

Pius ten Hacken

Wortbildung in elektronischen Lernerwörterbüchern

In diesem Beitrag werde ich versuchen, einige allgemeine Überlegungen zur Darstellung der Wortbildung in elektronischen Lernerwörterbüchern mit theoretischen Einsichten zum Wesen des Spracherwerbs, der Lexikographie sowie der Wortbildung zu verbinden.

1. Das Lernen von Sprachen

Lernerwörterbücher sind Wörterbücher, die speziell für Nicht-Muttersprachler entwickelt wurden. Gemäß Béjoint (2000, S. 66) wurde 1942 das erste Lernerwörterbuch im modernen Sinn publiziert, und zwar für Englischlerner in Japan. Seither hat sich das Genre des Lernerwörterbuchs stark entwickelt und verbreitet, speziell für Englisch (vgl. Bogaards 1996). Um das Konzept theoretisch einzubinden, sollten wir zuerst kurz die Frage behandeln, wie Lernende sich von Muttersprachlern unterscheiden.

1.1 Erstspracherwerb

Seit dem Aufkommen der generativen Grammatik hat der Spracherwerb eine zentrale Stellung in der Sprachwissenschaft erlangt. Obwohl Chomskys Theorie zum Erstspracherwerb (1981) nicht allgemein akzeptiert wird, hat sie dazu geführt, dass Linguisten, die diesen Vorschlag ablehnen, eine alternative Erklärung vorschlagen müssen.

Wie in Chomsky (1999) dargestellt, geht die generative Grammatik davon aus, dass sich die Sprache unter dem Einfluss genetischer und umgebungsbedingter Faktoren entwickelt. Ein Kind kommt auf die Welt mit einer Prädisposition, eine Sprache zu lernen. Diese Prädisposition wird 'Sprachinstinkt' (engl. *language faculty*) genannt. Sie ist ein Teil des Gehirns und entwickelt sich im Laufe des physischen und kognitiven Wachstums zu einem vollständig ausgeprägten Sprachmodul (engl. *competence*). Welche Sprache in diesem Sprachmodul erfasst wird, wird vom Input bedingt, dem das Kind im Laufe der Spracherwerbsperiode begegnet.

Eine wichtige Konsequenz dieser Sicht auf Spracherwerb ist, dass es keine unabhängig existierende Sprache als Objekt des Erwerbs gibt. Wie es Chomsky (1980, S. 134) ausdrückt, „in certain fundamental respects we do not re-

ally learn language; rather grammar grows in the mind“⁴. Das Kind hat keine Wahl, ob es eine Sprache lernen möchte oder nicht, und braucht keine spezielle Anstrengung zum Spracherwerb.

Die Möglichkeit, eine Sprache in dieser „natürlichen“ Weise als Muttersprache zu erwerben, nimmt mit zunehmendem Alter ab. Curtiss et al. (1974) beschreiben einen spektakulären Fall eines Kindes, Genie, das bis zum 13. Lebensjahr praktisch keinen sprachlichen Input hatte. Fälle wie Genie sind vergleichsweise selten. Viel häufiger kommt diese Art des fehlenden sprachlichen Inputs bei gehörlosen Kindern hörender Eltern vor (vgl. Mayberry 1994). Solche Kinder schaffen es nicht mehr, den Rückstand im Spracherwerb nachzuholen, wenn sie bis zum Ende der so genannten ‘kritischen Periode’ für den Spracherwerb keinen entsprechenden Input bekommen haben. Das Konzept der kritischen Periode hat eine neurologische Basis, die, wie von Hensch (2005) dargelegt, vor allem durch Experimente mit der Entwicklung des Sehens bei Katzen und Nagetieren erforscht wurde.

1.2 Zweitspracherwerb, Muttersprache und Norm

Wenn wir die generative Sicht auf Erstspracherwerb als Ausgangspunkt nehmen, um Lernerwörterbücher zu betrachten, so müssen zunächst einige Unterschiede klargestellt werden. Erstspracherwerb beruht auf der Fähigkeit des Kindes, unmoderierte sprachliche Äußerungen in der Umgebung zu benutzen, um Sprachkenntnisse aufzubauen. Lernerwörterbücher sind aber für Zweitspracherwerb gemeint, d.h. Spracherwerb nach der kritischen Periode. Es ist allgemein bekannt, dass Zweitspracherwerb alles andere als automatisch vor sich geht. Dies ist ein wichtiger Grund, wieso es überhaupt einen Bedarf an Lernerwörterbüchern gibt.

Eine Fremd- oder Zweitsprache wird im Unterschied zur Mutter- oder Erstsprache definiert. Der Begriff ‘Muttersprache’ ist aber keineswegs eindeutig. Franz Beckenbauer und Boris Becker haben beide Deutsch als Muttersprache. Das heißt, dass sie in einer deutschsprachigen Umgebung aufgewachsen sind. Die Bezeichnung ‘Deutsch’ wird hier in zweierlei Weise benutzt, einerseits als Klassifizierung der Sprachkenntnisse von Beckenbauer und Becker, andererseits als Klassifizierung des Sprachgebrauchs in der Umgebung, wo sie aufgewachsen sind. Chomsky (1965) benutzt „competence“ und „performance“ für diese zwei Sichten auf Sprache. Während Sprachkenntnisse und Sprachgebrauch empirische Entitäten darstellen, ist ‘Deutsch’ nur der Name einer nicht genau abgegrenzten Kategorie. Im Übergangsbereich zwischen Deutsch und Niederländisch gibt es beliebig viele Grenzfälle, wo es kaum möglich ist, ein-

wandfrei und ohne Willkür zu bestimmen, ob jemand einen deutschen oder niederländischen Dialekt spricht.

Der natürliche Spracherwerbsprozess, wie oben beschrieben, bezieht sich nur auf das Resultat der unmoderierten Verarbeitung der wahrgenommenen Umgebungssprache durch das Kind, nicht auf die normierte Schriftsprache. Die deutsche Standardsprache haben Beckenbauer und Becker in einer ähnlichen Weise gelernt, wie sie lesen und schreiben gelernt haben. Da ihre Muttersprache von der deutschen Standardsprache nicht weit entfernt ist, war der Lernaufwand zwar nicht groß, aber der Lernprozess war von einer ganz anderen Art als ihr Erstspracherwerb.

