

Sabina Ulsamer

Chancen und Probleme bei der automatischen Ermittlung von Wortbildungsprodukten für *elexiko* und bei ihrer Präsentation

1. Einleitung

Die Problematik der Darstellung von Wortbildungsbeziehungen im Wörterbuch wird seit vielen Jahren in der metalexikographischen Forschung reich diskutiert¹ und hat auch in der elektronischen Lexikographie nicht an Wichtigkeit verloren. Hier ergeben sich durch das nahezu unbeschränkte Platzangebot und durch Hyperlinks neue Möglichkeiten, um Wortbildungszusammenhänge im Wörterbuch sichtbar zu machen. Mehr und mehr wird die lexikographische Arbeit – besonders in der Internetlexikographie – durch automatische und computergestützte Methoden ergänzt und unterstützt (vgl. Klosa 2010). Neue Chancen eröffnen sich mit computerlinguistischen Werkzeugen wie z.B. Morphologie-Tools,² die Flexionsparadigmen automatisch generieren und gebildete Wörter in ihre Bestandteile zerlegen können.

Am Beispiel des Internetwörterbuchs *elexiko* soll in diesem Beitrag dargestellt werden, welche Möglichkeiten ein solches Morphologiewerkzeug – speziell Morphisto – in der Internetlexikographie für die Darstellung von Wortbildungsbeziehungen bietet, aber auch, welche Probleme damit einhergehen. Im Rahmen des Forschungsprojekts ‘Benutzeradaptive Zugänge und Vernetzungen in *elexiko*’ (BZ*Velexiko*)³ am Institut für Deutsche Sprache wurden Methoden entwickelt, wie aus einer Datenbank morphologisch analysierter Stichwörter alle in der *elexiko*-Stichwortliste vorhandenen Wortbildungsprodukte zu einem Lemma ermittelt werden können. Das Ziel ist, diese Wortbildungsprodukte zu extrahieren und online zu präsentieren, um eine stärkere Vernetzung der Daten innerhalb des Online-Wörterbuches *elexiko* zu erreichen. Die Methoden und dahinterstehenden Überlegungen sowie die Entwürfe zur Online-Präsentation der Wortbildungsprodukte sollen in diesem Beitrag beschrieben werden.

¹ Vgl. hierzu den Beitrag „Wortbildung in Wörterbüchern – Zwischen Anspruch und Wirklichkeit“ von Sabina Ulsamer in diesem Band.

² Vgl. hierzu den Beitrag „Finite-State-basierte Morphologie-Tools und ihre Stärken und Schwächen bei der maschinellen Wortbildungsanalyse“ von Christian Simon in diesem Band.

³ Zum Projekt BZ*Velexiko* vgl. <http://www.ids-mannheim.de/lexik/BZVelexiko/> (Stand: 08/2011).

Nach einer kurzen Vorstellung des Internetwörterbuchs *elexiko* in Kapitel 2.1 wird das Konzept zur Wortbildung in *elexiko* (Kap. 2.2) skizziert. Der Fokus des Beitrages liegt auf der Ermittlung der Wortbildungsprodukte und ihrer Darstellung in den *elexiko*-Wortartikeln. Im dritten Kapitel werden die Analysen von Morphisto beschrieben und es wird die Vorgehensweise bei der Extraktion der Wortbildungsprodukte aus der Datenbank geschildert. Kapitel 4 stellt die Überlegungen und Entwürfe zur Präsentation der automatisch ermittelten Wortbildungsprodukte vor. In beiden Kapiteln werden die jeweiligen Chancen, aber auch die Schwierigkeiten für die Lexikographie diskutiert. Das abschließende Kapitel 5 zieht ein Resümee.

2. Das Online-Wörterbuch *elexiko*

2.1 Konzeption

*elexiko*⁴ ist ein im Aufbau befindliches Online-Wörterbuch zur deutschen Gegenwartssprache, das sukzessive und kontinuierlich bearbeitet wird. Die 300 000 Stichwörter wurden aus einem eigens für *elexiko* zusammengestellten Korpus⁵ aus deutschsprachigen Zeitungen – sowohl bundesdeutsche, österreichische und schweizerische als auch Texte aus der ehemaligen DDR – frequenzbasiert ermittelt.⁶ In *elexiko* werden redaktionell erstellte mit (teil-)automatisch gewonnenen Angaben kombiniert. Alle Angaben werden korpusbasiert aus dem *elexiko*-Korpus ermittelt. Die lexikographische Bearbeitung der Stichwörter erfolgt nicht nach Alphabetstrecken, sondern in Teilwortschätzen (so genannten Modulen), die nach unterschiedlichen Kriterien ausgewählt werden. Auch alle automatisch ermittelten Angaben wie die Korpusbelege bei unbearbeiteten Artikeln und die im Folgenden beschriebenen Wortbildungsprodukte erfolgen nach bestimmten Kriterien.⁷

2.2 Wortbildung in *elexiko*

Neben der Gebildetheit eines Stichwortes möchte *elexiko* auch die Wortbildungsaktivität von Stichwörtern zeigen, d.h., in welche anderen Stichwörter

⁴ Zu *elexiko* vgl. Haß (2005), Klosa (2011) sowie http://www.owid.de/elexiko_/index.html (Stand: 08/2011).

⁵ Zum *elexiko*-Korpus vgl. Storzjohann (2005, S. 55-70).

⁶ Zur Ermittlung der *elexiko*-Stichwortliste vgl. Schnörch (2005, S. 71-90).

⁷ Zur Methodik in *elexiko* vgl. Haß (2005, S. 1-17), Klosa (2011, S. 9-26), sowie die *elexiko*-Projektseiten im Internet http://www.owid.de/elexiko_/Informationen.html und <http://www.ids-mannheim.de/lexik/elexiko/> (Stand: 08/2011).

diese eingehen.⁸ Mit diesem zweigleisigen Vorgehen soll Barz' (2001, S. 88f.) Vorschlag Rechnung getragen werden, von primären Wörtern ausgehend ihre Wortbildungsaktivität zu illustrieren, wodurch „man die Entfaltung der Wörter in der Wortbildung erfassen und in Wortnester gruppieren kann“ (ebd., S. 88). Von sekundären, d.h. gebildeten Wörtern ausgehend kann man den jeweils letzten Wortbildungsschritt rekonstruieren, wodurch man die unmittelbaren Motivationsbeziehungen ermittelt und zu Wortbildungstypen kommt (ebd., S. 89).

Bei den bereits bearbeiteten Stichwörtern wird in *ellexiko* die Gebildetheit beschrieben und damit Barz' zweite Forderung erfüllt. Die Art der Wortbildung wird angegeben und es wird verzeichnet, aus welchen Bestandteilen das Wort gebildet ist. Um auch Barz' ersten Vorschlag aufzugreifen und die Aktivität eines Stichwortes in der Wortbildung zu zeigen, sollen beispielsweise alle Wortbildungsprodukte zu *Jugend* angegeben werden, die selbst auch Stichwörter in *ellexiko* sind. Neben den Ableitungen *jugendlich* und *jugendhaft* gibt es in der *ellexiko*-Stichwortliste knapp 400 Komposita, die *Jugend* entweder als Bestimmungswort – wie in *Jugendbuch* – oder als Grundwort – wie in *Dorfjugend* – enthalten.

