE (2011ff.). Institut für Deutsche Sprache, Mannheim (Hrsg.). Online: <http://www.internetlexikografie.de> [Abruf: 2014-04-01].
- OED – Oxford English Dictionary. The definitive record of the English language. Online: <http://www.oed.com> [Abruf: 2014-04-01].
- OWID – Online-Wortschatz-Informationssystem Deutsch (2008ff.). Institut für Deutsche Sprache, Mannheim (Hrsg.). Online: <http://www.owid.de> [Abruf: 2014-04-01].
- REICHMANN, OSKAR (2012): *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen*. Berlin/New York: De Gruyter (Studia Linguistica Germanica. 111).

- SCHNÖRCH, ULRICH (2005): Der Autoren-Arbeitsplatz: Ein *lexiko*-Wortartikel entsteht. In: HAß (Hrsg.), 105–130.
- SCHRÖDER, MARTIN (1997): Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projekts „Deutsches Dialektwörterbuch“. In: Zeitschrift für Dialektologie und Linguistik 64/1, 57–66.
- SINCLAIR, JOHN MCHARDY (1982): Reflections on Computer Corpora in English Language Research. In: JOHANSSON, STIG (Hrsg.) (1982): Computer Corpora in English Language Research. Bergen: Norwegian Computing Centre for the Humanities, 1–6.
- SINCLAIR, JOHN MCHARDY (1991): Corpus, Concordance, Collocation. Oxford: Oxford University Press (Describing English Language).
- STORJOHANN, PETRA (2005a): Das *lexiko*-Korpus: Aufbau und Zusammensetzung. In: HAß (Hrsg.), 55–70.
- STORJOHANN, PETRA (2005b): Diachrone Angaben. In: HAß (Hrsg.), 315–322.
- STORRER, ANGELIKA (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: LEMBERG/SCHRÖDER/STORRER (Hrsg.), 53–69.
- STORRER, ANGELIKA/FREESE, KATRIN (1996): Wörterbücher im Internet. In: Deutsche Sprache 24, 97–153.
- WIEGAND, HERBERT ERNST/BEIßWENGER, MICHAEL/GOUWS, RUFUS H./KAMMERER, MATTHIAS/STORRER, ANGELIKA/WOLSKI, WERNER (Hrsg.) (2010): Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen [...]. Band 1: Systematische Einführung, A–C. Berlin/New York: De Gruyter.
- WIKTIONARY. Das freie Wörterbuch. Online: <http://de.wiktionary.org> [Abruf: 2014-04-01].

JULIA STEUBE

Wörterbücher und lexikografisches Wissen von Amateurlexikografen im Internet

1. Einleitung

Das digitale Medium, insbesondere das Internet, bringt in der Geschichte der Lexikografie – wie in vielen anderen Bereichen auch – einen Paradigmenwechsel mit sich, der geprägt ist von neuen Möglichkeiten wie auch Herausforderungen. Prominente Beispiele wie *Wikipedia*, aber auch unzählige kleinere Nachschlagewerke von Hobby-Lexikografen zeigen, dass nicht nur Wörterbuchverlage und akademische Wörterbuchprojekte das Internet als Publikationsplattform nutzen.¹ Mit seiner partizipativen Grundidee lädt es geradezu dazu ein, ein eigenes Wörterbuch-Projekt umzusetzen, denn es weist, anders als dies im Printbereich der Fall ist, kaum Publikationsbarrieren auf. Während sich die Wörterbuchforschung *Wikipedia* als enzyklopädisch-kollaborativem Kolossalwerk zumindest in einigen wenigen Beiträgen zuwendet, so z. B. in GUREVYCH/WOLF (2010), GUREVYCH/NIEMANN (2011), CÖLFEN (2012) oder MEDERAKE (in diesem Band)², blieben Werke, die von Einzelpersonen oder kleineren Gruppen angeboten werden, bisher unbeachtet. Gleichzeitig wird in der Wörterbuchforschung aber jüngst die Notwendigkeit erkannt, sich mit dem Phänomen des lexikografisch aktiven Laien im Internet auseinanderzusetzen, wie einige aktuelle Ar-

¹ Geschichtlich gesehen wurde Lexikografie in ihren Anfängen ebenso ‚dilettantisch‘ betrieben wie andere vorakademische ‚Wissenschaften‘ oder sich neu etablierende Handwerke.

² MEDERAKE, NATHALIE: Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive. Oder: Wenn das Leben dir Zitronen gibt ... (in diesem Band).

tikel zur Online-Lexikografie zeigen.³ In STEUBE (i. E.) habe ich mich aufgrund der bisher mageren Forschungslage dem ‚lexikografischen Amateur‘⁴ und seinen Nachschlagewerken zugewandt, um Bedingungen der Entstehung, wie Motivation und Wissen der Emittenten⁵, sowie Merkmale und beabsichtigte Funktion der Nachschlagewerke explorativ zu erschließen. Die Untersuchung beschränkt sich dabei auf sachbezogene Nachschlagewerke. Der vorliegende Beitrag fokussiert die Frage nach dem **lexikografischem Wissen** der Amateurlexikografen. Um Internet-Nachschlagewerke ‚auf eigene Faust‘ erstellen zu können, sind auf Seiten der Macher Kenntnisse, Fertigkeiten und Wissen auf unterschiedlichen Ebenen notwendig. Differenziert werden kann zwischen a) Wissen auf ‚fachlich-inhaltlicher‘ Ebene, d. h. Sachinformationen (und fachsprachlichem Wissen) zum Wörterbuchthema, b) lexikografischem Wissen, das Vorstellungen darüber beinhalten, wie Wörterbücher gestaltet und aufgebaut sein können, sowie c) ‚internettechnischem‘ Wis-

³ So werden beispielsweise in ABEL/KLOSA (2014a und Hrsg. 2014b) verschiedene Formen und Möglichkeiten der Benutzerpartizipation bei der Wörterbucherstellung und -erweiterung im Internet betrachtet und diskutiert. RAUTMANN (2014) und THIER (2014) zeigen Möglichkeiten der Beteiligung bei den Verlagswörterbüchern *Duden online* (www.duden.de [2014-07-15]) und *Oxford English Dictionary* (www.oed.com [2014-07-15]) auf, während LEW (2014) Formen des „user-generated content“ (UGC) in verschiedenen englischsprachigen Online-Wörterbüchern betrachtet. Das Konzept und verschiedene Ausprägungen der „Halbkollaborativität“ erläutert MELCHIOR (2014) am Beispiel des mehrsprachigen Online-Wörterbuchs *leo.org* (www.leo.org [2014-07-15]).

⁴ Zur Wahl dieser Bezeichnung etwa im Vergleich zum ‚lexikografischen Laien‘ siehe Abschnitt 2 dieses Beitrags.

⁵ Der in der Textlinguistik übliche Ausdruck des *Emittenten* wird hier in Anlehnung an GLINZ (1977, 17) anstelle von *Autor* oder *Anbieter* verwendet, da der Anbieter eines Nachschlagewerks nicht in allen Fällen auch der (alleinige) Autor ist. Die Bezeichnung bietet sich an, da sie Autorschaft nicht zwingend mit einbezieht, sondern allgemeiner auf den ‚Sender‘ eines Textes referiert.

sen, das benötigt wird um Inhalte online organisieren, präsentieren und ausbauen zu können.⁶

Nach einer terminologischen Klärung im folgenden Abschnitt werden in Abschnitt 3, 4 und 5 die folgenden Fragen anhand der in STEUBE (i. E.) erarbeiteten Ergebnisse behandelt:

- Über welche Formen des lexikografischen Wissens verfügen nicht-professionelle Lexikografen? (Mit welchen lexikografischen und medialen Merkmalen statten sie ihre Nachschlagewerke aus?)
- Woher stammt dieses Wissen?
- Was wissen sie über ‚Lexikografie‘ als professionelle Tätigkeit?

2. Terminologische Klärung: Warum nichtprofessionelle Lexikografen kein Laien sind

In der metalexikografischen Literatur wird das bisher kaum beachtete Phänomen der nichtprofessionellen Wörterbucharstellung zumeist als ‚Laienlexikografie‘ bzw. ihre Akteure als ‚Laienlexikografen‘ bezeichnet (siehe z. B. ENGELBERG/LEMNITZER 2009, 152, oder KEMMER 2010, 3). In der englischsprachigen Literatur ist dagegen meist von ‚amateur lexicographers‘ die Rede, so z. B. bei BERGENHOLTZ/TARP (1995, 14), GABRIELATOS (2009) oder LEW (2014, 25). Worum handelt es sich nun eigentlich – um *Laien*, *Amateure* oder gar *Dilettanten*?

⁶ Diese Art des Wissens ist insbesondere dann relevant, wenn kein Content Management System (CMS) zur Verfügung steht. Ein CMS ist eine „Software, die es dem Benutzer erlaubt, Inhalte im Web zu publizieren, ohne dass dafür HTML oder eine andere [Programmier-]Sprache erlernt werden müsste“ (ALBY 2008, 241). Technisches Wissen für das ‚Online-Stellen‘ oder Anordnen der Inhalte spielt damit kaum eine Rolle, der Nutzer kann sich weitestgehend auf die Inhalte, den ‚content‘, konzentrieren. Denkt man sich die nutzerfreundlichen Web-2.0-Plattformen weg und begibt sich auf den technischen Stand des ‚Web 1.0‘, so bedarf es dagegen gewisser HTML-Kenntnisse sowie Erfahrung im Umgang mit HTML-Editoren und FTP-Clients, um Internetseiten generieren, gestalten, online stellen und online verändern zu können (vgl. ALBY 2008, 25f.).

Mithilfe der folgenden fünf Bedeutungsdimensionen lassen sich die verschiedenen Lesarten der Lexeme erfassen und miteinander vergleichen:⁷

- **Bewertung.** Ist die jeweilige Lesart ‚negativ‘, ‚positiv‘ oder ‚neutral‘ konnotiert (s. WICHTER 1994, 56)?
- **Wissensstand.** Haben die jeweils bezeichneten Personen ‚geringes‘ oder kein ‚(null)‘ Wissen (s. WICHTER 1994, 56)?
- **Horizontalität.** Ist das Nicht-Wissen der jeweils bezeichneten Person auf eine bestimmte Wissensdomäne bezogen (s. WICHTER 1994, 56)?
- **Einstellung.** Wie steht die bezeichnete Person zum jeweiligen Sachgebiet oder Gegenstand ihrer Beschäftigung? Ist ihre Haltung ‚neutral‘ oder ‚positiv‘, was sich in „Interesse oder Bemühtheit“ (WICHTER 1994, 56) äußert?
- **Rolle.** Ist die Rolle der bezeichneten Person eher ‚rezipierend‘, ‚produzierend‘ oder ‚neutral‘ in Bezug auf den jeweiligen Gegenstand?

Für die semantische Analyse der drei Lexeme *Laie*, *Dilettant* und *Amateur* wurden die Bedeutungsbeschreibungen eines allgemeinsprachigen Wörterbuchs (DUW 2007) und ihre Verortung in einem onomasiologischen Wörterbuch (DORNSEIFF 1970 und 2004) herangezogen. Zusätzlich wurden Korpusrecherchen in der Kookkurrenzdatenbank (CCDB) des Instituts für Deutsche Sprache (BELICA 1995 und 2001ff.) durchgeführt. Bei der Analyse stellt sich heraus, dass sich zu jedem der drei Lexeme anhand der oben genannten Dimensionen drei bzw. vier Lesarten feststellen lassen.⁸ Die einzelnen Lexeme weisen somit recht heterogene Bedeutungen auf, was davon zeugt, „[...] in welchem Spannungsfeld sozialer, eigenperspektivischer Bewertung der Handelnden sich jene einfachere Wissensteilhabe bzw. Kompetenz befindet, die sich

⁷ WICHTER (1994) wendet vier dieser Dimensionen für die Bedeutungsanalyse der Wörter *Experte* und *Laie* in seiner Monografie „Experten- und Laienwortschätze“ an.

⁸ Für die detaillierte bedeutungsgeschichtliche wie auch gegenwartssprachliche Analyse der Wörter siehe STEUBE (i. E., 45ff.).

zuvörderst auf eine natürliche und nicht verwerfliche Weise aus dem unabweisbaren Umstand ergibt, daß das Expertenwissen auf Grund seines überindividuellen Umfangs selektiv verteilt sein muss“ (WICHTER 1994, 59).

Für die Frage, welches der drei Wörter *Laie*, *Amateur* und *Dilettant* den ‚nichtprofessionelle Lexikografen‘ nun am treffendsten zu bezeichnen vermag, sind vor allem die Dimensionen der ‚Rolle‘ und der ‚Bewertung‘ ausschlaggebend, da in den anderen Dimensionen schlicht keine Informationen vorliegen.⁹

Betrachtet man die drei Lexeme im Vergleich, so werden gegenwartssprachlich in Bezug auf die Dimension der ‚Rolle‘ Tendenzen deutlich, die sich auch bedeutungsgeschichtlich abzeichnen (vgl. STEUBE i. E., 45ff.): Während *Dilettanten* und *Amateure* eher **aktiv-produzierende** Personen sind, wird *Laie* vor allem im heutigen Sprachgebrauch eher für **rezipierende** Personen verwendet. *Amateur* und *Dilettant* unterscheiden sich von *Laie* außerdem darin, dass für beide Lexeme je eine der drei Lesarten eine thematische Spezifizierung aufweist (Dimension der Horizontalität): Der *Dilettant* wird gelegentlich in Bezug zu Musik gesetzt, während der *Amateur* zumindest in einer Lesart in Verbindung mit Sport, Fotografie oder Musik steht. In ihren übrigen Lesarten weisen *Dilettant* und *Amateur* keine thematische Spezifizierung auf, sondern werden neutral verwendet. Der Unterschied in der Verwendung der beiden Lexeme liegt allerdings darin, dass *Dilettant* in sehr ähnlichen Kontexten wie eine ganze Reihe negativ konnotierter Wörter wie *Selbstdarsteller*, *Demagoge*, *Opportunist* oder *Heuchler* u. Ä. verwendet wird.¹⁰ Dies lässt darauf schließen, dass auch das Lexem *Dilettant* häufig in pejorisierender Weise gebraucht wird, zumin-

⁹ In STEUBE (i. E.) bilden diese drei zu klärenden Dimensionen daher die Ausgangspunkte der Untersuchung zum lexikografischen Amateur und seinen Nachschlagewerken.

¹⁰ Unter den 31 Kookkurrenzprofilen der CCDB, die mit dem Profil von *Dilettant* am engsten verwandt sind, finden sich überwiegend negativ konnotierte Lexeme, wie die genannten. Diese Beobachtungen decken sich mit der Bedeutungsparaphrase im DUW (2007) wie auch mit der Einordnung im DORNSEIFF (1970; 2004).

dest aber in Kotexten, in denen es auf etwas Ähnliches referiert wie die genannten Wörter.

Betrachtet man die drei Lexeme nun hinsichtlich ihrer Angemessenheit für die Benennung des ‚nichtprofessionellen Lexikografen‘, so kommt der Ausdruck *Laien* aufgrund der in den Lesarten dominierenden Rolle des ‚Rezipienten‘ nicht ernsthaft in Frage. Gegen den *Dilettanten* spricht, dass dieser in seinen Lesarten, wie oben gesehen, negativer konnotiert ist als der *Dilettant*. Die Bezeichnung *Amateur* berücksichtigt dagegen einerseits, dass es sich um aktiv-produzierende Akteure handelt, andererseits bietet dieser Ausdruck die größere Neutralität als *Dilettant*. Vor dem Hintergrund der dargelegten Bedeutungsunterschiede erscheint es daher sinnvoll, im Zusammenhang mit nichtprofessioneller Lexikographie grundsätzlich vom *lexikografischen Amateur* (bzw. von *Amateurlexikografie*) zu sprechen.

3. Wissen über Wörterbücher: Lexikografische und mediale Merkmale amateurlexikografischer Nachschlagewerke

Methodisch ist das Wissen von Amateurlexikografen über Wörterbücher nicht ohne weiteres zu erheben. Eine Befragung der Emittenten müsste weitestgehend ohne linguistisch-lexikografische Fachterminologie auskommen, da nicht vorausgesetzt werden kann, dass diese dem Amateurlexikografen ausreichend bekannt ist. So soll stattdessen die Betrachtung ausgewählter lexikografischer und medialer Merkmale der Nachschlagewerke indirekt Anhaltspunkte geben, was Amateurlexikografen über Nachschlagewerke wissen bzw. welche (ggf. stereotypen) Vorstellungen sie über Wörterbücher haben.

Im Rahmen der empirischen Untersuchung (STEUBE i. E., 101ff.) wurden dafür neun ausgewählte lexikografische und mediale Merkmale an 75 sachbezogenen Internet-Nachschlagewerken erhoben. Die Auswahl der 75 Nachschlagewerke erfolgte mit Hilfe der Suchausdrücke ‚Fachwörterbuch‘ und ‚Fachwörterbuch online‘ sowie ‚Lexikon‘ und

„Lexikon online“ über die Metasuchmaschine MetaCrawler™.¹¹ Folgende Auswahlkriterien wurden zugrunde gelegt:

- **Von lexikografischen Amateuren erstellt.** In den Metatexten wie Impressum oder Startseite werden weder ein Verlag noch eine sich aus (haupt-)beruflich tätigen Mitarbeitern oder professionellen Lexikografen zusammensetzende Redaktion als Urheber des Nachschlagewerks genannt.
- **Sach- bzw. Fachbezogenheit des Nachschlagewerks.** Die Nachschlagewerke weisen Sach- bzw. Fachbezug auf, d. h. sie beinhalten einen domänenspezifischen Wortschatz und ihr Schwerpunkt liegt auf Sachinformationen.
- **Monolingualität des Nachschlagewerks.** Die Nachschlagewerke sind einsprachig deutsch.¹²
- **Freie Verfügbarkeit** des Nachschlagewerks. Aus pragmatischen Gründen wurden nur kostenlos und ohne vorherige Registrierung verfügbar Nachschlagewerke gewählt.

In der folgenden Tabelle sind die Merkmale und ihre möglichen Ausprägungen aufgeführt. Die Merkmale wurden zum Teil existierenden Wörterbuchtypologien entnommen und angepasst oder eigens für den Untersuchungsgegenstand aufgenommen. Sie wurden möglichst breit gefächert gewählt und reichen von allgemeinen Eigenschaften wie der thematisch-fachlichen Zuordnung über die Zugriffs- bzw. Makrostruktur, die Mediostruktur und Informationstypen bis hin zu medialen und technischen Aspekten. Die Merkmalsausprägungen orientieren sich am Untersuchungsmaterial und greifen damit nur Eigenschaften auf, die in den untersuchten Nachschlagewerken auch tatsächlich zu finden sind.

¹¹ Siehe www.metacrawler.de [2014-07-11]. Auf den Suchausdruck „Internet“ wurde verzichtet, da die Suchanfrage überwiegend Nachschlagewerke hervorbrachte, die das Internet (bzw. internetbezogenen Wortschatz) zum Gegenstand haben.

¹² Damit sind fremdsprachige Äquivalente in den Wörterbuchartikeln nicht ausgeschlossen, es wurde jedoch darauf geachtet, dass die Übersetzung nicht die zentrale Funktion des Nachschlagewerks darstellte.

Merkmal	Mögliche Merkmalsausprägungen
Wörterbuchgegenstand	<i>A) Allgemein</i>
	<i>B) Biologie</i>
	<i>C) Chemie und Pharmazie</i>
	<i>D) Geografie</i>
	<i>E) Geologie und Paläontologie</i>
	<i>F) Informatik und EDV</i>
	<i>G) Land- und Forstwirtschaft, Gartenbau, Fischereiwirtschaft, Hauswirtschaft</i>
	<i>H) Mathematik</i>
	<i>I) Medizin</i>
	<i>J) Philosophie</i>
	<i>K) Physik</i>
	<i>L) Psychologie</i>
	<i>M) Rechtswissenschaft</i>
	<i>N) Sport</i>
<i>O) Technik</i>	
<i>P) Theologie und Religionswissenschaft</i>	
<i>Q) Wirtschaft</i>	
Zugriffsangebot des Nachschlagewerks (vgl. STORRER/FREESE 1996, 125ff.)	<i>A) blätternnder Zugriff</i>
	<i>B) hypertextorientierter Zugriff</i>
	<i>C) retrievalorientierter Zugriff</i>
	<i>D) hypertextorientierter und retrievalorientierter Zugriff</i>
	<i>E) hypertextorientierter und blätternnder Zugriff</i>
	<i>F) blätternnder und retrievalorientierter Zugriff</i>
	<i>G) blätternnder, retrieval- und hypertextorientierter Zugriff</i>
Struktur des Datenangebots (vgl. FELBER/SCHAEDER 1999, 1738)	<i>A) alphabetisch</i>
	<i>B) nicht-alphabetisch</i>
	<i>C) alphabetisch und nicht-alphabetisch</i>
Verweise und ihre Hypertextualisierung (vgl. STORRER/FREESE 1996, 119ff.)	<i>A) mit Verweisen</i>
	<i>A1) hypertextualisierte Verweise</i>
	<i>A2) nicht hypertextualisierte Verweise</i>
	<i>B) ohne Verweise</i>

Merkmal	Mögliche Merkmalsausprägungen
Gewichtung der Informationstypen (vgl. KEMMER 2010, 13)	A) <i>Sachinformationen</i>
	B) <i>Sprach- und Sachinformationen</i>
Einbettungsverhältnis	A) <i>freies Nachschlagewerk</i>
	B) <i>scheinbar freies Nachschlagewerk</i>
	C) <i>integriert in Seite oder Portal</i>
Multimedialität (vgl. STORRER/FREESE 1996, 122ff.)	A) <i>nur Text</i>
	B) <i>Text und Bild</i>
	C) <i>Text, Bild und Ton</i>
	D) <i>Text, Bild und Film</i>
Web-2.0-Elemente	A) <i>Weblog</i>
	B) <i>Weblog und Wiki</i>
	C) <i>Forum</i>
	D) <i>Forum auf übergeordneter Webpräsenz</i>
	E) <i>keine Web-2.0-Elemente</i>
Abgeschlossenheit des Nachschlagewerks (vgl. STORRER/FREESE 1996, 117ff.)	A) <i>abgeschlossene Nachschlagewerke</i>
	B) <i>kontrollierter Aufbau</i>
	C) <i>unkontrollierter Aufbau</i>

Tab. 1: Lexikografische und mediale Merkmale amateurlexikografischer Nachschlagewerke

Im Folgenden werden die Merkmale und ihre Ausprägungen, falls nicht selbsterklärend, kurz erläutert sowie die Ergebnisse der Merkmalerhebung und häufig auftretende Merkmalskombinationen dargestellt.

1) Wörterbuchgegenstand. Für die Systematisierung der Wörterbuchgegenstände wurde die sogenannte ‚Regensburger Verbundklassifikation‘ (RVK)¹³, eine etablierte Aufstellsystematik für wissenschaftliche Bibliotheken, zugrunde gelegt.¹⁴

¹³ Siehe <http://rvk.uni-regensburg.de> [2014-07-14].

¹⁴ Themen, die in keine der Kategorien passen, wie z. B. Werke aus dem Bereich der Esoterik und Mythologie, wurden der Restklasse ‚Allgemein‘ zugeordnet. Thematische Bereiche, die als *eine* Thematik behandelt wurden, sind im Folgenden durch einen Schrägstrich aufgelistet.

Es stellte sich heraus, dass die meisten Werke aus dem Bereich *Medizin* stammen (20 %), gefolgt von *Informatik/EDV* (17 %), *Biologie* (17 %) und *Technik* (11 %). Die Bereiche *Chemie/Pharmazie*, *Land-/Forstwirtschaft*, *Mathematik*, *Philosophie*, *Psychologie* und *Wirtschaft* sind mit je rund 3 %, der Bereich *Physik* mit 4 % vertreten. *Geografie*, *Geologie/Paläontologie*, *Sport* und *Theologie/Religion* mit je rund 1 %. Knapp die Hälfte der Nachschlagewerke deckt damit medizinische oder naturwissenschaftliche Themen ab, 28 % beziehen sich auf technische Gebiete und den IT-Bereich. Geisteswissenschaftliche Fächer¹⁵ werden dagegen nur von rund 15 % der Nachschlagewerke behandelt.