Zweitspracherwerb nach der Kindheit unterscheidet sich in mehreren Hinsichten vom Erstspracherwerb: Die Ausgangslage, der Prozess, der Aufwand, das Lernziel und der Erfolg sind unterschiedlich. Ein wesentlicher Unterschied ist, dass ein Lerner einer zweiten Sprache schon eine Sprache beherrscht. Dies ändert nicht nur die Ausgangslage, sondern auch den Lernprozess, der ja diese erste Sprache benutzen kann. Außerdem fällt der Zweitspracherwerb nicht in der gleichen Weise mit einer kognitiven Entwicklung zusammen wie der Erstspracherwerb. Trotz dieser offensichtlichen Vorteile im Zweitspracherwerb ist der Aufwand generell größer und der Erfolg kleiner. Viele Sprachforscher sehen dies als ein weiteres Argument für eine kritische Periode für Spracherwerb. Wie aus der Übersicht in ten Hacken (2007, S. 300-317) hervorgeht, besteht aber innerhalb der generativen Zweitsprachenerwerbsforschung keine Einigung darüber, wie genau die Erstsprache und der genetisch bedingte Sprachinstinkt im Zweitspracherwerb verfügbar sind bzw. eingesetzt werden.

Es ist in diesem Kontext, dass Lernerwörterbücher benutzt werden. Obwohl Lernende sich sehr unterschiedliche Lernziele setzen, ist es häufig die Standardform der Sprache, deren Kenntnis angestrebt wird. Deshalb spielt die Standardsprache eine direktere Rolle im Zweitspracherwerb als im Erstspracherwerb.

2. Wörterbücher und Sprachbeschreibung

Wörterbücher werden oft als eine Beschreibung der jeweiligen Sprache dargestellt. Mit Hinblick auf die oben angeführte Sicht auf Sprache ist eine solche Darstellung allerdings problematisch. Es wird hier zunächst erklärt, was genau das Problem ist (Kap. 2.1) und wie man zu einer alternativen Interpretation des Wörterbuchs kommen kann (Kap. 2.2). Dann werden die Spezialfälle der Lernerwörterbücher (Kap. 2.3) und der elektronischen Wörterbücher (Kap. 2.4) behandelt.

2.1 Sprache und Sprachbeschreibung

Sprache kann auf drei Ebenen betrachtet werden, auf der individuellen Ebene, der Ebene der Sprachgemeinschaft und der Ebene des Menschen als Spezies. Im letzten Fall geht es um die genetische Spezifikation. In der generativen Grammatik werden die individuelle und die genetische Ebene als Basis für die wissenschaftliche Beschreibung des Phänomens Sprache angenommen, während Sprache auf der Ebene der Sprachgemeinschaft, wie Chomsky (1980) argumentiert, nicht als empirisches Objekt existiert. Für die lexikographische Arbeit ist die genetische Ebene nicht relevant, da die zu beschreibenden einzelnen Wörter auf dieser Ebene nicht existieren.

Im Selbstverständnis der Lexikographie wird die korpusbasierte Beschreibung des Sprachgebrauchs einer bestimmten Sprachgemeinschaft als Basis für das Wörterbuch genommen. Lexikographische Handbücher wie Atkins/Rundell (2008) und Svensén (2009) beschreiben zwei Methoden der Datensammlung für lexikographische Projekte: ein Leseprogramm und ein Korpus. Svensén (2009, S. 39f.) erwähnt auch die Introspektion als eine mögliche Methode der Datensammlung, hebt aber vor allem die Probleme eines darauf basierenden Verfahrens hervor.

Leseprogramme und Korpora unterscheiden sich durch die Art, wie die Sprachkenntnisse der Menschen die Auswahl der Daten beeinflussen. Im Leseprogramm bestimmen die Leser, welche Zitate „interessant“ sind, während eine korpusbasierte Methode auf einer Konkordanz über das gesamte Korpus beruht. In beiden Fällen ist die Quelle der Daten der Sprachgebrauch einer bestimmten Sprachgemeinschaft.

Aus chomskyanischer Sicht ist eine rein deskriptive Interpretation des Wörterbuches problematisch, da sich kein empirisches Objekt nachweisen lässt, das von einem Wörterbuch beschrieben werden kann. Wie in ten Hacken (2009a) ausführlicher dargestellt, eignet sich keine der möglichen Perspektiven auf Sprache als Objekt der Beschreibung. Es werden dort drei Perspektiven betrachtet: Sprachkompetenz, Sprachgebrauch und Standardsprache.

Die Sprachkompetenz ist der Zustand des Sprachmoduls nach dem Spracherwerb. Sie ist individuell und wird deshalb nicht als Basis für ein Wörterbuch genommen. Sogar dann, wenn die Introspektion als Datenquelle akzeptiert wird, ist das Ziel nicht, ein Wörterbuch eines einzelnen Sprechers zu produzieren.

Die zentrale Bedeutung der Korpora in der Lexikographie könnte man vielleicht als Zeichen dafür sehen, dass ein Wörterbuch den Sprachgebrauch beschreibt. Diese Sichtweise hat allerdings mehrere Nachteile. Erstens sehen viele Wörterbuchbenutzer die reine Beschreibung des Sprachgebrauchs nicht als die Aufgabe eines guten Wörterbuchs. Dies wird z.B. von der Kontroverse über die dritte Auflage des Webster-Wörterbuchs für Englisch belegt (vgl. Morton 1994). Zweitens ist die Abgrenzung des Sprachgebrauchs, der als Input für ein Wörterbuch dienen sollte, nicht ohne Weiteres vorgegeben, sondern erfordert dauernd Entscheidungen, was als „richtiges“ Deutsch, Englisch usw. akzeptiert wird. Diese Abgrenzung bezieht sich nicht nur auf Fehler, sondern auch auf Kontraste zwischen Dialekt und Standard, höheren und niedrigeren Sprachebenen, Akzeptanz von Lehnwörtern usw.

Die meisten Wörterbuchbenutzer würden wahrscheinlich erwarten, dass ein Wörterbuch die „richtige“ Form der Sprache beschreibt, d.h. den Standard. Ein solcher Standard ist aber kein empirisches Objekt, sondern eine Norm, die von bestimmten Leuten festgelegt wird. Deshalb ist ein rein deskriptiver Ansatz nicht mit der Beschreibung einer Norm vereinbar.