Im Unterschied zu Printwörterbüchern möchte *ellexiko* nicht nur eine Auswahl an Wortbildungsprodukten zu einem bestimmten Stichwort zeigen, sondern die gesamte Menge der zugehörigen Wortbildungsprodukte, die in der *ellexiko*-Stichwortliste enthalten sind. So sollen zu einem Nomen wie *Fisch* nicht nur Nomen-Nomen-Komposita wie *Fischmarkt*, sondern auch Zusammensetzungen mit anderen Wortarten, beispielsweise Adjektiven wie in *fischreich*, im Wortartikel *Fisch* präsentiert werden. Neben Komposita sollen auch die verschiedenen Derivate – hier unter anderem *Fischlein*, *fischig* und *fischen* – zu einem Lemma ermittelt und unter diesem aufgelistet werden. Damit sollen die Stichwörter innerhalb des Wörterbuches stärker miteinander vernetzt werden. Diese Vernetzungen sind nicht von semantischer oder paradigmatischer, sondern von morphologischer Art. Ausgehend von dem primären Wort soll sich ein Netz von Wortbildungen aufspannen, mit dem sich die Nutzer neue lexikologische Zusammenhänge erschließen können.

Gerade bei den redaktionell noch nicht ausgearbeiteten Artikeln möchte *ellexiko* auf diese Weise mehr Informationen bieten.

⁸ Vgl. hierzu den Beitrag „Wortbildung in *ellexiko*: Gegenwart und Zukunft“ von Annette Klosa in diesem Band sowie Klosa (2005).

3. Ermittlung der Wortbildungsprodukte

3.1 Morphologische Analyse der Stichwörter

Voraussetzung für die Ermittlung der Wortbildungsprodukte war eine morphologische Analyse aller Stichwörter. Sie mussten in ihre Bestandteile zerlegt und diese morphologisch bestimmt werden. Dazu wurde die gesamte *lexiko*-Stichwortliste mit dem Tool Morphisto morphologisch analysiert.⁹

Morphisto¹⁰ ist eine am IDS im Rahmen des TextGrid-Projektes¹¹ entwickelte, auf SMOR¹² basierende, frei zugängliche Computer-Morphologie für das Deutsche, die sowohl Flexionsparadigmen generieren als auch eine morphologische Analyse durchführen kann. Für *lexiko* wurde von den Morphisto-Entwicklern ein Verfahren ausgearbeitet, das aufgrund der morphologischen Analysen komplexen Wortbildungen binär verzweigende, hierarchisch strukturierte Bäume zuweist. Es wird in Zielinski/Simon/Wittl (2009, S. 71)¹³ skizziert. Mithilfe dieses Verfahrens wurde jedes gebildete Wort der *lexiko*-Stichwortliste in Komplement und Kopf zerlegt (vgl. Abb. 1). Als Kopf gilt dabei derjenige Teil des Wortes, der die grammatische und morphosyntaktische Kategorie des ganzen Wortes bestimmt. Nach dem Prinzip der Rechtsköpfigkeit – „Right Hand Head Rule“ (Williams 1981) – ist das immer der rechte Bestandteil eines Wortes. Bei Komposita ist das Zweitglied, d.h. das Grundwort, der Kopf der Zusammensetzung. Bei Suffixderivaten ist das Suffix der Kopf der Bildung. Das Komplement ist dann derjenige Teil, der den Kopf ergänzt oder näher bestimmt. Demnach gilt bei Komposita das Bestimmungswort als Komplement. Bei Suffixderivaten ist der Stamm, an den das Suffix angehängt wird, das Komplement. Bei Präfixderivaten übernimmt das Präfix die Aufgabe des Komplements und die Wortbildungsbasis die des Kopfes.¹⁴ Jedes Kom-

⁹ Zur Generierung der Analysen und Überführung in eine Datenbank für *lexiko* vgl. den Beitrag „Finite-State-basierte Morphologie-Tools und ihre Stärken und Schwächen bei der maschinellen Wortbildungsanalyse“ von Christian Simon in diesem Band.

¹⁰ Zu Morphisto vgl. <http://www.ids-mannheim.de/fi/projekte/textgrid.html> (Stand: 08/2011).

¹¹ Zum TextGrid-Projekt vgl. <http://www.textgrid.de> (Stand: 08/2011).

¹² Zu SMOR vgl. Schmid/Fitschen/Heid (2004).

¹³ Vgl. auch Kapitel 2.3.3 in dem Beitrag „Finite-State-basierte Morphologie-Tools und ihre Stärken und Schwächen bei der maschinellen Wortbildungsanalyse“ von Christian Simon in diesem Band.

¹⁴ Ein Problem für die „Right Hand Head Rule“ (RHR) sind Verben, die sich aus einem Präfix und einem Adjektiv oder Nomen als Stamm zusammensetzen, wie das englische *ennoble* und die deutschen Verben *beschweren*, *erblinden*, *entziffern* und *verarzten*. Williams selbst sowie Lieber (1981) und Selkirk (1984) fassen diese Fälle als Ausnahmen zur RHR auf und

plement und jeder Kopf wurde außerdem mit einem sog. Part-of-Speech-Tag, kurz POS-Tag, ausgezeichnet, also mit einer Kennzeichnung für seine jeweilige Wortart versehen. Dem POS-Tag des Kopfes ist ein Pluszeichen vorangestellt. Suffixe erhalten den POS-Tag für die jeweilige Wortart, die durch die Suffigierung entsteht. So wird ein Suffix wie *-heit*, das Nomen bildet, mit dem POS-Tag +NN ausgezeichnet.

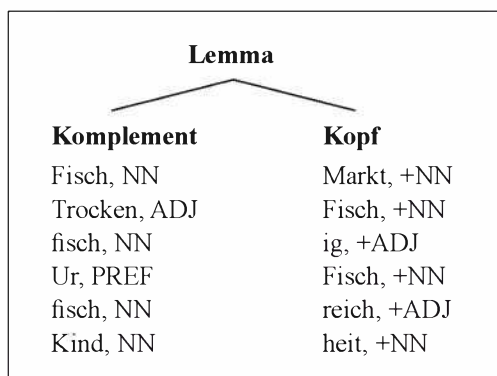


Abb. 1: Baumstruktur einiger Wörter mit ihrem jeweiligen POS-Tag

Ferner wurde mithilfe von Morphisto für jedes Stichwort die betreffende Wortbildungsregel ermittelt. Das Schema der Wortbildungsregeln ist immer $\langle +TAG \rangle \rightarrow \langle TAG \rangle \langle +TAG \rangle$. In zwei Spitzklammern eingeschlossen steht ein POS-Tag mit einem Pluszeichen, dann folgen ein Pfeil und darauf zwei weitere POS-Tags, die jeweils in Spitzklammern eingefasst sind. Die Wortbildungsregel für das Nomenkompositum *Fischmarkt* lautet $\langle +NN \rangle \rightarrow \langle NN \rangle \langle +NN \rangle$ und liest sich „ein Nomen besteht aus einem Nomen und einem Nomen“. Handelt es sich um Zusammensetzungen, so hat der zweite POS-Tag nach dem Pfeil, wie der vor dem Pfeil stehende Tag, ein vorangestelltes Pluszeichen. Das Plus symbolisiert den Kopf der Bildung. Die Zusammensetzung *fischreich* – ein Adjektiv aus dem Nomen *Fisch* und dem Adjektiv *reich* – erhält deshalb die Wortbildungsregel $\langle +ADJ \rangle \rightarrow \langle NN \rangle \langle +ADJ \rangle$. Das Adjektiv

