Thematic Part

Lexikographie im Internet 2010 – Einleitung

Editors of the Thematic Part: Ulrike Haß/Ulrich Schmitz

1	Wozu Internet-Lexikographie? (Vor- und Nachteile, Potenziale und Probleme)	2.1	Korpuslinguistischer impact
1.1	Lexikographie und Medientechnik	2.2	Fokus Zugriffsstrukturen
1.2	Neun Chancen online	2.3	Für eine neue Offensive zugunsten der linguistisch-lexikografischen Qualität
1.3	Neun Risiken online	3	Mehrsprachigkeit: Kleine und große Sprachen
1.4	Drei Herausforderungen (Finanzierung, medialer Mehrwert, Ergonomie)	4	Nutzersicht: mediale Wahrnehmungen
2	Was ist bisher realisiert worden und was nicht?	5	Überblick über den Inhalt des folgenden Thementails
		6	Literatur

Abstract

This introduction to the volume deals in general with media-related conditions and constraints of today's online-lexicography. Chances as well as problems and the challenges in this field are described, be they technical, cultural or economic. The authors outline the current achievements and point to what has not yet been accomplished. The users' perspective as well as the case of multilingual dictionaries are explicitly addressed. All in all, the contributions to this volume highlight some special focuses – the impact of corpus linguistics on lexicography, especially on online-lexicography, and the predominant interest in structures of access to information. Therefore some further developments have to be pointed out in order to emphasize their relevance for the quality of dictionaries. Finally, the contributions to the volume are characterized, and they are arranged in a reasonable structure.

1 Wozu Internet-Lexikographie? (Vor- und Nachteile, Potenziale und Probleme)

1.1 Lexikographie und Medientechnik

„Das Buch hat aufgehört, ein Mikrokosmos nach klassischer und abendländischer Art zu sein. [...] Findet die Stellen in einem Buch, mit denen ihr etwas anfangen könnt. Wir lesen und schreiben nicht mehr in der herkömmlichen Weise. Es gibt keinen Tod des Buches, sondern eine neue Art zu lesen. In einem Buch gibt's nichts zu verstehen, aber viel, dessen man sich bedienen kann. Nichts zu interpretieren und zu bedeuten, aber viel, womit man experimentieren kann. Ein Buch muß mit etwas anderem ‚Maschine machen‘, es muß ein kleines Werkzeug für ein Außen sein. [...] Die Kombinationen, Permutationen und Gebrauchsweisen sind dem Buch nie inhärent, sondern hängen von seinen Verbindungen mit diesem

oder jenem Außen ab.“¹ Was Deleuze und Guattari 1976, als niemand an das World Wide Web auch nur denken konnte, im Anschluss an Proust und Foucault hier über neuen Umgang mit Büchern generell sagen, galt nicht unbedingt theoretisch, aber doch in der Nutzungspraxis schon immer für Wörterbücher und Lexika: Innerhalb anderer Arbeitsprozesse wurden und werden sie vorrangig ad hoc zum Nachschlagen für einen bestimmten Zweck genutzt, der außerhalb des „Mikrokosmos“ des Buches liegt. Neue Techniken jenseits des Buchdrucks nun unterstützen und intensivieren derart selektive, anwendungsorientierte Nutzungen. Wenn die Informationen nicht mehr – wie beim Druck – in genau einer bestimmten und dann (mindestens bis zur nächsten Auflage) unveränderlichen Erscheinungsform dargeboten werden müssen, sondern – wie bei digitaler Aufbewahrung – von Fall zu Fall auf höchst variable Weise dem menschlichen Auge (zum Beispiel am Bildschirm) präsentiert werden können, können die Daten erstens jederzeit verändert und vermehrt und zweitens von vornherein so konstruiert werden, dass sie sich in ihrer Struktur, Fülle, Auswahl und Darbietung ad hoc höchst unterschiedlichen Nutzungen anzupassen in der Lage sind. Außerdem bietet die neue Hypertext-Technik, die ja beliebig viele und fortwährend veränderbare, intern und extern schnell realisierbare Querverweise erlaubt, die technisch bequeme Grundlage für rhizomatische Darstellungs- und Gebrauchsweisen. Weitere technische (insbesondere multimodale) Möglichkeiten kommen hinzu; sie lassen die Potenziale herkömmlicher Lexika besser verwirklichen und erheblich ausdehnen.

Wie weit und wie gut wurde von diesen Möglichkeiten bisher Gebrauch gemacht? Der thematische Schwerpunktteil des Jahrbuchs *Lexicographica* 2010 gibt einen Überblick über den gegenwärtigen Stand der Lexikographie (insbesondere Wörterbuchschreibung) im Internet.² In acht systematischen oder projektbezogenen Aufsätzen werden aktuelle theoretische, organisatorische, technische, medienspezifische und ergonomische Probleme, Lösungen, Leistungen, Potenziale und Perspektiven erörtert. Anschließend stellen neun regional orientierte Beiträge wichtige lexikographische Adressen im Web vor. Den Schluss bildet eine Sammlung wichtiger Linklisten zum Thema.

Schon immer hing die Arbeit von Lexikographen und in der Folge auch Lexikologen stark davon ab, mit welchen technischen Mitteln Daten erhoben, gesammelt, dokumentiert und publiziert werden konnten. Die Erfindung und Verbreitung des World Wide Web in den 1990er Jahren hat das gesamte Mediensystem ähnlich tief greifend, nur wesentlich schneller, revolutioniert wie ein halbes Jahrtausend zuvor der Buchdruck. Von diesen technischen Umwälzungen und ihren Folgen sind Lexika in besonderer Weise betroffen.

Lexika (im Folgenden insbesondere Wörterbücher) sammeln und präsentieren Wissen (im Folgenden geht es insbesondere um Wissen über Sprache und Sprachen) in möglichst zweckmäßiger, möglichst umfassender und möglichst nutzerfreundlicher Form. Vor allem die letzten beiden dieser drei Ziele stehen in Spannung untereinander; das führt zu vielfältigen Arten und Formen von Lexika bzw. Wörterbüchern. Je mehr und je detaillierter Material dokumentiert werden soll, desto länger dauert der Herstellungsprozess, desto weniger bequem ist das Ergebnis zu nutzen, und desto obsoleter können die Informationen mittlerweile geworden sein. Je stärker umgekehrt ein Lexikon auf bestimmte Zielgruppen und Anwendungen hin ausgerichtet werden soll, desto eingeschränkter und kurzlebiger mag es

¹ Deleuze/Guattari 1977, 40.

² Vgl. dazu auch Bergenholtz et al. (eds.) 2009; Kunze/Lemnitzer 2007; Gouws et al. (eds.) 2010, ch. xiv-xviii.

von Nutzen sein. Bei gedruckten Lexika muss man sich stets von Fall zu Fall für jeweils möglichst geeignete Varianten zwischen beiden Polen und also für Kompromisse entscheiden. Gedruckte Lexika können weder annähernd vollständig noch annähernd aktuell sein.

1.2 Neun Chancen online

Ganz anders – theoretisch jedenfalls— bei Online-Lexika und -Wörterbüchern. Sie werden nicht gedruckt, sondern digital im Internet zur Verfügung gestellt. Das bringt folgendes Bündel von Vorteilen mit sich:

- (1) Sie müssen nie abgeschlossen werden oder fertig sein, sondern können *jederzeit aktualisiert*, redigiert und erweitert werden.
- (2) Grundsätzlich können *unendlich umfangreiche Mengen* von Informationen (auch von Belegen) eingespeist werden.
- (3) Dazu gehören auch *multimodale Daten*, die im Druck entweder zu teuer oder gar nicht möglich wären: farbige Illustrationen (z.B. Zeichnungen oder Fotos von Gegenständen), bewegte Bilder (z.B. schematische Darstellungen von Bewegungsabläufen), Töne (z.B. gesprochene Sprache), Filme (z.B. Videos exemplarischen Sprachgebrauchs), automatisch erzeugte Infografiken (z.B. Statistiken, Kookkurrenzdiagramme) u.a.
- (4) Damit kann, wenn gewünscht, die herkömmliche Unterscheidung zwischen verschiedenen Typen von Wörterbüchern (z.B. ein- und mehrsprachig, ono- und semasiologisch, universal vs. fach- oder zweckspezifisch, etymologisch etc.) entfallen zugunsten einer *umfassenden Dokumentation*.
- (5) *Hypertextuelle Verweise* innerhalb eines Lexikons (einer Angebotsmenge) lassen die Nutzerin nicht zeitaufwendig hin- und herblättern, sondern führen sie per Mausklick unmittelbar auf die jeweils gesuchten Stellen.
- (6) *Externe Verweise* nach außerhalb des Lexikons müssen nicht unbedingt per Quellenhinweis zu anderen Büchern (und folglich Bibliotheksgängen), sondern können ggf. auch per Mausklick direkt zu anderen Adressen im WWW führen.
- (7) Damit brauchen Internet-Lexika nicht in sich abgeschlossen zu sein, sondern können gegenseitig aufeinander verweisen, miteinander verlinkt oder mit Textkorpora sowie mit nichtlexikographischen Angeboten im WWW verknüpft sein. Folglich können Nutzer sie gemeinsam mit anderen Angeboten in einem kontinuierlichen Arbeitsprozess jeweils ad hoc als Module oder *Fragmente innerhalb des gesamten (enzyklopädischen) WWW* verwenden.
- (8) So kann, wenn gewünscht, auch die traditionelle Unterscheidung zwischen Wörterbüchern und Enzyklopädien aufgegeben werden zugunsten *grenzenlos verflochtener Enzyklopädien*.
- (9) Anders als Buchprojekte können Online-Angebote auch so angelegt sein, dass Nutzerinnen und Nutzer sich *„interaktiv“* an deren Ausbau und Pflege beteiligen können.

1.3 Neun Risiken online

Diesen neuen und bei weitem noch nicht ausgeschöpften Potenzialen stehen Risiken und Nachteile gegenüber, die es zu bedenken und möglichst klein zu halten gilt. (A) Je umfangreicher, intern differenzierter und extern mit anderen Angeboten verflochtener ein digitales

Wörterbuch ist, desto schwerer fällt der Nutzerin oder dem Nutzer die Orientierung; sie können leicht darin verloren gehen (*lost in hyperspace*) und/oder die tatsächlichen Nutzungspotenziale übersehen. (B) Und je mehr eine Adresse nicht (allein) von einer (z.B. wissenschaftlich ausgewiesenen) Redaktion, sondern interaktiv von einer Vielzahl mehr oder weniger anonymer Informanten gefüllt wird, desto (im wörtlichen Sinne) fragwürdiger sind Zuverlässigkeit und *Seriosität* der Angaben – trotz erhoffter und unterschiedlich erfolgreicher Schwarmintelligenz wie etwa bei Wikipedia oder Wiktionary.

Im Einzelnen (gemäß obiger Nummerierung):

- (1) Wenn ein öffentlich zugängliches Wörterbuch in jedem Moment geändert werden kann, müssen zwar nicht, können jedoch seine Angaben *an Autorität verlieren*. Nutzer können verwirrt sein, wenn sie sich neuen Strukturen anpassen müssen oder auch nur gewohnte Angaben nicht mehr wiederfinden (und dann nicht ohne Weiteres zitieren können).
- (2) Je mehr Informationen angeboten werden, desto schwerer wird es dem Nutzer, genau die zu *finden*, die er braucht.
- (3) Je vielfältiger das multimodale Angebot, desto *ablenkungsreicher die Recherche*.
- (4) Je mehr unterschiedliche Funktionen ein Online-Wörterbuch erfüllt, desto *mehr Vorkenntnisse* sollte der Nutzer mitbringen und desto trennschärfer sollte seine Fragestellung sein.
- (5) Ein gedrucktes Lexikon liegt sinnlich spürbar in der Hand oder steht vollständig im Regal vor Augen. Online-Wörterbücher haben *keinen Anfang und kein Ende*; immer nur erscheint ein winziger Ausschnitt auf dem Bildschirm (wie ein vielfältig verknüpfter tomographischer Schnitt durch ein unüberschaubares Ganzes). Und wer lange am Bildschirm arbeitet, strapaziert seine Augen mehr als bei Büchern.
- (6) Externe Hyperlinks laden zu *unkontrolliertem Surfen* durch das gesamte Internet ein. Natürlich kann Flanieren statt zielstrebig Suchen sowohl angenehm als auch nützlich sein; bei der extrem heterogenen Vielfalt des Internet allerdings bedarf es, wenn man nicht süchtig oder kirre werden will, größerer Disziplin als in gedruckten Bücherwelten.
- (7) Arbeiten mit dem World Wide Web erfordert andere und teilweise erheblich *anspruchsvollere Kompetenzen der Nutzer*. Auch Arbeit mit Büchern gelingt umso besser, je umfassender die einschlägige Vorbildung der Leser ist. Erfolgreiche Web-Nutzung ist aber komplexer.
- (8) Die technisch realisierbare Utopie allgemein zugänglicher und vollständiger Dokumentation sämtlichen menschlichen Wissens kann individuelle Nutzer *überfordern*.
- (9) In dem Maße, wie jederfrau und jedermann sich an Pflege und Ausbau des Angebotes beteiligen kann, stehen *Qualität und Zuverlässigkeit der Informationen* in Frage, zumal wenn Quellen nicht überprüft werden können.

1.4 Drei Herausforderungen (Finanzierung, medialer Mehrwert, Ergonomie)

Die größte Aufgabe der Internet-Lexikographie besteht darin, diese neun Chancen und Risiken von Fall zu Fall in ein möglichst günstiges Verhältnis zu bringen. Das ist geistig, politisch und technisch alles andere als einfach und erfordert sowohl allgemein als auch bei jedem einzelnen Projekt aufs Neue viel grundsätzliche Überlegung, viel lexikologische Erkundung, viel Nutzerforschung, viel konzeptionelle Arbeit, viel alltäglich neue Entscheidungen, viel technisches Können und viel Fingerspitzengefühl. Mit wachsender Erfahrung,

die möglichst gut dokumentiert werden sollte, können lexikologische Arbeiten und lexikographische Angebote immer besser werden. Nach zwei Jahrzehnten Arbeit an digitalen Wörterbüchern stehen wir erst ganz am Anfang einer lexikographischen und in der Folge auch lexikologischen Epochenwende.

Drei Probleme werden bei der weiteren Arbeit an Online-Wörterbüchern und -Lexika im Mittelpunkt stehen: Geld, Nutzen und Ergonomie.

(I) Jeder Laie, der über einen Computer mit Internet-Anschluss und elementare technische Kenntnisse verfügt, kann selbst gebastelte Wörterbücher ins WWW stellen. Tatsächlich gibt es zahlreiche davon, und nicht alle sind schlecht oder überflüssig. Je seriöser, nutzbringender, umfangreicher und medial angemessener Online-Wörterbücher jedoch sein sollen, desto aufwendiger und entsprechend teurer wird auch ihre Erstellung und Pflege. Kostenpflichtige Angebote dürften zumindest in absehbarer Zeit wenig Chancen gegen gedruckte Wörterbücher haben. Für kommerzielle Anbieter dürften sie (außer als Werbeträger und/oder Eigenwerbung für andere Produkte) vorläufig kaum interessant sein. Kostenlos nutzbare Online-Wörterbücher hingegen – derzeit fraglos das Gros aller Angebote – leben entweder von freiwilliger Begeisterung bzw. Selbstaussbeutung oder aber müssen *teuer* bezahlt werden. Wer wird das finanzieren? Nachdem Akademien und ähnliche Institutionen herkömmliche Wörterbücher über Jahrzehnte mit teilweise erheblichen Summen überhaupt erst ermöglicht haben, ist nicht unbedingt zu erwarten, dass sie sich in größerem Umfang langfristig neuen Projekten gegenüber verpflichten.

(II) Wer Geld und oder Arbeitszeit investiert, möchte natürlich auch einen *Nutzen* sehen. Die hohen Nutzerzahlen (visits und hits), die viele Internet-Wörterbücher vorweisen können, sprechen für sich. Was angeboten wird und sich nicht versteckt, wird auch genutzt. Was bequem zu nutzen ist und bei Anfragen über Suchmaschinen weit oben in den Antwortlisten steht, wird oft genutzt und steigt in deren Algorithmen dadurch noch höher auf. Häufiger Gebrauch korreliert aber nicht unbedingt mit – wie auch immer – definierter Qualität.