Die Idee des Wörterbuchs als Beschreibung des Vokabulars einer Sprache scheidet somit an der Unmöglichkeit, eine Form der Sprache zu bestimmen, deren Vokabular gleichzeitig beschreibend betrachtet werden kann und sich als Objekt der Beschreibung in einem Wörterbuch eignet.

2.2 Das Wörterbuch als Hilfsmittel

Da sich keine widerspruchsfreie deskriptive Interpretation des Wörterbuchs finden lässt, wird in ten Hacken (2009a) vorgeschlagen, ein Wörterbuch als Hilfsmittel zur Problemlösung zu betrachten. Eine ähnliche Sicht ergibt sich aus der Arbeit von Henning Bergenholtz, z.B. Bergenholtz/Bergenholtz (2011).

In einer von der mentalistischen Sicht auf Sprache inspirierten Darstellung des Wörterbuchgebrauchs ist der Ort, wo lexikalische Probleme formuliert und gelöst werden, das Gehirn bzw. der Geist der Benutzer. Wie von Jackendoff (1983) dargestellt, hat jeder Mensch (wie auch andere Lebewesen) eine Vorstellung der Welt (‘projected world’) im Geist, die auf der Interpretation des sinnlichen Inputs basiert. Diese Vorstellung bildet das Wissen um unsere Umwelt sowie die Basis der Gedanken und der Kommunikation. Sie ist als Teil des Geistes notwendigerweise persönlich. Deshalb besteht die (erfolgreiche) Kommunikation darin, dass eine Person (A) einer anderen (B) die Informationen vermittelt, die es B erlauben, ihre Vorstellung der Welt in einer Weise anzupassen, die A gemeint hat.

Eine wesentliche Erkenntnis über Wörterbuchgebrauch in diesem Modell ist, dass das Wörterbuch nicht die Bedeutung, Flexion usw. eines Wortes enthält, sondern nur Informationen zu einem Wort bereitstellt, die von den Benutzern interpretiert werden müssen. Diese Interpretation geschieht in Hinblick auf die in der Vorstellung vorhandenen Informationen über die Welt und auf das Problem, wofür das Wörterbuch konsultiert wird. Ein Beispiel eines solchen Prozesses könnte sich auf (1) beziehen.¹

- (1) Dieser Rhythmus veranlasste Richard Wagner, die Sinfonie als „Apotheose des Tanzes“ zu bezeichnen.

Eine Leserin von (1), nennen wir sie Anna, hat ein Verständnisproblem. In einem ersten Schritt gelingt es ihr, dieses Problem mit dem Wort *Apotheose* zu verbinden. Dieser Schritt ist essenziell, damit das Problem in der Wirkungsdomäne eines Wörterbuchs liegt, aber er wird in keinerlei Weise von (1) bedingt. Es gehört zu Annas Verarbeitungsprozess von (1) und kann als Teil der Lösung des Verständnisproblems gesehen werden. Als Nächstes kann Anna das Wort *Apotheose* im Wörterbuch nachschlagen. Sie weiß, dass es ihr um die Bedeutung dieses Wortes geht, und findet Folgendes dazu:²

1. a. (bildungssprachlich) Erhebung eines Menschen zum Gott; Vergöttlichung eines Menschen: *die Apotheose Napoleons*
- b. (bildungssprachlich) Verherrlichung, Verklärung: *die Apotheose der modernen Naturwissenschaft*
- c. (bildungssprachlich, Kunst) Darstellung einer Apotheose (1a)
2. (Theater) wirkungsvolles [verherrlichendes] Schlussbild eines Bühnenstücks

Die zentrale Annahme ist jetzt, dass dieser Wörterbucheintrag nicht „die Bedeutung von *Apotheose*“ enthält, sondern Informationen, die Anna benutzen kann, um ihr Problem mit (1) zu lösen. Ein erster Schritt ist die Auswahl der passendsten Bedeutung. Anna könnte zum Beispiel das Thema des Textes, in dem (1) vorkommt, benutzen, um 1c und 2 auszuschließen. Die Kombination von *Apotheose* mit *Tanz* in (1) legt eine Bedeutung wie 1a nicht nahe. Deshalb scheint die beste Vorgehensweise, zu versuchen, 1b zu benutzen, um (1) zu verstehen. Der Erfolg einer Interpretation hängt natürlich stark von Annas Vorkenntnissen ab.

¹ Quelle: *7. Sinfonie (Beethoven)*. In: Wikipedia, [http://de.wikipedia.org/wiki/7._Sinfonie_\(Beethoven\)](http://de.wikipedia.org/wiki/7._Sinfonie_(Beethoven)) (Stand: 12/2011).

² Dieser Eintrag ist die Information unter *Bedeutung* im Duden-Online-Wörterbuch, <http://www.duden.de/rechtschreibung/Apotheose> (Stand: 12/2012).

Wenn ein solches Szenario die Betrachtung des Wörterbuchs bestimmt, ergibt sich daraus eine Reihe von Folgerungen, die sich zum Teil nicht mit der typischen lexikographischen Betrachtungsweise decken. Erstens ist ein gutes Wörterbuch nicht eines, das die (deutsche) Sprache beschreibt, sondern eines, das die Probleme der Benutzer löst. Deshalb können Lexikographen sich nicht darauf beschränken, das Wörterbuch in seinem Verhältnis zur (deutschen) Sprache zu sehen, sondern müssen sich bei der Erstellung von der Erwartung leiten lassen, welche Probleme auftreten werden und welche Informationen zur Lösung dieser Probleme beitragen. Das Korpus sollte deshalb auch nicht als eine absolute Autorität zur Sprache gesehen werden, sondern als eine Informationsquelle für Lexikographen. Die im Korpus enthaltenen Informationen zum Sprachgebrauch können als Inspirationsquelle einerseits für die möglichen Fragen, andererseits für die Formulierung des Wörterbuchartikels dienen.