lassen hier auch Präfixe als Köpfe zu, da es scheint, dass die Präfixe die Adjektive und Nomen zu Verben ableiten. Von Morphisto wurden diese Verben entweder gar nicht oder falsch analysiert und zerlegt. Um sie dennoch als Wortbildungsprodukte zu den jeweiligen Adjektiven oder Nomen ermitteln zu können, wurden sie von Hand korrigiert. Dazu wurde Scalises (1988, S. 240) Position übernommen: In einem (gedanklichen) Zwischenschritt wird durch Suffigierung ein mögliches aber nicht-existierendes Wort kreiert, das anschließend präfigiert wird. Auf diese Weise stellt Scalise sicher, dass Präfixe nie Köpfe sind, und die Präfixverben somit keine Ausnahme zur RHR darstellen.

als zweite Konstituente nimmt gemäß der „Right Hand Head Rule“ die Kopfposition der Bildung ein, weshalb die gesamte Zusammensetzung ein Adjektiv ist. In dem Adjektiv-Nomen-Kompositum *Trockenfisch* ist es umgekehrt: $\langle +NN \rangle \rightarrow \langle ADJ \rangle \langle +NN \rangle$. Hier ist das Nomen *Fisch* Kopf der Konstruktion, und damit ist auch die Zusammensetzung ein Nomen.

Ableitungen werden anders notiert. Bei Suffigierungen ist die zweite Tag-Konstituente in der Regel der POS-Tag $\langle SUFF \rangle$. Was das Resultat der Suffigierung ist, wird durch den Tag vor dem Pfeil ausgedrückt. Die Ableitung *Fischlein* hat die Regel $\langle +NN \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$, was bedeutet, dass es sich um ein Nomen handelt, das aus einem Nomen und einem (nomenbildenden) Suffix besteht. Das Adjektiv *fischig* wurde aus dem Nomen *Fisch* und dem Suffix *-ig* gebildet, weshalb die Regel $\langle +ADJ \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$ lautet. Das Suffix *-ig* macht aus dem Nomen *Fisch* ein Adjektiv. Analog zu $\langle SUFF \rangle$ lautet der POS-Tag für Präfixe $\langle PREF \rangle$, so dass *Urfisch* die Regel $\langle +NN \rangle \rightarrow \langle PREF \rangle \langle +NN \rangle$ erhält.¹⁵ In Tabelle 1 sind die in Abbildung 1 dargestellten Wörter mit ihrer jeweiligen Wortbildungsregel zusammengestellt.

Lemma	Wortbildungsregel
<i>Fischmarkt</i>	$\langle +NN \rangle \rightarrow \langle NN \rangle \langle +NN \rangle$
<i>Trockenfisch</i>	$\langle +NN \rangle \rightarrow \langle ADJ \rangle \langle +NN \rangle$
<i>fischig</i>	$\langle +ADJ \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$
<i>Urfisch</i>	$\langle +NN \rangle \rightarrow \langle PREF \rangle \langle +NN \rangle$
<i>fischreich</i>	$\langle +ADJ \rangle \rightarrow \langle NN \rangle \langle +ADJ \rangle$
<i>Kindheit</i>	$\langle +NN \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$

Tab. 1: Einige Wörter mit ihrer Wortbildungsregel

Ungebildete Stichwörter jeder Wortart werden mit einer Wortbildungsregel in der Form von $\langle +TAG \rangle \rightarrow$ Lemma notiert ($\langle +NN \rangle \rightarrow$ Kind; $\langle +ADJ \rangle \rightarrow$ grün; $\langle +V \rangle \rightarrow$ hören).

¹⁵ Für ein Präfixverb wie *erblinden* wurde in der manuellen Korrektur das Komplement *er-* mit dem Tag $\langle PREF \rangle$, der Head *blind* mit dem Head-Tag $\langle +V \rangle$ und die Regel $\langle +V \rangle \rightarrow \langle PREF \rangle \langle ADJ \rangle$ eingetragen. Um deutlich zu machen, dass das Adjektiv gemäß Scalise (1988) selbst nicht Kopf der Konstruktion ist, sondern in einem Zwischenschritt erst suffigiert wird, wird es in der Regel nicht mit einem Pluszeichen versehen. Die Regel für solche Präfixverben kann mit der Regel für Suffixableitungen verglichen werden. Die Tag-Konstituente lautet bei diesen $\langle SUFF \rangle$, aber $\langle SUFF \rangle$ ist nicht die Kategorie der gesamten Bildung. Mit dieser Notation wurde es möglich, das Verb *erblinden* als Ableitung zu dem Adjektiv *blind* zu extrahieren.

Die Analysen wurden in einer Tabelle einer relationalen Datenbank gespeichert. In den Spalten dieser Tabelle sind unter anderem das Lemma selbst, das Komplement, der POS-Tag des Komplements, der Kopf, dessen POS-Tag und die Wortbildungsregel gespeichert (vgl. Tab. 2).

Lemma	Komplement	Kompl_tag	Kopf	Kopf_tag	Regel
<i>Fisch</i>			<i>Fisch</i>	+NN	<+NN>→ <i>Fisch</i>
<i>Flussfisch</i>	<i>Fluss</i>	NN	<i>Fisch</i>	+NN	<+NN>→<NN> <+NN>
<i>Fischfabrik</i>	<i>Fisch</i>	NN	<i>Fabrik</i>	+NN	<+NN>→<NN> <+NN>
<i>Fischfabrikschiff</i>	<i>Fischfabrik</i>	NN	<i>Schiff</i>	+NN	<+NN>→<NN> <+NN>
<i>Trockenfisch</i>	<i>trocken</i>	ADJ	<i>Fisch</i>	+NN	<+NN>→<ADJ> <+NN>
<i>Urfisch</i>	<i>ur</i>	PREF	<i>Fisch</i>	+NN	<+NN>→<PREF> <+NN>
<i>fischig</i>	<i>Fisch</i>	NN	<i>ig</i>	+ADJ	<+ADJ>→<NN> <SUFF>
<i>fischen</i>	<i>Fisch</i>	NN	<i>en</i>	+V	<+V>→<NN> <SUFF>
<i>fischreich</i>	<i>Fisch</i>	NN	<i>reich</i>	+ADJ	<+ADJ>→<NN> <+ADJ>

Tab. 2: Vereinfachte Darstellung der Datenbanktabelle

Die Speicherung in dieser Basistabelle ermöglichte es, die Spalten der einzelnen Stichwörter nach bestimmten Mustern zu durchsuchen und so die verschiedenen Wortbildungsprodukte unterschiedlicher „Bauart“ zu ermitteln.