Als wesentliche Qualitätsmerkmale sollten zunächst – wie bei gedruckten Wörterbüchern – Zweckmäßigkeit, Menge, Differenzierung und Seriosität der Inhalte (i.e. Zuverlässigkeit der Angaben) gelten. Das rechtfertigt aber noch nicht die Existenz digitaler Wörterbücher. Sie hängt vielmehr allein davon ab, in welcher Weise die Potenziale an *medialem Mehrwert*, welche die Kommunikationsform World Wide Web bietet, ausgeschöpft werden. Praktisch alle Angebote bieten zwei der vier Merkmale, welche das Internet trivialerweise gegenüber gedruckten Erzeugnissen auszeichnen, nämlich schnellen Zugriff bei weltweiter Zugänglichkeit. Häufig, aber keineswegs immer, werden auch die beiden anderen Grundmerkmale realisiert, nämlich vergleichsweise große Informationsmenge und schnelle Aktualisierung. Darüber hinaus jedoch wurden die medienspezifischen technischen Möglichkeiten des WWW (vor allem was Hypertextstrukturen, Multimodalität und kooperative Arbeitsformen angeht) bisher erst in Ansätzen ausgenutzt. Das liegt zum Einen an institutioneller und individueller Trägheit: Auch Lexikographen geben bewährte Konventionen und gewohnte Arbeitsweisen nicht von heute auf morgen auf. Das ist verständlich und oft gut begründet. Zum Anderen aber bedeuten medienspezifische Umwälzungen erheblichen Aufwand und Kosten – siehe oben – viel Geld. Und schließlich werfen alle Innovationen, welche die neuen medientechnischen Potenziale erkunden und realisieren, schwierige Probleme bei der sachgerechten und nutzerfreundlichen Präsentation der Daten auf.

Womit wir – nach Geld und Nutzen – beim dritten genannten Problemkreis wären: der Ergonomie.

(III) Selbst und gerade wenn man darauf baut, dass es angesichts der großen lexikographischen Tradition und kontinuierlicher aktueller Weiterarbeit an mehr oder weniger qualitativem „Content“ nicht mangelt, stellen digitale Produktionen und hier insbesondere Online-Erzeugnisse völlig andere Anforderungen an die Aufbereitung und Präsentation der Materialien. Alle neun oben genannten Chance&Risiko-Paare können nur durch *intelligente Ergonomie* des Benutzerzugangs in eine optimale Balance gebracht werden:

- (1) Wenn Inhalte ständig verändert und vielfältig erweitert werden, müssen die Präsentationsstrukturen einerseits sich dem ständigen Wandel anpassen und andererseits dem Nutzer ebenso klare wie *verlässliche Orientierung* und ebenso überschaubare wie leistungsfähige Navigationsmöglichkeiten bieten.
- (2) Je stärker die Informationsmenge wächst, desto leichter muss es dem Nutzer gemacht werden, das für ihn *Brauchbare zu finden*.
- (3) Je mehr und je vielfältiger von anderen als rein schriftlichen Formen Gebrauch gemacht wird, desto intelligenter und transparenter muss deren *multimodales Zusammenspiel* gestaltet werden.
- (4) Je *hybrider* verschiedene herkömmliche Wörterbuchtypen vermischt und integriert werden, desto prägnanter müssen die unterschiedlichen Typen von Information als solche *erkennbar* sein.
- (5) Feste Makro-, Meso- und Mikrostrukturen gedruckter Wörterbücher werden in Online-Angeboten zugunsten komplexer *Hyperstrukturen* aufgelöst, die netzwerkartig organisiert sind, flexibel erweitert und verändert werden können, vor allem aber der Nutzerin oder dem Nutzer nicht offen vor Augen liegen. Die größte ergonomische Herausforderung bei der Konstruktion elektronischer Wörterbücher liegt darin, das Angebot so *transparent zu gestalten*, dass der Nutzer bei jedem einzelnen „Nachschlagevorgang“ (also jeder einzelnen Anfrage und ihrer maschinellen Antwort darauf) möglichst genau intuitiv weiß oder zumindest durch Nachfrage erfahren kann, an welcher Stelle innerhalb des Gesamtangebots er sich befindet.
- (6) Wenn *externe Verweise* auf andere Adressen im Web aufgenommen werden, muss erstens klar *ersichtlich* sein, dass bei deren Nutzung eine andere semiotische Landschaft mit anderen Konventionen betreten wird, und zweitens eine einfache *Rückkehrmöglichkeit* garantiert sein.
- (7) Die gesamte Gestaltung von Online-Wörterbüchern sollte auf deren prototypischen (häufigsten) Einsatz hin ausgelegt sein. Schon herkömmliche Wörterbücher dienten und dienen meistens einzelnen Ad-hoc-Abfragen innerhalb anderer Arbeitsprozesse. Integriertes Arbeiten mit unterschiedlichen Zielen, Erzeugnissen, Quellen, Techniken, Methoden, Kommunikationsformen, Produzenten und Adressaten(-gruppen) wird mit dem Web allerdings auf die Spitze getrieben. Dem sollte die Gestaltung der Angebote entgegenkommen. Der Weg vom klassisch-encyklopädischen Ganzheitsideal zur später angestrebten *Nutzungseffizienz innerhalb komplexer Arbeitsprozesse*, den gedruckte Lexika in ihrer mehrhundertjährigen Geschichte zurückgelegt haben, mag hier als wenngleich obsoletes Vorbild dienen.
- (8) Die Grenzen meines Angebots bedeuten die Grenzen meiner Ergonomie. Niemand kann und soll das gesamte Web strukturieren. Vielfalt heißt Kreativität, Kreativität Vielfalt. Doch auch ein einzelnes Wörterbuch oder Verbünde solcher Wörterbücher können ihre Zugänge von außen und Brücken nach außen ästhetisch und funktional gestalten. Auch hier sollte gelten: *form follows function*.

- (9) Wenn Wörterbuchprojekte teilweise oder vollständig ‚interaktiv‘ angelegt sind, sollte die Zuarbeit aller Personen technisch möglichst einfach und ergonomisch möglichst gleichartig gestaltet sein. Hier bieten sich gut durchdachte *Templates* an, also praktikable Eingabeformulare, die zu durchgehend gleichartig gestalteten Präsentationsformen führen. Für die Qualität der Inhalte allerdings ist der Designer nicht verantwortlich.

2 Was ist bisher realisiert worden und was nicht?

Die Internet-Lexikografie ist alles andere als die Fortsetzung der lexikografischen Tradition in einem neuen Medium, auch wenn die Innovationsrhetorik der Internet-Frühzeit in den 1990er Jahren und gelegentlich noch heute gern das Früher vom Jetzt abgrenzt und damit eine Art linearen Fortschritt suggeriert. Die ersten 15, 20 Jahre einer freien, die Möglichkeiten des Internets explorierenden und mit ihnen experimentierenden lexikografischen Praxis sind zu besichtigen; und weder müssen sie eine kritische Befragung scheuen, noch muss eine Kritik sich heute noch dem Verdacht des Traditionalismus und der Fortschrittsfeindlichkeit ausgesetzt sehen.

Nimmt man die Beiträge in diesem Band – ob Aufsatz oder Dokumentation – als exemplarisch für das, was die linguistische Forschung zur Internet-Lexikografie und die mit ihr verbundene Praxis im zu überblickenden europäischen Rahmen interessiert, was sie aktuell bewegt und umtreibt, und fragt man in guter wissenschaftstheoretischer Manier nach ihrer ‚Gegenstandskonstitution‘ und ihrem Fokus, dann lassen sich einige über die Beiträge dieses Bands hinausreichende, allgemeinere Einsichten gewinnen, die nachfolgend skizziert werden sollen.

2.1 Korpuslinguistischer impact

Den Beiträgen von Quasthoff, Lemnitzer, Strecker und Fisseni ist gemeinsam, dass sie von je unterschiedlichen Sprachdatensammlungen ausgehen und eine ganze Palette von lexikografisch nutzbaren und z.T. bereits genutzten Informationen zeigen, die mithilfe korpuslinguistischer bzw. statistischer Verfahren aus einer gegebenen Sprachdatenmenge gewonnen werden können. Sprachdaten liegen in Form mehr oder weniger definierter, abgeschlossener oder offener und dynamisch wachsender und in ihrer Zusammensetzung komponierter Textkorpora vor (so bei Quasthoff), aber auch das Internet selbst ist als unendliche Ressource entdeckt (Lemnitzer). Strecker geht hier noch einen Schritt weiter und greift mit dem grammatischen Wörterbuch auf einen selbst (bzw. von derselben Projektgruppe) verfassten Hypertext, hier gr@mmis, zurück, ähnlich wie man auch ein Sachregister zu einem Fachbuch als lexikografisches (Neben-)Produkt auffassen kann.