Eine Erklärung für die recht hohe Anzahl an Nachschlagewerken zum Thema *Informatik/EDV* könnte die Selbstreferenzialität zum Medium sein: Computerinteressierte und -versierte verfügen über das computerbezogene Fachwissen auf inhaltlicher Ebene, wie auch über das notwendige internettechnische Wissen. Das relativ hohe Vorkommen von Werken aus dem Bereich *Biologie* lässt sich z. B. dadurch erklären, dass Naturliebhaber und Amateurforscher ihre Beobachtungen in lexikografischer Form festhalten und einem interessierten Publikum präsentieren wollen (zu Amateurforschern und anderen Formen amateurhafter Tätigkeiten siehe ausführlicher STEUBE i. E., 59ff.).

Zugriffs- bzw. Makrostruktur

2) Zugriffsangebot (vgl. STORRER/FREESE 1996, 125ff.). Beim Zugriffsangebot wurde unterschieden, in welcher Form auf die Wörterbuchartikel zugegriffen werden kann. Ein Nachschlagewerk kann zum einen ‚blättern‘ angelegt sein, das heißt, alle Wörterbuchartikel sind auf einer Internetseite abgebildet und durch Scrollen erreichbar. Der Zugriff kann darüber hinaus ‚hypertextorientiert‘ über Links bzw. einen verlinkten Index erfolgen oder ‚retrievalorientiert‘, d. h. über die Texteingabe in ein Suchfeld. Die meisten Nachschlagewerke bieten mehr als eine Zugriffsmöglichkeit an, so dass zusätzlich verschiedene Kom-

¹⁵ Zu den geisteswissenschaftlichen Wörterbuchgegenständen wurden *Theologie/Religion*, *Philosophie*, *Psychologie*, *Wirtschaft* und *Rechtswissenschaft* gezählt.

binationstypen vorkommen (siehe Tabelle 1, Ausprägungen D bis G zum Zugriffsangebot).¹⁶

Die häufigste Kombination des Zugriffsangebots unter den betrachteten Nachschlagewerken ist mit 36 % die aus hypertextorientiertem und retrievalorientiertem Zugriff. Jeweils 24 % der Nachschlagewerke bieten die Kombination aus hypertextorientiertem und blätternem Zugriff an. Ebenso häufig kommt der rein hypertextorientierte Zugriff vor (24 %), seltener findet sich dagegen der rein blätternde Zugriff (11 %). Nur bei 5 % der Nachschlagewerke finden sich alle drei Zugriffsformen.

Betrachtet man die einzelnen Formen des Zugriffs separat, so stellt der hypertextualisierte die häufigste Form dar; er wird (teilweise neben anderen) in 50 % der 75 Nachschlagewerke angeboten, gefolgt vom retrievalorientierten Zugriff mit 36 %. Der blätternde Zugriff ist in insgesamt 35 % aller Werke zu finden. Es handelt sich dabei um die technisch am wenigsten aufwändige Art des Zugriffs, je nach Anzahl der Stichwörter und Umfang der Artikel ist sie aber nicht besonders benutzerfreundlich. HAB/SCHMITZ (2010, 9f.) sehen dennoch „Erhaltenswertes“ an dieser Art der Zugriffstruktur: Der Benutzer kommt mit „[...] einem sogar sinnlich, nämlich durch mehrfaches Scrollen als ‚große Menge‘ erfahrbaren Merkmal des Wörterbuchs“ in Berührung, ganz anders also als beim retrievalorientierten Zugriff, der den Umfang eines Nachschlagewerks nicht direkt sichtbar macht.

3) Struktur des Datenangebots (vgl. FELBER/SCHAEDER 1999, 1738). Dieses Merkmal erfasst die Makrostruktur im klassischen Sinne, die Anordnung der Artikel also, wie sie auch in Papierwörterbüchern vorzufinden wäre. Allerdings spielt sie bei digitalen Nachschlagewerken nur eine Rolle, wenn eine ‚blätternde‘ und bzw. oder ‚hypertextorientierte‘ Zugriffsmöglichkeit (s. Merkmal 2) gegeben ist, da beim allei-

¹⁶ Bei der Kombination ‚hypertextorientierter und blätternder Zugriff‘ ist nicht von parallelen, sondern von aufeinander folgenden Zugriffsmöglichkeiten auszugehen, also etwa einem Index, der auf eine Seite verlinkt, auf dem die Artikel zu einem Buchstaben blätternnd zu erreichen sind.

nigen Zugriff über ein Suchfeld (‚retrievalorientierter‘ Zugriff) keine feste Anordnung der Artikel gegeben ist.¹⁷

77 % der 75 Nachschlagewerke weisen eine alphabetische Anordnung der Artikel auf, 16 % eine Kombination aus alphabetischem und nicht-alphabetischem und 7 % einen onomasiologischen Aufbau. Diese Verteilung überrascht nicht, ist die alphabetische Struktur doch die in professionellen Nachschlagewerken am häufigsten vertretene und für den lexikografischen Amateur damit die vertrauteste Anordnung von Artikeln.

Auffällig hinsichtlich der Merkmalskombinationen ist, dass insbesondere Werke mit naturwissenschaftlichem Gegenstand¹⁸ häufiger eine nicht-alphabetische Anordnung aufweisen als andere: Von zwölf Nachschlagewerken bieten sieben (58 %) einen nicht-alphabetischen Zugriff an, davon drei ausschließlich, vier zusätzlich zur alphabetischen Struktur. Diese Verteilung lässt sich größtenteils darauf zurückzuführen, dass naturwissenschaftliches und insbesondere biologisches Wissen im Bereich der Fauna und Flora systematisch in Form einer ‚Morphologie‘ organisiert ist. Nachschlagewerke mit medizinischem Wörterbuchgegenstand weisen dagegen ausnahmslos eine alphabetische Anordnung der Artikel auf, was hinsichtlich der Strukturiertheit bzw. Strukturierbarkeit der fachlichen Inhalte verwundern mag, jedoch der medizinischen Wörterbuchtradition entspricht.

Mediostruktur: Verknüpfung von Informationen

4) Verweise und ihre Hypertextualisierung (vgl. STORRER/FREESE 1996, 119ff.). Insgesamt 88 % der Werke enthalten Verweise, in 76 % der Nachschlagewerke sind diese hypertextualisiert. Jeweils

¹⁷ Bei den kombinierten Merkmalsausprägungen des Merkmals ‚Zugriffsangebot‘ bezieht sich die Angabe zur Struktur des Datenangebots daher entweder auf die ‚blätternde‘ oder die ‚hypertextorientierte‘ Zugriffsstruktur (s. Tabelle 1, Ausprägungen D, F und G des Merkmals ‚Zugriffsangebot‘).

¹⁸ Zu den naturwissenschaftlichen Wörterbuchgegenständen werden hier *Mathematik, Physik, Biologie, Geologie/Paläontologie, Geografie* und *Chemie/Pharmazie* gezählt.

12 % der Werke enthalten entweder keine Verweise oder Verweise ohne Hypertextualisierung.

Mikrostruktur: Informationstypen auf Articlebene

5) Gewichtung der Informationstypen (vgl. KEMMER 2010, 13). Dieses Merkmal bezieht sich auf die in den Artikeln anzutreffenden Typen der Information. Da nicht davon auszugehen ist, dass Amateurlexikografen zwischen den diversen Informationsarten unterscheiden¹⁹, wurde lediglich zwischen Sach- und Sprachinformationen als elementaren Informationstypen differenziert, wie es auch KEMMER (2010, 13) handhabt. Für die Erhebung dieses Merkmals wurde zudem nach ‚überwiegender Gewichtung‘, nicht nach dem ausschließlichen Vorkommen der Informationstypen unterschieden.²⁰ Als Stichprobe wurden jeweils zehn Artikel pro Nachschlagewerk analysiert.²¹ Als sprachbezogen wurden solche Informationen eingestuft, die sich etwa auf die Phonetik, Orthografie, Etymologie oder Morphologie eines Ausdrucks beziehen. Bei der Unterscheidung von semantischer und enzyklopädischer Information (vgl. WIEGAND 1994) wurde im Zweifelsfall zugunsten des enzyklopädischen, sprich sachbezogenen Informationstypen entschieden, da eine bewusst intendierte semantische Erklärung unwahrscheinlicher ist als die Absicht der Emittenten, sachbezogene Informationen

¹⁹ Die metalexikografische Einteilung in Angabeklassen fällt sehr unterschiedlich aus. Während WIEGAND (vgl. 1989, 427) – Strukturanzeiger eingeschlossen – zwischen mehr als 200 Angabeklassen unterscheidet, fassen ENGELBERG/LEMNITZER (vgl. 2009, 157) die ihrer Ansicht nach wichtigsten Informationstypen in acht Gruppen zusammen.

²⁰ Dass ein Nachschlagewerk ausschließlich *einen* der beiden Informationstypen enthält, ist zwar theoretisch denkbar, war jedoch nicht überprüfbar, da dafür sämtliche Artikel eines Werks hätten analysiert werden müssen.

²¹ Die Stichprobe enthielt, sofern im Lemmabestand vorhanden, jeweils zwei substantivische, zwei verbale, zwei adjektivische Lemmata und falls möglich zwei Mehrwortlemmata und zwei nicht-indigene Stichwörter. Andernfalls wurde eine Auswahl getroffen, die dieser Verteilung möglichst nahekam.

zu liefern.²² Je nachdem, welche der drei Merkmalsausprägungen in der Stichprobe der zehn Artikel am häufigsten auftrat, wurde diese Ausprägung für das Nachschlagewerk angenommen.

Knapp zwei Drittel der Werke (64 %) enthalten sowohl sprach- als auch sachbezogene Informationen, bei den übrigen 36 % dominieren Sachinformationen. Nachschlagewerke, die ausschließlich oder überwiegend Sprachinformationen enthalten, kamen nicht vor.

Insbesondere bei Werken zur Thematik *EDV und Informatik* fällt auf, dass fast alle (12 von 13, also 92 %) eine Kombination aus Sprach- und Sachinformationen anbieten. Zu erklären ist dies sicher teilweise damit, dass der computerbezogene Fachwortschatz viele Anglizismen, Abkürzungen und Kurzworte enthält, deren Bedeutung, Herkunft oder Aussprache in den Artikeln erläutert werden.

Mediale Aspekte

6) Einbettungsverhältnis. Dieses Merkmal beschreibt, in welchem Kontext bzw. Rahmen das Nachschlagewerk angeboten wird. Zum einen kann es als ‚freie‘ Seite oder aber als ‚scheinbar freie‘ Seite angelegt sein, die eigentlich zu einer anderen Domain gehört, ohne dass dies aus der URL oder den Inhalten der Seite erkennbar wird.²³ Ein Nachschlagewerk kann darüber hinaus in eine Webpräsenz integriert sein; dabei kann es sich um eine ‚normale‘ Website oder ein komplexeres Portal handeln.

Auf 61 % der Werke trifft der letztgenannte Fall zu, d. h. sie sind in eine Seite oder ein Portal integriert. Bei 24 % handelt es sich um freie Nachschlagewerke und 15 % sind nur scheinbar frei, gehören aber tat-

²² Die Kriterien für die eindeutige Abgrenzung dieser Informationstypen werden selbst in der metalexikografischen Literatur diskutiert (vgl. z. B. ENGELBERG/LEMNITZER 2009, 12f.). Eine klare Unterscheidung von semantischer Bedeutungserklärung und enzyklopädischer Information kann oft nur vor dem Hintergrund(-wissen) semantischer Theoriemodelle wie dem der Prototypikalität oder der Merkmalssemantik getroffen werden (vgl. ENGELBERG/LEMNITZER 2009, 12).

²³ Die Zusammengehörigkeit einer Internetseite bzw. eines Portals mit dem Nachschlagewerk wurde i. d. R. über die Impressumsseiten aufgedeckt.

sächlich zu einer anderen Internetseite. Damit ist ein Großteil der Werke ‚nur‘ Teil einer größeren Einheit.

7) Multimedialität (s. STORRER/FREESE 1996, 122ff.). Mehr als die Hälfte der Nachschlagewerke (55 %) enthält nur Text und ist damit nicht multimedial angelegt. 39 % der Nachschlagewerke beinhalten über den Text hinaus Bilder, 5 % zusätzlich Tondateien. Diese Ergebnisse erstaunen angesichts der Möglichkeiten, die das Internet an multimedialen Gestaltungsmöglichkeiten bietet; der Verzicht auf multimediales Material ist aber sicher auch mit dem hohen Arbeitsaufwand beim Wörterbuchaufbau zu erklären.

Auffällige Merkmalskombinationen finden sich bei den zwölf Nachschlagewerken mit dem Wörterbuchgegenstand *Biologie*, von denen 83 % multimedial angelegt sind: Sieben der zwölf Werke enthalten zusätzlich zum Text Bilder, zwei darüber hinaus Ton. Auch das einzige Nachschlagewerk im Untersuchungskorpus, das Text, Bild und Film enthält, stammt aus dem Bereich der Biologie. Bei den elf Nachschlagewerken mit geisteswissenschaftlicher Thematik fällt dagegen auf, dass 82 % nur Text enthalten.

Des Weiteren enthalten 83 % aller Nachschlagewerke ohne Verweise oder mit nicht-hypertextualisierten Verweisen nur Text und sind nicht multimedial angelegt. Im Umkehrschluss bedeutet dies, dass Multimedialität und eine hypertextualisierte Mediostruktur häufig miteinander auftreten: Von 34 multimedial angelegten Werken enthalten 91 % hypertextualisierte Verweise. Diese Werke schöpfen die Möglichkeiten des elektronischen Mediums also in mehrfacher Hinsicht aus.

Web-2.0-Basiertheit²⁴

8) Web-2.0-Elemente. Mit diesem Merkmal wurde erhoben, ob die Nachschlagewerke Web-2.0-Elemente enthalten.

²⁴ Der auf O'REILLEY (2005) zurückgehende Begriff des ‚Web 2.0‘ ist nicht einheitlich definiert, umfasst aber im Wesentlichen das Phänomen des „[...] ‚ungehinderten‘ Austausches von Informationen, einer ‚schrankenlosen‘ Partizipation und einer ‚hierarchiefreien‘ Kollaboration [...]“ (REICHERT 2008, 8). So gehören nach ALBYs Definition auch Foren – anders als etwa Chatsysteme – zu den Web-2.0-Erscheinungen (ALBY 2008, 91).

8 % der Nachschlagewerke beinhalten ein Forum, bei weiteren 9 % bietet die übergeordnete Seite ein Forum an. Zwei von 75 Nachschlagewerken (je 1,3 %) bieten mit einem zusätzlichen Wiki bzw. einem Weblog alternative Web-2.0-Elemente an. Rund 80 % der 75 Nachschlagewerke enthalten allerdings keinerlei Web-2.0-basierte Zusatzangebote.

9) Abgeschlossenheit des Nachschlagewerks (s. STORRER/FREESE 1996, 117ff.). Über das Merkmal der Abgeschlossenheit wurde u. a. erhoben, ob es sich bei den Nachschlagewerken selbst um Web-2.0-Erscheinungen handelt, indem überprüft wurde, inwiefern sie eine Nutzerpartizipation ermöglichen. Knapp zwei Drittel der Werke (63 %) sind im Aufbau abgeschlossen und entsprechen damit nicht dem Web-2.0-Prinzip. 36 % der Werke bieten zwar einen kontrollierten Aufbau an, also beispielsweise das Vorschlagen von Lemmata oder Einsenden von Artikeln. Einen unkontrollierten Aufbau lässt aber lediglich ein Nachschlagewerk zu, und dies auch nur nach vorheriger Registrierung und mit weiteren Einschränkungen.

Unter den 27 Nachschlagewerken, die einen kontrollierten Aufbau anbieten, fällt zunächst auf, dass 78 % Sprach- wie auch Sachinformationen beinhalten. Eine weitere auffällige Kombination findet sich in Bezug auf die 13 Nachschlagewerke aus dem Bereich *EDV und Informatik*, von denen 77 % einen kontrollierten Aufbau zulassen. Dass gerade EDV-bezogene Nachschlagewerke diese Möglichkeit bieten, könnte einerseits damit zusammenhängen, dass ihre Emittenten besonders computeraffin sind und sich daher gern dauerhaft mit dem Ausbau und der Arbeit an ihrem Computer(-Nachschlagewerk) beschäftigen. Zum anderen verändert und erweitert sich der Wortschatz im Bereich *EDV* aufgrund technischer Entwicklungen relativ schnell und bedarf eher Aktualisierungen als Wortschätze anderer Domänen.

4. Quellen lexikografischen (Herstellungs-)Wissens

Um unter anderem etwas über die *Quellen lexikografischen Wissens* und das *Wissen über Lexikografie als professionelle Tätigkeit* (siehe

Abschnitt 5) von Amateurlexikografen zu erfahren, wurden die Emittenten der 75 Nachschlagewerke per Email befragt.²⁵ 29 Emittenten nahmen an der Befragung teil.²⁶ Damit lag die Beteiligung für eine freiwillige E-Mail-Befragung mit rund 40 % ausgesprochen hoch, was für eine hohe Motivation der Emittenten in Bezug auf ihr Projekt spricht. Die Befragung enthielt offene wie auch geschlossene Fragen mit vorgegebenen Antwortoptionen, die in einem zweiten Schritt eine freie Antwort zuließen. Ausgewertet wurden die freien Antworten zunächst qualitativ, indem aus dem Material heraus Kategorien durch sinnvolles Zusammenfassen ähnlicher Antworten bzw. Antwortbestandteile gebildet wurden.²⁷ Zudem wurde die Häufigkeit des Vorkommens der jeweiligen Antwortkategorien betrachtet. Aufgrund der relativ kleinen Datengrundlage können die folgenden Ergebnisse, insbesondere alle quantitativen Angaben, lediglich Tendenzen aufzeigen.

Als Vorüberlegung wurden verschiedene potenzielle Quellen lexikografischen Wissens für Amateurlexikografen in Betracht gezogen:

- a) existierende Nachschlagewerke, digital oder in Printform,

²⁵ Weitere Aspekte, die in der Befragung thematisiert wurden, waren die Motivation der Emittenten, ein Nachschlagewerk zu erstellen, und die von ihnen beabsichtigte Funktion der Werke (vgl. STEUBE i. E.).

²⁶ Da ein Befragungsteilnehmer der Emittent zweier im Untersuchungskorpus enthaltener Nachschlagewerke ist, liegen für insgesamt 30 der 75 Lexika Befragungsergebnisse vor. Im Folgenden wird jedoch die Anzahl der beantworteten Fragebögen (29) zugrunde gelegt.

²⁷ Dieses induktive Vorgehen lehnt sich an die Methode der qualitativen Inhaltsanalyse an (vgl. MAYRING 2008, 74f.), die allerdings i. d. R. mit größeren Datengrundlagen arbeitet. Gerade für eine explorative Untersuchung bietet sich diese Auswertungsmethode jedoch an, da das Datenmaterial nicht durch zuvor festgelegte Kategorien beeinflusst wird, sondern diese aus der Untersuchungsgrundlage erhoben bzw. entwickelt werden können und „[...] eine Erfassung des Gegenstands in der Sprache des Materials“ (MAYRING 2008, 75) ermöglichen. Grundsätzlich muss zur Wahl der Erhebungsmethode ‚Befragung‘ aber angemerkt werden, dass nur das ausgewertet werden kann, was von den Befragungsteilnehmern geantwortet wird.

- b) Wissensmanagement- bzw. Lexikon-Software²⁸ und
- c) lexikografische Ratgeber bzw. ‚Manuals‘ zur Erstellung von Nachschlagewerken.

Nach einer zunächst allgemeinen Frage nach generellen Informationsquellen zur Erstellung von Nachschlagewerken wurden die drei oben genannten potenziellen Quellen abgefragt. Im Folgenden werden die an die Emittenten gerichteten Fragen, die aus dem Antwortmaterial gebildeten Kategorien und die Anzahl ihrer Nennungen (in Klammern) aufgeführt. Berücksichtigt wurden alle Nennungen im Antwortmaterial, so dass mehrere Nennungen pro Befragungsteilnehmer möglich waren.

Frage 1. *Haben Sie sich im Vorfeld über die Erstellung von (Online-)Lexika informiert? Wenn JA: Bitte nennen Sie nachfolgend Ihre Informationsquelle wie z. B. Internetforen, andere Informationsquellen im Internet, Ratgeberbücher, Hilfe durch Personen mit bestimmten Vorkenntnissen usw. Wenn möglich, nennen Sie bitte konkrete Quellen.*

Zwei Emittenten bejahten diese Frage, 27 verneinten sie. Einige Emittenten gaben zusätzliche Kommentare dazu, warum sie sich nicht über die Erstellung von Online-Nachschlagewerken informieren konnten. Genannt wurden ein Mangel an Online-Nachschlagewerken zum Entstehungszeitpunkt ihres Werks und das Fehlen von Ratgebern. Die in der Befragung später abgefragten lexikografischen Ratgeber (s. u., Frage 4) scheinen den Emittenten nicht bekannt zu sein.

Folgende Antwortkategorien zu den allgemeinen Wissensquellen zur Erstellung von Nachschlagewerken konnten gebildet werden: *Gedruckte Nachschlagewerke* (2), *andere gedruckte ‚Bücher‘* (1), *Internetrecherche* (1), *Wikipedia* (1), *Informationen über dritte Personen* (1), *Literatur zum Thema EDV online und offline* (1).

²⁸ Bei dieser Art von Software handelt es sich um Programme oder Erweiterungen von Programmen, die verschiedene Arbeitsschritte des Herstellungsprozesses vorgeben und teilweise automatisiert durchführen, z. B. die Stichwortauswahl anhand eines Textkorpus oder Erstellung von Verweisen (vgl. STEUBE i. E., 86f.).

Frage 2. *Haben Sie sich beim Aufbau ihres Nachschlagewerks an existierenden Lexika orientiert? Wenn JA: Handelt es sich um gedruckte, CD-Rom- oder Online-Lexika? Bitte nennen Sie wenn möglich konkrete Werke.*

Diese Frage bejahten neun und verneinten 20 Befragungsteilnehmer, acht der neun Emittenten machten weitere Angaben, die zu folgenden Antwortkategorien zu den zur Orientierung dienenden Typen von Werken führten: *Gedruckte Nachschlagewerke* (3), *Online-Nachschlagewerke* (2), *Fachliteratur* (1).

Frage 3. *Haben Sie eine bestimmte Software für die Lexikon-Erstellung genutzt? Wenn JA: Welche?*

Von den 18 Teilnehmern, die diese Frage mit ‚ja‘ beantworteten, machten 13 nähere Angaben. Folgende Kategorien zu verwendeter Software lassen sich aus den Antworten ableiten: *Text- oder HTML-Editor* (13), *eigene Programmierung* (6), *Content Management System (CMS)* (2), *CMS-Erweiterung für Glossare* (1), *Blogging-Software* (1).

Bei Betrachtung der Antwortkategorien wird deutlich, dass es sich fast ausschließlich um Software handelt, die zur technischen Umsetzung, nicht aber zur lexikografischen Unterstützung dient. Lediglich ein Emittent gibt an, die Erweiterung eines Content Management Systems zur Glossar-Erstellung eingesetzt zu haben.²⁹

Als weitere potenzielle Quelle lexikografischen Wissens wurde lexikografische Ratgeberliteratur in Betracht gezogen. Bei der Recherche nach entsprechender Literatur wurden ausschließlich englischsprachige Werke ausfindig gemacht, die in der Frage an die Emittenten aufgeführt wurden (für einen detaillierteren Überblick zu den Ratgebern siehe STEUBE i. E., 87ff.). Zudem wurde die Möglichkeit gegeben, andere ggf. verwendete Ratgeber zu nennen.

²⁹ Es handelt sich dabei um das Content Management System ‚Typo3‘ und dessen Erweiterung ‚Content parser and tagger (Glossary)‘ (s. <http://typo3.org/extensions/repository/view/contagged> [2014-07-14]).

Frage 4. *Kennen Sie einen der folgenden Ratgeber?*

- ATKINS, Beryl T. S./RUNDELL, Michael (2008): *The Oxford Guide to Practical Lexicography*. Oxford [u. a.]: Oxford Univ. Press.
- BERGENHOLTZ, Henning/TARP, Sven (1995): *Manual of Specialised Lexicography*. Amsterdam [u. a.]: Benjamins.
- SVENSÉN, Bo (1993): *Practical lexicography: principles and methods of dictionary-making*. Oxford [u. a.]: Oxford Univ. Press.
- VAN STERKENBURG, Piet (2003): *A practical guide of Lexicography*. Oxford [u. a.]: Oxford Univ. Press.
- ZGUSTA, Ladislav (1971): *Manual of Lexicography*. Prague [u. a.]: Academia [u. a.].