3.2 Chancen und Probleme der automatischen morphologischen Analyse

Der Vorteil einer automatischen morphologischen Analyse liegt darin, dass innerhalb kurzer Zeit eine große Menge an Wörtern – die *lexiko*-Stichwortliste enthält ca. 300 000 Einträge – analysiert werden kann. Von Hand ist eine solche Aufgabe nicht in angemessener Zeit zu leisten.

Auf der anderen Seite gibt es aber auch Probleme mit einer automatischen Analyse. Es kann vorkommen, dass Wörter nicht oder falsch analysiert sind. Das liegt daran, dass Morphisto diese Wörter nicht „kannte“ und somit auch keine Analyse und morphologische Zerlegung vornehmen konnte. Das trifft vor allem auf Fremdwörter wie *Toxoplasmose* oder sehr komplexe Wörter wie *Staatssekretärsrunde* zu. Neben fehlenden Zerlegungen kommen auch falsche

Zerlegungen vor, beispielsweise wurde *Kindergartensaal* in das Komplement *Kindergarten*, das Fugenelement *-s-* und den Kopf *Aal* zerlegt. Eine andere Problemstelle sind falsche Wortartauszeichnungen in den Tags. Ein Beispiel für eine falsche Auszeichnung und falsche Zerlegung ist das Wort *Herbstlaub*, das in der Datenbank als Zusammensetzung aus dem Adjektiv *herb* und dem Nomen *Laub* gespeichert ist.

3.3 Vorgehensweise zur Ermittlung der Wortbildungsprodukte

Ausgangspunkt der Ermittlung der Wortbildungsprodukte war für jede Wortart jeweils zunächst die Ermittlung der Simplizia. Das Ziel war, eine Datenbanktabelle zu erstellen, in der jedem Simplex alle in der Datenbank gespeicherten Komposita und Derivate der verschiedenen Wortarten zugeordnet sind. Von den Simplizia ausgehend sollten die Wortbildungsprodukte, die mit den Simplizia gebildet werden, aus der Datenbank ermittelt werden. Für die nominalen Simplizia wurde eine vereinfachte Arbeitsdefinition von Simplizia festgelegt. Nach dieser gelten alle Nomen, die nicht zerlegt werden können, als Simplizia. Deshalb fallen auch Konversionen wie *Tanz* oder *Druck* unter diese Definition. Exemplarisch soll hier die Ermittlung der Wortbildungsprodukte zu den nominalen Simplizia erläutert werden. Die Extraktion der Wortbildungsprodukte zu den adjektivischen und verbalen Simplizia erfolgte analog.

Die Simplizia zeichnen sich dadurch aus, dass sie kein Komplement haben. Diese Spalte ist also in der Datenbank bei Simplizia leer. Außerdem entspricht der Kopf dem Lemma, d.h., die Einträge in den Spalten 'Kopf' und 'Lemma' sind identisch. Nominale Simplizia haben den Kopf-Tag +NN, ihre Wortbildungsregel hat immer die Form <+NN> → Lemma, d.h., auf den Pfeil folgt keine weitere Spitzklammer, sondern direkt das jeweilige Simplex. Eine neue Spitzklammer würde ein gebildetes Wort bedeuten. Die Datenbanktabelle wurde deshalb auf diese Merkmale hin abgefragt und alle Stichwörter, auf die diese Bedingungen zutrafen, wurden ausgewählt und extrahiert. Die so ermittelten knapp 5 600 nominalen Simplizia wurden in einer eigenen Tabelle, der Simplextabelle, gespeichert.

Zur anschließenden Ermittlung der Nomenkomposita, die das Simplex enthalten, musste zunächst eine Fallunterscheidung vorgenommen werden: Das Simplex ist entweder Bestimmungswort, also die erste Konstituente, oder das Simplex ist Grundwort, also die zweite Konstituente. Im ersten Fall – das Simplex ist Bestimmungswort – sind die Simplizia aus der Simplextabelle in der Spalte Komplement der Basistabelle gespeichert, wie in *Fischfabrik* und *Fisch-*

markt zu *Fisch*. Im zweiten Fall, in dem das Simplex Grundwort ist, sind folglich die Simplizia aus der Simplextabelle in der Spalte ‘Kopf’ der Basistabelle, wie in *Brotfisch* und *Flussfisch* (vgl. Abb. 2).

Simplextabelle		Basistabelle			
Lemma	Lemma	Kompl	Kopf	Regel	
<i>Aal</i>	<i>Aal</i>		<i>Aal</i>	<+NN>→ <i>Aal</i>	
...	<i>atmen</i>		<i>atmen</i>	<+V>→ <i>atmen</i>	
<i>Brot</i>	<i>Beutfisch</i>	<i>Beute</i>	<i>Fisch</i>	<+NN>→<NN> <+NN>	
...	<i>bremsen</i>		<i>bremsen</i>	<+V>→ <i>bremsen</i>	
<i>Ei</i>	<i>Brotfisch</i>	<i>Brot</i>	<i>Fisch</i>	<+NN>→<NN> <+NN>	
...	<i>dunkel</i>		<i>dunkel</i>	<+ADJ>→ <i>dunkel</i>	
<i>Fisch</i>	<i>Fischfabrik</i>	<i>Fisch</i>	<i>Fabrik</i>	<+NN>→<NN> <+NN>	
...	<i>Fischmarkt</i>	<i>Fisch</i>	<i>Markt</i>	<+NN>→<NN> <+NN>	
<i>Haus</i>	<i>Floß</i>		<i>Floß</i>	<+NN>→ <i>Floß</i>	
...	<i>Flussfisch</i>	<i>Fluss</i>	<i>Fisch</i>	<+NN>→<NN> <+NN>	

Abb. 2: Vergleich von Simplex- und Basistabelle zur Ermittlung der Komposita am Beispiel *Fisch*

Dazu wurde eine Anfrage geschrieben, die die Simplextabelle mit der Basistabelle kombiniert. Genauer gesagt wurden die Spalten ‘Komplement’ bzw. ‘Kopf’ der Basistabelle mit der Spalte ‘Lemma’ der Simplextabelle verglichen. Sind sie identisch, dann liegen Determinativkomposita vor. In der Anfrage für Komposita mit dem Simplex als Grundwort muss neben der Gleichheit von Head in der Basistabelle und Lemma der Simplextabelle in einer zusätzlichen Bedingung ausgeschlossen werden, dass die Einträge der Spalte ‘Lemma’ der Basistabelle mit den Einträgen der Spalte ‘Head’ der Basistabelle identisch sind. Lediglich auf Gleichheit von Head und Lemma zu prüfen, würde nämlich – wie in der oben genannten Anfrage – in Simplizia resultieren. Die auf diese Weise ermittelten etwa 106 000 Komposita mit Simplex als Grundwort und die knapp 88 000 Komposita mit Simplex als Bestimmungswort wurden in zwei eigenen Tabellen gespeichert und jede Zusammensetzung wurde ihrem Simplex zugeordnet.