Die ‚Wörterbuchbasis‘ ist also mächtig in Fluss geraten!³ Galt es ‚früher‘ als wesentliches Qualitätsmerkmal eines Wörterbuchs, dass die Beschreibung des Wortgebrauchs auf

³ Ein weiteres Beispiel ist das Deutsche Fremdwörterbuch, 6. Band, 2008 (vgl. den Rezensionsteil

genau diejenige(n) Varietät(en), denjenigen Diskurs, Zeitraum usw. bezogen werden kann, den das „Quellenverzeichnis“ belegte, so scheint es vielen heute eine uneingeschränkte Bereicherung, dass das Meer der Quellen, aus denen lexikologische Informationen gewonnen werden, unendlich ist. Zunächst ist zu fragen, wie die klassische Forderung „Definiere deine Wörterbuchbasis!“ begründet war. Sie gründete im allgemein-wissenschaftlichen Prinzip der Beleg- oder Nachweispflicht einer These und in der Forderung, dass Experimente bzw. Verfahren wiederholbar sein und dass deshalb Verfahren und Materialbasis benannt werden und zugänglich sein müssen. Angaben zu Wortbedeutung und Wortgebrauch sind und bleiben ‚Thesen‘, die mittels linguistischer Verfahren, seien sie maschinell und/oder intellektuell, aufgestellt werden. Hinzu kam das sprachwissenschaftliche Wissen, dass Wörter und ihre Eigenschaften und natürlich auch Wortfelder und semantische Relationen sich von Varietät zu Varietät, Diskurs zu Diskurs und Zeitraum zu Zeitraum deutlich unterscheiden können und dass allgemeingültige Angaben mit einem eine gesamte Einzelsprache umfassenden Skopus deshalb nur in den seltensten Fällen plausibel sind. Wenn also nahezu jede Wörterbuchbasis selektiv ist, muss man die Frage nach den Selektionskriterien beantworten können.

Es ist klar: Diese Ansprüche sind nicht länger haltbar, wenn ein Wörterbuch aus dem digitalen Ozean des Internets bzw. der anderweitig verfügbaren digitalen Sprachdaten schöpft, wenn die Wörterbuchbasis sich von Tag zu Tag ändert, ohne dass angegeben werden könnte, wie; wenn die Verfahren der Gewinnung lexikologischer Informationen intersubjektiv nicht mehr transparent und wiederholbar sind. Doch was gewinnt man dafür? Erstens den Vorteil der schieren Masse. Korpusstatistische Verfahren, wie sie von Quasthoff in diesem Band eindrucksvoll ausgebreitet werden, führen zu umso besseren Ergebnissen, je größer die Menge der Sprachdaten ist, auf denen sie operieren; und es gibt keinerlei Grund, die Menge morgen zu begrenzen, wenn sie übermorgen noch größer ist. Kookkurrenz- bzw. Kollokationsanalysen sind das mittlerweile herausragende Mittel für die Gewinnung von Wortinformationen unterschiedlichster Arten. Man könnte auch sagen: Mithilfe der automatischen Analyse von ‚very large scale corpora‘ ist die Unterscheidung zwischen usuellen und okkasionellen Eigenschaften von Lexemen und Lexemrelationen sicherer möglich denn je. Dass dies überaus positive Folgen etwa für die Makrostruktur bzw. Lemmaauswahl eines Wörterbuchs hat und in Verbindung mit Korpusannotationen eine Reihe weiterer Vorteile bringt, soll hier nur angedeutet werden.

Manche Information älterer und aus definierter Wörterbuchbasis heraus entstandener Wörterbücher erweist sich als korrekturbedürftig, wo neue Quellenmassen zugänglich werden. In der historisch-diachronen Lexikografie spielt(e) der ominöse „Erstbeleg“ eine ungehörlich große Rolle und führte bei Nutzern nicht selten zu der irrigen Vorstellung, etwa Luther oder Goethe habe ein bestimmtes Wort ‚geprägt‘ (weil die Schriften dieser beiden in fast allen deutschsprachigen Wörterbuchbasen des 19. und 20. Jahrhunderts überproportional großen Raum einnahmen). Die zunehmende Digitalisierung historischer Texte und ihre Verfügbarkeit im Internet bringt nun immer öfter frühere ‚Erstbelege‘ zum Vorschein, so dass die Vorstellung der individuellen sprachgewaltigen Worterfinder zunehmend relativiert werden kann.

in diesem Band), das die ursprüngliche Wörterbuchbasis eines Zettelkatalogs seit etlichen Jahren um Belege aus den Mannheimer Korpora und neuerdings auch aus dem Internet, vor allem den dort zugänglichen digitalisierten historischen Quellen bereichert.

Dennoch ist die Verletzung der o. g. wissenschaftlichen Grundprinzipien mit den zuletzt genannten Vorteilen nicht kompensiert – die beiden Aspekte sind zu unterschiedlich. Eine Möglichkeit, wie man die Offenlegung der lexikologisch wesentlichen Eigenschaften einer Wörterbuchbasis mit ihrer dynamischen Unabgeschlossenheit verbinden kann, ist der Ausbau bzw. die Weiterentwicklung des Annotationspotenzials der jeweiligen Einheiten eines Korpus (Texte, Teiltex te, Sätze) nach allen interessierenden (sozio-, historio-, textlinguistischen usw.) Merkmalen. Wenn es gelänge, jede Korpuseinheit in dem Moment automatisch zu annotieren, in dem sie Teil der dynamischen Wörterbuchbasis wird, ließe sich im Prinzip jede einzelne Angabe in jedem Wortartikel mit einer Matrix verlinken, aus der sich etwa die prozentuale Verteilung von Jahrzehnten, Varietäten u.v.m. in dem der Angabe zugrundeliegenden Korpus ablesen lässt. Daran ließen sich weitere Vorteile anschließen, von denen der Vergleich verschiedener Korpusschnitte und daraus resultierender Angaben nur einen darstellt.

2.2 Fokus Zugriffsstrukturen

In der Internet-Lexikografie ist immer wieder vom „Mehrwert“ die Rede⁴, die das neue Medium in diversen Aspekten gegenüber dem alten mit sich bringt. Die damit gemeinten vielfältigeren Nutzungsmöglichkeiten digitaler Nachschlagewerke sind, so zeigen nicht nur die Beiträge dieses Bands (v.a. Mann, Lemnitzer, Mihm, Klein/Geyken), ein deutlich erkennbarer Schwerpunkt des metalexikografischen Interesses. Es sind lexikografisch gesprochen die „Zugriffsstrukturen“ und die Meso- (Verlinkungs)-Strukturen, die bzw. deren Optimierung extensiv wie intensiv diskutiert und auch kritisiert werden. Die Frage, wie man in einem Online-Nachschlagewerk suchen kann – fehlertolerant, trunkiert, kombiniert, mittels Boolescher Operatoren, mono- oder multimodal, ob ‚nur‘ Lemmata angesteuert werden können oder gar einzelne Positionen in der Mikrostruktur – diesem Komplex von Fragen scheint ein Großteil des aktuellen lexikografischen Forschungsinteresses gewidmet. Und diese Frage steht aus der Nutzerperspektive in der Tat ganz im Vordergrund, denn hier liegen der Zugang, der Schlüssel und auch die Schwelle, über die NutzerInnen überhaupt in den Informationsraum hinein gelangen.