2.3 Lernerwörterbücher

Der Unterschied zwischen allgemeinen Wörterbüchern und Lernerwörterbüchern findet sich primär im Zielpublikum. Die Benutzer eines Lernerwörterbuchs sind Leute, die die Sprache des Wörterbuchs als Zweitsprache lernen. Die Sprachkompetenz in einer Zweitsprache wird 'Interimsprache' (engl. *interlanguage*) genannt. Selinker (1992) gibt eine Übersicht der Entwicklungen, die zur Erkenntnis geführt haben, dass diese Interimsprache als Teil des Sprachmoduls (engl. *competence*) im Gehirn der Lernenden betrachtet werden sollte. Die Aussage, dass Lernende einer Zweitsprache (z.B. Deutsch) eine niedrigere Ebene der Sprachkompetenz haben als Muttersprachler der gleichen Sprache, kann somit folgendermaßen paraphrasiert werden: Die Sprachnorm (z.B. für Deutsch) ist auf der Basis der Sprachkompetenz einer ausgewählten Gruppe Muttersprachler festgelegt. Die Interimsprache kann in der gleichen Weise mit der Sprachnorm verglichen werden wie die muttersprachliche Sprachkompetenz, aber die Interimsprache weist im Allgemeinen größere Unterschiede zur Norm auf.

Da eine Interimsprache kein grundsätzlich anderes Phänomen ist als eine muttersprachliche Sprachkompetenz, ist es sehr wohl vorstellbar, dass es einen fließenden Übergang zwischen den beiden gibt. Wenn wir uns auf den Wortschatz beschränken, so ist dieser Übergang leicht wahrnehmbar. Ein Franzose, nennen wir ihn Bernard, der seit langer Zeit als Professor an einer deutschen Hochschule arbeitet, hat wahrscheinlich einen Wortschatz, der sich in Umfang und Korrespondenz zur Norm kaum von dem einer Muttersprachlerin wie

Anna unterscheiden lässt, und wenn dies trotzdem der Fall wäre, kann man das nicht primär aus dem Unterschied zwischen Lerner und Muttersprachlerin erklären, sondern eher aus den Umständen, in denen Anna und Bernard leben. Lernerwörterbücher richten sich aber nicht an Lernende wie Bernard, sondern an solche, die gerade das Sprachniveau erreicht haben, das es ihnen erlaubt, ein einsprachiges Wörterbuch zu benutzen.

Ein wesentlicher Unterschied zwischen Lernenden, die zum Zielpublikum eines Lernerwörterbuchs gehören, und Muttersprachlern, die ein allgemeines Wörterbuch benutzen, ist, dass die Lernenden neben dem allgemeinen Ziel, im Wörterbuch Informationen zu finden, mit denen sie ihre Probleme lösen können, auch Hilfe beim weiteren Spracherwerb erwarten. Ein Lernerwörterbuch ist nicht nur eine Informationsquelle über den Wortschatz, sondern auch ein Lehrmittel zur gezielten Erweiterung des Wortschatzes.

Bei der Erstellung eines Lernerwörterbuchs sollten somit zwei Eigenschaften des Zielpublikums berücksichtigt werden, die in der Erstellung eines allgemeinen Wörterbuchs keine Rolle spielen. Erstens ist der Wortschatz des Zielpublikums kleiner, zweitens ist neben der allgemeinen Problemlösung auch die Erweiterung des Wortschatzes ein Ziel der Benutzer. Der ersten Eigenschaft kann mit speziellen Definitionen Rechnung getragen werden. Zwei Ansätze, die dieses Ziel verfolgen, sind die Beschränkung des Definitionsvokabulars, wie z.B. im LDOCE (1978), und die Idee, ganze Sätze als Definition zu benutzen, wie z.B. von Hanks (1987) erklärt. Die zweite Eigenschaft bedingt ein grundsätzliches Umdenken in der Gestaltung des Wörterbuchs. Ein Lernerwörterbuch sollte nicht nur Informationen geben, die die Benutzer suchen, sondern auch solche, die unabhängig von einem bestimmten Problem nützlich sind. Idealerweise sollte das Wörterbuch nicht nur das anbieten, was gesucht wird, sondern auch das, wovon die Benutzer gar nicht wissen, dass sie es brauchen. Ich werde diese beiden Perspektiven hier als die der Lernenden und die der Lehrenden bezeichnen.

2.4 Elektronische Wörterbücher

Während Lernerwörterbücher sich von allgemeinen Wörterbüchern in ihrem Zielpublikum unterscheiden, werden elektronische Wörterbücher nur durch die Weise, wie sie verfügbar sind, charakterisiert. Sie unterscheiden sich also nicht von allgemeinen, sondern von Papierwörterbüchern. Interessanterweise ist eines der ersten Wörterbücher, dessen elektronische Version breit erforscht wurde, ein Lernerwörterbuch, LDOCE (1978). Wie die Übersicht von Boguraev/Briscoe (Hg.) (1989) bezeugt, wurde zunächst die Benutzung in der

Computerlinguistik als bevorzugtes Arbeitsgebiet betrachtet. Inzwischen ist es weitgehend üblich, dass in Buchform publizierte Wörterbücher mit einer CD-ROM versehen sind, die von den (menschlichen) Benutzern am Computer benutzt werden kann. Außerdem sind viele Wörterbücher mit oder ohne Bezahlung im Internet zugänglich.

Wegen der breiten Erfahrung mit elektronischen Wörterbüchern ist die Einsicht, dass die Vorteile der elektronischen Darstellung nur dann voll zutage treten, wenn die Struktur der Darstellung dem elektronischen Medium gerecht wird, jetzt Gemeingut (vgl. van Sterkenburg 2003a, S. 5f.). Viele Beiträge in Boguraev/Briscoe (Hg.) (1989) beschreiben die mühselige Aufarbeitung dessen, was ursprünglich nur zum Drucken gemeint war, für die elektronische Form. Inzwischen werden viele Wörterbücher zunächst als lexikalische Datenbanken konzipiert, und die gedruckte Form ist nur ein Derivat. Die neuen Möglichkeiten bestehen nicht nur in der Form neuer Suchmechanismen, wie van Sterkenburg (2003b) sie für eine onomasiologische Suche beschreibt, sondern auch im fast unbeschränkten Platz, wie von de Caluwe/Taeldeman (2003) als Argument für die Aufnahme regelmäßiger Wortbildungen erwähnt. Bei den Lernerwörterbüchern nehmen die Benutzer, wie Nesi (1999) erwähnt, vor allem die schnelle Suche wahr.

Die Einsicht, dass nur die Kombination der kreativen Exploration der Möglichkeiten mit der Berücksichtigung des Bedarfs zu einem Produkt führt, das das gedruckte Wörterbuch an Benutzerfreundlichkeit übersteigt, bleibt aber weiterhin wichtig.