Auch Nomen-Adjektiv-Komposita, Adjektiv-Nomen-Komposita und nominale Derivate wurden nach diesem Muster ermittelt. Die Simplextabelle und die Basistabelle wurden kombiniert, wobei die Bedingungen immer die Gleichheit von Head bzw. Komplement der Basistabelle und dem Lemma der Simplexta-

belle und eine spezifische Wortbildungsregel erforderten. Im Falle der Komposita mit einem Nomen als Erstglied und den Suffixableitungen wurde die Gleichheit von Komplement der Basistabelle und Lemma der Simplextabelle erfragt und es wurden die Wortbildungsregeln $\langle +ADJ \rangle \rightarrow \langle NN \rangle \langle +ADJ \rangle$ (*fischreich*) bzw. $\langle +NN \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$ (*Fischlein*), $\langle +ADJ \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$ (*fischig*) oder auch $\langle +ADV \rangle \rightarrow \langle NN \rangle \langle SUFF \rangle$ (*löffelweise*) gefordert. Im Falle von Präfixableitungen und Komposita mit dem Nomen als Zweitglied lauteten die Regelbedingungen $\langle +NN \rangle \rightarrow \langle PREF \rangle \langle +NN \rangle$ (*Urfisch*) bzw. $\langle +NN \rangle \rightarrow \langle ADJ \rangle \langle +NN \rangle$ (*Trockenfisch*). Die verschiedenen Wortbildungen mit Adjektiv- und Verbsimplizia lassen sich ganz analog aus der Datenbank erfragen.

Außerdem wurde die Korpusfrequenz zu jedem Produkt ermittelt und in einer eigenen Spalte notiert. Die so genannte Wortbildungsrolle zeichnet jedes Simplex hinsichtlich seiner Rolle in dem Wortbildungsprodukt aus. Das Produkt *Fischabfall* zu *Fisch* erhält die Rolle ‘compound-c-nn’, da es sich um ein Kompositum aus zwei Nomen handelt, in dem das Simplex *Fisch* Bestimmungswort, d.h. Komplement ist. Entsprechend wird *Beutefisch* zu *Fisch* mit ‘compound-h-nn’ ausgezeichnet; das gleiche Kompositum erhält aber ‘compound-c-nn’, wenn es *Beute* zugeordnet ist. Ein Adjektiv wie *fischig* wird mit der Rolle ‘deriv-c-nadj’ versehen, denn das nominale Simplex nimmt in diesem Derivat die Rolle des Komplements ein. Das Resultat der Ableitung ist ein Adjektiv.

Lemma	Produkt	Produkt_Frequenz	Wb-Rolle
<i>Fisch</i>	<i>Anglerfisch</i>	6	compound-h-nn
<i>Fisch</i>	<i>Aquarienfisch</i>	30	compound-h-nn
<i>Fisch</i>	<i>Beutefisch</i>	23	compound-h-nn
<i>Fisch</i>
<i>Fisch</i>	<i>Fischabfall</i>	32	compound-c-nn
<i>Fisch</i>	<i>Fischadler</i>	93	compound-c-nn
<i>Fisch</i>
<i>Fisch</i>	<i>fischen</i>	3670	deriv-c-nv
<i>Fisch</i>	<i>fischig</i>	49	deriv-c-nadj
<i>Fisch</i>	<i>Fischindustrie</i>	68	compound-c-nn
<i>Fisch</i>

Tab. 3: Ausschnitt aus der Tabelle der Wortbildungsprodukte mit den Wortbildungsprodukten zum Lemma *Fisch*

Die Daten der einzelnen Tabellen werden in eine Gesamttabelle (vgl. Tab. 3) eingefügt, sodass jedem simplizischen Lemma alle Komposita und Derivate zugeordnet werden können, die in der *lexiko*-Stichwortliste enthalten sind.

Aus dieser Gesamttabelle der Wortbildungsprodukte können die Wortbildungsprodukte pro Stichwort extrahiert werden und stehen anschließend für die Online-Darstellung in einem passenden Format zur Verfügung.

3.4 Chancen und Probleme der automatischen Ermittlung der Wortbildungsprodukte aus einer Datenbank

Die Chancen der Speicherung von morphologisch analysierten Stichwörtern in einer Datenbank liegen darin, dass man mit einer einzigen Anfrage nach allen Wortbildungsprodukten suchen kann, die ein bestimmtes Lemma, z.B. hier *Fisch*, enthalten, und eine Liste wie in Tabelle 3 erhält. Lemmata, die durch Wortbildung umgelauteet werden, wie beispielsweise *ärztlich*, werden in der Datenbank zu ihrem Basislemma in Beziehung gebracht; im Falle von *ärztlich* also *Arzt*. Wo eine rein zeichenbasierte Suche umgelauteete Fälle nicht erfassen kann, ist dies durch die spezielle Speicherung in dieser Datenbank möglich, vorausgesetzt sie wurden zuvor von Morphisto korrekt segmentiert. Durch die Kombination von Tabellen können die vorhandenen Daten neu zusammengestellt und gruppiert werden. Die ermittelten Wortbildungsprodukte mit ihrer Zuordnung zu einem Lemma, den Frequenzinformationen und der jeweiligen Wortbildungsrolle sind in der eigens angelegten Datenbanktabelle für weitere Verarbeitungsschritte zugänglich. Auf diese Weise ist es möglich, die Wortbildungsprodukte in die einzelnen *lexiko*-Artikel online zu integrieren.

Die Schwierigkeiten und Probleme der Datenbank liegen vorrangig in ihrem Datenbestand. Die Datenbank ist nur so gut wie die Daten in ihr. Fehler, die auf Morphisto zurückgehen, resultieren in fehlerhaften Abfrageergebnissen. Wenn *Kindergartensaal* von Morphisto in *Kindergarten* und *Aal* zerlegt ist, dann wird dieses auch als Kompositum zu *Aal* ermittelt. Wurde *ärztlich* nicht korrekt in *Arzt* und *-lich* segmentiert, sondern vielleicht gar nicht analysiert, ist es nicht möglich, *ärztlich* als Adjektivableitung zu *Arzt* zu ermitteln. Dieses Wortbildungsprodukt würde demnach in der Liste der Wortbildungsprodukte zu *Arzt* fehlen. Um einen ersten Eindruck von der Fehlerzahl zu bekommen, wurden 88 verschiedene Nomensimplizia aus unterschiedlichen Frequenzbereichen ausgewählt und es wurde die Anzahl ihrer Komposita gezählt. Von den so ermittelten 6652 Komposita waren etwa 792 fehlerhaft analysiert. Das

sind 12% oder 9 falsch analysierte Komposita pro Nomensimplex. Bei anderen Wortbildungsprodukten war die Fehlerzahl noch größer. So waren 15% von 2 762 Nomen-Adjektiv-Komposita falsch zerlegt, von 5 866 Adjektiv-Nomen-Komposita waren 19% fehlerhaft. Eine Fehlerquelle lag bei den Präfixen *haupt-* und *vorder-*, die von Morphisto fälschlicherweise als Adjektive analysiert worden waren. Auch homographische Konstituenten und falsche Wortbildungsregeln trugen zu der hohen Fehlerzahl bei. Unter den 8 594 Komposita aus einem Verb und einem Nomen waren 22% inkorrekt. Auch hier waren Homographen die Hauptursache für falsche Zerlegungen. Die 1 262 Adjektivsimplizia waren mit 38 Fehlern auf 100 Wörter die Gruppe mit den meisten Fehlern. Das Hauptproblem waren fälschlich als Simplizia analysierte komplexe Adjektive. Soweit möglich wurden die fehlerhaften Zerlegungen und falschen Wortbildungsregeln von zwei studentischen Hilfskräften innerhalb von 14 Monaten manuell korrigiert. Aufgrund der automatischen morphologischen Analyse und der automatischen Ermittlung der Wortbildungsprodukte ist außerdem keine Zuordnung der Wortbildungsprodukte zu der jeweiligen Lesart der Ausgangslexeme möglich. Das hat Konsequenzen für die Art der Darstellung der Wortbildungsprodukte.