Manch einer (so Mann in diesem Band) geht gar so weit, das Vorhandensein makrostruktureller Elemente, also z.B. Lemma(-teil-)strecken, aus denen Nutzer das Gesuchte scrollend auswählen, als „printorientierte, wenig innovative Darstellung“, also als Relikt einer überholten Epoche zu bewerten. Doch wird hier wohl etwas Wichtiges und Erhaltenswertes übersehen: Beispielsweise enthält „elexiko“ (in www.owid.de) neben den o.g. neuen Zugriffsmöglichkeiten auch eine „Stichwortliste“ zum Anklicken, die in sich ‚portioniert‘ ist. Dass die Strecke von *Halbglatze* bis *Hallendach* 300 Lemmata mit Wortartikeln umfasst, wird hier zu einem sogar sinnlich, nämlich durch mehrfaches Scrollen als ‚große Menge‘ erfahrbaren Merkmal des Wörterbuchs. Die Menge der rechtserweiternden Komposita mit *Halb/halb* lässt einen staunend den Umfang des Wörterbuchs ahnen und regt zu allen mög-

⁴ Es wäre interessant zu untersuchen, ob dieser Mehrwert in Anlehnung an Marx oder eher in Analogie zum Marketinginstrument der Kundenbindung durch Mehrwert oder in völlig neuer Weise verstanden wird.

lichen Fragen an.⁵ Dass es zu *Halbzeit* allein zehn Rechtserweiterungen gibt (und bei diesem Wörterbuch weiß man, dass es nicht okkasionelle, sondern durchaus gebräuchliche Bildungen als Lemma ansetzt), lässt nicht zu Unrecht den Schluss zu, dass das Thema Fußball im zugrundeliegenden Korpus eine nicht gerade kleine Rolle spielen wird. Das Anklicken der Strecke ab *Fußball* bestätigt diesen Verdacht eindrucksvoll. Entdeckungen wie die genannten erlauben gedruckte Wörterbücher auch, aber im Internet werden Art und Umfang von Lemmastrecken sogar noch deutlicher erkennbar.

Die Darstellung der Makrostruktur ist und bleibt eine wichtige Informationsquelle, die Laien schlicht Spaß machen kann, die großes (sprach-)didaktisches Potenzial besitzt (Schulklassen stellen ein Fußballwörterbuch zusammen, Studierende ein Wirtschaftswörterbuch.) und die nicht zuletzt wissenschaftlichen Wert hat.

Die neue Vielfalt der Suchmöglichkeiten im Wörterbuch wird zu Recht aus Nutzersicht gelobt, ihre Optimierung aus Nutzersicht diskutiert. Doch scheint auch hier ein wenig Skepsis angebracht. Das Spektrum der Suchmöglichkeiten hat sich seit Beginn des Internets kaum verändert, denn es beruht auf den Möglichkeiten von Mathematik, Logik und Informatik, die in allen Internet-Suchmaschinen mehr oder weniger gleich funktionieren. Nutzerinteressen, wie die schreibungstolerante Suche, die den Nutzern ja Rechtschreibschwäche und/oder mangelnde Sorgfalt unterstellt, wurden den gegebenen maschinellen Möglichkeiten wohl eher nachträglich zugeordnet. Dass sich die inzwischen bei den Suchmaschinen herausgebildeten Standards von Trunkierung und Suche mittels Boolescher Operatoren auch in der Internet-Lexikografie weithin durchgesetzt haben, ist allerdings ein großer Vorteil.

Das Thema Zugriffsstrukturen im Internet-Wörterbuch hängt auch mit den Möglichkeiten der Vernetzung durch Hyperlinks zusammen. Hyperlinks stellen gewissermaßen eine zweite oder sekundäre Zugriffsebene auf Angaben dar, die sich in der Regel nicht von einer Makro-, sondern von einer Mikrostruktur aus eröffnet. Eines der Probleme, die Nutzer mit gedruckten Wörterbüchern haben, liegt in einer komplexen und deshalb von vielen für ‚unübersichtlich‘ gehaltenen Mikrostruktur; dies gilt umso mehr, je schneller der Informationsbedarf gedeckt werden soll. Wo nun elektronische Wörterbücher ihre ‚Manuskripte‘ in Form einer Datenbank anlegen, wird es möglich, die Mikrostruktur zu modularisieren und NutzerInnen die Informationen nicht nur übersichtlicher, sondern gerade auch zur selektiven Rezeption anzubieten, etwa nach dem Motto: ‚Wozu semantische Angaben lesen, wenn man nur nach einer etymologischen (orthografischen, grammatischen usw.) Information sucht?‘ Noch sind Internet-Wörterbücher mit selektiv rezipierbarer Mikrostruktur selten, denn ihre Herstellung erfordert von Grund auf andere konzeptionelle Überlegungen sowie Arbeitsweisen.⁶

Die zusätzliche Attraktivität, die ein Nachschlagewerk durch Verknüpfungen zwischen Wortartikeln bzw. zwischen Elementen der Mikrostruktur erhält, die weit über das traditionelle Verweissystem hinausgehen, ist unbestreitbar. Ganz besonders in diesem Punkt hat sich für die Lexikografie durch das Internet eine neue Welt aufgetan, die dann aber auch einer wissenschaftlichen, kritischen Reflexion bedarf: Welche Relationen, linguistisch gesprochen, werden mit Hyperlinks realisiert, und welche Informationen oder Verständ-

⁵ Sind Wörterbücher ausschließlich dazu da, Fragen zu beantworten? Nein, sie sollten auch zu Fragen anregen, solange sie doch auch in irgendeinem Zusammenhang mit Bildung stehen.

⁶ Zur Modularität in der Internet-Lexikographie vgl. die Konzeption des IDS-Projekts *elexico* (unter www.owid.de), dargelegt in Haß (Hg.) 2005; ähnlich Klein/Geyken in diesem Band.

nishilfen erhalten Nutzer über die Art der jeweiligen Relation? Vernetzung ist ‚immer gut‘, aber sie ist offensichtlich auch komplex und stellt hohe Anforderungen an die Nutzer.

Es kann wohl nicht bestritten werden, dass die Vielfalt der elektronischen Verweise nicht aus einem artikulierten oder sonstwie erkennbaren Bedarf der Nutzer heraus entstanden ist. Vielmehr wird bei den Links derzeit überwiegend das umgesetzt, was Maschinen automatisch erzeugen können, d.h. dass Links vor allem von formativgebundenen, ausdrucksseitigen Merkmalen von Lexemen abhängen (vgl. Haß 2005). Wortbildungsmorphologische und orthografische Aspekte sowie Kollokationsrelationen der in einem Wörterbuch behandelten Lexik rücken damit automatisch viel weiter in den Vordergrund der dargestellten Vernetzung, als es in klassischen Bedeutungswörterbüchern der Fall ist. Bekanntlich ist es sehr schwer oder kaum möglich, von Wortbildungsregeln bzw. Kollokationsbeziehungen auf semantische Regeln zu schließen. Lexikalisierung und Idiomatisierung lassen regelbasierte Schlüsse (etwa von *machen* auf *sich breit machen*) nicht zu, doch wer klärt die Nutzer darüber auf?

Um das Gemeinte nur kurz an einem aktuellen Beispiel zu veranschaulichen⁷, sei ein Blick in das vor einigen Jahren noch indiskutable deutschsprachige Wiktionary geworfen. In dieses Internet-Wörterbuch muss mittlerweile einige intellektuelle und lexikografisch fundierte Arbeit eingeflossen sein; dennoch hat man den Eindruck, dass unter der Rubrik „Unterbegriffe“ ausschließlich linkerweiternde Komposita (s.v. *Schule: Abendschule Tanzschule*) aufgezählt werden, die automatisch erzeugt werden können. Unter der Rubrik „Abgeleitete Begriffe“ scheinen sich nahezu ausschließlich rechterweiternde Lexeme zu finden (<http://de.wiktionary.org/wiki/Schule>).⁸ Die laiensprachliche Vagheit des Ausdrucks *Begriff* legt hier eine semantische Ableitungsrelation fatal nahe.

Ein Schritt in die richtige Richtung ist hier sicherlich die Praxis des DWDS, wo den entsprechenden Links die Überschrift „Automatisch berechnete semantische Relationen aus dem WDG“ bzw. „Automatisch berechnete Kollokationen aus dem DWDS Kerncorpus“ vorangestellt wird. Die Angabe wird also relativiert, doch ist es mehr als fraglich, ob nicht-(korpus-)linguistisch versierte Nutzer diese Relativierung auch als solche verstehen.