Alle 29 Teilnehmer beantworteten diese Frage mit ‚nein‘. Weder war ihnen einer der genannten Ratgeber bekannt, noch nannten sie alternative Werke.

5. Wissen über und Bewusstsein von Lexikografie als professioneller Tätigkeit

Die Arbeit an den Nachschlagewerken, so wurde in Abschnitt 4 deutlich, fand größtenteils ohne Informationsbeschaffung auf lexikografischer Ebene statt. Diese Gegebenheit und die Tatsache, dass keiner der Befragungsteilnehmenden lexikografische Ratgeberliteratur kannte, geschweige denn eingesetzt hat, legt die Frage nahe, welches Wissen bzw. was für Vorstellungen die Befragten eigentlich zur (professionellen) Tätigkeit der Lexikografie haben. Folgende Frage wurde ihnen im Rahmen der Befragung dazu gestellt:

Frage 5. *Ist Ihnen das Berufsbild des Lexikografen bekannt? Wenn JA: Beschreiben Sie kurz, was Ihrer Meinung nach die Tätigkeit eines Lexikografen/einer Lexikografin ausmacht.*³⁰

³⁰ DIEKMANN (vgl. 2011, 482) zufolge handelt es sich bei einer so formulierten Frage um eine sogenannte ‚heikle Frage‘, da hier direkt Wissen abge-

28 Emittenten beantworteten die Frage mit ‚nein‘, ein Emittent beantwortete sie nicht. Obwohl sie die Frage verneinten, gaben vier Personen eine Einschätzung hinsichtlich der Tätigkeit von Lexikografen ab: „*Erstellung und Konzeption von Lexika*“, „*Mundgerechtes Aufbereiten von Texten*“, „*die etwas erklären (sollen)?!*“, „*Kürze und Verständlichkeit*“ bzw. „*Recherchen zu Begriffen, die in ein Lexikon gehören und ggf. erstellen eigener Definitionen*“.

Als mögliche Tätigkeitsfelder von Lexikografen werden also zum einen die Herstellung von (*Text-*)*Verständlichkeit* (2) und zum anderen die *Lexikonerstellung* (2) vermutet.

Die Kontrollfrage *Haben Sie für diese Frage eine Recherche durchgeführt?* wurde von 26 Emittenten mit ‚nein‘ beantwortet, drei Befragungsteilnehmer gaben keine Antwort. Somit kann davon ausgegangen werden, dass es sich bei den oben formulierten Antworten tatsächlich um Einschätzungen ohne vorherige Recherche handelt.

Auch wenn der Emittent eines psychologischen Nachschlagewerks angibt, bereits an mehreren Fachlexika mitgewirkt zu haben, sprechen die Ergebnisse insgesamt dafür, dass es den Befragten nicht nur an Wissen über die konkrete Tätigkeit eines Lexikografen mangelt, sondern dass vielen ein grundsätzliches Bewusstsein für die professionelle Beschäftigung mit Wörterbüchern fehlt – sowohl in Form der praktischen Lexikografie und erst recht im Rahmen einer wissenschaftlichen Auseinandersetzung mit Wörterbüchern. So merkt der Emittent eines EDV-bezogenen Werks zu einer anderen Frage an:³¹ „[...] *Es erschrickt*

fragt wird. Im Falle einer Negativantwort kann beim Befragungsteilnehmer das Gefühl unzulänglichen Wissens entstehen, das er oder sie versucht z. B. durch eine Recherche auszugleichen. Um diese Möglichkeit zu minimieren oder aber zu erfassen, wurde mit Hilfe der Folgefrage überprüft, ob Hilfsmittel herangezogen wurden. Falschangaben sind bei dieser Kontrollfrage zwar nicht auszuschließen, sie müsste allerdings bewusst falsch beantwortet werden.

³¹ Es handelte sich um die Frage, worauf die Emittenten bei ihrer Recherche zur Lexikon-Erstellung keine Antwort gefunden haben, und welche Punkte sie in Bezug auf die Erstellung von Nachschlagewerken ggf. noch interessieren würden.

mich fast, dass das ein wissenschaftliches Gebiet sein soll/ist. Über das Erstellen [von Nachschlagewerken] habe ich noch nie recherchiert! Nur über die Begriffe... [...]“.

Auch Beiträge in Internetforen, in denen Fragen zur Erstellung von Online-Nachschlagewerken aufgeworfen werden, spiegeln ein mangelndes Bewusstsein für die Komplexität des lexikografischen Prozesses bei der professionellen Wörterbucheerstellung wider (vgl. STEUBE i. E., 85f.). Diese Ergebnisse und Beobachtungen werfen die Frage nach einem öffentlichen Bewusstsein für und Wissen über die Lexikografie als professionelle Tätigkeit sowie Konsequenzen für Lexikografie und Wörterbuchforschung auf.

6. Zusammenfassung, Diskussion der Ergebnisse und Ausblick

Die in Abschnitt 3 behandelten Merkmale der Nachschlagewerke sollten einen ersten Eindruck davon vermitteln, **was Amateurlexikografen über Wörterbücher wissen** bzw. welche wörterbuchbezogenen Vorstellungen sie haben und in ihren eigenen Nachschlagewerken umsetzen. Zwei Erkenntnisse können bezüglich dieser Zielsetzung aus den Ergebnissen abgeleitet werden.

Hinsichtlich der Merkmale lässt sich feststellen, dass die folgende Merkmalskombination unter den betrachteten Werken die am häufigsten vertretene darstellt und damit gewissermaßen einen (Proto-)Typ des amateurlexikografischen Nachschlagewerks umreißt: Von den acht Nachschlagewerken mit blätterndem Zugriff sind sieben in ihrem Aufbau abgeschlossen, keines enthält Web-2.0-Elemente oder multimediale Bestandteile, und alle sind in eine übergeordnete Webpräsenz integriert. Bei den betreffenden Nachschlagewerken handelt es sich um kleinere, einfach angelegte ‚Glossare‘ im Rahmen privater oder themenspezifischer Internetpräsenzen, die zusätzlich zu anderen Informationsangeboten, z. B. Foren, zur Verfügung gestellt werden.

Betrachtet man das Vorkommen der Merkmale einzeln, so liegt es nahe, über stereotype Vorstellungen der Amateurlexikografen zu Nachschlagewerken nachzudenken. Allerdings kann allein anhand der Häu-

figkeit einer Merkmalsausprägung noch keine Aussage über stereotype Vorstellungen getroffen werden, da es sein kann, dass Merkmale aus rein pragmatischen Gründen anderen vorgezogen wurden. Erst dort, wo zu einem bestimmten lexikografischen Merkmal aus Sicht des Amateurlexikografen keine ernsthafte Alternative besteht, könnte man mit PÜSCHEL (1989, 128) von „lexikographischen Stereotypen“ sprechen. In den Antworten zur Befragung lassen sich einige wenige Äußerungen finden, die solche Überzeugungen zu Tage bringen, zum einen in Bezug auf die alphabetische Anordnung der Artikel und zum anderen in Bezug auf die Hypertextualität von Online-Nachschlagewerken. So gibt ein Befragungsteilnehmer auf die Frage, an welcher Art von Nachschlagewerk sich die Emittenten gegebenenfalls bei der eigenen Wörterbucherstellung orientiert haben (Frage 2), an, „den *allgemeinen Regeln für gedruckte Lexika (alphabetische Sortierung, übersichtliches Format, möglichst kurze Texte) und für Hypertexte (anklickbare Verlinkungen)*“ gefolgt zu sein. Ein anderer Emittent gibt das „*lateinische Alphabet*“ als Orientierung an und fügt hinzu, kein „[...] *konkretes Werk dieser Art [zu kennen], das das nicht täte*“. Da beide Merkmalsausprägungen zudem besonders häufig im Untersuchungsmaterial vorkommen (s. Abschnitt 3), könnte man hier also von einem einerseits lexikografischen und einem andererseits medienbezogenen Stereotyp in Bezug auf die Gestaltung von Internet-Wörterbüchern sprechen.³²

Hinsichtlich der **Quellen lexikografischen Wissens** wurde deutlich, dass sich immerhin ein Drittel der Emittenten an existierenden Nachschlagewerken orientiert. In den freien Antworten werden *gedruckte Nachschlagewerke* am häufigsten als Informationsquellen genannt (insgesamt fünfmal von unterschiedlichen Emittenten). Daneben dienten auch zweimal *Online-Lexika* und einmal *Wikipedia* als Orientierung für die Erstellung ihres Nachschlagewerks.³³ Nur ein Emittent

³² Um weitere Stereotype aufdecken zu können, müsste man in einer weiteren Befragung auf Seiten der Emittenten mehr Aussagen wie die zitierten provozieren.

³³ Darüber hinaus können diese Quellen natürlich auch zur Klärung fachlicher Fragen (fachlich-inhaltliche Wissens Ebene) gedient haben.

verwendete eine *Software*-Erweiterung zur Unterstützung bei der Lexikonerstellung. Keinem der Befragungsteilnehmer war eines der genannten fünf lexikografischen ‚Manuals‘ bekannt; auch wurden keine alternativen Ratgeber genannt.

Insgesamt fällt auf, dass nur wenige Quellen und Hilfsmittel für lexikografische Fragen von den Amateurlexikografen verwendet (bzw. genannt) werden. Bei den wenigen Quellen, die angeführt wurden, mag die Tendenz zur Orientierung an gedruckten Publikationen erstaunen, seien es Nachschlagewerke oder andere (Fach-)Bücher. Möglicherweise haben die Emittenten eher zuverlässig wirkende Print- als gegebenenfalls unsichere Internetquellen gewählt oder nennen diese bewusst bei der Befragung.

Die Ergebnisse zum amateurlexikografischen **Wissen über Lexikografie als professioneller Tätigkeit** und die diesbezüglichen Aussagen einiger Amateurlexikografen deuten auf einen offensichtlichen Mangel an Bewusstheit für die professionelle praktische wie auch die wissenschaftliche Beschäftigung mit Wörterbüchern hin. Während WICHTER (1994, 28f.) „anerkannte“ und „bestrittene“ Vertikalität bei Mitgliedern einer Wissensdomäne unterscheidet, kann für die befragten Amateurlexikografen von ‚unbewusster Vertikalität‘ gesprochen werden. Wissen über die Existenz einer ausgewiesenen Expertise in der Domäne der Lexikografie einerseits und über das Fehlen an eigenem Wissen andererseits ist bei den befragten Amateurlexikografen nicht oder kaum zu beobachten. Als Ursache für diesen ‚blinden Fleck‘ kann zum einen die Komplexität der Wörterbucharstellung eine Rolle spielen. Gerade bei der Erstellung eines sachbezogenen Nachschlagewerks treten lexikografische Fragen vor fachlich-inhaltlichen und internettechnischen in den Hintergrund. Zurückzuführen ist das fehlende lexikografische Bewusstsein auf Seiten der Amateurlexikografen aber zweifelsohne auch auf fehlende Kontaktpunkte mit Personen oder Einrichtungen, in denen professionelle Wörterbucharstellung betrieben oder über Wörterbücher geforscht wird. Die Befragungsergebnisse zum Berufsbild des Lexikografen fallen dabei noch deutlicher aus als diejenigen in BARTH-WEINGARTENS und METZGERS (2004) Untersuchung

zum gesellschaftlichen Wissen über und Vorstellungen von Sprachwissenschaft. Hinsichtlich des gesellschaftlichen Bekanntheitsgrads muss daher der Tatsache ins Auge gesehen werden, dass die Lexikografie – nimmt man die Befragungsergebnisse ernst – als Stiefkind der Linguistik gelten kann bzw. muss (s. STEUBE i. E., 192).

Umgekehrt scheint aber auch auf Seiten der Lexikografie ein verzerrtes Bild von Amateurlexikografen zu bestehen, wenn Wörterbuchforscher lexikografische Ratgeber verfassen, die als Zielgruppe unter anderem lexikografisch interessierte Laien ansprechen sollen (siehe dazu STEUBE i. E., 88f.). Das soziologische Konzept des „imaginier-te[n] Laien“ (GISLER u. a. 2004, 7) erfasst diese Situation treffend; die Autoren beziehen sich damit auf eine nur in der Vorstellung von Experten existierende ‚Idee‘ des Laien (hier: Amateurs), seine kognitiven Voraussetzungen und sein Wissen zum betreffenden Gegenstand.

Zugespitzt formuliert ist das Verhältnis zwischen (Meta-)Lexikografie und Amateurlexikografie somit geprägt von gegenseitigem Nicht-Wissen. Gerade als anwendungsbezogene Teildisziplin der Sprachwissenschaft sollte der Lexikografie wie auch der praktisch orientierten Wörterbuchforschung aber daran gelegen sein, Kontakt zu den Zielgruppen und Abnehmern ihrer Produkte herzustellen, zumindest aber sich in deren Bewusstsein zu rücken.

Eine Möglichkeit, dies in Zukunft umzusetzen, könnte u. a. der verstärkte Einbezug von Nutzern bzw. potenziellen Amateurlexikografen beim Auf- und Ausbau professioneller Nachschlagewerke sein, wie es von einigen etablierten Wörterbüchern im Internet bereits gehandhabt wird (vgl. RAUTMANN 2014 und THIER 2014). Eine Nutzerbeteiligung – in welchem Ausmaß dies auch sei – auch an fachbezogenen Nachschlagewerken würde lexikografisch ambitionierten und fachlich z. T. sehr versierten Personen die Möglichkeit geben, sich lexikografisch zu betätigen und gleichzeitig ein Bewusstsein für Wörterbucherstellung als komplexen Prozess und professionelle Tätigkeit zu schaffen.

Darüber hinaus ist weitere Forschung im Bereich der Benutzerbeteiligung wie auch der eigenständigen Amateurlexikografie anzustre-

ben, um für die praktische Lexikografie angemessene und ‚professionelle‘ Lösungen im Umgang mit Amateuren im Internet zu erarbeiten.

Nachschlagewerke

- DORNSEIFF, FRANZ (1970): Der Deutsche Wortschatz nach Sachgruppen. 7., unveränderte Auflage. Berlin/New York: De Gruyter.
- DORNSEIFF, FRANZ (2004): Der deutsche Wortschatz nach Sachgruppen. 8., völlig neu bearbeitete Auflage und mit einem vollständig alphabetischen Zugriffsregister versehene Auflage. Berlin/New York: De Gruyter.
- DUW (2007): Duden, Deutsches Universalwörterbuch. 6., überarbeitete und erweiterte Auflage. Hrsg. vom Wissenschaftlichen Rat der Dudenredaktion: KUNKEL-RAZUM, KATHRIN/SCHOLZE-STUBENRECHT, WERNER/WERMKE, MATHIAS. Mannheim [u. a.]: Dudenverlag.

Literatur

- ABEL, ANDREA/KLOSA, ANNETTE (2014a): Einleitung: „Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess“. In: ABEL/KLOSA (Hrsg.), 3–7.
- ABEL, ANDREA/KLOSA, ANNETTE (Hrsg.) (2014b): Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“. Mannheim: Institut für Deutsche Sprache (OPAL – Online publizierte Arbeiten zur Linguistik. 4/2014). Online: <http://pub.ids-mannheim.de/laufend/opal/opal14-4.html> [Abruf: 2014-07-15].
- ALBY, TOM (2008): Web 2.0: Konzepte, Anwendungen, Technologien. 3., überarbeitete Auflage. München: Hanser.
- BARTH-WEINGARTEN, DAGMAR/METZGER, MARKUS (2004): „Sprachwissenschaft? Was ist das?“ Über einen praxisrelevanten Gegenstand und eine unbekannte Disziplin – eine Pilotumfrage. In: WICHTER, SIGURD/STENSCHKE, OLIVER (Hrsg.): Theorie, Steuerung und Medien des Wissenstransfers. Frankfurt am Main [u. a.]: Peter Lang (Transferwissenschaften. 2), 57–87.
- BELICA, CYRIL (1995): Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analysemethoden. Mannheim: Institut für Deutsche Sprache.
- BELICA, CYRIL (2001ff.): Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs. Mannheim: Institut für Deutsche Sprache. Online: <http://corpora.ids-mannheim.de/ccdb/> [Abruf: 2014-07-14].

- BERGENHOLTZ, HENNING/TARP, SVEN (1995): *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: Benjamins.
- CÖLFEN, HERMANN (2012): Wikipedia. In: HAß, ULRIKE (Hrsg.): *Große Lexika und Wörterbücher Europas*. Berlin/New York: De Gruyter, 509–523.
- ENGELBERG, STEFAN/LEMNITZER, LOTHAR (2009): *Lexikographie und Wörterbuchbenutzung*. 4., überarbeitete und erweiterte Auflage. Tübingen: Stauffenburg.
- FELBER, HELMUT/SCHAEDER, BURKHARD (1999): Typologie der Fachwörterbücher. In: HOFFMANN, LOTHAR/KALVERKÄMPER, HARTWIG/WIEGAND, HERBERT ERNST (Hrsg.): *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin/New York: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft. 14.2), 1725–1743.
- GABRIELATOS, COSTAS (2009): From Dictionary User to Amateur Lexicographer: Possibilities of on-line searches [Transcript of the lecture held on March 23rd 2009 in the Look it up yourself! A History of the Dictionary series at Gresham College, UK]. Online: <http://www.gresham.ac.uk/lectures-and-events/from-dictionary-user-to-amateur-lexicographer-possibilities-of-on-line-searches> [Abruf: 2014-07-14].
- GISLER, PRISKA/GUGGENHEIM, MICHAEL/MARANTA, ALESSANDRO/NOWOTNY, HELGA/POHL, CHRISTIAN (2004): Über die Schultern von ExpertInnen. In: GISLER, PRISKA/GUGGENHEIM, MICHAEL/MARANTA, ALESSANDRO/NOWOTNY, HELGA/POHL, CHRISTIAN (Hrsg.): *Imaginierte Laien. Die Macht der Vorstellung in wissenschaftlichen Expertisen*. Weilerswist: Velbrück, 7–39.
- GLINZ, HANS (1977): *Textanalyse und Verstehenstheorie*. 2., verbesserte Auflage. Wiesbaden: Athenäum.
- GUREVYCH, IRYNA/NIEMANN, ELISABETH (2011): The People's Web meets Linguistic Knowledge: Automatic Sense Alignment of Wikipedia and WordNet. In: BOS, JOHAN/PULMAN, STEPHEN (Hrsg.): *Proceedings of the Ninth International Conference on Computational Semantics (IWCS)*, Oxford, January 12th–14th 2011, 205–214. Online: <http://www.let.rug.nl/bos/pubs/BosPulman2011.pdf> [Abruf: 2014-07-14].
- GUREVYCH, IRYNA/WOLF, ELISABETH (2010): Aligning Sense Inventories in Wikipedia and WordNet. In: AKBC 2010 (Hrsg.): *First Workshop on Automated Knowledge Base Construction (AKBC)*, Grenoble, France, May 17th–19th 2010. Grenoble, 24–28.
- HAß, ULRIKE/SCHMITZ, ULRICH (2010): Lexikographie im Internet 2010 – Einleitung. In: *Lexicographica* 26, 1–17.
- HAUSMANN, FRANZ JOSEF/REICHMANN, OSKAR/WIEGAND, HERBERT ERNST/ZGUSTA, LADISLAV (Hrsg.) (1989): *Wörterbücher. Ein internationales*

- Handbuch zur Lexikographie. 1. Teilband. Berlin/New York: De Gruyter. (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.1)
- KEMMER, KATHARINA (2010): Online-Wörterbücher in der Wörterbuchkritik. Ein Evaluationsraster mit 39 Bewertungskriterien. Mannheim: Institut für Deutsche Sprache (OPAL – Online publizierte Arbeiten zur Linguistik. 2/2010). Online: <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2010-2.pdf> [Abruf: 2014-07-14].
- LEW, ROBERT (2014): User-generated content (UGC) in English online dictionaries. In: ABEL/KLOSA (Hrsg.), 8–26.
- MAYRING, PHILIPP (2008): Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim/Basel: Beltz.
- MELNICHOR, LUCA (2014): Ansätze zu einer halbkollaborativen Lexikographie. In: ABEL/KLOSA (Hrsg.), 27–46.
- O'REILLY, TIM (2005): What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. 09/30/2005. Online: <http://www.oreilly.de/artikel/web20.html> [Abruf: 2014-07-14].
- PÜSCHEL, ULRICH (1989): Wörterbücher und Laienbenutzung. In: HAUSMANN u. a. (Hrsg.), 128–135.
- RAUTMANN, KARIN (2014): Duden online und seine Nutzer. In: ABEL/KLOSA (Hrsg.), 49–62.
- REICHERT, RAMÓN (2008): Amateure im Netz. Selbstmanagement und Wissenstechnik im Web 2.0. Bielefeld: transkript.
- STEUBE, JULIA (i. E.): „Wie erstelle ich ein elektronisches Wörterbuch?“ Nichtprofessionelle Lexikografen und ihre Nachschlagewerke im Internet. Eine explorative Untersuchung. Duisburg: Universitätsverlag Rhein-Ruhr.
- STORRER, ANGELIKA/FREESE, KATRIN (1996): Wörterbücher im Internet. In: Deutsche Sprache 24/2, 97–153.
- THIER, KATRIN (2014): Das Oxford English Dictionary und seine Nutzer. In: ABEL/KLOSA (Hrsg.), 63–70.
- WICHTER, SIGURD (1994): Experten- und Laienwortschätze. Umriss einer Lexikologie der Vertikalität. Tübingen: Niemeyer (Reihe Germanistische Linguistik. 144)
- WIEGAND, HERBERT ERNST (1989): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: HAUSMANN u. a. (Hrsg.), 409–462.
- WIEGAND, HERBERT ERNST (1994): Zur Unterscheidung von semantischen und enzyklopädischen Daten in Fachwörterbüchern. In: SCHAEDELER, BURKHARD/BERGENHOLTZ, HENNING (Hrsg.): Fachwissen und seine Repräsentation in Wörterbüchern. Tübingen: Narr (Forum für Fachsprachen-Forschung. 23), 103–132.

NATHALIE MEDERAKE

Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive

Oder: Wenn das Leben dir Zitronen gibt ...

1. Einleitung

That knowledge is a property of the network means more than that crowds can have a type of wisdom in certain circumstances [...]. Rather, the change in the infrastructure of knowledge is altering knowledge's shape and nature. (WEINBERGER 2011, XIII)

Die Beobachtung kollaborativer Prozesse von Wissensauswahl und Wissensorganisation in einem enzyklopädisch-lexikalischen Informationssystem, hier Wikipedia, bedient nicht nur ein lexikografisches Interesse. Verschwimmende Muster des Zugriffs auf und der Darstellung von Inhalte(n) in einem Wiki berühren viele Bereiche in der Linguistik, nicht zuletzt Gesichtspunkte, die in einer interdisziplinären Textlinguistik diskutiert werden. Zu klären ist aus diesen beiden Perspektiven einerseits das Artikelwesen der Wikipedia¹, andererseits die Möglichkeit der strukturierten Beobachtung von Artikeln, die dem „Informationsverarbeitungssystem Buch“ (BOLZ 1993, 203) entwachsen sind. Die Ergebnisse zeigen, dass traditionelle Theorien nicht überholt sind, sondern vielmehr der Untersuchungsgegenstand als solcher diesen Theorien neue Aspekte abverlangt, aus denen zukünftige lexikografische

¹ Bemerkung zur Zitierkonvention: Sämtliche Webadressen, die sich auf Namensräume der Wikipedia beziehen, werden der Übersichtlichkeit halber abgekürzt durch das Namensraumpräfix (WIKIPEDIA), Doppelpunkt und Seitentitel (z. B. ARTIKEL). <http://de.wikipedia.org/wiki/Wikipedia:Artikel> löst sich damit auf in WIKIPEDIA:ARTIKEL.

Projekte oder auch Informationssysteme allgemeinerer Art, die sich der sozialen Medien bedienen, wiederum ihren Nutzen ziehen können.