4. Online-Darstellung der Wortbildungsprodukte

4.1 Konzeptuelle Überlegungen

Für die Darstellung der Wortbildungsprodukte im *ellexiko*-Stichwortartikel stellen sich die Fragen,

- wo die Wortbildungsprodukte platziert werden bzw. wo der Hinweis auf sie platziert wird,
- wie deutlich wird, dass sie automatisch ermittelt wurden,
- wie sie gruppiert/sortiert werden,
- wie viele gezeigt werden.¹⁶

Auf die automatisch ermittelten Wortbildungsprodukte soll auf der Artikeleingangsseite hingewiesen werden. Da sie – wie bereits erwähnt – nur lesartenübergreifend angezeigt werden können, muss die entsprechende Überschrift, wie die Orthographie, zu den lesartenübergreifenden Angaben gesetzt werden. Sowohl bei bearbeiteten (vgl. Abb. 3, oberer Teil) als auch bei nicht bearbei-

¹⁶ Zu Überlegungen für die grafische Darstellung von Wortbildungsbeziehungen vgl. den Beitrag „Visualisierung von Wortbildungsbeziehungen im elektronischen Wörterbuch“ von Peter Meyer und Carolin Müller-Spitzer in diesem Band.

teten Artikeln (vgl. Abb. 3, unterer Teil) wird die Überschrift an dieser Stelle platziert. Analog zur Überschrift ‘Belege (automatisch ausgewählt)’ – wie sie in unbearbeiteten Artikeln steht – lautet die Überschrift für die Wortbildungsprodukte ‘Wortbildungsprodukte (automatisch ermittelt)’.



Abb. 3: Artikeleingangsseiten des bearbeiteten Stichworts *Fisch* (oben) und des unbearbeiteten Stichworts *Haus* (unten) mit dem Hinweis auf die automatisch ermittelten Wortbildungsprodukte

Ein so genannter Info-Knopf neben dem Hinweis führt zu einem Informationstext im Glossar. Der Informationstext erläutert in knappen Worten die morphologische Analyse mit Morphisto und wie die Wortbildungsprodukte ermittelt wurden. In diesem Text wird besonderer Wert darauf gelegt, herauszustellen, dass die Wortbildungsprodukte automatisch ermittelt wurden und bereits auf einer automatischen morphologischen Analyse beruhen. Hier wird auch auf die daraus resultierenden Probleme hingewiesen. Im Unterschied zu den automatisch ausgewählten Korpusbelegen werden die Wortbildungsprodukte jedoch nicht auf der Eingangsseite gezeigt, sondern der Link „weiter“ öffnet eine neue Ansicht, in der die Wortbildungsprodukte präsentiert werden.

In der neuen Ansicht werden die Wortbildungsprodukte getrennt nach ihrer jeweiligen Wortbildungsart auf verschiedene Registerkarten verteilt angezeigt, d.h., es gibt eine Registerkarte für Komposita, eine für Derivate und eine für weitere Wortbildungsprodukte. Innerhalb der Registerkarten für die einzelnen Wortbildungsarten werden die Wortbildungsprodukte nach ihren Bestandteilen bzw. nach dem Resultat ihrer Wortbildung sortiert. Das heißt, in der Registerkarte ‘Komposita’ (vgl. Abb. 4) gibt es für simplizische Nomen die Rubriken ‘Nomen-Nomen-Komposita’, ‘Adjektiv-Nomen-Komposita’, ‘Nomen-Adjektiv-Komposita’ und ‘Verb-Nomen-Komposita’. Die Nomen-Nomen-Komposita werden zusätzlich in zwei Spalten präsentiert, in der einen Spalte die Komposita mit dem Simplex als Bestimmungswort, in der anderen die Komposita mit dem Simplex als Grundwort.

Fisch
Wortbildungsprodukte

zur Übersichtsseite

Komposita Derivate Andere Wortbildungsprodukte

Nomen und Nomen als zweiter Bestandteil		Nomen und Nomen als erster Bestandteil	
alphabetisch ▲	Häufigkeit	alphabetisch ▲	Häufigkeit
Anglerfisch	14	Fischabfall	47
Aquarienfisch	86	Fischabteilung	34
Austernfisch	2	Fischadler	301
Beutfisch	31	Fischangebot	57
Brotfisch	41	Fischart	1984
Diskusfisch	36	Fischauge	60
	mehr >		mehr >

Adjektiv und Nomen		Nomen und Adjektiv	
alphabetisch ▲	Häufigkeit	alphabetisch ▲	Häufigkeit
Blindfisch	151	fischähnlich	29
Edelfisch	292	fischarm	6
Frischfisch	250	fischreich	308
Jungfisch	494		
Kleinfisch	128		
Plattfisch	102		
Trockenfisch	56		
Weißfisch	292		

Verb und Nomen	
alphabetisch ▲	Häufigkeit
Backfisch	1024
Brätfisch	114
Köderfisch	130
Nutzfisch	15
Räucherfisch	257
Saugfisch	3
Wanderfisch	351

Informationen zu Wortbildungsprodukten

Abb. 4: Ansicht für die automatisch ermittelten Wortbildungsprodukte zum Lemma *Fisch* mit geöffneter Registerkarte für Komposita

Innerhalb der Registerkarte ‘Derivate’ (vgl. Abb. 5) werden die Ableitungen nach der resultierenden Wortart, also Nomen (beispielsweise *Kinderei* zu *Kind*), Adjektive (*kindlich*, *kindisch*) und Verben (*fischen* zu *Fisch*), gruppiert.

The screenshot shows the 'Kind' register card in the 'ellexiko' interface. The 'Derivate' tab is active, displaying frequency data for various word formation products. The data is organized into three sections: Nomen with prefix/suffix, Nomen with suffix, and Adjektive. Each section lists words and their absolute frequencies in the 'ellexiko' corpus.