Ein anderer und deutlich innovativerer Typ von Vernetzung zeigt sich in der Integration verschiedener Makrostrukturen unter einer Suchmaske. Die aktuellen Bestrebungen, einzelne Nachschlagewerke (oft Digitalisate gedruckter Wörterbücher und Lexika) in Verbünden und über gemeinsame Portale und Suchmasken zugänglich zu machen, stellt ein herausragendes Merkmal der gegenwärtigen lexikografischen Praxis im Internet dar (dazu Mihm sowie Klein/Geyken in diesem Band). Die dadurch entstehenden neuen Möglichkeiten der Informationsgewinnung sind sicherlich noch längst nicht ausgeschöpft, auch wenn man sie wohl eher oder gar ausschließlich auf Expertenebene suchen muss. Nutzer erhalten viele, möglicherweise auch zu viele kommentarlos nach einander gelistete Antworten auf die Frage nach einem eingegebenen Stichwort. Was haben die Einträge miteinander zu tun, in welchem Verhältnis stehen sie zu einander? Ein Problem kann besonders dann entstehen, wenn die vernetzten Wörterbücher ganz unterschiedliche Varietäten beschreiben, etwa den mittelhochdeutschen Wortschatz und den Goethewortschatz, den trotz seines Umfangs sehr selektiven Wortschatz des Grimmschen Wörterbuchs und den Wortschatz des WDG. Da in

⁷ In Haß 2005 wird das Gesagte breiter exemplifiziert.

⁸ Noch deutlicher s.v. *Hund*: Als Hyponyme werden nur Komposita mit *-hund* als rechtem Element aufgeführt; von *Pudel* und *Dackel* keine Spur. <http://de.wiktionary.org/wiki/Hund>

den letzten Jahren erfreulicherweise etliche historische Wörterbücher und Sachlexika bzw. Enzyklopädien digitalisiert und einer gezielten Abfrage zugänglich gemacht wurden – die urheberrechtlich weniger problematische Lage bei historischen Werken mag hierbei eine Rolle gespielt haben –, ist das Problem sogar besonders virulent. Für diejenigen NutzerInnen, die die konzeptionellen Besonderheiten der Werke und die lexikologischen Merkmale des beschriebenen Wortschatzes bereits kennen – und das betrifft auch unter den LinguistInnen nur eine spezialisierte Gruppe –, eröffnen sich mit diesen Verbünden fantastische Möglichkeiten. ‚Normale‘ Germanistik-Studierende jedoch haben gewaltige Schwierigkeiten mit dem Verständnis nicht nur der dargebotenen Informationsfülle, sondern erst recht mit den vielen zusätzlichen, in ihrer Art nicht identifizierten Verweisen. Manche wehren ab: „Das will ich doch alles gar nicht wissen“.

Eine Aufgabe der nächsten Jahre wird es daher sein, die Quantität der Vernetzung qualitativ zu bereichern und die dargebotenen Klassifikationen von Hyperlinks (z.B. „Unterbegriffe von“) linguistisch besser oder überhaupt erst zu fundieren und sie den NutzerInnen auch angemessen zu erläutern. Die Möglichkeiten dafür zeichnen sich bereits ab: So genannte Ontologien und semantische Thesauri eröffnen Wege, semantische Relationen zwischen Lexemen unabhängig von ausdrucksseitigen Ähnlichkeiten zu erstellen und in Hyperlinks zu überführen. Doch solange das Material für Ontologien seinerseits, was oft der Fall ist, aus Wörterbüchern stammt, die selbst nicht korpusbasiert sind, ist Vorsicht geboten. Streng genommen müsste eine Ontologie aus demselben Korpus heraus entwickelt werden, das die Basis des betreffenden Wörterbuchs darstellt.

2.3 Für eine neue Offensive zugunsten der linguistisch-lexikografischen Qualität

Zusammengefasst lässt sich festhalten, dass der Einsatz korpusanalytischer Verfahren, die Realisierung ausdifferenzierter Zugriffsstrukturen und die ebenso ausdifferenzierten Möglichkeiten der Verlinkung, d.h. die ‚Vernetzung‘, gegenwärtig sowohl die Praxis als auch die wissenschaftliche Reflexion (‚Metalexikografie‘) der Internetlexikografie prägen. Es sind also die innovativen Merkmale der Lexikografie im Internet, die im Fokus der Aufmerksamkeit stehen. Es scheint, als seien die metalexikografischen Diskussionen der Epoche der gedruckten Nachschlagewerke uninteressant geworden oder zu einem guten und endgültigen Abschluss gekommen. Die 1980er und 1990er Jahre waren von Diskussionen geprägt, die sich exemplarisch in den Artikeln des HSK-Bandes ‚Wörterbücher – Dictionaries – Dictionnaires (1989, 1990, 1991), in der Reihe ‚Studien zur neuhochdeutschen Lexikographie‘ und in der Reihe ‚Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger Lexikographischen Kolloquium‘ (ab 1989) finden. Ohne das gesamte Spektrum der hier verhandelten Themen wiederholen zu wollen, sei doch daran erinnert, was ‚seinerzeit‘ im Fokus der lexikografischen Diskussionen stand: Die Frage, wie jede einzelne der lexikografischen Informationsarten sowie ihre Anordnung innerhalb der Mikrostruktur dem Stand des sprachwissenschaftlichen Wissens gemäß, das strukturalistische, pragmlinguistische und kognitionslinguistische Ansätze integrierte, und den Bedürfnissen der Nutzer folgend zu gestalten sei. Hinzu kam ein starkes Interesse am sprachlich-sozialen Kontext der produktiven wie der rezeptiven Seite der Lexikografie; hierzu gehörte auch das Interesse an Lexikografiegeschichte.

Es wird niemand bestreiten wollen, dass all diese Fragen nicht wirklich abschließend beantwortet, dass noch wissenschaftlicher Klärungsbedarf besteht und nicht alle Nutzerbe-

dürfnisse befriedigt sind. Dennoch spielen sie in der aktuellen Internet-Lexikografie anscheinend keine Rolle – der Fokus hat sich deutlich verlagert. Einen optimistischen Blick in die Zukunft zu werfen, bedeutet, nahezu alle klassischen Fragen und Probleme der Lexikografie mit Bezug auf die neuen Möglichkeiten des Internets und der Sprachtechnologie neu zu formulieren und zu diskutieren. Nur wenige Punkte hierzu seien genannt:

Wörterbuchtypologie – dieses Kapitel wird in absehbarer Zeit ganz neu zu schreiben sein. Einer immer größer werdenden Ausdifferenzierung von Wörterbucharten und Wörterbuchmerkmalen steht die Tendenz zum Vielzweck-Wörterbuch gegenüber, das sprachliche und enzyklopädische Informationen über ein und dieselbe Suchmaske anbietet; und genauso könnten Produktions- und Rezeptionswörterbücher, spezial- und allgemeinsprachliche, historisch-diachrone und gegenwartsbezogene, einsprachige und mehrsprachige Wörterbücher konvergieren.

- Die Kontextualisierung eines Internet-Wörterbuchs ist nicht mehr allein gesellschaftlicher und (bildungs-)historischer Natur, sondern hat jetzt auch eine weitere, mediale Dimension, denn neue Kontexte sind nur einen Mausklick weit entfernt: Verbünde, Portale, Konzern-Webseiten, öffentliche, private usw. usf., aber auch Ressourcen und Datensammlungen anderer Art (Korpora, Terminologiedatenbanken, Fachdatenbanken). Welche Folgen hat es für Nutzer, wenn sie den ‚Übertritt‘ vom Wörterbuch A zur Datensammlung B nicht mehr deutlich genug erkennen? Dies ist im Sinne einer kritischen Mediennutzung nicht unerheblich.
- Wörterbuchvorworte und Umtex te werden transformiert in „Hilfe“-Texte einerseits und Verlinkung mit Projektwebseiten andererseits. Angaben zu den konzeptionellen Eckdaten, z.B. zum Veröffentlichungsdatum, zur Wörterbuchbasis und anderem scheinen bei Internet-Wörterbüchern schwerer auffindbar als bei gedruckten.
- Probleme der Lemmaselektion stellen sich im Zusammenhang mit offenen, dynamischen Wörterbuchkorpora (s.o.) in neuer Weise.
- Erstaunlicherweise sind Formen der Textverdichtung aus Internet-Wörterbüchern keineswegs verschwunden. Handelt es sich hierbei um Relikte oder um Textsortenmerkmale, die mit typografischen Strukturanzeigern zusammen für schnelle Rezipierbarkeit und den Eindruck von ‚Übersichtlichkeit‘ sorgen?
- Die „Theorie des lexikografischen Beispiels“ ist neu zu schreiben, wo automatische Korpusbelege bzw. die Verknüpfung mit Korpusdaten wie KWICs und Kookkurrenzlisten eingesetzt werden und wo Druckraumbegrenzung entfällt. Beispiele und Belege können gruppiert und diversen Funktionen (nutzungssituationsspezifisch etwas veranschaulichen, etwas nachweisen, etwas Neues zeigen usw.) zugeordnet werden.
- Die Darstellung der Wortbildungszusammenhänge, der semantischen Relationen und der syntagmatischen Verbindungen (Phraseologismen, Kollokationen) sollte, wie oben gezeigt, auf Dauer nicht der Maschine überlassen bleiben.
- Verändertes Nutzerverhalten und andere Nutzerbedürfnisse werden die Diskussionen um die seit langem zentralen Fragen nach einer ‚guten‘ Bedeutungserläuterung und nach der angemessenen Kombination der Bedeutungserläuterung mit den anderen Informationsarten neu beleben, ohne dass hierbei wesentlich Neues erfunden werden müsste.