3. Wortbildung

Die Wortbildung ist ein Bereich, wofür in Lernerwörterbüchern eine klar unterschiedliche Behandlung erforderlich ist als in allgemeinen Wörterbüchern. Um dies zu erklären, werde ich hier zunächst einige allgemeine Betrachtungen zur Funktion der Wortbildung in der Sprache (Kap. 3.1) und im Lexikon (Kap. 3.2) formulieren. Dann werde ich kurz die daraus resultierende Sicht auf morphologische Produktivität darstellen (Kap. 3.3).

3.1 Die Bildung neuer Wörter

Die Wortbildung ist ein regelbasierter Mechanismus zur Erweiterung des Wortschatzes. Von einem Wortbildungsprodukt wie *unweit* können die Form, die syntaktischen Eigenschaften sowie die Bedeutung weitgehend aus den entsprechenden Eigenschaften des Wortstamms *weit* und der Präfigierungsre-

gel, die *un-* anhängt, abgeleitet werden. Allerdings kann die Bedeutung nicht immer vollständig auf der Basis der Form vorhergesagt werden. Die Beziehung zu *Arzt* ist in *Augenarzt* und *Hausarzt* nicht dieselbe. Ebenso lässt sich der Unterschied zwischen der Bedeutung des Suffixes *-er* in Bezug auf eine Maschine (z.B. *Anrufbeantworter*) oder eine Person (z.B. *Fußballspieler*) nicht aus der Form ableiten. Wie in ten Hacken (2010) ausführlicher dargestellt, wäre es deshalb nicht richtig, die Wortbildung als regelbedingt (engl. *rule-driven*) zu bezeichnen, wie es die Syntax ist.

Während die Beziehungen zwischen Form und Bedeutung eine verwirrende Vielfalt aufweisen, wenn wir sie aus der Sicht der Form betrachten, klärt sich die Lage merklich, wenn wir die Wortbildung als Lösung eines Benennungsproblems sehen. In dieser Perspektive ist der Ausgangspunkt ein Konzept, das noch keinen Namen hat. Eine Sprecherin, die das Konzept bezeichnen möchte, versucht einen Namen zu benutzen, der von ihren Zuhörern verstanden wird. Der Grund, wieso *Anrufbeantworter* ein Gerät und nicht eine Person bezeichnet, ist, dass für dieses Gerät ein Name gesucht wurde. In ähnlicher Weise ist die Bildung von *Hausarzt* zunächst die Lösung eines Benennungsproblems. Bei der Bildung spielt die Existenz einer Reihe von Wörtern vom Typ *X-arzt*, wo *X* ein Organ oder einen Körperbereich bezeichnet, keine direkte Rolle. Im Falle von *unweit* ist interessant, dass es nicht das übliche Antonym von *weit* ist. Da dafür schon *nahe* existierte, konnte *unweit* nur aus dem Bedarf entstehen, ein neues Konzept, das sich von *nahe* unterscheidet, zu benennen.

Bei der Benennung neuer Konzepte stehen Sprechern außer der Wortbildung auch andere Mechanismen zur Verfügung: einerseits die Bedeutungserweiterung existierender Wörter, z.B. durch metaphorische oder metonymische Prozesse, andererseits die Entlehnung, entweder direkt, z.B. *Computer* im Deutschen, oder als Lehnübersetzung, z.B. *Rechner* für das gleiche Gerät. Man könnte diese Prozesse als Konkurrenten der Wortbildung betrachten, aber Konkurrenz impliziert eine Absicht, die man der Wortbildung nicht zudichten sollte. Nur Sprecher und Hörer haben Absichten, nicht Prozesse wie Wortbildung.

3.2 Wortbildung und die Struktur des Lexikons

Unter ‘Lexikon’ verstehe ich hier den Teil der Sprachkompetenz eines Sprechers bzw. einer Sprecherin, der sich auf den Wortschatz bezieht. Eine Sprecherin des Deutschen als Muttersprache, nennen wir sie Claudia, hat ein Lexikon, das es ihr (zusammen mit Syntax- und anderen Regeln) ermöglicht, Deutsch zu reden und zu verstehen. Die Hauptfunktion der Wortbildungsre-

geln, wie wir sie oben gesehen haben, ist die Bildung von neuen Benennungen zur Bezeichnung neuer Konzepte. Claudia wird aber in ihrem Leben nur relativ selten grundsätzlich neue Konzepte bilden. Wenn wir den Wortschatz einer Sprache wie Deutsch auf einige Hunderttausend Wörter und die Zahl der Deutschsprechenden auf ungefähr 100 Millionen schätzen, wäre es ein großer Zufall, wenn Claudia überhaupt ein einziges neues Wort in ihrem Leben bilden würde. Dieses Bild bedarf aber einer Korrektur, denn es beruht auf einer Kombination von Zahlen, die nicht miteinander vergleichbar sind.

Der Wortschatz einer Sprache wie Deutsch ist eine Konstruktion, die von der Kompetenz ausgewählter Sprecher abgeleitet werden kann, aber nur indem jedes Mal separat beschlossen wird, ob ein bestimmtes Wort bzw. eine bestimmte Bedeutung eines Wortes zur deutschen Sprache gehört. Der Wortschatz ist keine empirische Entität, sondern das Resultat einer Reihe von Klassifikationsentscheidungen. Das Lexikon von Claudia hingegen ist Teil ihrer Sprachkompetenz und somit eine wirklich bestehende, empirische Entität. Die Anwendung einer Wortbildungsregel kommt jedes Mal zum Zuge, wenn Claudia die Regel braucht, um ein entsprechendes neues Wort zu ihrem Lexikon hinzuzufügen, sei es, weil sie ein Konzept benennen möchte, für welches es in ihrem Lexikon noch kein Wort gibt, sei es, weil sie ein entsprechendes Wort hört, das neu für sie ist. Dies ist weit häufiger als der schlecht definierte Vorgang der Bildung eines neuen Wortes im Wortschatz des Deutschen.

Neben der Anwendung der Wortbildungsregeln zur Erweiterung des Lexikons spielen die Regeln auch eine wichtige Rolle in der Strukturierung desselben. Es wäre kaum vorstellbar, dass das Lexikon als eine alphabetisch geordnete Liste gestaltet ist. Wenn das der Fall wäre, wäre es leichter, das nächste Wort im Alphabet hervorzurufen, als eines, das semantisch oder phonologisch ähnlich ist. Viel wahrscheinlicher ist es, dass Wörter im Lexikon miteinander in vielen unterschiedlichen Weisen verbunden sind, darunter auch durch Wortbildungsregeln.