2. Lexikografische und textlinguistische Aspekte in einem Wiki

2.1 Ein Nachschlageformat in Wiki-Funktion

Wikis sind web-basierte Anwendungen, die ein einfaches Editieren von WWW-Seiten erlauben, welche durch Hyperlinks verbunden sind (vgl. LEUF/CUNNINGHAM 2001). Hierbei greift sowohl das Konzept des Hypertexts als auch das des interaktionsorientierten Schreibens (STORRER 2013). Nachdem sich der Bekanntheitsgrad von Wikis bis Ende der 1990er-Jahre eher in Grenzen gehalten hat, gehören sie seit Anfang 2000 neben der sogenannten Blogosphäre zu den treibenden Kräften des Web 2.0-Phänomens (vgl. BRUNS 2008, 102).²

Nichts weniger als eine logische Folge dürfte für die geistigen Väter der Wikipedia, Jimmy Wales und Larry Sanger, der Weg von der Wiki-Software mit ihren Möglichkeiten des öffentlichen Austauschs und Bearbeitens von Daten zu einer offenen, internetbasierten Enzyklopädie gewesen sein (vgl. FIEBIG 2005, 22f.). Die gewachsene Popularität von Wikis, insbesondere die der seit 2001 existierenden Wikipedia, zeugt zudem nicht zuletzt davon, dass hier eine überaus geeignete Funktion³ für die Form kollaborativer Nachschlagewerke gefunden

² Für die Begriffsetablierung *Web 2.0* ist offenbar der 2005 von Tim O'Reilly veröffentlichte Artikel „What is Web 2.0“ (<http://oreilly.com/web2/archive/what-is-web-20.html>) mit verantwortlich, wobei sich das Phänomen – ihm folgend – wohl am besten unter dem Aspekt der ‚Nutzbarmachung kollektiver Intelligenz‘ fassen lässt (vgl. <http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html> [beide 2014-05-12]).

³ *Funktion* kann hier durchaus, in Anlehnung an TARP (z. B. 2013), in der Verbindung von lexikografisch relevanten Bedürfnissen des Nutzers und den Unterstützungsoptionen eines Nachschlagewerks als kulturelles Erzeugnis verstanden werden. Allerdings gilt es zu beachten, dass sich die extra- und intra-lexikografischen Elemente des Funktionenkonzepts TARPs (ebd.) bei Wikis stark, wenn nicht sogar unauflösbar, ineinanderfügen, da

wurde. Vermutet wird, dass insbesondere bei diesem Projekt die Möglichkeit der passiven Nutzung bei gleichzeitiger aktiver Mitgestaltung dem neuen Weg einer digitalen Wissensverbreitung⁴ entspricht, der die gemeinschaftliche und vernetzte Herstellung eines Wissensprodukts vorsieht (vgl. PSCEIDA 2010, 332). Wikipedia gilt derzeit als das umfangreichste Lexikon bzw. die bedeutendste Online-Enzyklopädie in gleich mehreren Sprachen (vgl. CÖLFEN 2012, 512f.).

Das für Wikipedia zum Einsatz kommende technische Hilfsmittel⁵ ist ein Wiki, für das – wie o. g. – das Hypertextkonzept konstitutiv ist. Die technische Umsetzung von Wikis erfolgt in einem Elementverbund von datenbankgestütztem Content-Management-System und kollektivem Schreibwerkzeug (vgl. etwa MÖLLER 2006). „[F]or building a repository of knowledge in bits and pieces“ (BARRETT 2009, 5) wird für Wikipedia gegenwärtig die Software *MediaWiki* als entsprechende Verwaltungsgrundlage genutzt. Auch braucht es eine auf die Benutzung vieler Mitwirkender ausgelegte Auszeichnungssprache (*Wikitext*), die die Bearbeitung von Informationstexten und audio-/visuellen Komponenten in einem Wiki-Artikel ermöglicht. Dazu gehört etwa das verein-

sich etwa Nutzerbedürfnisse direkt auf die lexikografische Unterstützung (insbesondere in Form von Daten, aber auch Zugriffsstrukturen) abbilden können.

⁴ Eine gerade anfangs geführte, scheinbar nicht enden wollende Diskussion drehte sich um die öffentliche Autorschaft und die flüchtige Qualität der Wikipedia-Artikel (z. B. KOHL/LIEBERT (2004) und CÖLFEN (2012, 509)). Einschränkend muss wohl betont werden, dass bei der Menge der Wikipedia-Einträge zunächst nicht zwangsläufig von qualitativ hochwertigen Ergebnissen oder Informationen ausgegangen werden konnte. Zweifelsohne entwickelte sich jedoch über die Jahre ein recht widerstandsfähiges System: Es scheint, dass gerade die Kombination aus hohen Beitragszahlen mit niedrighelligem Zugang für Autoren und wenig restriktiven Editierprozessen dazu führte, dass sich eine große Anzahl von Mitwirkenden fand, die letztendlich zum Gelingen und auch zur qualitativen Verbesserung der Wikipedia beigetragen und dem Projekt in der Folge zu einer gewissen Eigendynamik verholfen haben.

⁵ Vgl. in diesem Zusammenhang auch den Medien-Begriff bei HABSCHIED (2000, 136).

fachte Auszeichnen und Hervorheben bestimmter Elemente (z. B. von Hyperlinks), das durch bestimmte Symbol- und Zeichensetzung (wie Klammern und Anführungszeichen) dem in einem Artikel enthaltenen Text jeweils eine bestimmte Formatvorlage zuweist; der Speicherung der Inhalte dient dann eine relationale Datenbank.⁶ Die für diese mediale Spezifik erforderliche Handhabung ist nicht mehr unbedingt mit der analoger Nachschlagewerke vergleichbar und wirkt sich aufgrund der aktiven und passiven Beteiligung vieler Nutzer maßgeblich auf Kommunikations- und Interaktionsprozesse bezüglich Herstellung, Darstellung und Verbreitung von Wissen aus (vgl. PSCEIDA 2010, 87).

Für die Beschäftigung mit Wikipedia ist also hervorzuheben, dass eine wichtige Unterscheidung zwischen der Form (Enzyklopädie) und der Funktion (ermöglicht durch das Wiki) getroffen werden muss: Es ist zudem nicht allein die Wiki-Software, die sich gut für eine Online-Enzyklopädie eignet; vielmehr wird die Funktion und hierüber zwangsläufig der Benutzer zu ihrem Hauptleistungsträger. Die Eckpunkte Form, Funktion und insbesondere der Nutzer, der auf die Funktion maßgeblich einwirkt, müssen damit bei der Eruierung lexikografischer und textlinguistischer Aspekte vorbehaltlich einer kontinuierlichen Be- und Erarbeitung von Wiki-Inhalten berücksichtigt werden.

Zunächst ist deshalb festzustellen, dass Wikis von der Funktionsweise eines potenziell kollaborativen Werkzeugs ausgehend durch einfache Auszeichnungssprache und laufende Datenbankversionierung ein recht dynamisches (allerdings auch nachvollziehbares) Inventar von Informationstexten und weiteren medialen Komponenten in einem Artikel erlauben, was gerade im Hinblick auf die Beachtung des Wiki-Prinzips eine wichtige Rolle spielt.⁷ In den abgebildeten Domänengebieten können ferner recht unterschiedliche Schwerpunkte gesetzt werden, da die

⁶ Ausführliche Ausführungen dazu z. B. bei BARRETT (2009).

⁷ Dagegen ist Dynamik in Bezug auf den Hypertext eine wichtige, jedoch eher optionale Eigenschaft (vgl. STORRER 2008, 321), da dieser auch von statischen bzw. archivierten Inhalten gespeist wird.

Beitragenden jeweils in ihrem Kenntnisbereich⁸ arbeiten. Das Textinventar eines Wikis erfüllt zudem durch die über Hyperlinks geschaffenen Schnittstellen den Gesichtspunkt der Interaktivität (STORRER 2008, 322) und eröffnet dadurch Vernetzungsmöglichkeiten in ein erweitertes Artikel- bzw. Themenspektrum. Parallel, etwa über die Diskussionsseiten, ist außerdem ein begleitendes Textinventar⁹ zu beobachten, das das dynamisch-kollaborative Artikel- und Textwesen in einem Wiki vervollständigt (s. a. STVILIA u. a. 2005, 17). Das bedeutet: Einträge können beliebig oft überarbeitet werden, wodurch sich Form und Umfang des Dargestellten ständig an einen neuen Informationsstand anpassen. Zugleich wird jeder Überarbeitungsstand archiviert, was den Artikelprozess durchsichtig und rekonstruierbar macht.

2.2 Aspekte der lexikografischen Praxis

Zwangsläufige Überschneidungen, die sich für Sprach- und Sachwörterbücher ergeben¹⁰, gelten auch für die lexikografische und enzyklopädische Praxis. Grundlegende Aufgaben der Lexikografie (z. B. SCHLAEFER 2009, 71) sind deswegen disziplinenübergreifend aufzufassen. Zu ihnen gehören:

- (1) die Dokumentation und Erschließung des lexikalischen oder enzyklopädischen Wissens, je nach Ausrichtung des Nachschlagewerks,
- (2) die je nach Art des Informationsprofils festzustellenden inhaltlichen und methodischen Grundlagen der Bezugsdisziplinen sowie
- (3) die aus einem jeweils aktuellen Diskussionsstand abzuleitenden Erkenntnis- und Bezugsinteressen.

⁸ Die Art und Weise der Beteiligung etwa bei Wikipedia ist nicht festgelegt: Es besteht sowohl Bedarf, neue Artikel anzulegen bzw. bestehende (auch in Form von Bildern und Grafiken) zu erweitern als auch an Redaktionsvorgängen teilzunehmen und dgl. (WIKIPEDIA:BETEILIGEN [2014-04-28]).

⁹ In etwa vergleichbar mit der Erarbeitungsreihe von WICHTER (2011, 272).

¹⁰ Eine kritische Beobachtung dazu bei WIEGAND (1998); zu Abgrenzung und Wechselwirkungen von Sprach- und Sachlexikografie auch HAß-ZUMKEHR (2001, 291–296). Einen funktionstheoretischen Überblick über elektronische Enzyklopädien gibt FUERTES-OLIVERA (2013).

Eine sich in diesem aufgespannten Rahmen bewegende lexikografische bzw. enzyklopädische Arbeitsweise gilt nicht zuletzt für Nachschlagewerke, die sich der Wiki-Funktionsweise bedienen.¹¹

Ferner sind bei Wikipedia einerseits viele Merkmale einer typischen Enzyklopädie aufzufinden (vgl. FANDRYCH/THURMAIR 2011, 89–113), die grundsätzlich auf einem intertextuellen Aspekt¹² fußen, und zwar dahingehend, dass sie bereits in anderen Quellen vorliegende Informationen zusammenfassend und systematisierend aufbereiten sowie durch Fußnoten belegen. Die noch bei analogen Nachschlagewerken aufzufindende starke Selektion und Kondensation, um das jeweils Relevante eines Gegenstandes zu fokussieren¹³ tritt andererseits bei digitalen Inhalten immer mehr in den Hintergrund. Dessen ungeachtet gilt sowohl für traditionelle Lexikon- als auch für gegenwärtige Wikipedia-Artikel das Bemühen der einführenden Orientierung in einen Sachverhalt sowie eine ausgewogene und angemessene Themendarstellung (WIKIPEDIA:ARTIKEL#UMFANG). Darüber hinaus finden basale Prinzipien der Textarchitektur und Textstruktur einer Enzyklopädie (FANDRYCH/THURMAIR 2011, 93–103) bzw. des Informationsaufbaus eines Wörterbuchs (vgl. SCHLAEFER 2009, 80–92), wie beispielsweise die neben dem Artikelteil existierenden Einführungs- und Benutzungshinweise und die mikrostrukturelle Gliederung der Artikel in Stichwort, Bedeutungsteil und ggf. weiterführende Informationen, auch bei Wikipedia Verwendung.¹⁴

¹¹ Eine u. a. metalexikografische Perspektive in Bezug auf *Wiktionary* als Beispiel kollaborativer Lexikografie verfolgt die aufschlussreiche Arbeit von MEYER (2013).

¹² Auf die verschiedenen Ausprägungen des Komplexes der speziellen bzw. aktuellen Intertextualität im Sinne von KRAUSE (2000, 62–66) kann hier durchaus Bezug genommen werden. Für Wikipedia-Artikel gilt darüber hinaus grundsätzlich eine Belegpflicht (WIKIPEDIA:BELEGE [2014-04-28]).

¹³ Wobei sich insbesondere bei Sprachwörterbüchern „der funktionelle Wert eines bestimmten Textteils aus der textsyntaktischen Position ergibt“ (vgl. SCHLAEFER 2001, 306).

¹⁴ WIKIPEDIA:WIE_SCHREIBE_ICH_GUTE_ARTIKEL#AUFBAU_EINES_ARTIKELS [2014-04-28].

Diese Elemente der Wissensorganisation (vgl. SCHLAEFER 2009, 75f.) können, müssen aber nicht zwingend als stabile Einheiten bzw. Strukturelemente der Wikipedia¹⁵ auftreten: Da die – häufig euphorisch postulierte – ‚Weisheit der Vielen‘ (vgl. SUROWIECKI 2004) die Wissensbasis bestimmt, ist diese zwangsweise Änderungen unterworfen.¹⁶ Auch die Komponente der Wissensprüfung bzw. Wissensermittlung ist von individuellen oder auch gruppenspezifischen Erkenntnissen und Interessen abhängig (vgl. STEGBAUER 2009). Das bedeutet, dass im Zuge des sogenannten lexikografischen Prozesses (WIEGAND u. a. 2010, 8–21) der vorhandene Wissensstand ständig überprüft und modifiziert wird, wobei die individuelle bzw. gruppenspezifische Wissensprüfung ausgewertet wird und Veränderungen bezüglich der behandelten Gegenstände berücksichtigt werden (s. a. MEYER 2013, 49–53). Das heißt: Sowohl Lemmaauswahl, Artikeldisposition und Artikelinhalte als auch deren Grundlagen stehen ständig auf dem Prüfstand. Für die Phasen der Artikeltextproduktion bedeutet das entsprechend, dass Planung, Formulierung und Überprüfung niemals abgeschlossen sind sowie aufgrund von Diskussionen stagnieren oder überlagert sein können. Bei der Artikelerstellung wird darüber hinaus ein anfänglich subjektives und intuitives Vorgehen, das oft zu kurz greift und erst durch kontinuierliche Überarbeitung einen ausgewogenen Artikel entstehen lässt, durchaus einkalkuliert (WIKIPEDIA:ARTIKEL#UMFANG). Die Artikelresultate unterliegen damit der Aushandlung und sind stark vom Charakter des Vorläufigen geprägt.

¹⁵ Wikipedia-Vorgaben bezüglich der inhaltlichen (WIKIPEDIA:WIE_SCHREIBE_ICH_GUTE_ARTIKEL) und formalen Artikelstruktur (WIKIPEDIA:WIE_GUTE_ARTIKEL_AUSSEHEN) entwickeln sich projektbegleitend auf Informations- bzw. Hilfeseiten. Die Formatvorgaben können überdies für bestimmte Fachbereiche recht unterschiedlich sein (WIKIPEDIA:FORMATVORLAGE) [alle 2014-04-15].

¹⁶ Die in diesem Zusammenhang häufig diskutierte Intelligenz von Schwärmen beruht auf den fundamentalen Konzepten Stigmergie und Emergenz. Ihr Zusammenspiel führt Vertretern dieser Theorie zufolge dazu, dass die Fähigkeiten und Eigenschaften eines Gesamtsystems die eines Individuums übersteigen (vgl. KRAMER 2009, 41f.).

2.3 Dynamisch-kollaboratives Textwesen

Das dynamisch-kollaborative Textwesen innerhalb eines Wikis ist keinesfalls bestimmt von einem lediglich alternierenden Wechsel von Zugriff und Darstellung. Es wird vielmehr unablässig von einem Produser¹⁷ beeinflusst und stellt sich damit ungleich fluider¹⁸ dar, als es bei linear organisierten, traditionell gedruckten oder E-Texten (z. B. STORRER 2008, 323) der Fall sein könnte, deren Erarbeitung vor der Veröffentlichung abgeschlossen wird. Bei der im Prinzip kontinuierlich möglichen Überarbeitung der Textoberfläche und der Textinhalte steht die zeichenhafte Materialität im gleichen Maße wie die inhaltliche Zielsetzung der Artikeltexte beständig auf dem Prüfstand.¹⁹

Wagt man den Vergleich, dann wird mit der Wiki-Funktionsweise im Prinzip der Proklamation eines prozeduralen Textmodells (DE BEAUGRANDE/DRESSLER 1981) in all seiner Plastizität Ausdruck verliehen, bei dem es im Kern darum geht, Sprache als interaktives System zu betrachten sowie um die Frage, welche Operationen es sind, die die Einheiten und Muster während der Verwendung von sprachlichen Systemen regeln. In Bezug auf die Textgestaltung ergibt sich folglich, auch

¹⁷ BRUNS (2008, 2) argumentiert, dass für viele konsumierte Inhalte der Web-2.0-Generation (z. B. Weblogs, Wikipedia etc.) die Nutzer selbst verantwortlich seien und sie somit gleichzeitig Produzenten ihrer Wissensbasis werden, unabhängig davon, ob sie sich dieser Rolle bewusst sind. Das Portmanteau *Producer* ebenso wie das dazugehörige Konzept der Produsage für sogenannte nutzer-generierte Inhalte geht über die noch bei PSCHIDA gezogene Grenze von passiver Nutzung und aktiver Mitgestaltung (s. o.) hinaus.

¹⁸ Sowohl was den Wechsel von Produzenten- und Rezipientenrolle angeht als auch in Bezug auf die Aktualisier- und Ersetzbarkeit von Hypertextmodulen und Links: „This [process] blends previously separated textual creation efforts and their outcomes into a continuous, ongoing, unfinished process“ (BRUNS 2008, 139).

¹⁹ Für die Phasen der Textproduktion gilt ein Blick auf die grundlegenden Komponenten Planung, Übersetzung und Überprüfung (zusammengefasst bei WROBEL 2000), die sich unter Wiki-Bedingungen jedoch überlagern und wiederholen (s. o.).

unter Berücksichtigung des Ansatzes nach HEINEMANN/VIEHWEGER (1991), dass Texte zu bestimmten Zwecken und in bestimmten Situationen wissenschaftlich produzierte Äußerungen sind, „deren Sinn und Struktur mitbestimmt sind durch diese Zwecke und Situationen“ (GANSSEL/JÜRGENS 2009, 113).

Gemein ist diesen Ansätzen, dass sie die statische Auffassung von Text zu überwinden suchen, wobei es durchaus – und gerade in Bezug auf eine enzyklopädische Arbeitsweise – sinnvoll erscheinen mag, zusätzlich zur Prozesshaftigkeit der Textgestaltung eine thematische Strukturierung (vgl. BRINKER 2010, 65–69) und Orientierung anzunehmen. HELLWIG (1984, 14) stellt außerdem ohne jegliche Zweifel fest, „daß ein Thema im normalsprachlichen Sinne des Wortes etwas Fragliches ist, zu dem in einem Text eine Lösung gesucht wird“.

Bei der Stichwortauswahl und der Artikelerstellung wird damit im Prinzip bereits im Vorfeld ein Problem des Rezipienten antizipiert und folglich versucht, durch den Artikeltext bzw. seine Textteile eine Lösung anzubieten. Führt man sich jedoch vor Augen, wie in digitalen Medien Inhalte aufgerufen werden können, nämlich über die Eingabe bestimmter Stichwörter in eine Suchmaschine, könnte diese Annahme noch wie folgt konkretisiert werden: Der Artikeltext ist eine Antwort (oder das Explanans) auf eine über die Suchmaschine gestellte Frage (Explanandum). Das Suchergebnis liegt dann in Form eines Artikels vor, der in seiner Versionierungsgestalt als jeweilige Textfassung²⁰ eines Artikels konkretisiert werden kann.

²⁰ An dieser Stelle lohnt m. E. ein Aus- und Rückblick in die Editionsphilologie, die die über variante Stadien verlaufende Entwicklung eines in der Regel literarischen Werks zu erfassen sucht. Textfassungen stellen dabei nicht nur arbeitstheoretische Größen, sondern greifbare, vollendete oder nicht vollendete Ausführungen eines Werks dar, die durch Textidentität aufeinander beziehbar und durch Textvarianz voneinander unterscheidbar sind (vgl. SCHEIBE 1971, 17). Selbiges gilt für die Artikelversionen der Wikipedia, die nach jeder Änderung eines Wikipedia-Autors archiviert werden und über Zeit, Datum, Benutzername bzw. IP-Adresse des Bearbeiters identifizierbar sind.

Dieses Textverständnis ergänzt o. g. integrativ orientierte Textbegriffe, erfordert jedoch gleichzeitig eine neue Perspektive auf diese. Der Vorschlag von BRINKER (2010, 17), etwa für den Text eine begrenzte Folge von sprachlichen Zeichen anzunehmen, die in sich kohärent ist und die als Ganzes eine erkennbare kommunikative Funktion signalisiert, erfordert unter Berücksichtigung des explikativen Vertextungsmusters nach der Funktionsweise eines Wikis eine Neuorientierung. Kohärenz, Struktur und Funktion von Texten müssen neu diskutiert werden, insbesondere, wenn diese als elektronische Erzeugnisse vorkommen, die in den Händen vieler liegen (vgl. auch STORRER 2013).

2.4 Wo alle Fäden zusammenlaufen: die Aushandlungskohärenz

Der Text- bzw. Artikelprozess bei Wikipedia kann durch einen laufenden und bisweilen nicht mehr zu definierenden Wechsel von Produktion und Rezeption, von Schreib- und Verstehensvorgängen charakterisiert werden. Die von der analogen Arbeitsweise (z. B. SCHLAEFER 2009, 75f.) abzuleitenden Aufgaben für die Artikelerstellung, die etwa darin bestünden,

- (a) eine Wissensbasis durch Rezeption aufzubauen,
- (b) eine korpusgestützte und objektbereichsspezifische Wissensermittlung und Wissensüberprüfung durchzuführen,
- (c) die Wissensorganisation in einer spezifischen Form vorzunehmen und
- (d) die damit eng verbundene benutzerbezogene Wissensvermittlung zu erfüllen,

sind nicht mehr klar voneinander trennbar. Ein anzunehmender regelhafter Kreislauf, nach dem der Artikelprozess nach einer eingeleiteten Revision erneut beginnt, wird ebenfalls unterbrochen durch den ständigen Austausch zwischen einer Objektebene, also der Ebene, auf der die Artikelrepräsentation stattfindet, und einer Metaebene, auf der sich ex-

terne Faktoren wie der Produser, zeitlich-thematische Entwicklungen u. Ä. positionieren lassen.²¹

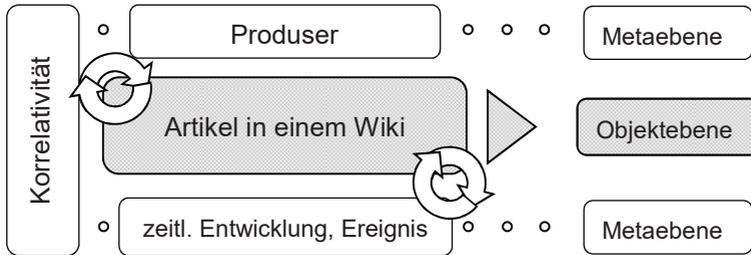


Abb. 1: Modellvorstellung Artikelprozesse in einem Wiki

Darüber hinaus existieren jedoch noch weitere Einflussfaktoren, wie die auf einer Metaebene zu situierende Etablierung und Diskussion etwa politischer Themen und deren Auswirkungen auf ein themenentsprechendes Artikelcluster der Wikipedia, die ggf. berücksichtigt werden müssten; hierauf sollte jedoch in einem anderen Rahmen eingegangen werden.²²

Der Umstand, dass die auf dieser so definierten Objektebene ablaufenden Textprozesse mit hinzuziehenden Einflüssen einer Metaebene tendenziell zu einer inhaltlichen Sinnkontinuität zusammenfinden, kann als Aushandlungskohärenz beschrieben werden. Bezeichnet wird damit der Prozess, der Artikelstruktur und -funktion laufend dadurch neu bestimmt, dass Kohärenz zwar als dominantes, aber nicht fixierbares

²¹ In der Wikipedia-Terminologie bezeichnet *Metabereich* bereits alle Bereiche außerhalb des Artikelnamensraums, d. h. Seiten mit entsprechendem Präfix, wie Hilfeseiten (z. B. HILFE:NAMENSRAUM) und Projektdiskussionen (DISKUSSION:ZITRONENPRESSE); vgl. HILFE:GLOSSAR#METABEREICH [2014-04-28].