3 Mehrsprachigkeit: Kleine und große Sprachen

Der vorliegende Band soll auch einen Eindruck davon geben, in welchem Ausmaß und auf welche Weise die Internet-Lexikografie mehrsprachig ist bzw. auf Mehrsprachigkeit Bezug nimmt. Dabei scheint sich einerseits eine Tendenz zu bestätigen, die schon für die Welt der gedruckten Wörterbücher gilt: Große Sprachen dominieren, allen voran das Englische. Bei zwei- und mehrsprachigen Internet-Wörterbüchern scheint Englisch geradezu weltweite Referenzsprache zu sein. Auf der anderen Seite stehen die Chancen kleinerer Sprachen auf lexikografische Repräsentation im Internet besser als je zuvor, wenn sich enthusiastische Freiwillige oder auch öffentliche Stellen finden, die bereit sind, in lexikografische Arbeiten zu investieren. Dies ist regelmäßig dort der Fall, wo Ausbau und Sicherung einer Nationalsprache in einem engen Zusammenhang mit staatlicher Souveränität und nationaler Identitätsbildung gesehen wird, wie dies für viele ehemalige Kolonialländer und für die Länder der ehemaligen Sowjetunion der Fall ist (vgl. die Beiträge von Katinas und Prinsloo in diesem Band). Nicht zufällig lässt sich beobachten, dass Länder wie Litauen, Lettland, Estland oder auch Tschechien und die Slowakei Wörterbucherstellung von vornherein mit dem Aufbau eines nationalsprachlichen Korpus verbinden. Diese Aktivitäten sind überwiegend ‚nach innen‘ in die eigene Sprachgemeinschaft gerichtet und in der internationalen lexikografischen Fachwelt schwer wahrzunehmen.

Weniger gut ist die Repräsentation kleinerer Sprachen in mehrsprachigen Internet-Wörterbüchern. Die Europäische Union fördert die europäischen „less widely used less taught“- oder LWULT-Sprachen besonders, denn es handelt sich bei ihnen um gleichberechtigte Amtssprachen der EU. Das mehrsprachige Internet-Wörterbuch „Interactive Terminology for Europe“ (<http://iate.europa.eu>) bezeichnet sich zwar selbst als Terminologie, lässt sich aber durchaus als 22 Sprachen abdeckendes Wörterbuch zu allen irgendwie europarelevanten Themen begreifen.⁹ Es scheint rein maschinell aus Textkorpora aller 22 europäischen Sprachen erstellt und präsentiert parallele Syntagmen semantisch disambiguiert. Doch erkennt man rasch, dass die zugrundeliegenden Textkorpora zu Deutsch, Englisch, Französisch und auch Italienisch weit umfangreicher sein müssen als etwa diejenigen zu Lettisch, Litauisch, Estnisch und weiteren LWULT-Sprachen, sodass viele Suchanfragen keine Einträge zu diesen Sprachen anzeigen. Das Beispiel zeigt, dass hinter einer geringen lexikografischen Repräsentation kleinerer Sprachen ein grundlegendes Problem liegen kann: Sprachausbau und entsprechende Textproduktion in all den öffentlichen Domänen, die für die großen Sprachen selbstverständlich sind, die einige der kleineren sich aber erst ‚erobern‘ müssen.

⁹ Unter den deutschen Quellen befindet sich z.B. das Metzler-Lexikon Sprache, hg. von H. Glück.

4 Nutzersicht: mediale Wahrnehmungen

Die Wörterbuchbenutzungsforschung ist einer der wenigen lexikografischen Bereiche, in denen das Internet ganz neue Konzeptionen erfordert, und wäre wohl ein gesondertes Lexicographica-Thema wert. Nur Weniges soll hier dazu gesagt werden.

Man kann zu Recht annehmen, dass Internet-Wörterbücher und -Lexika von viel mehr Menschen (gelegentlich) genutzt werden als gedruckte Wörterbücher vor Etablierung des Internets. Auch wird ein Individuum im Internet mehr als das eine Wörterbuch oder Lexikon nutzen, das meist einsam in seinem Bücherregal steht. Der Zugang scheint niederschwellig und schnell, doch weiß die Nutzerforschung sicher noch zu wenig darüber, wie schnell und warum eine lexikografische Webseite wieder verlassen wird. Der Umgang mit Studierenden lehrt, dass ihnen die Unterscheidung zwischen Laien- und Qualitätswörterbüchern und das Erkennen wissenschaftlicher Seriosität bzw. die Beurteilung der Verlässlichkeit der Informationen schwerfällt und dass sie gelernt werden muss.

Dass Wörterbücher dereinst etliche Regalmeter in Bibliotheken füllten, wie schwer ein einzelner Band in der Hand wiegt (ähnlich wie das großstädtische Telefonbuch früherer Zeiten), wie eine Druckseite gestaltet ist und dass es erläuternde Vorworte oder Nutzerhinweise gibt, das alles sind für die meisten jüngeren Internet-Nutzer mehr oder weniger unbekannte Erfahrungen. Insbesondere der physische Eindruck hat Auswirkungen auf Erwartungen an den Inhalt; dies fehlt dem Internet-Wörterbuch.

Ganzheitlich oder fragmentarisch ist auch die Wahrnehmung auf Wortartikel-Ebene, denn Bildschirme sind eher klein (Tablet-PCs, Palms und iPhones), Inhalte modular aufgeteilt und dadurch u.U. unübersichtlicher als im gedruckten Wörterbuch. Andererseits bilden sich bei denen, die mit dem Internet aufgewachsen sind, ganz andere Rezeptionshaltungen und -muster heraus, als sie für den Umgang mit gedruckten Wörterbüchern üblich waren und auch dort noch immer nötig sind, wo gedruckte Wörterbücher als Digitalisate im Internet angeboten werden.

Der Erwerb eines gewissen Know-hows auf Seiten der Nutzer ist unerlässlich; neben einem medienunabhängigen Anteil an Wissen über Funktionen und Strukturen von Wörterbuchinformationen im Allgemeinen geht es auch um medienspezifische oder gar produktspezifische Fähigkeiten. Insbesondere will es gelernt sein, Kohäsion und Kohärenz im Wörterbuch zu erkennen bzw. durch angemessene Nutzung zu erzeugen, Nützliche und unnütze Zufallsfunde gibt es beim Blättern ebenso wie beim Surfen (serendipity). Im Buch wird der Bereich des Sinn-Freien durch die Ordnung des Alphabets gewährleistet, das etwa *Glibber* mit *Gräzist* (eine der Bandgrenzen im Grimmschen Wörterbuch) zu verknüpfen vermag. Umso bedauerlicher wäre das völlige Verschwinden makrostruktureller Zugriffsmöglichkeiten über Stichwortlisten im Internet. Heiterkeit erregen in heutigen Internet-Wörterbüchern aus den o.g. Gründen manche rein maschinell erzeugten Links, aber das ist ja nicht gewollt. Je perfekter die internen und externen Verlinkungen in Internet-Wörterbüchern werden, desto weniger wird man dort zu lachen haben.

In Zukunft wird es stark darauf ankommen, dass neue technische Entwicklungen genutzt werden, doch weder Wörterbuchproduzenten noch Wörterbuchnutzer sich davon abhängig machen oder ihnen gar hinterherlaufen. Nicht jedes Wörterbuch muss iPhone-, iPod- oder iPad-fähig gemacht werden, nicht jedes muss GPS-gerecht auf dem Navi abgerufen werden können – und was sonst noch alles denkbar scheint oder noch undenkbar ist. Wörterbücher

sollen nicht vorführen, was technisch alles möglich ist, sondern sinnvolle Nutzungen mit technischen Möglichkeiten verbessern.