Für einen Lerner des Deutschen als Fremdsprache, nennen wir ihn David, ist die Wortbildung womöglich noch wichtiger als für Claudia. Während der Lexikonerwerb für Claudia recht anstrengungsfrei vor sich geht, ist er für David ein wichtiger Teil des Zweitsprachenerwerbs. Mit seinem kleineren Lexikon braucht er häufiger Wörter, die nicht im Lexikon vorhanden sind, sowohl aktiv als passiv. Beim Sprechen und Schreiben kann David sich oft mit einer Paraphrase retten, aber beim Zuhören und Lesen sind die Wörter einfach vorgegeben. Außerdem ist die Strukturierung des Lexikons für David viel wichtiger als Mittel, um die verstandenen Wörter im Lexikon verfügbar zu halten und neue Wörter nicht gleich wieder zu vergessen.

3.3 Produktivität

Ein wichtiger Grund, weshalb Wortbildungsregeln im Zweitspracherwerb relevant sind, ist ihre Produktivität. Das Konzept der Produktivität wird allerdings sehr unterschiedlich verstanden. Einerseits gibt es eine Strömung in der Linguistik, die die Produktivität als eine binäre Eigenschaft betrachtet, z.B. Schultink (1961). Andererseits ist die Auffassung verbreitet, Produktivität sei grundsätzlich graduell, wie z.B. von Baayen (1992) vertreten. Der Vorteil von Corbins (1987) Analyse der Produktivität ist, dass beide Auffassungen gleichermaßen berücksichtigt werden und die Beziehung zwischen ihnen klargestellt wird.

Corbin (1987, S. 176-78) unterscheidet drei Konzepte, die gemeinsam der Produktivität entsprechen, *régularité* ('Regelmäßigkeit'), *disponibilité* ('Verfügbarkeit') und *rentabilité* ('Rentabilität'). Regelmäßigkeit ist eine Eigenschaft, die im Prinzip jede Regel hat. Im Bereich der Wortbildung kann nur eine bestimmte Bildung unregelmäßig sein, nicht die Regel, die sie bildet. Allerdings hat die Existenz eines unregelmäßigen Resultats einer Wortbildungsregel einen negativen Effekt auf die Vorhersehbarkeit der Form und Bedeutung weiterer Bildungen mit dieser Regel. Aronoff (1976, S. 40-43) zeigt auf, wie z.B. die englische Nominalisierungsregel mit dem Suffix *-ity* weniger produktiv ist nach *-ulous* als nach *-acious*, weil im letzten Fall die Form eindeutig ist (z.B. *audacity*), während es für den ersten zwei mögliche Resultate gibt (vgl. *credulity* und *fabulosity*).

Das Konzept der Rentabilität stimmt ziemlich genau mit Baayens (1992) Konzept der Produktivität überein. Es geht hier um die quantifizierbare Größe, die die Häufigkeit von neuen Bildungen aufgrund der betreffenden Regel ausdrückt. Das Problem mit der Definition dieser Größe ist, dass berücksichtigt werden sollte, wie groß der mögliche Anwendungsbereich einer bestimmten Regel ist. Da Deutsch mehr Nomina als Verben hat, gibt es mehr Möglichkeiten, eine Regel anzuwenden, die Verben aus Nomina bildet, als eine, die Nomina aus Verben bildet. Die Frage dabei ist, wie diese Erkenntnis im Ausdruck der Größe der Rentabilität quantifiziert wird.

Das Konzept der Verfügbarkeit schließlich drückt aus, ob eine Regel überhaupt für die Bildung von neuen Wörtern angewandt werden kann. Es handelt sich hier um ein binäres Konzept, aber nicht das von Schultink (1961), der nur Regeln, die ohne bewusste Absicht angewandt werden, als produktiv bezeichnet.

Corbin (1987, S. 177) benutzt ihre Analyse der Produktivität als Begründung, wieso sie sich nur mit der Verfügbarkeit beschäftigt. Nur die Verfügbarkeit ist eine Eigenschaft der Sprachkompetenz, und die Verfügbarkeit einer Regel ist eine logische Voraussetzung für die Betrachtung ihrer Rentabilität und der Regelmäßigkeit ihrer Wortbildungsergebnisse. Ich glaube aber, dass die Regelmäßigkeit und die Rentabilität zwar der Verfügbarkeit untergeordnet sind, gleichzeitig aber auch für das Verständnis der Wortbildung nützliche Konzepte darstellen, wie ich weiter unten nachzuweisen versuchen werde.

4. Darstellung der Wortbildung in Lernerwörterbüchern

Nachdem wir in den bisherigen Abschnitten Überlegungen zum Zweitspracherwerb, zum (elektronischen) Wörterbuch und zur Wortbildung formuliert haben, können wir uns jetzt der Kombination dieser Themen zuwenden, die die Behandlung der Wortbildung in elektronischen Lernerwörterbüchern erfordert.

4.1 Bedeutung der Wortbildung im Zweitspracherwerb

Ein Lernerwörterbuch ist ein Wörterbuch, das den Zweitspracherwerb unterstützt. Die Benutzer eines solchen Wörterbuchs haben eine Interimsprachkompetenz, mit der sie ein einsprachiges Wörterbuch benutzen können, aber keine, die ihre Bedürfnisse denen von Muttersprachlern angleicht.

Im Bereich der Wortbildung werden die Bedürfnisse vom Niveau der Interimsprachkompetenz bedingt. In Kapitel 3.2 haben wir David als Deutschlermer kennengelernt. Seine Benutzung der deutschen Wortbildungsregeln kann sich auf drei unterschiedliche Perspektiven beziehen. Erstens kann er die Regeln als Generalisierungen über ihm bekannte Wörter betrachten. Oft ist diese Perspektive nur unbewusst oder halbunbewusst vorhanden, sodass es einen Trigger braucht, um die Existenz dieser Generalisierungen bewusst zu machen. Die Generalisierungen spielen eine wichtige Rolle im Aufbau des Lexikons als Teil der Interimsprachkompetenz und in der Aktivierung der richtigen Einträge dieses Lexikons im Sprachgebrauch.