²² Es sollte in diesem Zusammenhang darauf hingewiesen werden, dass Wikipedia per se keine Diskurse darstellen kann. Zu überprüfen bleibt jedoch, ob sich hier Diskurse kondensieren (vgl. etwa GIRNTH 2002, 52) oder Artikel und Diskussionen eine Art Transportfunktion erfüllen können.

Prinzip zugrunde liegt. Dies betrifft sowohl die Relation von Hyper-
textkonstituenten auf lokaler Ebene als auch die im Rahmen globaler
Kohärenz einzuordnende übergreifende thematische Gesamtvorstellung
und die Artikeltextfunktion in einem größeren Kommunikationszu-
sammenhang (vgl. STORRER 2004). Das veränderte Interagieren sprach-
licher und außersprachlicher Faktoren dürfte sich ferner sowohl auf
Kohärenzbildung als auch -planung (ebd.) auswirken, was die abzu-
grenzenden Phasen und Wechsel von einzusetzenden Mitteln und zu er-
zielender Wirkung anbelangt. Die angesprochene Sinnkontinuität kann
unter Vorbehalt des dynamisch-kollaborativen Textwesens deshalb
niemals als statisch, sondern immer nur als vorläufig und stetig aus-
und verhandelbar beschrieben werden.

Kommunikationstheoretisch sollte zudem davon ausgegangen wer-
den, dass die Artikelerstellung einem Aufmerksamkeitszyklus unter-
liegt, aufgrund dessen es vermutlich zu qualitativen Unterschieden der
Artikel kommt. Anzunehmen sind Konjunkturzyklen, ähnlich wie bei
der Verhandlung von Risikothemen in den Medien (z. B. VASTERMAN
2005), wonach durch ein Auslöseereignis²³ ein gesteigertes Hand-
lungs- und Problembewusstsein in Bezug auf die Artikel Tätigkeit existi-
ert, bevor die Aufmerksamkeit nach einem kürzeren oder längeren
Zeitraum schwindet bzw. erlahmt.²⁴ Natürlich ist ebenso das Wiki
selbst Gegenstand eines Konjunkturzyklus, wonach dessen Funktions-
weise anfangs eher eine Art Nischendasein im Internet führte. Erst nach
einer Popularisierungsphase zu Beginn der 2000er setzte sich ein Ver-
ständnis für die Vorteile, praktische Umsetzung und Grenzen der An-
wendbarkeit von Wikis durch.

²³ Die Spannweite der Artikelrelevanz (WIKIPEDIA:RELEVANZKRITERIEN [2014-04-15]) reicht dabei von aktuellen politischen Ereignissen bis zum jüngsten Sieger einer Castingshow. Ein Auslösefaktor dafür, einen bereits bestehenden Artikel zu verbessern, kann allerdings auch ein Schreibwettbewerb sein (WIKIPEDIA:SCHREIBWETTBEWERB [2014-04-15]).

²⁴ Vgl. auch die quantitativen Analysen bei STVILIA/GRASSER (2008).

3. Erfassung der Artikel-Text-Dynamik – ein Vorschlag

3.1 Explanans und Framesystematik der Zitronenpresse

Der in der Lexikologie und linguistischen Diskursanalyse etablierte Frame-Ansatz (z. B. KONERDING 1993 und ZIEM 2008) erwies sich bislang als probate Methode, um die Konstitution eines Gegenstands oder Konzepts über den Wortgebrauch im jeweiligen kontextuellen Zusammenhang festzustellen und die jeweils aktualisierten Wissensrahmen, auch über eine diachrone Anlage, zu erschließen. Entsprechend dieses Vorgehens wird hier die exemplarische Anwendung auf einen Wikipedia-Artikel in einer begrenzten Zeitspanne veranschaulicht, um die für Wikipedia charakteristische Textdynamik, insbesondere die thematisch-strukturelle Entfaltung über mehrere Artikelversionen hinweg, greifbar zu machen. Das Vorgehen soll als Anregung zu verstehen sein, auch weitere Artikel framesystematisch zu erschließen, um darüber Merkmale einer Aushandlungskohärenz zu erfassen. Angemerkt sei jedoch, dass inhaltliche Sinnkontinuitäten sich aus verschiedenen – hier nicht in Gänze diskutierten – Parametern zusammensetzen, die immer nur zu einem vorläufigen Ergebnis führen. Letztendlich dürfte sich auch hier das alltägliche Kreuz des Lexikografen zeigen, nämlich, dass am Ende jeder Artikel und jede Artikelentwicklung in gewisser Hinsicht ein Eigenleben besitzt. Nichtsdestoweniger bietet m. E. insbesondere die Setzung einer frameorientierten Perspektive optimale Vergleichsmomente, da sie sowohl die Entwicklung der Artikelstruktur als auch der zugeordneten Elemente in Form von slot-filler-Kombinationen erfassen kann.

Zu diesem Zweck wird im Folgenden ein Frameinstrumentarium auf den Artikel *Zitronenpresse* angewandt, das sich an den Matrixframegedanken KONERDINGS (1993) anlehnt. Ausgangspunkt ist das Artikellemma (*Zitronenpresse*) mit der dahinter zu vermutenden Frage bzw. dem Explanandum (*Was ist eine Zitronenpresse?*), das auf die

Dimension des anzuwendenden Matrixframes²⁵ hinweist. Das Explans kann je nach Artikelaufbau bestimmten Teilaspekten zugeordnet werden, die Präzisierungen des Frameinstrumentariums zulassen. So können zum einen etwa die Entwicklungen der Artikelstruktur verfolgt werden, zum anderen die der Artikelinhalte:

Frame: <i>Zitronenpresse</i> (Wikipedia-Artikel)			
Teilaspekt	Leerstelle	Füllwert in	
		Artikelversion A	Artikelversion B
Allgemein	funktionale Rolle	<i>Eine Zitronenpresse ist ein Küchengerät, mit dem man <u>Saft aus Zitrusfrüchten</u> wie Zitronen oder Limonen herauspresst.</i>	<i>Eine Zitronenpresse ist ein Küchengerät zum Herauspressen von <u>Saft aus Zitronen</u> oder <u>Limonen</u>. Zitruspressen sind universell einsetzbare Geräte [...], mit denen sich auch größere Früchte wie <u>Orangen</u>, <u>Grapefruits</u> bzw. <u>Pampelmusen</u> auspressen lassen.</i>
Geschichte	Entstehung	<i>Vermutlich funktionieren Zitronenpressen nach dem Modell von <u>Kartoffelpressen</u>. <u>Zitronensaft</u> löste in der mittelalterlichen Küche den <u>Verjus</u>, einen Saft aus unreifen Trauben, als Säuerungsmittel ab.</i>	<i>Vermutlich funktionieren Zitronenpressen nach dem Modell von <u>Kartoffelpressen</u>.</i> -----
	funktionaler Zusammenhang	-----	<i>Zangenartige Geräte, die, wie moderne <u>Knoblauchpressen</u>, die Früchte mit einem Stempfen in einen Zylinder mit Löchern drücken.</i>

²⁵ Mittels Hyperonymtypenreduktion ließe sich *Zitronenpresse* auf das höchste Hyperonym *Artefakt* zurückführen; damit entsprechen die Leerstellen des Zitronenpresse-Frames dem übergeordneten Matrixframe zu ‚Artefakt‘ (KONERDING 1993, 309–314), aus dem vorliegend einige wenige Leerstellen herausgegriffen sind.

Frame: <i>Zitronenpresse</i> (Wikipedia-Artikel)			
Teilaspekt	Leerstelle	Füllwert in	
		Artikelversion A	Artikelversion B
Funktionsprinzip	Funktionsteile	<i>Der austretende <u>Fruchtsaft</u> wird mit einem <u>Filter</u> abgetrennt.</i>	<i>Der austretende <u>Fruchtsaft</u> wird mit einem <u>Filter</u> abgetrennt.</i>
	besondere Eigenschaften	<i>Dabei nimmt der Saft zusätzlich <u>ätherische Öle</u> aus der Schale auf.</i>	<i>Dabei nimmt der Saft zusätzlich <u>ätherische Öle</u> aus der Schale auf.</i>

Tab. 1: Framesystematik für Wikipedia-Artikel²⁶

Im vorgestellten Beispiel ist der Fokus auf (durch Unterstreichungen gekennzeichnete) Hyperlinks und die sie umgebende Proposition in Bezug auf die Artikelstruktur, hier im Auftreten von gliederungsbedingten Teilaspekten, eingeschränkt. Dies hat folgende Gründe: Zunächst ist die Artikelstruktur ein Parameter für die Struktur, die, nimmt man das Explanandum als Ausgangspunkt, in verschiedene Teilfragen aufgeteilt werden kann. Ergänzend hierzu erfüllen Hyperlinks in ihrer hervorgehobenen Auszeichnung gleich mehrere Aufgaben im Artikel; sie sind sowohl Aufmerksamkeitslenker (vgl. RÖSSLER/BECK 2001, 143), indem sie die Textoberfläche zusätzlich auszeichnen, als auch Organisationselement, indem sie Schnittstellen für weitere Artikel und damit Wissensbereiche schaffen.

3.2 Verfestigung von Standardwerten

Entwicklungen über verschiedene Artikelversionen hinweg können durch vergleichbare Artikeltextsituationen, wie in Tab. 1, deutlich gemacht werden. Für die zahlreichen Teilaspekte und Leerstellen wird

²⁶ Die Artikelversionen A und B dieser Übersicht sind zur Veranschaulichung eines framesystematischen Zugriffs teilweise idealisiert und stehen exemplarisch für Versionen des Wikipedia-Artikels *Zitronenpresse* zwischen dem 29.12.2006 und dem 31.7.2007; der Zugriff auf diese erfolgt über die Versionsgeschichte: <http://de.wikipedia.org/w/index.php?title=Zitronenpresse&action=history> [2014-04-29].

dabei eine Auswahl in der Darstellung getroffen. Das oben angegebene Beispiel etwa zeigt für die angeführten Artikelversionen gleichbleibende Leerstellenbesetzungen für den Teilaspekt **Funktionsprinzip** an, wohingegen der Teilaspekt **Geschichte der Zitronenpresse** für Artikelversionen A und B unterschiedlich in Erwägung gezogen wurde. Kann nun über mehrere Artikelversionen hinweg beobachtet werden, dass für die gleiche Leerstelle eines Teilaspekts der Füllwert nicht geändert wird, ist die Verfestigung eines Standardwertes naheliegend (vgl. ZIEM 2008, 347).

Ebenso ist es möglich, dass eine Leerstelle stetig mit veränderten Füllwerten bedient wird und somit eine hohe Type-Frequenz vorliegt (ebd.): Für den Frame *Zitronenpresse* kann etwa der Wissensaspekt der **funktionalen Rolle** relevant sein, der darin liegt, *Zitrusfrüchte* auszupressen. Ob zu diesen nur *Zitronen* und *Limonen* oder auch größere Früchte, wie *Orangen*, *Pampelmusen* bzw. *Grapefruits*, gehören, muss nicht entschieden sein. Es zeigt sich also die Verfestigung der Leerstelle bzw. des Wissensaspekts **funktionale Rolle** eines Frames *Zitronenpresse*, nicht jedoch die der Füllwerte. Eine weitere Beobachtung besteht darin, dass ein anzunehmender Füllwert, wie *Saft*, über mehrere Artikelversionen hinweg an unterschiedlichen Positionen des Artikels auftaucht, doch keiner Leerstelle eindeutig zugeordnet werden kann. Ob es etwa primäre Eigenschaft bzw. Merkmal einer Zitronenpresse ist, seitdem *Zitronensaft den Verjus in der Küche abgelöst hat* (Entstehung), *Saft herauszupressen* (funktionale Rolle) oder *durch einen Filter abzutrennen* (Funktionsteil) scheinen auf den ersten Blick triviale Punkte zu sein. Sie zeugen jedoch davon, dass ein Token *Saft* sich verfestigt, wobei eine mögliche zugehörige Leerstelle gleichzeitig stark in den Hintergrund rückt. – Wie ZIEM (2008, 269) betont, bleiben weder Füll- noch Standardwerte zeitlich unbegrenzt stabil; sie sind jederzeit veränder- und austauschbar, was nicht zuletzt in einem Wiki deutlich zu machen ist.

4. Fazit und Ausblick

In diesem Beitrag wurde versucht, einige wenige grundsätzliche Eigenschaften der Wikipedia in ihrem lexikografischen und textlinguistischen Radius darzustellen. Die Beschäftigung mit einem enzyklopädisch-lexikalischen Informationssystem, das gleichzeitig der Social-Media-Technologie zugeordnet wird, zeigt sich als ein vielschichtiger Gegenstand, der dem Forschenden einige bekannte Grundprinzipien vor Augen führt, die jedoch nur Teilfaktoren eines wesentlich komplexeren Gesamtsystems sind, dessen umfangreiche Beschreibung den vorliegenden Rahmen sprengen würde.

Die vorgestellte Methodik der frameorientierten Erfassung der Artikelentwicklung stellt deshalb nur einen Weg von vielen möglichen dar. Sie zeigt in dieser beispielhaften Ausführung etwa Unzulänglichkeiten darin, dass wesentliche mediale Komponenten wie Bild, Ton und Video für das Nachschlageformat in der Funktionsweise eines Wikis nicht in die Überlegungen einbezogen wurden. Diese müssten bei weiteren und tiefergehenden Erforschungen des Wikipediamodus, in dem oft mit diversen medialen Möglichkeiten gearbeitet wird, unbedingt berücksichtigt werden.

In den vorliegenden Ausführungen ging es jedoch nicht (nur) darum, wie auch der Titel vermuten lässt, einer von Wikis dominierten Situation für Linguistik und Lexikografie das Beste abzugewinnen. Wie das angeführte Beispiel zeigte, handelte es sich zunächst um den Versuch, einen trivialen Allerwelts-Gegenstand in einen enzyklopädisch relevanten Artikel umzusetzen. Aus der Vogelperspektive betrachtet, ermöglichte dieses Experiment interessante Einblicke in die Artikelwelt der Wikipedia. Ein typisches Vorgehen, auf das beständig hinzuweisen ist, ist das der ständigen Artikel-Evaluierung, wie es auch die Artikelversionierung offenbart. Daneben erlaubt der jeden Artikel begleitende Diskussionsbereich einen Einblick in die Entwicklungsstationen und Bewertungsprozesse. Bezüglich der Diskussion, warum einem Artikel wie *Zitronenpresse* so viel Aufmerksamkeit geschenkt wurde, sei etwa noch folgender Kommentar angeführt:

Das Lemma mag trivial erscheinen (Ding [sic!] mit dem man gelbes Zeug auswringt), der zugehörige Artikel zeigt, dass dem nicht so ist und wie man sich einem (Artikel-)Gegenstand nähern kann. Man sollte ihn aber auch lesen ... (DISKUSSION:ZITRONENPRESSE [2014-04-15]).

Nicht nur als Artikel-Experiment aus Sicht der Wikipedia-Autoren kann *Zitronenpresse* als gelungen gewertet werden. Es offenbart als einfaches, aber prägnantes Beispiel das Artikeltextwesen, das hinter jeder bei Wikipedia stattfindenden Artikelentwicklung steht, als nicht auflösbare Verbindung von enzyklopädischer Form, Funktionsweise eines Wikis und Nutzerengagement.

Wenig überraschend zeigt sich schließlich vor dem Hintergrund Wikipedias, wie sich Kulturtechniken der Rezeption und Produktion vor dem Hintergrund derjenigen Phänomene, die das WWW hervorbringt, verändern. Dieses erfolgt jedoch weder losgelöst von analoger Motivation oder Reflexivität, noch ist die Veränderung von digitaler Demenz – wie die krude Formulierung lautet – geleitet. Möglich scheint hingegen, dass in Gegenwart einer fluiden Artikelrealität ein Umgang mit vorläufigen Resultaten alltäglich wird. Womit Inhalte in den Vordergrund drängen, die einer Aushandlungskohärenz unterliegen und sich dadurch in Struktur und Funktion stetig ändern können. Veränderungen in einer daran ausgerichteten Artikelpraxis, die zudem die aktive Rolle des Nutzers stark in den Mittelpunkt rücken oder sogar auf den Nutzer im Hinblick auf die Artikelerstellung angewiesen sind, werden damit weiterhin grundsätzlich für digitale Wörterbuchprojekte diskutiert werden müssen.

Literatur

- BARRETT, DANIEL J. (2009): MediaWiki. Wikipedia and beyond. Beijing.
- BEAUGRANDE, ROBERT DE/DRESSLER, WOLFGANG U. (1981): Einführung in die Textlinguistik. Tübingen (Konzepte der Sprach- und Literaturwissenschaft. 28).
- BOLZ, NORBERT (1993): Am Ende der Gutenberg-Galaxis: die neuen Kommunikationsverhältnisse. München.

- BRINKER, KLAUS (2010): *Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden.* 7., durchgesehene Auflage. Berlin (ESV basics. 29).
- BRUNS, AXEL (2008): *Blogs, Wikipedia, Second Life, and beyond. From production to produsage.* New York (Digital formations. 45).
- CÖLFEN, HERMANN (2012): *Wikipedia.* In: ULRIKE HAß (Hrsg.): *Große Lexika und Wörterbücher Europas. Europäische Enzyklopädien und Wörterbücher in historischen Porträts.* Berlin/Boston (De Gruyter Lexikon), 509–523.
- FANDRYCH, CHRISTIAN/THURMAIR, MARIA (2011): *Textsorten im Deutschen. Linguistische Analysen aus sprachdidaktischer Sicht.* Tübingen (Stauffenburg Linguistik. 57).
- FIEBIG, HENRIETTE (2005): *Wikipedia. Das Buch.* Original-Ausgabe, 2. Auflage. Berlin (Wikipress. 1).
- FUERTES-OLIVERA, PEDRO A. (2013): *Electronic Encyclopedias.* In: GOUWS u. a. (Hrsg.), 1069–1081.
- GANSEL, CHRISTINA/JÜRGENS, FRANK (2009): *Textlinguistik und Textgrammatik. Eine Einführung.* Neuausgabe der 2. überarbeiteten und ergänzten Auflage. Göttingen (UTB Sprachwissenschaft. 3265).
- GIRNTH, HEIKO (2002): *Sprache und Sprachverwendung in der Politik. Eine Einführung in die linguistische Analyse öffentlich-politischer Kommunikation.* Tübingen (Germanistische Arbeitshefte. 39).
- GOUWS, RUFUS H./HEID, ULRICH/SCHWEICKARD, WOLFGANG/WIEGAND, HERBERT ERNST (Hrsg.) (2013): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary volume: Recent Development with special focus on computational lexicography.* Berlin/Boston (Handbücher zur Sprach- und Kommunikationswissenschaft. 5.4).
- HABSCHEID, STEPHAN (2000): *Medium in der Pragmatik. Eine kritische Bestandsaufnahme.* In: *Deutsche Sprache* 2, 126–143.
- HAB-ZUMKEHR, ULRIKE (2001): *Deutsche Wörterbücher – Brennpunkt von Sprach- und Kulturgeschichte.* Berlin (De Gruyter Studienbuch).
- HEINEMANN, WOLFGANG/VIEHWEGER, DIETER (1991): *Textlinguistik. Eine Einführung.* Tübingen (Reihe Germanistische Linguistik Kollegbuch. 115).
- HELLWIG, PETER (1984): *Titulus oder über den Zusammenhang von Titeln und Texten. Titel sind ein Schlüssel zur Textkonstitution.* In: *Zeitschrift für germanistische Linguistik* 12, 1–20.
- KOHL, CHRISTIAN/LIEBERT, WOLF-ANDREAS (2004): *Selbstorganisation der Wissenschaftsvermittlung. Quellentransparenz, Kontroversität und Qualitätssicherung in der Internet-Enzyklopädie Wikipedia.* In: *Fachsprache* 3–4, 134–148.

- KONERDING, KLAUS-PETER (1993): Frames und lexikalisches Bedeutungswissen. Tübingen (Reihe Germanistische Linguistik. 142).
- KRAMER, OLIVER (2009): Computational Intelligence. Eine Einführung. Berlin (Informatik im Fokus).
- KRAUSE, WOLF-DIETER (2000): Textsorten. Kommunikationslinguistische und konfrontative Aspekte. Frankfurt am Main (Sprache. 33).
- LEUF, BO/CUNNINGHAM, WARD (2001): The Wiki Way: Quick Collaboration on the Web. Boston.
- MEYER, CHRISTIAN M. (2013): Wiktionary: The Metalexigraphic and Natural Language Processing Perspective. Darmstadt.
- MÖLLER, ERIK (2006): Die heimliche Medienrevolution. Wie Weblogs, Wikis und freie Software die Welt verändern. 2., erweiterte und aktualisierte Auflage. Hannover (Telepolis Magazin der Netzkultur).
- PSCHEDA, DANIELA (2010): Das Wikipedia-Universum. Wie das Internet unsere Wissenskultur verändert. Bielefeld (Kultur- und Medientheorie).
- RÖSSLER, PATRICK; BECK, KLAUS (2001): Aufmerksamkeitskalküle bei verschiedenen Modi der Online-Kommunikation. In: BECK, KLAUS/SCHWEIGER, WOLFGANG (Hrsg.): Attention please! Online-Kommunikation und Aufmerksamkeit. München (Internet research. 1), 141–158.
- SCHIEBE, SIEGFRIED (1971): Zu einigen Grundprinzipien einer historisch-kritischen Ausgabe. In: MARTENS, GUNTER/ZELLER, HANS (Hrsg.): Texte und Varianten. Probleme ihrer Edition und Interpretation. München, 1–44.
- SCHLAEFER, MICHAEL (2001): Wissenstransfer und Allgemeinwörterbuch. In: WICHTER, SIGURD/ANTOS, GERD (Hrsg.): Wissenstransfer zwischen Experten und Laien. Umriß einer Transferwissenschaft. Frankfurt am Main [u. a.] (Transferwissenschaften. 1), 301–314.
- SCHLAEFER, MICHAEL (2009): Lexikologie und Lexikographie. Eine Einführung am Beispiel deutscher Wörterbücher. 2., durchgesehene Auflage. Berlin (ESV basics. 40).
- STEGBAUER, CHRISTIAN (2009): Wikipedia. Das Rätsel der Kooperation. Unter Mitarbeit von ALEXANDER RAUSCH. Wiesbaden (Netzwerkforschung).
- STORRER, ANGELIKA (2004): Kohärenz in Hypertexten. In: Zeitschrift für germanistische Linguistik 31/2, 274–292.
- STORRER, ANGELIKA (2008): Hypertextlinguistik. In: JANICH, NINA (Hrsg.): Textlinguistik. 15 Einführungen. Tübingen (Narr-Studienbücher), 211–227.
- STORRER, ANGELIKA (2013): Neue Text- und Schreibformen im Internet: Das Beispiel Wikipedia. In: FEILKE, HELMUTH (Hrsg.): Textkompetenzen für die Sekundarstufe II. Freiburg, 277–304.
- STVILIA, BESIKI/GRASSER, LES (2008): An activity theoretic model for information quality change. In: First Monday 13/4.

- STVILIA, BESIKI/TWIDALE, MICHAEL B./GRASSER, LES/SMITH, LINDA C. (2005): Information quality discussions in Wikipedia. Technical Report ISRN UIUCLIS--2005/2+CSCW.
- SUROWIECKI, JAMES (2004): The wisdom of crowds. Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. New York.
- TARP, SVEN (2013): Lexicographic functions. In: GOUWS u. a. (Hrsg.), 460–468.
- VASTERMAN, PETER L. M. (2005): Media-Hype: Self-Reinforcing News Waves, Journalistic Standards and the Construction of Social Problems. In: *European Journal of Communication* 20, 508–530.
- WEINBERGER, DAVID (2011): Too big to know. Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room. New York.
- WICHTER, SIGURD (2011): Kommunikationsreihen aus Gesprächen und Textkommunikaten. Zur Kommunikation in und zwischen Gesellschaften. Berlin/Boston (Reihe Germanistische Linguistik. 294).
- WIEGAND, HERBERT ERNST (1998): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband. Berlin/New York.
- WIEGAND, HERBERT ERNST/BEIßWENGER, MICHAEL/GOUWS, RUFUS H./KAMMERER, MATTHIAS/STORRER, ANGELIKA/WOLSKI, WERNER (Hrsg.) (2010): Wörterbuch zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Mit englischen Übersetzungen der Umtexte und Definitionen sowie Äquivalenten in neun Sprachen [...]. Band 1: Systematische Einführung, A–C. Berlin/New York.
- WROBEL, ARNE (2000): Phasen und Verfahren der Produktion schriftlicher Texte. In: BRINKER, KLAUS/BURCKHARDT, ARMIN/UNGEHEUER GERHARD (Hrsg.): Text- und Gesprächslinguistik. Linguistics of Text and Conversation. Ein internationales Handbuch zeitgenössischer Forschung. Berlin/New York (Handbücher zur Sprach- und Kommunikationswissenschaft. 16.1), 458–472.
- ZIEM, ALEXANDER (2008): Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz. Berlin (Sprache und Wissen. 2).