Nomen mit Präfix/Konfix		Nomen mit Suffix	
alphabetisch	Häufigkeit	alphabetisch	Häufigkeit
Schwiegerkind	2319	Kinderei	229
Stiefkind	2159	Kindheit	32313
		Kindschaft	17

Adjektive	
alphabetisch	Häufigkeit
kinderlos	4319
kindgemäß	577
kindhaft	85
kindisch	1581
kindlich	9457

Abb. 5: Die Ansicht für die automatisch ermittelten Wortbildungsprodukte zum Lemma *Kind* mit geöffneter Registerkarte für Derivate

Zu jedem Wortbildungsprodukt wird seine Frequenz im *ellexiko*-Korpus angegeben. Die Problematik von absoluten Häufigkeiten bei einem stetig wachsenden Korpus wurde diskutiert. Relative Häufigkeiten geben jedoch ebenfalls nur den Stand zu einem bestimmten Zeitpunkt an und sind bei einem Korpus von ca. 2,8 Milliarden Textwörtern außerdem sehr klein: *Stiefkind* aus Abbildung 5 hat eine relative Häufigkeit von etwa $3,76e^{-7}$, d.h. 0,000000375714285714285; eine Zahl, die wohl für die meisten Nutzerinnen und Nutzer nicht mehr zu erfassen ist. Eine andere Alternative zu absoluten Häufigkeiten ist die Einordnung der Wortbildungsprodukte in Frequenzklassen. Die Wortbildung *Stiefkind* würde nach der *ellexiko*-Klassifikation der Frequenzschicht VII angehören.¹⁷ Um diese Information allerdings interpretieren zu können, müssen Nutzerinnen und Nutzer eine weitere Nachschlagehandlung durchführen. Aus diesen Gründen wurde entschieden, absolute Häufigkeiten anzugeben. Die Produkte können sowohl alphabetisch als auch absteigend nach Frequenz sortiert werden. In einer Benutzerstudie zeigte sich eine Präferenz für die alphabetische

¹⁷ Für die Definition der Frequenzschichten in *ellexiko* vgl. <http://www.owid.de/wb/ellexiko/glossar/Frequenzschichten.html> (Stand: 08/2011).

Sortierung,¹⁸ weshalb diese als Default-Einstellung gewählt wurde. Alle Wortbildungsprodukte sind mit ihren jeweiligen Stichworteinträgen verlinkt, wo weitere Informationen verfügbar sind. Es handelt sich bei dieser Darstellung jedoch nicht um eine Nest- oder Nischenanordnung von Sublemmata, wie sie in manchen Printwörterbüchern zur Darstellung von Wortbildungszusammenhängen gewählt wird. In der für *elexiko* gewählten Präsentation der Wortbildungsprodukte werden die als Hyperlinks realisierten Lemmata nach Gesichtspunkten der Wortbildung gruppiert; man kann also am ehesten von wortbildungsbestimmten Verweislemma-Gruppen (Engelberg/Lemmitzer 2009, S. 150) sprechen.

4.2 Chancen und Probleme bei der Darstellung der automatisch ermittelten Wortbildungsprodukte

Die gewählte Darstellung zeigt die Wortbildungsaktivität eines Stichwortes. Die Nutzer erfahren so, welches Potenzial das betreffende Stichwort in der Wortbildung hat. Mit der Trennung in die beiden Wortbildungsarten Komposition und Derivation und dort jeweils mit der Gruppierung gemäß ihren Bestandteilen bzw. in die unterschiedlichen Wortbildungsergebnisse werden die vielfältigen Möglichkeiten zur Wortbildung sichtbar, die sich mit diesem Stichwort ergeben. Die Auflistung aller in der Stichwortliste vorhandenen Wortbildungen, in die das nachgeschlagene Stichwort eingeht, zeigt nicht nur, wie aktiv dieses Stichwort in der Wortbildung ist, sondern macht besonders auch Zusammenhänge zwischen den Stichwörtern deutlich, die sonst nicht hervortreten würden. Die jeweils angegebene Korpusfrequenz liefert zugleich erste Informationen über das jeweilige Produkt, aber auch über das nachgeschlagene Stichwort selbst. Die Korpusfrequenzen der Wortbildungsprodukte zeigen nicht nur, welche Wortbildungsprodukte das Stichwort besonders häufig bildet, sondern sie können gerade bei Komposita einen Einblick in die Diskurse geben, in denen das betreffende nachgeschlagene Stichwort in einer bestimmten Lesart verwendet wird. Folgen die Nutzer den Hyperlinks, gelangen sie zu den Wortartikeln der Wortbildungsprodukte und erhalten dort (im Falle von bearbeiteten Stichwörtern) ausführliche Informationen zur Bedeutung des Stichworts, typische Verwendungsmuster, lexikalische Mitspieler, sinnverwandte Wörter und Grammatikangaben. Die Verlinkung ist gerade für Lerner sehr nützlich, da sie unmittelbar Antworten auf ihre spezifischen Fragen, wie beispielsweise nach dem Genus des Produkts und dessen morphologischem Bauplan, finden.

¹⁸ Vgl. hierzu den Beitrag „Die Wortbildungsangaben im Online-Wörterbuch und wie Nutzer sie beurteilen – eine Umfrage zu *elexiko*“ von Antje Töpel in diesem Band.

Allerdings führen Fehler, die auf Morphisto beruhen, – wie gezeigt – zu fehlerhaften Abfrageergebnissen, was wiederum zu Fehlern und Mängeln in der Darstellung führt. Wurde die falsche Zerlegung von *Kindergartensaal* nicht korrigiert, wird dieses Kompositum in der Registerkarte ‘Komposita’ unter der Rubrik ‘Nomen-Nomen-Komposita’ in der Spalte ‘als Grundwort’ im Artikel des Simplexes *Aal* angezeigt statt korrekt unter *Saal*. Wurde eine Wortbildung wie *ärztlich* gar nicht analysiert oder statt *Arzt* in das fiktive Komplement *Ärzt* und *-lich* zerlegt, so fehlt *ärztlich* in der Registerkarte ‘Derivate’ unter der Rubrik ‘Adjektive’ im Artikel *Arzt*. Ein weiteres der bereits erwähnten technischen Probleme setzt sich auch auf Darstellungsebene fort: die Zuordnung der Wortbildungsprodukte zu den Lesarten ihrer jeweiligen Ausgangslexeme. So werden die Komposita *Zugreise*, *Charakterzug*, *Musikzug*, *Atemzug*, *Schachzug* und *Seilzug*, die das Lemma *Zug* in jeweils unterschiedlichen Lesarten verwenden, alle lesartenübergreifend im Artikel *Zug* präsentiert. Die Zuordnung zu den Lesarten müssen die Nutzerinnen und Nutzer selbst vornehmen. Bei der Zusammensetzung *Zugspitze* kann es sich sowohl um die Spitze eines Zuges in der Lesart ‘Bahn’, um die Spitze eines Zuges der Lesart ‘Gruppe’ als auch um einen Eigennamen, nämlich den Namen von Deutschlands höchstem Berg, handeln.

5. Fazit

In diesem Beitrag wurde mit einem Anwendungsbeispiel eines automatischen morphologischen Analyse-Tools in der Internetlexikographie eine neue Methode zur Darstellung von Wortbildungsbeziehungen im elektronischen Wörterbuch vorgestellt. Ein computerlinguistisches Werkzeug wurde genutzt, um eine sprachliche Ressource zu erstellen, die neue Möglichkeiten in der Wörterbucherstellung eröffnet. Mit der morphologisch analysierten Stichwortliste wurden die Wortbildungsprodukte zu simplizischen Nomen, Adjektiven und Verben durch Datenbankabfragen systematisch und umfassend ermittelt, um sie online in die Wortartikel der jeweiligen Simplizia integrieren zu können. In jedem Bearbeitungsschritt lagen Chancen und Probleme, die hier angesprochen wurden.