Eine zweite Perspektive ist die der Regeln als Mittel zur Erweiterung des Lexikons. Wie oben erklärt, geht es dabei nicht um die Bildung neuer Wörter als Teil des deutschen Wortschatzes, sondern um die Bildung neuer Wörter im Lexikon, das David als Teil seiner Interimsprachkompetenz des Deutschen hat. Da David als Lerner ein relativ kleines Lexikon hat, wird er ziemlich häufig

diese Perspektive der Wortbildungsregeln brauchen, um Texte und gesprochene Sprache zu verstehen. Sie ist am wirkungsvollsten bei geschriebener Sprache, wo er das Rezeptionstempo bestimmen kann. Inwiefern David die Wortbildungsregeln auch im aktiven Sprachgebrauch anwendet, hängt stark von seiner Persönlichkeit ab, daneben auch vom Stand seiner Interimsprachkompetenz und von der kommunikativen Situation. Diese sog. Neubildungsperspektive ist den Lernern meist viel bewusster als die der Generalisierungen, aber das impliziert keineswegs, dass sie die Regeln explizit formulieren könnten.

Schließlich gibt es die Perspektive der explizit formulierten Wortbildungsregel. Für Muttersprachler spielt diese Perspektive kaum eine Rolle, außer sie beschäftigen sich mit morphologischer Forschung. Für den Spracherwerb ist die explizite Formulierung aber durchaus nützlich, um die Entwicklung der Generalisierungs- und Neubildungsperspektiven zu beschleunigen und die richtige Anwendung zu unterstützen.

4.2 Darstellung der Wortbildung

Für die Besprechung der Darstellung der Wortbildung können wir als Beispiele die deadjektivischen Nominalbildungen *Finsternis* und *Flüssigkeit* nehmen. Ein erster Informationstyp, den wir für diese Wörter festhalten können, ist, wie sie gebildet sind. Für *Finsternis* heißt das, dass das Wort als Bildung von *finster* mit dem Suffix *-nis* ausgewiesen wird. Diese Informationen sind häufig schon im Wörterbuch vorhanden,³ obwohl es oft den Benutzern überlassen wird, die implizite Beziehung explizit zu machen. In einem elektronischen Wörterbuch besteht aber kein Grund, in dieser Weise Platz zu sparen. Außerdem sollte ein Hyperlink eingefügt werden, der gleich zum Eintrag für *finster* führt. Diese Information ist wichtig, um das Wort *Finsternis* optimal im Lexikon der Interimsprachkompetenz einzubinden. In dieser Hinsicht verhält sich *Flüssigkeit* ähnlich wie *Finsternis*.

Ein zweiter Informationstyp ist die Klasse der Wörter, die in gleicher Weise gebildet wurden. Es gibt mehrere Dimensionen der Gleichheit, die dabei als Ausgangspunkt dienen können. *Finsternis* gehört zur Klasse der Wörter, die mit *-nis* gebildet werden, ein Adjektiv als Basis haben (im Gegensatz z.B. zu *Erkenntnis*) und weiblich sind (im Gegensatz z.B. zu *Geheimnis*). Das Wort gehört auch zur Klasse der Wörter, die *finster* als Basis haben. Außerdem könnte man alle deadjektivischen Nominalbildungen als eine Klasse betrach-

³ Einen Überblick über die Praxis der Wortbildungsangaben in gedruckten wie elektronischen Wörterbüchern findet sich im Beitrag „Wortbildung in Wörterbüchern – Zwischen Anspruch und Wirklichkeit“ von Sabina Ulsamer in diesem Band.

ten, die in dem Fall auch *Flüssigkeit* einschließt. Diese Information ist selten im Wörterbuch auffindbar, ist aber besonders wichtig, um eine gute Einschätzung der Wortbildungsregeln herbeizuführen. Die Suffigierung mit *-nis* produziert weibliche und sächliche Ableitungen von Adjektiven, aber Canoo.net (2011) gibt nur je 11 davon. Für die Suffigierung mit *-keit* hingegen gibt Canoo.net (2011) 1888 Beispiele. Die Canoo.net (2011) unterliegenden Datenbanken sind im *Word Manager*-System, beschrieben von ten Hacken (2009b), erfasst worden. Es ist natürlich eine weitere Frage, wie die für ein Lernerwörterbuch relevanten Informationen zu diesen Suffigierungsregeln optimal ausgewählt und dargestellt werden sollten.

Ein dritter Informationstyp ist die explizite Formulierung von Wortbildungsregeln. Hier geht es darum, Informationen zu den Suffixen zu geben, die häufig in Wörterbüchern als eigene Einträge aufgeführt werden. So findet man *-keit* (aber nicht *-nis*) als Eintrag in Wahrig (1997). Der Unterschied zum zweiten Informationstyp besteht darin, dass dort die Regel im Prinzip implizit bleibt und nur von den Benutzern auf der Basis der Beispiele in ihrem Lexikon erstellt wird. Es wäre aber sinnvoll, die Informationen des zweiten und dritten Typs in einem Lernerwörterbuch so zusammenzubringen, dass falsche Generalisierungen vermieden und richtige gefördert werden. Der Eintrag für *-keit* in Wahrig (1997) gibt eine genaue Regel, wann *-keit* statt *-heit* gebraucht wird. Für ein Lernerwörterbuch wäre es aber auch wünschenswert, *-nis* zu behandeln, mit der Absicht, das Fehlen einer klaren Regel zu erklären.

Es ist interessant festzustellen, wie die drei Informationstypen sich zu den drei Arten der Produktivität, die Corbin (1987) unterscheidet, verhalten. Die Regelmäßigkeit ist zunächst eine Eigenschaft einzelner Bildungen, wie sie im ersten Informationstyp erfasst werden. Die Rentabilität wird am leichtesten wahrgenommen, wenn Regeln als Klassen dargestellt werden, wie im zweiten Informationstyp, denn die Klassengröße gibt über die Rentabilität Aufschluss. Die Verfügbarkeit ist eine absolute Eigenschaft einer Regel, die am deutlichsten im dritten Informationstyp hervortritt. Allerdings geben die Klassengrößen auch interessante Auskünfte über die Regelmäßigkeit. Es ist durchaus plausibel, dass die geringe Klassengröße der Bildungen mit *-nis* in einer Weise mit der Unvorhersehbarkeit des Wortgeschlechts zusammenhängt, die vergleichbar ist mit dem in Kapitel 3.3 erwähnten Beispiel von *-ity*.