SUSANNE DYKA / EVA SCHARF

Vergleichende Analyse von Artikelstrukturen deutsch-englischer Onlinewörterbücher

1. Einleitung

Gespräche mit Studierenden ergeben immer wieder, dass sie kaum noch Printwörterbücher benutzen, sondern fast ausnahmslos frei zugängliche und damit leicht erreichbare Onlinewörterbücher bevorzugen. Dies steht neben dem einfachen und kostenlosen Zugriff möglicherweise auch in Zusammenhang mit den grundsätzlichen Problemen, die bei zweisprachigen Wörterbüchern häufig festgestellt werden. Hierzu zählt zum einen der begrenzte Platz, der aufgrund der Gegenüberstellung der beiden Sprachen, aber auch aufgrund des Mediums nicht genügt, um die komplexen Gefüge, die zwischen zwei Sprachen herrschen, darzustellen.¹ Zum anderen ist – wohl zum Teil auch auf das Platzproblem zurückzuführen, aber nicht nur darauf – die unzureichende Kennzeichnung der verschiedenen Informationstypen² ein oft diskutiertes Manko von (zweisprachigen) Printwörterbüchern. Insgesamt dauert der Nachschlagevorgang zu lange, der Benutzer muss häufig nicht nur eine Stelle im Wörterbuch suchen, sondern mehrere.³

Da erscheint das Onlinewörterbuch⁴ mit unbegrenztem Platz, anderen Suchmöglichkeiten, insbesondere der Volltextsuche, und anderen

¹ Hinsichtlich des Platzproblems vgl. u. a. LEW (i. Dr.), hinsichtlich der komplexen Gefüge vgl. HERBST/KLOTZ (2003, 103f.).

² Siehe DYKA/SCHARF (2013), HERBST/KLOTZ (2003, 160).

³ Zum Nachschlagevorgang im Printwörterbuch siehe ENGELBERG/LEMNITZER (2009, 97ff.).

⁴ Für eine Einordnung der Onlinewörterbücher in eine Typologie elektronischer Wörterbücher vgl. LEHR (1996, 315) und darauf aufbauend DE SCHRUYVER (2003, 148).

Darstellungsformen das geeignete Medium zu sein, diese Probleme zu überwinden.⁵ Es soll deshalb im Folgenden untersucht werden, ob genau dies einer Auswahl der zurzeit im Internet kostenlos verfügbaren zweisprachigen deutsch-englischen Onlinewörterbücher⁶ gelingt. Dabei werden zuerst die klassischen Makro- und Mikrostrukturen des zweisprachigen Printwörterbuchs mit den Möglichkeiten des Onlinewörterbuchs verglichen. Anschließend werden fünf deutsch-englische Onlinewörterbücher exemplarisch auf Eigenheiten und Probleme in ihren Artikelstrukturen untersucht und in einem anschließenden Fazit bewertet.

2. Untersuchte Onlinewörterbücher

Für die vorliegende Untersuchung wurden fünf deutsch-englische Onlinewörterbücher ausgewählt, bei deren Auswahl vor allem ein hoher Bekanntheitsgrad und eine möglichst breite Streuung in Bezug auf die lexikografische Basis eine Rolle spielten.⁷ Alle fünf deutsch-englischen

⁵ Vgl. DE SCHRYVER (2003) für eine Zusammenfassung von Vor- und Nachteilen der elektronischen Wörterbücher, darunter auch speziell der Onlinewörterbücher, gegenüber den Printwörterbüchern. Vgl. auch HAB/SCHMITZ (2010, 3–4), die neun Chancen (Aktualisierbarkeit der Informationen, Aufhebung des Platzproblems, Multimodalität der Daten, Verwendung von Hyperlinks (auch zu externen Seiten), interaktiver Charakter von Onlinewörterbüchern etc.) und neun Risiken von Onlinewörterbüchern (Autoritätsverlust, Probleme bei der Informationsfindung, Korrelation zwischen größerer Zahl an Funktionen eines Wörterbuchs und größeren Kompetenzerfordernissen auf Seiten des Nutzers, geringere Zuverlässigkeit der Informationen bei Beteiligung der Nutzer an der Datengenerierung etc.) herausarbeiten.

⁶ Aus Gründen der besseren Lesbarkeit wird hier immer von ‚deutsch-englischen‘ Onlinewörterbüchern gesprochen. Soweit nicht explizit anders dargestellt, sollen hierunter Onlinewörterbücher für das Sprachenpaar Englisch/Deutsch und somit beide Wörterbuchteile (Englisch–Deutsch und Deutsch–Englisch) verstanden werden.

⁷ In der auf alexa.com angebotenen Rangliste der beliebtesten Wörterbücher im Internet finden sich die fünf Wörterbücher auf folgenden Rängen: dict.cc – Rang 6; Leo – Rang 7; Linguee – Rang 16; Beolings – Rang 30; Pons –

Wörterbücher sind auf den unten genannten Websites in größere Wörterbuchportale⁸ integriert, die allesamt noch Wörterbücher für andere Sprachenpaare sowie eine Reihe weiterer (zum Teil unterschiedlicher) Elemente (wie Diskussionsforen, Vokabeltrainer) enthalten. Im Folgenden sollen die ausgewählten deutsch-englischen Onlinewörterbücher kurz in Bezug auf ihre Größe, ihre Datenbasis und die angebotenen Sprachrichtungen vorgestellt werden.

Beim ersten der fünf Wörterbücher handelt es sich um *Leo*, das laut Angabe auf der Internetseite 793.060 Einträge⁹ enthält. Das Wörterbuchprojekt hat seinen Ausgangspunkt im Bereich der Informationstechnik an der TU München und wurde seit seiner Entstehung im Jahr 1995 bis heute kontinuierlich weiterentwickelt und erweitert. Zu dem anfangs ausschließlich aus Informatikern und freiwilligen Helfern bestehenden Wörterbuchteam zählen seit der Jahrtausendwende auch einige Übersetzer.¹⁰ Der in diesem Wörterbuch enthaltene Wortschatz geht laut Angaben auf der Website neben den vom Wörterbuchteam selbst entwickelten Wortlisten unter anderem auf diverse ‚Wortschatzschenkungen‘¹¹ zurück. *Leo* bietet darüber hinaus seinen Nutzern auch die Möglichkeit einzelne neue Einträge vorzuschlagen. Dies geschieht

Rang 3132 (Stand vom 02.06.2014). Die Auswahl basiert hier vor allem auf dem zweiten Kriterium einer möglichst unterschiedlichen lexikografischen Basis der zu untersuchenden Wörterbücher.

⁸ Vgl. KLOSA/LEMNITZER/NEUMANN (2008) für eine Definition von ‚Wörterbuchportal‘. Sie verstehen darunter Webanwendungen, die lexikalisches Wissen zugänglich machen und unterschiedliche Informationen und lexikalische Ressourcen integrieren. Sie weisen außerdem darauf hin, dass auch die Zuverlässigkeit der im Wörterbuch enthaltenen Informationen sowie das damit oft korrelierende Charakteristikum der kostenlosen Verfügbarkeit ein wichtiges Merkmal darstellen.

⁹ http://dict.leo.org/ende/index_de.html [2014-05-26]

¹⁰ Die Geschichte von *Leo* findet sich auf http://dict.leo.org/pages/about/en/de/history_de.html#/page=1 [2014-05-26].

¹¹ http://dict.leo.org/pages/about/ende/contributions_de.html [2014-05-26]

in den meisten Fällen über das Diskussionsforum, das über Links mit den entsprechenden Lemmata ins Wörterbuch integriert ist.¹²

Als zweites zu untersuchendes Onlinewörterbuch wurde *dict.cc* ausgewählt, das sich noch klarer als *Leo* (aufgrund der Möglichkeit, dass Benutzer aktiv eigene Vorschläge einbringen können) der Kategorie der *community-based* Wörterbücher¹³ zuordnen lässt. Auf der Webseite von *dict.cc* findet sich die Angabe, dass 1.002.806 Deutsch-Englisch-Übersetzungen enthalten sind.¹⁴ Laut Aussage der Betreiber der Website enthält das Wörterbuch einerseits Übersetzungen von der TU Chemnitz und aus *Mr Honey's Business Dictionary (Englisch/Deutsch)* und basiert andererseits vor allem auf dem Konzept der „freien Weitergabe von Wissen“:

dict.cc möchte es seinen Benutzern ermöglichen, ihr Wissen mit anderen zu teilen. Wenn eine bestimmte Englisch-Deutsch-Übersetzung noch nicht im Wörterbuch enthalten ist, kann sie von jedem Benutzer eingetragen werden. Bevor die Übersetzung für alle sichtbar wird, muss sie von mehreren anderen Beitragenden geprüft werden. Der dadurch entstehende Wortschatz kann jederzeit im Download-Bereich (Extras) heruntergeladen werden. (<http://www.dict.cc/> [2014-03-26])

Im Sinne dieser freien Weitergabe von sprachlichem Wissen bietet *dict.cc* seinen Nutzern vier verschiedene Möglichkeiten sich aktiv an der Weiterentwicklung des Onlinewörterbuchs zu beteiligen, indem sie 1) neue Übersetzungen hinzufügen, 2) bestehende Übersetzungen von anderen überarbeiten, 3) Audioaufnahmen von Wörtern im Wörterbuch machen, um die Computeraufnahmen zu ergänzen, sowie 4) Angaben zur Flexion hinzufügen beziehungsweise überarbeiten. Laut Angabe auf

¹² Vgl. http://dict.leo.org/pages/help/ende/faq_de.html#forum_newEntry [2014-05-26].

¹³ Vgl. auch CARR (1997, 214), der hier von „bottom-up editing“ (im Gegensatz zu „top-down editing“ bei den Verlagswörterbüchern) spricht: „new types of references are evolving upward from readers directly onto the Net“. Vgl. dazu auch DE SCHRIVER (2003, 160) und SVENSÉN (2009, 449f.).

¹⁴ Siehe <http://www.dict.cc/deutsch-englisch-woerterbuch.php> [Beitrag vom 2013-11-06; Abruf: 2014-05-26].

der Website gilt ein Eintrag dann als korrekt, wenn er von zehn verschiedenen (anonymen) Nutzern in der vorliegenden Form bestätigt wurde beziehungsweise zehn Nutzer den gleichen Korrekturvorschlag gemacht haben. Es wird dabei explizit darauf hingewiesen, dass es sich hierbei um ein fehlertolerantes System handelt, also kein Expertenwissen für die Beteiligung im oben beschriebenen Sinne erforderlich ist (<http://contribute.dict.cc/?action=show-introduction> [2014-05-26]).

Ein drittes Wörterbuch, das hier mitberücksichtigt werden soll, ist *Beolingus*, das im Unterschied zu den zwei vorangegangenen Projekten eine deutlichere linguistische beziehungsweise lexikografische Basis aufweist. Die Betreiber der Website geben einen Umfang von „**über 600.000 Wortübersetzungen Deutsch–Englisch**, mit Beispielen zur Verwendung, Synonymen und grammatikalischen Angaben“ an. Darüber hinaus wird darauf verwiesen, dass das Wörterbuch Definitionen und Angaben zu Synonymen und Antonymen englischer Wörter auf der Basis von *WordNet* und anderen *DICT*-Wörterbüchern sowie Synonyme zu deutschen Wörtern basierend auf dem *OpenThesaurus*-Projekt verzeichnet. Zudem enthält das Wörterbuch laut Angabe auf der Website über 1.115.000 deutsch-englische Beispielsätze. Auch bei diesem Wörterbuch haben Nutzer die Möglichkeit, eigene ‚Wortvorschläge‘ über ein Formular auf der Website an das Redaktionsteam zu senden. Die Aufnahme des Vorschlags ist damit jedoch noch nicht garantiert (<http://dict.tu-chemnitz.de/de-en/newtranslation.de.html> [2014-05-26]).

Beim vierten zu untersuchenden Onlinewörterbuch handelt es sich um das *Pons*-Wörterbuch, hinter dem mit der *PONS GmbH* ein Verlag mit langjähriger Erfahrung im Bereich der Erstellung von Printwörterbüchern steht. Auch dieses Onlinewörterbuch bietet dem Nutzer die Möglichkeit, einen eigenen Eintrag zu schreiben und diesen direkt über die Website an die Wörterbuchredaktion zu senden. Zur Größe dieses Onlinewörterbuchs finden sich auf der Website keine Informationen.

Das letzte zu betrachtende Onlinewörterbuch ist *Linguee*. Bei dieser Website handelt es sich um eine „Kombination aus einem redaktionellen Wörterbuch und einer Suchmaschine“. ¹⁵ Laut Angabe der Betreiber

¹⁵ <http://www.linguee.de/deutsch-englisch/page/about.php> [2014-05-26]

werden hier 100 Millionen zweisprachige Texte durchsucht; für bereits redaktionell verarbeitete Suchergebnisse werden dann auf der Ergebnis-seite in der linken Spalte die Inhalte des ‚redaktionellen Wörterbuchs‘ angezeigt, während rechts Satzbeispiele aus den durchsuchten Texten präsentiert werden. Das *Linguee*-Team sieht in Umfang und Art der dargestellten Informationen einen wesentlichen Unterschied zu anderen Onlinewörterbüchern:

Im Vergleich zu traditionellen Online-Wörterbüchern enthält *Linguee* etwa 1.000 mal mehr übersetztes Material, jeweils in vollständigen Sätzen. Dadurch finden Sie auch Übertreffungen für Ausdrücke wie "schnelles Auto" oder "schnelle Auffassungsgabe" sowie für seltenes Fachvokabular. (<http://www.linguee.de/deutsch-englisch/page/about.php> [2014-05-26])

Ähnlich wie *Leo* und *dict.cc* beinhaltet auch *Linguee* ein *community-based* Element, da es explizit das ‚Mitmachen‘ der Nutzer erwünscht.¹⁶ Diese haben die Möglichkeit, neue Lemmata vorzuschlagen, bestehende Lemmata und Beispielsätze zu bewerten sowie Verbesserungsvorschläge für bestehende Lemmata einzureichen. Laut Angaben auf der Website werden die Nutzervorschläge nach Überprüfung durch das *Linguee*-Team akzeptiert und übernommen.

Allen fünf Onlinewörterbüchern ist gemeinsam, dass sie die Möglichkeit bieten, entweder eine spezifische Suchrichtung auszuwählen, d. h. vom Deutschen ins Englische oder vom Englischen ins Deutsche, oder unabhängig von der Suchrichtung eine Anfrage zu starten. Letztere Option entspricht in allen fünf Wörterbüchern der Grundeinstellung beim ersten Aufruf der jeweiligen Website. Da davon auszugehen ist, dass Benutzer der Onlinewörterbücher, die ja meist von den schnellen Zugriffsmöglichkeiten des Mediums profitieren wollen, innerhalb dieser Grundeinstellung suchen, wird diese auch meist in der vorliegenden Untersuchung gewählt.

Ein weiterer Aspekt, der für alle fünf Onlinewörterbücher zutrifft, ist das Fehlen einer Angabe zum intendierten Benutzerkreis. Hierzu finden sich weder Metainformationen auf den Websites noch Aus-

¹⁶ <http://www.linguee.de/deutsch-englisch/page/contribute.php> [2014-05-26]

wahlmöglichkeiten für den Benutzer selbst (z. B. beim Starten einer Anfrage).

3. Vergleich der Wörterbuchstrukturen von zweisprachigen Printwörterbüchern und Onlinewörterbüchern

3.1 Makrostruktur¹⁷

Bei Printwörterbüchern herrscht normalerweise eine alphabetische Struktur der Lemmaliste vor. Dies garantiert schnellstmöglichen Zugriff auf die gesuchten Daten, erfordert von dem Benutzer aber auch die grundsätzliche Kompetenz der Beherrschung des Alphabets.¹⁸ Diese alphabetische Struktur wird bei *dict.cc* (als Leiste mit Buchstaben, die über Hyperlinks zu alphabetischen Wortlisten führen) und *Beolingus* (Nachbarwörter im Alphabet in einem kleinen Fenster) ebenfalls noch angeboten. Es ist nicht zu erwarten, dass der Benutzer diese alphabetische Suche bei Onlinewörterbüchern nutzt, da alle Onlinewörterbücher die Möglichkeit einer direkten Sucheingabe aufweisen.¹⁹ Anders als im Printwörterbuch, das nur eine Möglichkeit – nämlich das Suchen in einer alphabetischen Liste – bietet, gibt es bei Onlinewörterbüchern Möglichkeiten und Hilfen, die sich folgendermaßen gestalten: Eine erste Möglichkeit besteht darin, dass man bei der Suche im Normalfall die Suchrichtung einstellen kann. Des Weiteren gibt es verschiedene Suchroutinen, wie z. B. über ein einfaches Eingabefeld, die Auswahl aus einem alphabetischen Index oder die Möglichkeit der Verwendung von *wild cards* (HERBST/KLOTZ 2003, 257–259).²⁰ In Tab. 1 werden die

¹⁷ Zur Thematik des Vorhandenseins einer Makrostruktur im Onlinewörterbuch siehe MANN (2010, 22) und HAB/SCHMITZ (2010, 9f.). Hier soll Makrostruktur als Zugriffsstruktur verstanden werden.

¹⁸ Vgl. dazu ENGELBERG/LEMNITZER (2009, 97).

¹⁹ Besonders die bei LEW (2013, 19f.) genannte „Stage 3: Locating entry information“ scheint hier wichtig.

²⁰ Vgl. zu den Suchoptionen in Onlinewörterbüchern auch LEW (2013, 23–28), ENGELBERG/LEMNITZER (2009, 101–112), HAB/SCHMITZ (2010, 10ff.).

vorhandenen Suchmöglichkeiten der untersuchten Internetwörterbücher dargestellt.

	Leo	dict.cc	Beolingus	Pons	Linguee
einfaches Eingabefeld	✓	✓	✓	✓	✓
alphabetischer Index	✓	✗	✗	✓	✓
wild cards	✗	✓	✗	✗	✗

Tab. 1: Suchmöglichkeiten in den Onlinewörterbüchern

3.2 Mikrostruktur

Die klassische Mikrostruktur in Printwörterbüchern enthält nach SVENSÉN (2009, 345) für monoseme Lemmata folgende Bestandteile: *lemma section*, *formal section*, *semantic-pragmatic section*, *contextual section*.²¹ Dies bedeutet, dass sich der Angabe des Lemmas üblicherweise die *formal section* mit Informationen zur Aussprache, zur Wortklasse u. a. anschließt. Die *semantic-pragmatic section* besteht aus den Äquivalenten und differenzierenden Angaben. Die *contextual section* enthält Konstruktionen, Kollokationen (SVENSÉN 2009, 345ff.) und Restriktionen (HERBST/KLOTZ 2003, 144). In Printwörterbüchern werden einzelne Bedeutungen meist nicht in einer neuen Zeile angegeben, sondern sind eingebettet in einen sehr verdichteten und durchaus nicht leicht zu interpretierenden Text.²² Bei polysemen Lemmata lässt sich diese Struktur für jede Bedeutung annehmen, mit Ausnahme von Informationen, die für alle Bedeutungen gelten und daher häufig vor der polysemen Bedeutungsstruktur (Nummerierung) stehen.

²¹ Zur Artikelstruktur des zweisprachigen Wörterbuchs vgl. auch BAUNEBJERG HANSEN (1990) und HAUSMANN/WERNER (1991).

²² Vgl. zur Diskussion der „Wörterbuchartikel als Textsorte“ HERBST/KLOTZ (2003, 169f.).

Onlinewörterbücher unterscheiden sich in Bezug auf die Mikrostruktur erheblich von Printwörterbüchern, da die klassische Eintragsstruktur der Printwörterbücher in vielen Fällen größtenteils aufgehoben ist. Dies liegt sicherlich vor allem an den grundsätzlichen Unterschieden zwischen dem Print- und dem digitalen Medium. Zu den wesentlichen Unterschieden zählt hier einerseits das Platzangebot, das im Printmedium zu zahlreichen Kompromissen bei der Art und Weise der lexikografischen Angaben und dem Umfang der enthaltenen linguistischen Informationen führt. Andererseits verfügt das Onlinemedium – als spezieller Fall des digitalen Mediums und im Unterschied nicht nur zu Printwörterbüchern, sondern auch zu CD-ROM-Versionen – über eine dynamische Struktur, das heißt Veränderungen und Ergänzungen sind prinzipiell technisch jederzeit möglich. Diesen Aspekt machen sich die Onlinewörterbücher vor allem durch ihre *community-based* Elemente zunutze.

Die genannten Unterschiede werden bereits sehr deutlich in Hinblick auf die formale Struktur der Einträge in Onlinewörterbüchern: Im Gegensatz zu der dichten Eintragsstruktur im Printwörterbuch findet sich hier meist eine tabellenartige Aufstellung der Suchtreffer. Einige der hier untersuchten Onlinewörterbücher zeigen dabei immer – unabhängig von der tatsächlichen Suchrichtung – die englischen Angaben auf der linken Seite und die deutschen auf der rechten (vgl. *Leo* und *dict.cc*) beziehungsweise in umgekehrter Reihenfolge (vgl. Suchen innerhalb der Grundeinstellung bei *Beolingus*). Dies steht in deutlichem Kontrast zum Printwörterbuch, in dem das gesuchte, ausgangssprachliche Lemma grundsätzlich am Beginn des Eintrags steht (also links) und die zielsprachlichen Äquivalente rechts davon zu finden sind, worauf die Unterteilung in deutsch-englische beziehungsweise englisch-deutsche Wörterbuchteile begründet ist. Hier zeigt sich auch ein zweiter, wesentlicher Unterschied zwischen Print- und Onlinewörterbüchern in Bezug auf die formale Struktur. Im Printwörterbuch enthält die *lemma section* das Lemma und hierauf bezogene Informationen wie etwa die Silbentrennung. Die Angabe des Lemmas erfolgt dabei normalerweise einmal zu Beginn des Eintrags. Das Lemma wird zumeist durch

Fettdruck und möglicherweise eine andere Schriftart, -größe und/oder -farbe deutlich vom Wörterbuchtext abgehoben. Dies ist notwendig, um die alphabetische Zugriffsstruktur zu verdeutlichen. Bei Onlinewörterbüchern hingegen wird das Lemma in jeder Zeile wiederholt. Das bedeutet, dass in vielen Fällen jedes Äquivalent, jede phraseologische Angabe und zum Teil auch Valenzangaben in einer neuen Zeile eingetragen werden, in der dann jeweils auch das Lemma wiederholt wird. Ausnahme ist hier *Linguee*, das in der neuen Ansicht das Äquivalent in rotem Fettdruck in der Regel einmal, für verschiedene Konstruktionen bzw. für verschiedene Wortklassen jedoch mehrmals angibt.

Die *formal section*, die nach SVENSÉN (2009, 346) Informationen zur Aussprache und Wortklasse enthält, stellt sich bei den Onlinewörterbüchern folgendermaßen dar: Es gibt keine IPA-Angaben für die englischen Lemmata, aber bei allen fünf Wörterbüchern sind Audiofiles zum Anhören der Aussprache vorhanden (bei *Beolingu*s erfolgt das nicht immer, nur für das erste Äquivalent). Einige der Wörterbücher geben die britische Aussprache (*Leo*, *Linguee*), andere wiederum die amerikanische (*Beolingu*s, *Pons*) an. Bei *dict.cc* ist es möglich, zwischen den beiden Aussprachen zu wählen, indem man auf die entsprechende Flagge klickt. Die Angabe der Wortklasse erfolgt bei allen fünf Wörterbüchern, wenn auch in unterschiedlicher Weise wie z. B. als Abkürzung (*Beolingu*s) oder als Überschrift der (Teil-)Tabelle (*Leo*). Wie bereits angedeutet, wird die *semantic-pragmatic section* (SVENSÉN 2009, 346) in den Onlinewörterbüchern immer als (einzeilige) Lemma-Äquivalent-Einheit dargestellt, die äquivalentdifferenzierende Angaben erhalten kann. Die Angaben zu Konstruktionen, Kollokationen und Idiomen, die typischerweise in der *contextual section* (SVENSÉN 2009, 347f.) angeboten werden, erfolgen auch in den Onlinewörterbüchern – normalerweise ebenfalls in einer Zeile, in der die entsprechende Konstruktion, Kollokation oder auch ein Beispiel plus die entsprechende Übersetzung angeboten wird. Sehr häufig sind diese Angaben nicht auf ein Lemma bzw. Äquivalent adressiert.