Die Umsetzung aller genannten Weiterentwicklungen erfordert auch und gerade in der Internet-Lexikografie erheblich mehr Zeit und Personal, als in den bekannten nationalen und internationalen Förderprogrammen pro Projekt vorgesehen ist. Gute Wörterbücher entstehen nicht automatisch, sondern werden von Expertinnen und Experten im Grunde immer noch ‚geschrieben‘, auch wenn ihre Arbeit in vielfacher Weise computerunterstützt ist. Die oft als bloßes Handwerk unterschätzte Lexikografie hat es bei Fördermittelgebern bzw. bei der Budgetverteilung besonders schwer. Verlage scheinen sich in denjenigen Ländern aus der Internet-Lexikografie verabschiedet zu haben, in denen sie mit kostenpflichtigen Angeboten nicht neben den kostenfreien Angeboten reüssieren konnten. Leider würden konzentrierte Finanzierungsaktionen mehrerer Fördermittelgeber vermutlich ein einzelnes Wörterbuchprojekt je Land bzw. Nationalsprache monopolisieren statt Vielfalt zu ermöglichen.

5 Überblick über den Inhalt des folgenden Thementails

Die folgenden acht Beiträge eröffnen mit einer aktuellen Lagebeschreibung von *Michael Mann* (Universität Erlangen). In einer systematischen Untersuchung vergleicht er klar definierte Merkmale von Struktur, Multimodalität, Datenpräsentation, Datenzugriff und ‚Interaktivität‘ von 88 nach transparenten Kriterien ausgewählten besonders beliebten ein- und mehrsprachigen elektronischen Wörterbüchern (insbesondere Fachwörterbüchern) unterschiedlichster Art, doch mit einem Schwerpunkt auf Englisch als einer beteiligten Sprache im Sommer 2009. Aus identifizierten Eigenschaften und Mängeln leiten sich Wünsche für die Zukunft ab, nämlich differenziertere und mächtigere Zugriffsstrukturen und -möglichkeiten sowie stärkere Nutzung dynamischer Präsentationsformen und multimodaler Inhalte.

Am Beispiel des Projektes „Deutscher Wortschatz“ (wortschatz.uni-leipzig.de) erörtert *Uwe Quasthoff* (Universität Leipzig) Schritt für Schritt, wie Rohdaten aus maschinell verfügbaren großen Korpora (hier mit mehreren Hundert Millionen Sätzen) für elektronische Wörterbücher aufbereitet und ausgewertet werden können, und zwar mit vorwiegend automatischen Mitteln. Über statistische Berechnungen ist es dabei möglich, Unterschiede zwischen verschiedenen Textgenres zu erkennen sowie typologische Abweichungen zu messen.

Lothar Lemnitzer (Berlin-Brandenburgische Akademie der Wissenschaften Berlin) legt dar, in welchem Maße die zeitgenössische Neologismenlexikographie von der Nutzung des Internet als Ressourcenquelle und als Publikationsplattform profitiert. Am Schluss des Beitrages zeigt er, wie vor allem durch die Vernetzung der vorhandenen lexikographischen Projekte der Wert dieser Ressourcen für die Benutzer noch erhöht werden kann.

Wolfgang Klein (Max Planck Institut für Psycholinguistik Nijmegen) und *Alexander Geyken* (Berlin-Brandenburgische Akademie der Wissenschaften Berlin) erläutern Hintergrund, Ziele und derzeitigen Stand des Projekts „Das Digitale Wörterbuch der Deutschen Sprache (DWDS)“. Von 2007 bis 2025 sollen die Daten des Grimmschen Wörterbuchs und seiner Neubearbeitung sowie des *Wörterbuchs der deutschen Gegenwartssprache* (WDG) und des *Etymologischen Wörterbuchs des Deutschen* sowie eines neu erstellten umfangrei-

chen Korpus der deutschen Gegenwartssprache in einem einheitlichen digitalen Wörterbuchsystem zusammengeführt werden.

Hildegard Mihm (Universität Duisburg-Essen) rezensiert das „Wörterbuch-Netz“ (germazope.uni-trier.de/Projects/WBB) aus Nutzersicht. Sie stellt dessen Gesamtstruktur sowie den Aufbau der darin verbundenen einzelnen Wörterbücher vor, würdigt Leistungen und Grenzen kritisch, erörtert die Mehrwert-Frage und formuliert Verbesserungswünsche.

Auf empirischer Grundlage erörtern *Larissa M. Bieler* und *Maren Runte* (Zürcher Hochschule für Angewandte Wissenschaften) lexikografische Probleme der Bedeutungsbeschreibung von Geschmacksadjektiven.

Bruno Strecker (Institut für Deutsche Sprache Mannheim) stellt das Glossar grammatischer Fachbegriffe im grammatischen Informationssystem „grammis“ (hypermedia.ids-mannheim.de/grammis) vor.

Bernhard Fisseni (Universität Duisburg-Essen) berichtet über seine Erfahrungen als linguistischer Experten-Nutzer mit diesem Glossar.

Der anschließende Dokumentationsteil gibt einen Überblick über wichtige internetlexikographische Adressen, sortiert nach Regionen. *Vincent Ooi* (National University of Singapore) stellt englischsprachige Angebote vor, *Angelika Storrer* (Universität Dortmund) deutschsprachige, *Kevin Masalon* (Universität Duisburg-Essen) französisch- und italienischsprachige, *Bernhard Pöll* (Universität Salzburg) die Iberoromania. *Vladimir Bokov* und *Daria Bobrovskaya* (Barnaul bzw. Samara, Russland) skizzieren russischsprachige Adressen, *Daumantas Katinas* (Vilniaus Universitetas) das Baltikum und die übrige slawische Welt. *Danie Prinsloo* (University of Pretoria, Department of African Languages) widmet sich afrikanischen Internet-Wörterbüchern, *Hitoshi Yamashita* (Osaka University) japanischen und *Robert Blust* (University of Hawai'i) austronesischen. Dass nicht alle Gebiete der Welt (z.B. China, Skandinavien, die arabischen Länder) abgedeckt werden, war kurzfristig nicht zu ändern. Deshalb und der Vollständigkeit halber werden abschließend einige Linklisten zur Lexikographie im Internet genannt, mit deren Hilfe wohl das gesamte relevante weltweite Angebot recherchiert werden kann.

6 Literatur

- Bergentholtz et al. (eds.) 2009 = Bergenholtz, Henning/ Nielsen, Sandro/ Tarp, Sven (eds.): *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Bern et al.: Lang, 2009.
- Deleuze/Guattari 1977 = Deleuze, Gilles/ Guattari, Félix: *Rhizom* [frz. 1976]. Berlin: Merve, 1977.
- Kunze/Lemnitzer 2007 = Kunze, Claudia/ Lemnitzer, Lothar: *Computerlexikographie. Eine Einführung*. Tübingen: Narr, 2007. (E-Book) (accompanying website www.lemnitzer.de/lothar/CoLex).
- Gouws et al. (eds.) 2010 = Gouws, Rufus H./ Heid, Ulrich/ Schweickard, Wolfgang/ Wiegand, Herbert Ernst (eds.): *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography (HSK 15.4)*. Berlin, New York: de Gruyter, 2010 [forthcoming].
- Haß 2005 = Haß, Ulrike: Verlinkungen in Internet-„Lexika“ und Begriffsrelationen. In: *Lexicographica* 21/2005, 108–121
- Haß (Hg.) 2005 = Haß, Ulrike (Hg.): *Grundfragen der elektronischen Lexikographie. elexiko – das Online-Informationssystem zum deutschen Wortschatz*, Berlin (de Gruyter) 2005.

Lexikographie im Internet 2010 – Einleitung

Haß, Ulrike; Schmitz, Ulrich

Dieser Text wird über DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt.

Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

DOI: <https://doi.org/10.1515/9783110223231.1.1>.

URN: <urn:nbn:de:hbz:464-20190408-172642-3>

Link: <https://duepublico.uni-duisburg-essen.de:443/servlets/DocumentServlet?id=48554>

Rechtliche Vermerke:

Ulrike Haß dankt dem De Gruyter Verlag für die freundliche Genehmigung, diesen Text online veröffentlichen zu dürfen. Zudem dankt Frau Haß den Herausgebern für die Aufnahme des Beitrags in die Zeitschrift.

Quelle: Erschienen in: Lexicographica. International annual for lexicography. Bd. 26 (2009). S. 1-18.