5. Schlussfolgerungen

Zum Schluss möchte ich hier die aus dem Vorgegangenen hervorgehenden Erkenntnisse zur Wortbildung in Lernerwörterbüchern kurz zusammenfassen. Wie in Kapitel 2 argumentiert, sind Wörterbücher nicht als Sprachbeschreibung zu betrachten, sondern als Mittel, die die Benutzer brauchen, um Sprachprobleme zu lösen. Die Wörterbücher stellen dazu Informationen bereit, die die Benutzer interpretieren müssen, um die spezifischen Probleme zu lösen. Im Falle von Lernerwörterbüchern sind die Probleme nicht nur einzelnen Wörtern zuzuordnen, sondern auch dem allgemeinen Spracherwerbsprozess. Lerner brauchen Informationen über die Wortbildung (größtenteils unbewusst) als Unterstützung der Lexikonbildung. Deshalb ist die Darstellung der Wortbildung in Lernerwörterbüchern noch wichtiger als in allgemeinen Wörterbüchern.

Um die Wortbildung optimal verfügbar zu machen, sollte sie in drei Perspektiven präsentiert werden, erstens als Analyse einzelner Bildungen, zweitens als Klassen von ähnlichen Bildungen und drittens als explizite Regel. Die erste und dritte Perspektive sind auch in Papierwörterbüchern repräsentierbar, aber die zweite erfordert zu viel Platz. Im elektronischen Lernerwörterbuch spielt diese Platzbeschränkung keine Rolle. Wichtig ist aber eine gute Verbindung zwischen den Perspektiven mittels Hyperlinks. Zudem sollte die Darstellung größerer Mengen von Daten so gestaltet werden, dass die Benutzer nicht vom Überfluss der Informationen überwältigt werden.

6. Literaturverzeichnis

6.1 Wörterbücher

Canoo.net – Deutsche Wörterbücher und Grammatik. Internet: <http://www.canoo.net> (Stand: 12/2011).

Duden Online. Internet: <http://www.duden.de/woerterbuch> (Stand: 12/2012).

LDOCE (1978): Dictionary of Contemporary English. Hrsg. von Paul Procter. Harlow.

Wahrig, Gerhard (1997): Deutsches Wörterbuch. 6. Aufl. Hrsg. von Renate Wahrig-Burfeind. Gütersloh.

6.2 Forschungsliteratur

Aronoff, Mark H. (1976): Word formation in generative grammar. Cambridge, MA.

Atkins, B.T. Sue/Rundell, Michael (2008): The Oxford guide to practical lexicography. Oxford.

- Baayen, Harald (1992): Quantitative aspects of morphological productivity. In: Booij, Geert/van Marle, Jaap (Hg.): Yearbook of morphology 1991. Dordrecht, S. 109-149.
- Béjoint, Henri (2000): Modern lexicography: An introduction. Oxford.
- Bergenholtz, Henning/Bergenholtz, Inger (2011): A dictionary is a tool, a good dictionary is a monofunctional tool. In: Fuertes-Olivera, Pedro/Bergenholtz, Henning (Hg.): e-Lexicography: The internet, digital initiatives and lexicography. London, S. 187-207.
- Bogaards, Paul (1996): Dictionaries for learners of English. In: International Journal of Lexicography 9, S. 279-320.
- Boguraev, Bran/Briscoe, Ted (Hg.) (1989): Computational lexicography for natural language processing. London.
- Chomsky, Noam (1965): Aspects of the theory of syntax. Cambridge, MA.
- Chomsky, Noam (1980): Rules and representations. New York.
- Chomsky, Noam (1981): Lectures on government and binding. Dordrecht.
- Chomsky, Noam (1999): On the nature, use, and acquisition of language. In: Ritchie, William C./Bhatia, Tej K. (Hg.): Handbook of child language acquisition. San Diego, S. 33-54.
- Corbin, Danielle (1987): Morphologie dérivationnelle et structuration du lexique. Tübingen.
- Curtiss, Susan/Fromkin, Victoria/Krashen, Stephen/Rigler, David/Rigler, Marilyn (1974): The linguistic development of Genie. In: Language 50, S. 528-554.
- De Caluwe, Johan/Taeldeman, Johan (2003): Morphology in dictionaries. In: van Sterkenburg (Hg.), S. 114-126.
- ten Hacken, Pius (2007): Chomskyan linguistics and its competitors. London.
- ten Hacken, Pius (2009a): What is a dictionary? A view from Chomskyan linguistics. In: International Journal of Lexicography 22, S. 399-421.
- ten Hacken, Pius (2009b): Word Manager. In: Mahlow, Cerstin/Piotrowski, Michael (Hg.): State of the art in computational morphology. Berlin, S. 88-107.
- ten Hacken, Pius (2010): Synthetic and exocentric compounds in a parallel architecture. In: Linguistische Berichte, Sonderheft 17, S. 233-251.
- Hanks, Patrick (1987): Definitions and explanations. In: Sinclair, John M. (Hg.): Looking up. An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English Language Dictionary. London/Glasgow, S. 116-136.
- Hensch, Takao K. (2005): Critical period plasticity in local cortical circuits. In: Nature Reviews Neuroscience 6, S. 877-888.
- Jackendoff, Ray S. (1983): Semantics and cognition. Cambridge, MA.

- Mayberry, Rachel I. (1994): The importance of childhood to language acquisition: Evidence from American sign language. In: Goodman, Judith C./Nusbaum, Howard C. (Hg.): *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA, S. 57-90.
- Morton, Herbert C. (1994): *The story of Webster's third: Philip Gove's controversial dictionary and its critics*. Cambridge.
- Nesi, Hilary (1999): A users' guide to electronic dictionaries for language users. In: *International Journal of Lexicography* 12, S. 55-66.
- Schultink, Henk (1961): Produktiviteit als morfologisch fenomeen. In: *Forum der Letteren* 2, S. 110-125.
- Selinker, Larry (1992): *Rediscovering interlanguage*. Harlow.
- van Sterkenburg, Piet (2003a): The dictionary: Definition and history. In: van Sterkenburg (Hg.), S. 3-17.
- van Sterkenburg, Piet (2003b): Onomasiological specifications and a concise history of onomasiological dictionaries. In: van Sterkenburg (Hg.), S. 127-143.
- van Sterkenburg, Piet (Hg.) (2003): *A practical guide to lexicography*. Amsterdam.
- Svensén, Bo (2009): *A handbook of lexicography: The theory and practice of dictionary-making*. Cambridge.