Trotz der in der automatischen morphologischen Analyse und in der automatischen Ermittlung der Wortbildungsprodukte begründeten Probleme bietet sich auf diese Weise die Möglichkeit, die Daten des Wörterbuchs stärker miteinander zu vernetzen. Die Nutzer erhalten mit der gewählten Darstellung einen Einblick in das Wortbildungssystem und haben damit Material zur eigenen Wortschatzerweiterung zur Hand. Der von Holly (1986, S. 205) beklagten „Atomisierung“ des Wortschatzes im Wörterbuch und „in den Köpfen der

Benutzer“ wird mit der im vierten Kapitel vorgestellten Präsentation der Wortbildungsprodukte entgegengewirkt. Die „totale Herrschaft des Alphabets“ (Wiegand 1977, S. 102) wird zumindest teilweise aufgehoben. Das zeigt sich interessanterweise gerade in der alphabetischen Sortierung der Bestimmungswortkomposita: In der am linken Rand stets sichtbaren striktalphabetischen Stichwortliste von *elexiko* liegen zwischen den Komposita *Zugabteil* und *Zuganschluss* das Stichwort *Zugang* und seine 35 Komposita und Derivate, sodass die Wortbildungsaktivität von *Zug* nicht direkt hervortritt. Im Reiter „Komposita“ des Simplexes *Zug* wird in der alphabetischen Sortierung der Bestimmungswortkomposita die Liste der Komposita mit *Zug* als Bestimmungswort nicht unterbrochen. Auf *Zugabteil* folgen direkt *Zuganschluss* und die anderen Komposita mit *Zug* als Bestimmungswort. Die Verlinkung zu den jeweiligen Stichworteinträgen lässt die angegebenen Wortbildungsprodukte nicht „unkommentiert“ (van der Colff 1998, S. 199), denn die Nutzer gelangen über den Link zu den gerade für Lerner wichtigen morphosyntaktischen und grammatischen Angaben. Erst so dienen die Wortbildungsprodukte der Textproduktion.

Die Präsentation aller automatisch ermittelten Wortbildungsprodukte, in die ein Lemma eingeht, macht nicht nur das Wortbildungspotenzial eines Lemmas sichtbar, sondern offenbart auch Wortbildungszusammenhänge, die in einem gedruckten Wörterbuch nicht hervortreten (können). In einem Online-Wörterbuch wie *elexiko* bietet sich dagegen die Möglichkeit, auch auf Wortbildungsebene die Stichwörter miteinander zu vernetzen.

6. Literatur

6.1 Wörterbücher

elexiko (2003ff.). In: Institut für Deutsche Sprache (Hg.): OWID – Online-Wortschatz-Informationssystem Deutsch. Mannheim. Internet: <http://www.elexiko.de>
Stand: (08/2011).

6.2 Forschungsliteratur

Barz, Irmhild (2001): Wortbildungsbeziehungen im einsprachigen Bedeutungswörterbuch. In: Korhonen, Jarmo (Hg.): Von der mono- zur bilingualen Lexikografie für das Deutsche. Frankfurt a.M., S. 85-100.

Engelberg, Stefan/Lemnitzer, Lothar (2009): Lexikographie und Wörterbuchbenutzung. 4., überarb. u. erw. Aufl. Tübingen.

Haß, Ulrike (Hg.) (2005): Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York.

- Haß, Ulrike (2005): *elexiko* – Das Projekt. In: Haß (Hg.), S. 1-17.
- Holly, Werner (1986): Wortbildung und Wörterbuch. In: *Lexicographica* 2, S. 195-213.
- Klosa, Annette (2005): Wortbildung. In: Haß (Hg.), S. 141-162.
- Klosa, Annette (2010): On the combination of automated information and lexicographically interpreted information in two German online dictionaries. In: Granger, Sylviane/Paquot, Magali (Hg.): *eLexicography in the 21st century. New challenges, new applications. Proceedings of eLex 2009*. Louvain-la-Neuve, S. 157-163.
- Klosa, Annette (Hg.) (2011): *elexiko* – Erfahrungsberichte aus der lexikografischen Praxis eines Internetwörterbuchs. (= Studien zur Deutschen Sprache 55). Tübingen.
- Lieber, Rochelle (1981): *On the organization of the lexicon*. Bloomington.
- Scalise, Sergio (1988): The notion of ‘head’ in morphology. In: Booij, Geert/van Marle, Jaap (Hg.): *Yearbook of morphology*. Dordrecht, S. 229-245.
- Schmid, Helmut/Fitschen, Arne/Heid, Ulrich (2004): SMOR: A German computational morphology covering derivation, composition, and inflection. In: *Proceedings of the IVth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon, Portugal, S. 1263-1266. Internet: www.lrec-conf.org/proceedings/lrec2004/pdf/468.pdf (Stand: 11/2011).
- Schnörch, Ulrich (2005): Die *elexiko*-Stichwortliste. In: Haß (Hg.), S. 71-90.
- Selkirk, Elisabeth (1982): *The syntax of words*. Cambridge, MA.
- Storjohann, Petra (2005): Das *elexiko*-Korpus: Aufbau und Zusammensetzung. In: Haß (Hg.), S. 55-70.
- van der Colff, Adri (1998): Die Komposita in Langenscheidts Großwörterbuch Deutsch als Fremdsprache. In: Wiegand, Herbert Ernst (Hg.): *Perspektiven der pädagogischen Lexikographie des Deutschen I*. (= *Lexicographica. Series Maior* 86). Tübingen, S. 193-207.
- Wiegand, Herbert Ernst (1977): Nachdenken über Wörterbücher: Aktuelle Probleme. In: Drosdowski, Günther/Henne, Helmut/Wiegand, Herbert Ernst (Hg.): *Nachdenken über Wörterbücher*. Mannheim, S. 51-102.
- Williams, Edwin (1981): On the notions ‘lexically related’ and ‘head of word’. In: *Linguistic Inquiry* 12, 2, S. 245-274.
- Zielinski, Andrea/Simon, Christian/Wittl, Tilman (2009): Morphisto: Service-oriented open source morphology for German. In: Mahlow, Cerstin/Piotrowski, Michael (Hg.): *SFCM 2009, CCIS 41*. Heidelberg, S. 64-75.

6.3 Internetressourcen

BZ*Velexiko* – Benutzeradaptive Zugänge und Vernetzungen in *elexiko*. Internet: <http://www.ids-mannheim.de/lexik/BZVelexiko/> (Stand: 08/2011).

elexiko. Internet: <http://www.ids-mannheim.de/lexik/elexiko/> (Stand: 08/2011).

Morphisto. Internet: <http://www.ids-mannheim.de/lexik/TextGrid/morphisto.html> (Stand: 08/2011).

TextGrid. Internet: <http://www.textgrid.de> (Stand: 08/2011).