4. Vergleich der Artikelstrukturen der Onlinewörterbücher *Leo*, *dict.cc*, *Beolingus*, *Linguee* und *Pons*

4.1 Leo

Bei der Analyse einzelner Einträge in *Leo* fällt zunächst – wie bereits erwähnt – auf, dass englische Angaben unabhängig von der Suche immer auf der linken Seite, deutsche Angaben immer auf der rechten Seite angezeigt werden. Neben dieser Einstellung lässt sich außerdem die Unterteilung der Suchtreffer nach Wortklassen in Frage stellen. Bei der Suche nach einem englischen Äquivalent für *bad*²³ beispielsweise werden die Treffer in voneinander abgetrennten Abschnitten mit den Überschriften ‚Substantive‘ und ‚Verben‘ dargestellt. Diese Wortartenzuweisung ist insofern problematisch, als die Einträge unter ‚Verben‘ nicht der Wortart ‚Verb‘, sondern den Kollokationen zuzuordnen sind. Hier erscheint also sowohl die Kategorisierung der Treffer als auch die fehlende Zuordnung zu den Äquivalenten im Abschnitt ‚Substantive‘ fragwürdig. Warum wird beispielsweise die Kollokation *to take a bath* nicht direkt auf das weiter oben angegebene Äquivalent *bath* adressiert? Denn die Angabe der Kollokation ist zweifellos aus Nutzersicht sehr sinnvoll, ob der Nutzer sie allerdings im Abschnitt ‚Verben‘ sucht und in jedem Fall richtig zuordnen kann, ist fraglich. Auch bei der Suche nach englischen Äquivalenten für *baff* werden die Ergebnisse in ‚Adjektive‘ und ‚Verben‘ unterteilt. Im Abschnitt zu Adjektiven wird hier lediglich *gobsmacked* verzeichnet, im Verbabschnitt finden sich hingegen für *baff machen* das Äquivalent *to bamboozle s.o.*; *to be flabbergasted*, *to be gobsmacked* für *baff sein*, sowie *to be weirded out* für *völlig baff sein*. Die Unterteilung ist hier aus zweierlei Sicht fragwür-

²³ Die Suchrichtung wurde hier auf Deutsch → Englisch festgelegt. Bei der Grundeinstellung (d. h. Suche im deutsch-englischen und englisch-deutschen Wörterbucheil) werden außerdem für engl. *bad* noch die Abschnitte ‚Adjektive/Adverbien‘, ‚Phrasen‘ und ‚Beispielsätze‘ angezeigt. Auch hier erscheint die Kategorisierung problematisch, da idiomatische Wendungen (z. B. *That's not half bad.*), für die die Kategorie ‚Phrasen‘ vorgesehen zu sein scheint, auch im Abschnitt ‚Beispielsätze‘ auftauchen.

dig: Zum einen handelt es sich größtenteils wiederum um Kollokationen (des Typs Adjektiv + Verb), zum anderen impliziert die Angabe von *gobsmacked* und *to be gobsmacked* die Möglichkeit des attributiven und prädikativen Gebrauchs, die weder für das Deutsche noch für das Englische gegeben ist.

Offen bleibt auch die Frage nach der Struktur innerhalb der Wortartenabschnitte. In manchen Fällen liegt offensichtlich eine alphabetische Ordnung der Äquivalente vor (wie z. B. bei den Äquivalenten für *bahnbrechend*), in anderen Fällen jedoch nicht. Hier lässt sich zum Teil eine semantische Struktur vermuten wie im Fall von *Bad*, wo (im Abschnitt ‚Substantive‘) erst allgemeine, dann technische Bedeutungen verzeichnet werden.

Darüber hinaus erfolgt nur eine lückenhafte und unsystematische Äquivalentendifferenzierung.²⁴ So wird beispielsweise bei den englischen Äquivalenten *bath*, *bathroom*, *bagnio*, *baths*, *spa*, *swim* und *washing* für dt. *Bad* ganz auf eine Äquivalentendifferenzierung verzichtet. In anderen Fällen, wie zum Beispiel bei *baff*, findet sich die Angabe ‚coll.‘ bei den englischen Äquivalenten *to bamboozle s.o.*, *to be flabbergasted* und *to be gobsmacked*, wobei gleichzeitig das deutsche Lemma *baff machen* für die beiden letztgenannten Äquivalente mit ‚ugs.‘ markiert ist. Die doppelte Kennzeichnung erscheint hier äußerst fragwürdig. Bei den Suchergebnissen für *Backenzahn* ist ein Teil der englischen Äquivalente mit ‚anat.‘ und ein anderer Teil mit ‚med.‘ gekennzeichnet. Inwieweit diese Unterscheidung sinnvoll ist, ist ebenfalls fraglich. Hier zeigt sich außerdem ein weiteres Problem: Der für ‚anat.‘ hinterlegte Link läuft ins Leere.

4.2 dict.cc

Wie *Leo* zeigt auch *dict.cc* die deutschen Angaben grundsätzlich auf der rechten Seite, die englischen auf der linken. Auch bei der Gliederung der Treffer zeigen sich hier Parallelen zu *Leo* – mit einem Zusatz:

²⁴ Zur Differenzierung von Äquivalenten in zweisprachigen Wörterbüchern siehe z. B. DYKA (2005).

Bei *dict.cc* erfolgt die Gliederung zuerst gemäß der Zahl an Wörtern in der Ausgangssprache; bei gleicher Wortzahl wird dann jedoch außerdem noch nach Wortarten untergliedert. Diese Struktur erscheint aus mehreren Gründen problematisch, die im Folgenden kurz diskutiert werden sollen.

Grundsätzlich stellt sich die Frage nach dem generellen Nutzen der Angabe der Wortanzahl, da der Nutzer nach Bedeutungen sucht und wohl in den wenigsten Fällen darüber nachdenkt oder bereits wissen kann, nach wie vielen Wörtern er sucht.

Darüber hinaus ist nicht auf den ersten Blick erschließbar, auf welche Sprache sich die angegebene Zahl der Wörter bezieht. Dass sie sich auf die Ausgangssprache bezieht, ist besonders bei der Suche nach Äquivalenten für ein deutsches Lemma schwer erkennbar, da die Zahlenangabe über den immer links angezeigten englischen Treffern und nicht über den (rechts angezeigten) deutschen Angaben steht.

Das dritte Problem zeigt sich – wie bei *Leo* – in der Verwendung der Wortartenangaben als Untergliederungshilfe. Denn wie bei *Leo* werden auch hier Kollokationen und freie Kombinationen beispielsweise unter ‚Verben‘ (vgl. *to have a bath* bei der Suche nach dt. *Bad*) oder unter ‚Substantive‘ (vgl. *cold bath* bei derselben Suche) gefasst. Ähnlich wie bei *Leo* (‚Phrasen‘, ‚Beispiele‘) gäbe es auch bei *dict.cc* eine Kategorie (‚Andere‘), unter der solche Treffer eher zu erwarten wären beziehungsweise zu denen diese richtigerweise zuzuordnen sind. Die Kollokation *to have a bath* in der Trefferliste zu der Suche nach dt. *Bad* zeigt exemplarisch, zu welchen Problemen die Struktur nach Zahl der Wörter und die problematische Wortartenzuweisung zum Teil führen. Die Kollokation ist einerseits im Abschnitt „3 Wörter: Andere“ in der Form *sb. is having a bath* (dt. *jd. nimmt ein Bad*) verzeichnet. Im Abschnitt „3 Wörter: Verben“ findet sie sich ebenfalls, hier in der Form *to have a bath* (dt. *ein Bad nehmen*). Schließlich ist sie ein drittes Mal weiter unten in der Trefferliste enthalten: im Abschnitt „4 Wörter: Andere“ ist sie in der Form *He is having a bath.* (dt. *Er nimmt ein Bad.*) verzeichnet. Da sowohl das erste und das dritte Vorkommen der Kollokation Beispielsätze darstellen (im letzteren Fall mit lexikalischer Fül-

lung des Platzhalters *sb.* im ersten Fall), liegt hier eine unnötige Redundanz vor. Darüber hinaus würde der Nutzer die Kollokation und den dazugehörigen Beispielsatz sicherlich schneller finden, wenn die Kollokation direkt auf das entsprechende Äquivalent und der Beispielsatz direkt auf die Kollokation adressiert wären.

Was die Reihenfolge der Angaben innerhalb eines Abschnittes in *dict.cc* betrifft, greift das Wörterbuch bei einigen (aber nicht bei allen) Wörtern auf Daten aus den Statistiken des ebenfalls auf diesem Portal verfügbaren Vokabeltrainers zurück. Daraus lässt sich schließen, dass die hier am häufigsten gebrauchte Übersetzung (mit Angabe der Frequenz) in der Trefferliste oben steht.²⁵ Dies ist insofern fragwürdig, als die auf diese Art gewonnenen Häufigkeitsangaben sicherlich nicht unbedingt mit den tatsächlichen Frequenzen übereinstimmen. Vor allem bei niedrigfrequenten Wörtern sind mehr oder weniger starke Verzerrungen zu erwarten.

Die Suche nach dt. *Bad* zeigt außerdem, dass *dict.cc* – zumindest an dieser Stelle – äquivalentdifferenzierende Mittel in der Form von deutschen Glossen verwendet und sich hiermit klar von *Leo* unterscheidet. Bei anderen Suchen, z. B. nach dt. *Ball*, zeigen sich hingegen auch bei *dict.cc* Lücken.

4.3 Beolingu

Im Unterschied zu *Leo* und *dict.cc* präsentiert *Beolingu* bei Suchen innerhalb der Grundeinstellung (d. h. Englisch ↔ Deutsch) in der tabellarischen Präsentation der Treffer die deutschen Angaben auf der linken Seite, die englischen auf der rechten. Spezifiziert der Nutzer seine Suche außerdem hinsichtlich Ausgangs- und Zielsprache, werden – ebenfalls im Unterschied zu den anderen beiden Onlinewörterbüchern – die ausgangssprachlichen Einheiten immer links angezeigt, die entsprechenden Äquivalente in der rechten Spalte. Dies entspricht nicht nur der klassischen Präsentation in Printwörterbüchern, sondern auch der aus

²⁵ Vgl. <http://contribute.dict.cc/guidelines/> [2014-06-06].

der Textlinguistik stammenden Beobachtung, dass üblicherweise die bekannte Information, also das Lemma, zuerst gegeben wird, und erst anschließend die neue Information, also das Äquivalent.²⁶

Einen weiteren Unterschied zu *Leo* und *dict.cc* stellt die klare semantische Struktur der Trefferliste in *Beolingus* dar. Hier sind die einzelnen Bedeutungsabschnitte durch , ▼ ‘ gekennzeichnet. Anders als in den oben beschriebenen Wörterbüchern werden hier folglich auch phrasologische Angaben der relevanten Bedeutung direkt zugeordnet und sind so sicherlich leichter und schneller zu finden beziehungsweise überhaupt als relevant zu erkennen.

Im Unterschied zu den anderen Wörterbüchern enthält eine Zeile in *Beolingus* zum Teil mehr als eine Angabe pro Sprache. Dies ist beispielsweise bei *baff* der Fall – hier werden die Äquivalente *to be flabbergasted* und *to be gobsmacked*²⁷ in einer Zeile angegeben. Sind mehrere Äquivalente ohne weitere äquivalentdifferenzierende Angaben nebeneinandergestellt, impliziert dies, dass die Äquivalente in einem Großteil von infrage kommenden Kontexten synonym verwendet werden können. Die Äquivalentdifferenzierung scheint hier eher über die Angabe von deutschen Synonymen auf der Lemmaseite vorgenommen zu werden. Auch die Verwendung von abwechselnd grauen und weißen Zeilenhinterlegungen scheint zum Teil zum Zweck der Äquivalentdifferenzierung durch Kennzeichnung unterschiedlicher Bedeutungsabschnitte zu erfolgen.

Die Struktur betreffend fällt außerdem auf, dass die durch , ▼ ‘ eingeleiteten Bedeutungsabschnitte feste Blöcke darstellen, die bei der Suche nach darin enthaltenen Stichwörtern immer in gleichem Umfang und gleicher Form angezeigt werden. Dies ist beispielsweise bei *spre-*

²⁶ Vgl. zur Diskussion von Thema-Rhema-Struktur und dem Konzept von *given/new* in der englischen Textlinguistik z. B. BROWN/YULE (1983, 125–133, 153–189).

²⁷ Anders als *Leo* und *dict.cc* verzichtet *Beolingus* außerdem auf die isolierte Angabe von *baff* (stattdessen gibt es, neben weiteren Synonymen, „*baff sein*“ an) und signalisiert so implizit, dass die Verwendung des deutschen Adjektivs sowie der englischen Entsprechungen nur prädikativ, nicht aber attributiv möglich ist.

chen der Fall, wo im ersten Block für das Lemma *sprechen* (näher spezifiziert durch die Synonyme *reden* und *sich unterhalten*) das Äquivalent *talk* zusammen mit einer Reihe von grammatischen Angaben (z. B. Partizipformen), phraseologischen Angaben sowie Beispielsätzen zu *talk* und den jeweiligen deutschen Entsprechungen verzeichnet ist. Dieser *talk*-Block wird genauso bei der Suche nach *reden* oder *sich unterhalten* ausgegeben. Diese Praxis garantiert zwar eine gewisse Konstanz in Umfang und Präsentation der relevanten Äquivalente, erscheint aber insofern fragwürdig, als darin auch immer einige deutsche Entsprechungen von z. B. phraseologischen Angaben mit *talk* verzeichnet sind, die das vom Nutzer gesuchte Lemma (z. B. *sprechen*) überhaupt nicht enthalten. Zwar ist es sicherlich sinnvoll, diese Informationen ebenfalls – auch an dieser Stelle – zu präsentieren, es ist allerdings zu überlegen, ob weitere Einrückungen in der Struktur oder eine andere Darstellungsmöglichkeit für solche Fälle nicht angemessener wären. Hinzu kommt, dass die Fettschreibung der Vorkommen von *sprechen* eigentlich dazu beitragen soll, dass der Nutzer die relevanten Treffer schneller findet; dies ist aber insofern nicht ganz zufriedenstellend gelöst, als nur die Infinitivform des Verbs, nicht aber die flektierten Formen fett gedruckt sind.

Wie zu Beginn dieses Abschnitts erläutert, geht *Beolingu* zwar stärker auf den Nutzer ein als *Leo* und *dict.cc*, indem es bei Spezifizierung der Suchrichtung das ausgangssprachliche Material in der linken Spalte präsentiert und die Struktur insofern flexibel hält. Die Anpassung der präsentierten Informationen an die Nutzerbedürfnisse ist jedoch durch die oben beschriebene blockartige, starre Darstellung von Bedeutungsabschnitten wiederum stark eingeschränkt. So stellt beispielsweise die (in diesen Blöcken für Ausgangs- und Zielsprache immer enthaltene) Angabe der Flexionsformen eines gesuchten Verbs oder der Steigerungsformen eines deutschen Adjektivs für den deutschen Muttersprachler überflüssige Information dar, die zudem die Übersichtlichkeit des Eintrags stark einschränkt.

Positiv zu bewerten ist, dass *Beolingu* Kollokationslisten verwendet. In diesen Listen wird einer Bedeutung des Lemmas – der Kolloka-

tionsbasis – eine Reihe von relevanten Kollokationen zugeordnet. Eine solche Liste findet sich zum Beispiel bei *Ball*, wo eine Liste von Adjektiv + Substantiv-Kollokationen (*flacher/hoher/kurzer/langer Ball* auf der Lemmaseite und *low/high/short/long ball* auf der Äquivalentseite) verzeichnet ist.

4.4 Linguee

Bei *Linguee* gibt es zwei grundsätzliche Ansichten, die neue und die alte. Die alte Ansicht entspricht eher einer klassischen Wörterbuchansicht mit dem Lemma links und dem entsprechenden Äquivalent rechts. Für jedes neue Äquivalent gibt es eine neue Zeile, wobei das Lemma jedes Mal wiederholt wird.

Die Grundeinstellung scheint allerdings die neue Ansicht zu sein. Dort erfolgt auf der linken Seite das „redaktionelle Wörterbuch“ und rechts „Übersetzungsbeispiele“ (z. B. *aufheben*). Innerhalb des redaktionellen Wörterbuchs findet sich das Lemma links oben rot und fett gedruckt, die Äquivalente erfolgen danach weitgehend ohne differenzierende Angaben. Die Reihenfolge ist frequenzbasiert, wobei die Häufigkeit durch Tortendiagramme angezeigt wird. Über den darunter liegenden Hyperlink ist die Information leicht zugänglich (z. B. „sprechen‘ (Verb, intransitiv) wird selten mit ‚talk‘ übersetzt (7 %)“²⁸). Dabei wird keine linguistische Terminologie, sondern z. B. *manchmal/häufig/selten* verwendet. Oft erfolgt die Frequenzangabe nur für die ersten Äquivalente, darunter wird dann die Liste weiterer Übersetzungsbeispiele, ohne Frequenzangaben weitergeführt. Unklar ist, woraus die Frequenz berechnet wird.

Die fehlenden differenzierenden Angaben sind problematisch, da z. B. bei *baggern* als erstes Äquivalent *dredge* gegeben wird, und neben den sicherlich üblichen Äquivalenten *dig* und *excavate* auch noch *pull* aufgelistet wird. Diese vier Angaben erfolgen ohne Differenzierung,

²⁸ <http://www.linguee.de/deutsch-englisch/search?source=auto&query=sprechen>
[2014-06-04]

was es dem Benutzer praktisch unmöglich macht, das richtige Äquivalent zu finden. Insbesondere da hier auch keine Beispielsätze erfolgen.

Unter der Äquivalentliste gibt es häufig Beispielsätze (vgl. *aufheben*), die zum Teil den Gebrauch von oben angeführten Äquivalenten verdeutlichen (z. B. *Schutz aufheben – suspend cover*) oder aber Restriktionen darstellen (z. B. *eine Steuer aufheben – abandon a tax*).

In der neuen Ansicht gibt es neben dem redaktionellen Wörterbuch viele Übersetzungsbeispiele aus fremden Quellen (die aus dem Korpus stammen), wobei es unklar ist, was die Bezeichnung „aus fremden Quellen“ eigentlich genau bedeutet. In der alten Ansicht gibt es außerdem zahlreiche ‚ungesicherte Beispiele‘, die aus dem Internet²⁹ stammen. Ebenso scheint es unklar, ob die Übersetzungsbeispiele vom Englischen ins Deutsche übersetzt wurden oder umgekehrt oder möglicherweise beides? Dies hätte natürlich Auswirkungen auf die Authentizität und somit Brauchbarkeit der Beispielsätze.

Ein weiteres Problem, das bei der Suche aufgefallen ist: Sucht man z. B. *Bündchen*, gibt es sehr viele Treffer für ‚Gisele Bündchen‘. Dies würde man in einem Wörterbuch nicht unbedingt erwarten, hier scheint eine fehlende Aussortierung von Eigennamen vorzuliegen.

4.5 Pons

Bei *Pons* fällt zuerst die semantische Struktur auf. Ebenso wie die anderen Onlinewörterbücher werden auch hier die Ergebnisse tabellenartig angegeben. Die Angabe des Lemmas erfolgt auf der linken Seite, die Äquivalente werden rechts gegeben. Semantisch zusammengehörende Äquivalente und Phrasen werden farblich in einem Block angegeben, sodass nicht die Zeilen abwechselnd farbig dargestellt sind, sondern zusammenhängende Blöcke klar erkennbar sind. Wendungen erfolgen am Ende des Artikels. Die Darstellung ist daher etwas übersichtlicher

²⁹ Vgl. <http://www.linguee.de/deutsch-englisch/search?source=auto&layout=2&query=sprechen> [2014-06-04]. Dort erscheint in der alten Ansicht, wenn man die Maus auf „Ungesicherte Beispiele“ fährt, dass diese Beispiele aus dem Internet stammen und ungeprüft sind.

als bei den anderen Onlinewörterbüchern. Allerdings erkennt diese farbliche Unterscheidung möglicherweise nicht jeder Benutzer.

Neben der Anführung des Lemmas in jeder Zeile erfolgt hier aber für jede semantische Bedeutung eine Glosse oder Differenzierung. Wobei bezweifelt werden dürfte, dass die Verwendung von Markierungen³⁰ wie *sl* zum Beispiel bei *baggern* jedem Nutzer deutlich macht, welche Bedeutung gemeint ist.

Als erste Aufteilung des Wörterbuchartikels dient die Angabe der Wortklasse (z. B. bei *aufheben*: VERB trans, VERB refl, SUBSTANTIV). Das *Pons*-Onlinewörterbuch verwendet ebenso wie häufig die Printwörterbücher von *Pons* die Wortklassenmarkierungen beziehungsweise indirekten Grammatikangaben ‚transitiv‘ und ‚intransitiv‘. Dies ist problematisch, weil nicht allen Benutzern die Bedeutung dieser Trennung klar sein dürfte und dieses Vorgehen zum anderen eine unnötige Dopplung von Äquivalenten zur Folge hat.³¹

Für Verben, aber auch für Adjektive, erfolgt systematisch die Angabe der Valenz (siehe *aufheben* oder *baff*). Bei *baff* erfolgt die Valenzangabe sogar in redundanter Weise zum einen als Angabe zur Wortart ADJ pred und dann als phraseologische Angabe beziehungsweise Patternillustration³² *baff sein*, die den prädikativen Gebrauch aufnimmt. Dazu gehörende Beispiele erfolgen gleich darunter innerhalb des farblich einheitlichen Blocks. Allerdings gibt es nicht zu jeder Lemma-Äquivalent-Einheit entsprechende Beispielsätze. Auch Kollokationsangaben werden auf eine Lemma-Äquivalent-Einheit adressiert. Häufig dienen Kollokationsangaben als Restriktionsangaben (d. h. sie beschreiben Fälle, in denen das vorher angegebene Äquivalent nicht ver-

³⁰ Zu Markierungen siehe z. B. HAUSMANN (1977, 112–143).

³¹ Siehe HERBST/KLOTZ (2003, 180ff.) zu einer ausführlicheren Diskussion der Problematik dieser Angaben.

³² Zur Darstellung von grammatischen Angaben in zweisprachigen Wörterbüchern siehe HERBST/KLOTZ (2003, 136f.). Als am besten geeignete Form der Grammatikangabe wird hier die Patternillustration angesehen, da sie das Verb und die entsprechende Konstruktion in einer verallgemeinerbaren Weise darstellt – anders als in einem Beispiel, wo es für den Benutzer häufig unklar ist, ob die Konstruktion verallgemeinerbar ist.

wendet werden kann; dies ist z. B. bei *banal* der Fall, wo mit *eine banale Angelegenheit/Ausrede, eine banale Bemerkung, ein banales Thema* drei Kollokationen angegeben werden, bei denen *banal* nicht mit dem angegebenen Äquivalent *banal* übersetzt werden kann³³) und dienen somit auch der Differenzierung der Äquivalente.

Insgesamt lässt sich bei *Pons* feststellen, dass dieses Onlinewörterbuch deutlich näher an Artikelstruktur und Angaben im Printwörterbuch ist.

4.6 Suche nach phraseologischen Angaben in den fünf Onlinewörterbüchern

In den vorangegangenen Abschnitten wurden vor allem die Struktur beziehungsweise einzelne Aspekte der Struktur der Artikel untersucht, die bei der Suche nach einem Einzelwort angezeigt werden. Da das elektronische Medium im Gegensatz zum Printmedium die Möglichkeit einer Volltextsuche bietet, werden im Folgenden außerdem die Ergebnisse einer Suche nach bestimmten phraseologischen Angaben präsentiert. Der Suche liegen die phraseologischen Angaben zu den Lemmata *Bad* und *Backe* zugrunde, die in drei großen deutsch-englischen Printwörterbüchern³⁴ verzeichnet sind (vgl. DYKA/SCHARF 2013). Dazu wurden die Angaben einerseits in ihrer Grundform, wie sie in den Printwörterbüchern auftauchen, und andererseits im Rahmen eines Satzes (in den Tabellen in der Spalte ‚Phraseologische Angabe‘ in Klammern angegeben), in dem das Verb im Imperfekt steht, gesucht.³⁵ Dabei

³³ Siehe HERBST/KLOTZ (2003, 143f.).

³⁴ Die in Tab. 2 verwendeten Abkürzungen beziehen sich auf folgende Wörterbücher: DudOx = Duden-Oxford Großwörterbuch Englisch. Deutsch–Englisch, Englisch–Deutsch. (2005). LCGwE = Langenscheidt Collins Großwörterbuch Englisch. Englisch–Deutsch, Deutsch–Englisch. (2008). Pons = Pons Großwörterbuch Englisch. Englisch–Deutsch, Deutsch–Englisch. (2005).

³⁵ Es handelt sich bei den hier präsentierten Ergebnissen um das Resultat einfacher Sucheangaben. Die Eingabe mithilfe des Boole’schen Operators UND führt bei *Leo* zu den gleichen Ergebnissen wie bei der einfachen Su-

zeigen Tab. 2 und 3 die Zahl der relevanten Treffer. Die Angaben in Klammern weisen darauf hin, wie viele relevante Treffer für den in Klammern angegebenen Satz gefunden wurden. Die grauen Angaben zeigen an, dass zwar die entsprechende Zahl an Treffern gefunden wurde, nicht aber an erster Stelle in der Trefferliste. Auf Probleme bei der Darstellung in den einzelnen Onlinewörterbüchern wird im Anschluss an die Tabellen eingegangen.

Phraseologische Angabe	Quelle	Leo	dict.cc	Beolingus	Linguee	Pons
<i>ein ~ nehmen</i> (<i>er nahm ein Bad</i>)	LCGwE, DudOx, Pons	3 (3)	2 (2)	2 (2)	2 (-)	6 (1)
(<i>sich</i>) <i>ein ~ einlaufen</i> <i>lassen</i> (<i>er ließ ein Bad einlaufen</i>)	LCGwE, DudOx, Pons	- (-)	1 (1)	- (-)	- (-)	1 (1)
<i>~ in der Menge</i>	LCGwE, Pons	-	2	-	1	3
<i>türkisches ~</i>	LCGwE	-	-	1	1	1
<i>jdm/sich ein ~ einlassen</i> (<i>er ließ ein Bad ein</i>)	Pons	4 (4)	3 (3)	3 (-)	1 (-)	4 (1)

Tab. 2: Verzeichnetsein der phraseologischen Angaben zu *Bad* in den fünf Onlinewörterbüchern

che, *dict.cc*, *Pons* und *Beolingus* erkennen die Eingabe der Operatoren offensichtlich nicht als solche, die Suche liefert hier zahlreiche irrelevante Treffer; bei *Linguee* finden sich keine Treffer. Keine Unterschiede finden sich bei den Trefferlisten im Fall der Suche nach der phraseologischen Angabe in Anführungszeichen.

Phraseologische Angabe	Quelle	Leo	dict.cc	Beolingus	Linguee	Pons
<i>mit vollen ~n kauen</i> (<i>er kaute mit vollen Backen</i>)	LCGwE, Pons	- (-)	- (-)	- (-)	- (-)	1 (1)
<i>au ~</i>	LCGwE, Pons, DudOx	-	3	-	-	1
<i>auf einer ~</i>	LCGwE	-	-	-	-	1
<i>etw auf einer ~ absitzen</i> (<i>er saß es auf einer Backe ab</i>)	Pons	- (-)	- (-)	- (-)	- (-)	1 (1)

Tab. 3: Verzeichnetsein der phraseologischen Angaben zu *Backe* in den fünf Onlinewörterbüchern

Die Suche nach den phraseologischen Angaben in *Leo* ergibt zunächst, dass die sehr stark idiomatischen Wendungen zu *Backe* überhaupt nicht enthalten sind. Dies gilt auch für die Wendung *Bad in der Menge*. Bei der Suche nach den Kollokationen *ein Bad nehmen* und *ein Bad einlaufen lassen* finden sich hingegen relevante Treffer. Es fällt jedoch negativ auf, dass *Leo* hier erstens neben den gesuchten Treffern eine lange Liste von größtenteils vollkommen irrelevanten Treffern anzeigt. Zwar erscheint es sinnvoll, dass bei der Suche nach *ein Bad nehmen* auch auf Kollokationen wie *ein Bad einlassen* oder *ein Sonnenbad nehmen* verwiesen wird, der Nutzen von Treffern wie *ein Einstellgerät außer Betrieb nehmen* oder der Liste aller möglichen Äquivalente für *Bad*, auf die ohnehin der für *Bad* hinterlegte Hyperlink führt, darf jedoch stark angezweifelt werden. Der knappen Anzeige von ausschließlich relevanten Treffern steht zweitens entgegen, dass *Leo* offensichtlich nicht automatisch erkennt, dass es sich bei den Bestandteilen einer Kollokation wie *ein Bad nehmen* ausschließlich um deutsche Wörter handelt. Auf die Anzeige langer Trefferlisten für engl. *bad* könnte damit – im Sinne des Nutzers – verzichtet werden. Die Satzsuche nach *er nahm ein Bad*

liefert zwar zuerst eine Trefferliste für ‚Substantive‘, in der vor allem Äquivalente für *Bad* verzeichnet sind, aber auch – und noch dazu an erster Stelle – die Wendung *ein Einstellgerät außer Betrieb nehmen*. Im zweiten Abschnitt (‚Verben‘) und damit noch relativ weit oben werden aber bereits die Übersetzungen *to bathe*, *to take a bath* und *to have a bath* für die Grundform der Kollokation (*ein Bad nehmen*) gelistet. Wie auch bei der Suche nach den Grundformen bestimmter Kollokationen schließt sich diesen relevanten Treffern wiederum eine lange Liste irrelevanter Treffer an. Gleiches gilt für die Suche nach *er ließ ein Bad ein*.

Das Onlinewörterbuch *dict.cc* verzeichnet zwar mehr Wendungen aus dem obigen Katalog als *Leo*, offenbart aber ähnliche Probleme in der Darstellung. Fragwürdig ist hier vor allem auch die (seitenlange) Auflistung irrelevanter Treffer im Anschluss an die relevanten Treffer. Wie *Leo* erkennt auch *dict.cc* nicht automatisch, dass es sich nur um eine Suche vom Deutschen ins Englische handeln kann; die Auflistung von engl. *bad* mit seinen deutschen Äquivalenten ist also auch hier wiederum überflüssig. Bei der Kollokation *ein Bad einlassen* kommt außerdem hinzu, dass für *sich ein Bad einlassen* das Äquivalent *draw a bath* (ohne die Angabe von *oneself* als Entsprechung für *sich*) verzeichnet ist, während für *jdm. ein Bad einlassen* ein anderes Verb angegeben wird, nämlich *run*, und hier auch für *jdm.* eine Entsprechung verzeichnet ist (*to run a bath for sb.* und *to run sb. a bath*). Für den nicht-fortgeschrittenen Nutzer könnte dies implizieren, dass bei Verwendung der Kollokation mit *jdm.* ausschließlich das Verb *run* gewählt werden kann, bei Verwendung mit *sich* ausschließlich das Verb *draw*. Bei den Satzsuchen nach *er nahm ein Bad* und *er ließ ein Bad ein* werden die relevanten Treffer erst sehr weit unten in der Trefferliste angezeigt, sodass ein schnelles Auffinden nicht möglich scheint.

Auch *Beolingus* zeigt Lücken bei den gesuchten Wendungen und liefert vielfach lange Listen irrelevanter Treffer. Aufgrund der bereits angesprochenen (starren) semantischen ‚Blockstruktur‘ innerhalb der Trefferlisten in *Beolingus* muss der Nutzer in einigen Fällen auch die relevanten Blöcke nach der eingegebenen Kollokation durchsuchen. Bei der Kollokation *ein Bad nehmen* wird der Nutzer zwar relativ

schnell fündig, da diese im Eintrag von *Wannenbad/Bad* bereits in der dritten Zeile erfasst ist und darüber hinaus die Suchwörter in allen Zeilen durch Fettmarkierung hervorgehoben sind, bei *ein Bad einlassen* hingegen finden sich in allen drei verzeichneten Bedeutungsblöcken (zu *einlassen/einfüllen*, *Wannenbad/Bad* und *Badewanne*) relevante Suchtreffer jeweils am Ende jedes Blocks. Größere Schwierigkeiten zeigen sich außerdem bei Satzsuchen. Hier werden zum Teil – wie bei *er ließ ein Bad ein* – lange Listen irrelevanter Treffer angezeigt, oder der relevante Treffer erscheint erst sehr weit unten in der Ergebnisliste – im Fall von *er nahm ein Bad* zum Beispiel erst im 43. Bedeutungsblock. Ohne die Suchfunktion im Internetbrowser erscheint das Auffinden durch den Nutzer relativ aussichtslos. Damit diese Suche aber erfolgreich ist, muss der Nutzer wiederum die Grundform der Kollokation eingeben (also *ein Bad nehmen*), oder – in diesem Fall – genau die innerhalb eines Beispielsatzes ebenfalls verzeichnete finite Form des Verbs (*hat genommen*) suchen.

Die Suche nach den phraseologischen Angaben zu *Bad* in ihrer Grundform ist zwar bei *Linguee* in vier von fünf Fällen erfolgreich, wie bei *Leo* und *Beolingus* findet sich jedoch keine der Angaben zu *Backe*. Im Gegensatz zu den anderen Wörterbüchern bleiben bei *Linguee* auch die Satzsuchen allesamt ergebnislos.

Im verlagsbasierten Onlinewörterbuch *Pons* finden sich alle Wendungen bei Eingabe in ihrer Grundform. Im Gegensatz zu *Leo*, *dict.cc* und *Beolingus* ist hier positiv hervorzuheben, dass sich die Anzeige der Treffer auf die tatsächlich relevanten beschränkt. Zudem zeigt *Pons* an, aus welchem Wörterbuchteil die Treffer stammen (z. B. „Übersetzungen für ein bad nehmen im Englisch >> Deutsch-Wörterbuch“). Auf diese Hinweise könnte jedoch verzichtet werden, wenn die Informationen aus den beiden Wörterbuchteilen auf doppelte Vorkommen überprüft und redundante Informationen gegebenenfalls herausgenommen würden. Größere Probleme zeigen sich aber bei den Trefferlisten für Satzsuchen. Wie bei *Beolingus* finden sich relevante Treffer erst sehr weit unten in der Liste und sind ohne die Suchfunktion im Internetbrowser meist nur nach sehr langem Suchen auszumachen. Auch hier

wird dafür vom Wörterbuchbenutzer wieder verlangt, dass er die gesuchte Wendung auf ihre Grundform zurückführen kann und diese dann insgesamt oder in Einzelteilen über die Browser-Suchfunktion sucht.

5. Fazit

Insgesamt lässt sich zum einen bei einigen der untersuchten Onlinewörterbücher eine (zumindest in Teilen fehlende) linguistische und lexikografische Basis feststellen, die sich vor allem in der unsystematischen Verwendung von Glossen, der mangelhaften Zuordnung und Unterscheidung von Kollokationen, freien Kombinationen und Beispielsätzen zu bestimmten Kategorien und der rigiden tabellenartigen Struktur der Treffer innerhalb eines Abschnitts in den Trefferlisten niederschlägt. Dies ist vor allem bei den Onlinewörterbüchern der Fall, deren Ursprünge rein im Bereich der Informatik liegen und die den Nutzer – weitgehend unabhängig von seiner linguistischen Kompetenz – stark miteinbeziehen (d. h. bei den *community-based* Wörterbüchern). Diese Wörterbücher unterscheiden sich in diesem Punkt deutlich von einem verlagsbasierten Onlinewörterbuch wie *Pons*. Zum anderen fällt bei allen untersuchten Wörterbüchern auf, dass sie die Möglichkeiten des Onlinemediums bei Weitem nicht ausschöpfen und damit Probleme aufweisen, die sich zwar von denen des Printwörterbuchs auf den ersten Blick deutlich unterscheiden, teilweise aber vergleichbare Folgen haben und daher ähnlich schwer wiegen. Die Ergebnisse der Diskussionen zu den einzelnen Wörterbüchern sollen im Folgenden noch einmal zusammengefasst werden.

In Bezug auf die Äquivalentdifferenzierung hat sich gezeigt, dass die Äquivalente manchmal alphabetisch, manchmal wohl auch semantisch geordnet sind. Sehr häufig lässt sich jedoch feststellen, dass eine genaue und vollständige Äquivalentdifferenzierung kaum vorhanden ist. Grundsätzlich scheint eine eher zufällige Markierungspraxis vorzuliegen. Es gibt zwar durchaus regionale Markierungen beziehungsweise Register-, Fachbereichs- oder Stilangaben (z. B. *Leo*: **baff** machen – to bamboozle so. [*coll*]), oder auch Kollokationspartner (*Beolingus*: **bag-**

gern (Fluss, Fahrerinne; Schlamm)), die durchaus auch als äquivalent-differenzierende Angaben dienen können. Manchmal finden sich auch traditionelle Glossen (*dict.cc*: **baggern** [sich jdm. in sexueller Absicht nähern (bes. Mann)]), eine grundsätzliche Systematik ist jedoch nicht erkennbar. Daher liegt eine große Menge von Äquivalenten ohne weitere Angaben zur Hilfe bei der Auswahl des richtigen Äquivalents vor.

Ein ähnliches Phänomen lässt sich auch bei Grammatikangaben feststellen. Die Angabe der Wortart erfolgt zumeist, jedoch nicht immer, für jede Lemma-Äquivalent-Einheit, manchmal (etwa bei *dict.cc*) auch übergeordnet als Überschrift. *Beolingus* gibt als einziges der untersuchten Wörterbücher die unregelmäßigen Formen aller Lemmata und aller Äquivalente an, was die Übersichtlichkeit jedoch deutlich gefährdet. Die anderen Onlinewörterbücher verhalten sich weniger systematisch (zumeist erfolgt die Angabe der unregelmäßigen Formen für die deutschen Wörter, wie z. B. bei *Leo* und *dict.cc*). Valenzangaben lassen sich durchaus manchmal finden, so zum Beispiel im *dict.cc*: *to bathe sb*. Erfreulicherweise erfolgt die Angabe zumeist in Form von Patternillustrationen, was ja die wünschenswerte Form der Darstellung ist. Doch auch hier vermisst man eine systematische und vollständige Darstellung der entsprechenden Valenzangaben. Dabei kann kaum das Argument hervorgebracht werden, dass kein Platz dafür vorhanden ist, da in den Zeilen immer Platz für derartige Information wäre. Im Hinblick auf grammatische Angaben ist außerdem positiv zu bewerten, dass auf die Verwendung von ‚transitiv/intransitiv‘ weitgehend verzichtet wird (Ausnahme ist hier *Pons*). Problematisch ist hingegen die Angabe der Wortart als Gliederungsmittel bei *dict.cc* und *Leo*, da die Angabe der Wortart nicht immer richtig zugeordnet ist und ihre Relevanz für den Nutzer außerdem fraglich erscheint.

Darüber hinaus fällt auf, dass es zum Teil zahlreiche Beispiele gibt, wobei sich über die Sinnhaftigkeit der Auswahl der Beispielsätze in einigen Fällen natürlich diskutieren lässt. Häufig sind die Beispielsätze außerdem keiner Valenzangabe, keinem Äquivalent oder keiner Kollokationsangabe zugeordnet, sodass die übliche und durchaus sinnvolle Struktur der Zuweisung eines Beispielsatzes zu einer bestimmten In-

formationseinheit nicht gegeben ist und vom Nutzer deshalb vermutlich überhaupt nicht als relevant für seine Suche wahrgenommen werden wird. Eine Ausnahme bilden hier das *Pons*-Onlinewörterbuch und teilweise *Beolingus*.

Wie eingangs bereits erwähnt, lässt sich außerdem feststellen, dass die Möglichkeiten, die das Medium Internet bietet, aktuell von den Onlinewörterbüchern nicht voll ausgeschöpft werden. Dabei fällt die fehlende Integration von Korpora (Ausnahme: *Linguee*) ebenso wie die zwar lockerere Darstellung als in den Printwörterbüchern, aber trotzdem starre und nutzerunabhängige Präsentation der Treffer in den tabelleartigen Listen auf. Die Möglichkeit der Integration dynamischer Elemente bleibt hier weitestgehend ungenutzt. Hierzu zählt auch die Anzeige der deutschen Treffer in der rechten Spalte unabhängig von der Suchanfrage, wie dies bei einigen Wörterbüchern vorkommt, und im Widerspruch zu dem Nutzerkreis deutscher Muttersprachler steht. Denn intuitiv wird die bekannte Information (also das Lemma) immer zuerst, also in der linken Spalte erwartet, und die neue Information in der rechten Spalte. Die tabellarischen Listen selbst sind aufgrund ihrer Länge oft im gleichen Maße unübersichtlich wie die gedruckten Einträge in vielen Printwörterbüchern. Daraus entsteht ein doppeltes Problem: Zum einen stellt sich die Frage, ob die Listenstruktur für alle Angabetypen geeignet ist. Und zum anderen wären die Listen in vielen Onlinewörterbüchern wohl viel kürzer, wenn nur relevante Informationen darin enthalten wären. Denn die erweiterten Platzmöglichkeiten im Vergleich zum Printmedium werden vor allem so ausgenutzt, dass oft eine sehr große Zahl an nicht relevanten Treffern mitangezeigt wird (Ausnahme: *Pons*, evtl. auch *Linguee*). Das führt natürlich auch dazu, dass sich der Nutzer, der durch die Benutzung eines Onlinewörterbuchs vermeintlich Zeit spart, häufig durch sehr viele ungeordnete Äquivalente (ohne differenzierende Angaben) und möglicherweise viele unbrauchbare, weil auf seinen Fall nicht zutreffende Beispielsätze „kämpfen“ muss, bis er die gesuchte Information findet. Durch die oft fehlende Hilfe kann sich der Nachschlagevorgang dadurch erheblich verlän-

gern. Dies hat – ähnlich wie beim Printwörterbuch – zur Folge, dass er eher zu einem der ersten Treffer greift als weiterliest.

Was die Zuverlässigkeit der Onlinewörterbücher betrifft, so ist oft nicht (unmittelbar) ersichtlich, um welche Informationen es sich handelt beziehungsweise aus welcher Quelle sie stammen (dies ist v. a. bei Informationen, die durch Nutzer ungeprüft ins Wörterbuch eingetragen werden können, relevant). Die Nutzung des Onlinewörterbuchs ist hier im Grunde nur in dem Maße sinnvoll, dass der Nutzer sich Anregungen holt, die Vorschläge aber selbst noch verifizieren muss (z. B. in einschlägigen Verlagswörterbüchern, in Terminologiedatenbanken, in einsprachigen Wörterbüchern, über einen Paralleltextvergleich, etc.). Insofern sind die Onlinewörterbücher zwar für den sprachrezeptiven Gebrauch von Nutzen, haben aber für den sprachproduktiven Zweck oft nur eingeschränkten Wert.

Literatur

- BAUNEBJERG HANSEN, GITTE (1990): Artikelstruktur im zweisprachigen Wörterbuch. Tübingen: Niemeyer.
- BROWN, GILLIAN/YULE, GEORGE (1983): *Discourse Analysis*. Cambridge: Cambridge University Press.
- CARR, MICHAEL (1997): Internet Dictionaries and Lexicography. In: *International Journal of Lexicography* 10/3, 209–230.
- DE SCHRYVER, GILLES-MAURICE (2003): Lexicographers' Dreams in the Electronic-Dictionary Age. In: *International Journal of Lexicography* 16/2, 143–199.
- DYKA, SUSANNE (2005): *Probleme der Differenzierung im deutsch-englischen Wörterbuch für Deutsche*. Tübingen: Niemeyer.
- DYKA, SUSANNE/SCHARF, EVA (2013): Phraseologische Angaben im produktionsorientierten zweisprachigen Wörterbuch – eine empirische Analyse. In: BÜRCEL, CHRISTOPH/SIEPMANN, DIRK (Hrsg.): *Sprachwissenschaft – Fremdsprachendidaktik: Neue Impulse*. Baltmannsweiler: Schneider Verlag Hohengehren, 21–42.
- ENGELBERG, STEFAN/LEMNITZER, LOTHAR (2009): *Lexikographie und Wörterbuchbenutzung*. 4., überarbeitete und erweiterte Auflage. Tübingen: Stauffenburg.

- HAB, ULRIKE/SCHMITZ, ULRICH (2010): Lexikographie im Internet 2010 – Einleitung. In: *Lexicographica* 26, 1–18.
- HAUSMANN, FRANZ JOSEF (1977): Einführung in die Benutzung der neufranzösischen Wörterbücher. Tübingen: Niemeyer.
- HERBST, THOMAS/KLOTZ, MICHAEL (2003): *Lexikografie*. Paderborn: Schöningh.
- KLOSA, ANNETTE/LEMNITZER, LOTHAR/NEUMANN, GERALD (2008): Wörterbuchportale – Fragen der Benutzerführung. In: KLOSA, ANNETTE (Hrsg.): *Lexikographische Portale im Internet*. Mannheim: Institut für Deutsche Sprache (OPAL-Sonderheft. 1/2008), 5–35. Online: <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2008-1.pdf> [Abruf: 2014-06-02].
- LEHR, ANDREA (1996): Zur neuen *Lexicographica*-Rubrik ‚Electronic dictionaries‘. In: *Lexicographica* 12, 310–317.
- LEW, ROBERT (2013): Online dictionary skills. In: KOSEM, IZTOK/KALLAS, JELENA/GANTAR, POLONA/KREK, SIMON/LANGEMETS, MARGIT/TUULIK, MARIA (Hrsg.): *Electronic lexicography in the 21st century: Thinking outside the paper*. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 16–31. Online: http://eki.ee/elex2013/proceedings/eLex2013_02_Lew.pdf [Abruf: 2014-06-04].
- LEW, ROBERT (i. Dr.): Space restrictions in paper and electronic dictionaries and their implications for the design of production dictionaries. In: BAŃSKI, PIOTR/WÓJTOWICZ, BEATA (Hrsg.): *Issues in Modern Lexicography*. München: Lincom Europa (im Druck). Auch online: http://www.staff.amu.edu.pl/~rlew/pub/Lew_space_restrictions_in_paper_and_electronic_dictionaries.pdf [Abruf: 2014-06-04].
- MANN, MICHAEL (2010): Internet-Wörterbücher am Ende der „Nullerjahre“: Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien unter besonderer Berücksichtigung der Fachlexikographie. In: *Lexicographica* 26, 19–45.
- SVENSÉN, BO (2009): *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Wörterbücher

- DudOx = Duden-Oxford Großwörterbuch Englisch. Deutsch–Englisch, Englisch–Deutsch. Herausgegeben von der Dudenredaktion/bearbeitet von BRIGITTE ALSLEBEN u. a. Mannheim: Dudenverlag, 2005.

LCGwE = Langenscheidt Collins Großwörterbuch Englisch. Englisch–Deutsch, Deutsch–Englisch. Herausgegeben von Collins in Zusammenarbeit mit der Langenscheidt-Redaktion/bearbeitet von LORNA SINCLAIR-KNIGHT. Berlin/München: Langenscheidt, 2008.

PGW = Pons Großwörterbuch Englisch. Englisch–Deutsch, Deutsch–Englisch. Bearbeitet von EVELYN AGBARIA. Barcelona [u. a.]: Klett Sprachen, 2005.

Internetquellen

Beolingus = Beolingus – Ihr Online-Wörterbuch. Online:

<http://dict.tu-chemnitz.de>

dict.cc = dict.cc Deutsch-Englisch-Wörterbuch. Online: <http://www.dict.cc>

Leo = Leo Englisch–Deutsch. Online: <https://dict.leo.org>

Linguee = Linguee. Online: <http://www.linguee.de>

Pons = Pons Online-Wörterbuch. Online: <http://de.pons.com>

Alexa. Online: <http://www.alexa.com>