

Wissen über Wörter – Die Mikrostruktur als DTD*

Sonja Müller-Landmann

Zusammenfassung

Das hypermediale Informationssystem *Wissen über Wörter* des Instituts für Deutsche Sprache basiert auf der Neuentwicklung einer Wörterbuchstruktur speziell für die digitalen Medien. Mit einer linguistisch-lexikologisch fundierten XML-Mikrostruktur werden die quantitativen und qualitativen Möglichkeiten der computergestützten Lexikografie, u. a. Speicherung, Vernetzung, Recherchierbarkeit, genutzt und gleichzeitig die Inhalte in einer beliebigen (partiellen) Detailliertheit und Ausführlichkeit – weit über das hinaus, was in Printwörterbüchern zu finden ist – für Lexikografen bearbeitbar wie für Rezipienten verfügbar sein. Dieser Beitrag zeigt exemplarisch Details der Mikrostruktur anhand eines Lexikonartikels für eine Mehrworteinheit.

3.1. Einleitung

Die für das lexikalisch-lexikologische, korpusbasierte Informationssystem des Instituts für Deutsche Sprache *Wissen über Wörter* (WiW) entwickelte XML-Dokumenttypdefinition (DTD) modelliert die Mikrostruktur der Wörterbuchartikel speziell für die Erarbeitung, Speicherung und Darstellung in den digitalen Medien. Sie fungiert als Schnittstelle zwischen lexikografischem Arbeitsplatz und Datenbank sowie zwischen Datenbank und Internet-Browser. Dabei bietet XML¹ in Kombination mit einer hypermedialen Benutzeroberfläche alle Vorteile einer Vernetzung des lexikalischen Materials durch typisierte, hybride (X)Links sowohl innerhalb der WiW-Mikro- und Makrostruktur als auch von und zu externen Ressourcen, wie multimedialen Erläuterungen, lexikologischen Abhandlungen, dynamischen Prozessen (z. B. Korpusrecherchen), externen Informationssystemen etc.

Um die Informationsauswahl und -tiefe sowohl bei der Recherche durch die Adressaten von WiW als auch bei der Bearbeitung durch die Lexikografen interaktiv und flexibel zu gestalten, ermöglicht die erarbeitete Mikrostruktur die abgestufte Navigation in einem mehrdimensionalen Suchraum: Von der Suche nach einzelnen lexikalischen Einheiten bis zur Kollektion von Untermengen des Wortschatzes, die durch frei definierbare (z. B. onomasiologische oder semasiologische) Kombinationen von Merkmalen und deren Ausprägungen konstituiert werden.

Dementsprechend ist die in WiW abzubildende Lemmastrecke zunächst in Einträge zu Einwort-, Mehrwort- und Wortelementlemmata unterteilt, mit einer jeweils partiell eigenständigen Informationsstruktur. Die gewählte Perspektive ist allerdings die der Lesart: Deskriptive Angaben

* Erschienen in: *Proceedings der GLDV-Frühjahrstagung 2001*, Henning Lobin (Hrsg.), Universität Gießen, 28.–30. März 2001, Seite 31–40. <http://www.uni-giessen.de/fb09/ascl/gldv2001/>

¹ <http://www.w3.org/TR/REC-xml>

werden zu denjenigen lexikalischen Einheiten gemacht, die sich durch ihre Position in unterschiedlichen Paradigmen bzw. Syntagmen/Kollokationsbezügen voneinander abgrenzen lassen.

Die lexikalisch-lexikologischen Informationen zu jeder Lesart bündelt WiW in vier strukturierten Dimensionen:

- **Bedeutung & Verwendung** mit Angaben zu Semantik (Bedeutungsparaphrase, semantische Klasse, Argumentstruktur, ontologische Referenz, Paradigmatik) und Pragmatik (Diatopik, Domäne, Textspezifik, Fachsprache, Diskursspezifika, Situation, Sprechereinstellung)
- **Grammatik** mit Angaben zu Flexion, Valenz, Argumentstruktur, Wortbildung und syntaktischem Verhalten
- **Geschichtliches & Sachliches** mit Angaben zu Etymologie und Enzyklopädischem
- **Kritisches & Normatives** mit Angaben zu sprachlichen Normen/sprachkritischen Bewertungen und Kategorisierungen in den Bereichen Semantik, Pragmatik, Orthographie, Stil etc.

Dem Lemma bzw. Formativ zugeordnet sind die Angaben der fünften Informationsdimension **Schreibung & Aussprache** (v. a. Schreibung, Silbentrennung und Aussprache) sowie Listen signifikanter **Kookkurrenzpartner** und die **Polysemieangabe**, die die konzeptuellen Beziehung zwischen den Lesarten des Lemmas diskutiert.

Weitere Informationen zum informationstechnischen Hintergrund befinden sich in Müller-Landmann (2000) und Haß-Zumkehr (2001), Ausführungen zu Mehrworteinheiten in WiW in Steyer (2000), Allgemeineres und Aktuelles zum Projekt ist unter <http://www.ids-mannheim.de/wiw> und in Fraas und Haß-Zumkehr (1998) zu finden. Im Internet kann auch die hinsichtlich der linguistischen Aspekte vollständige Wörterbuch-DTD heruntergeladen und somit alle Unterscheidungen, Weiterverzweigungen, Attribut-Wertemengen etc. nachvollzogen werden, die im Rahmen dieses Beitrags natürlich nicht vollständig erwähnt werden können.

Im Folgenden werde ich die lexikografische Bearbeitung einiger zentraler Angaben zur Mehrworteinheit „Bären aufbinden“ exemplarisch vorführen.

3.2. Lemmaselektion

Die Auswahl und die Beschreibung der Lemmata und Lesarten in WiW orientiert sich maßgeblich am Sprachgebrauch, wie er sich in den elektronischen Korpora des IDS manifestiert.

Um die Aktualität des Datenmaterials zu gewährleisten, werden die Angaben auf der Grundlage der Textkorpora des IDS – von z. Z. ca. 860 Millionen in elektronischer Form erfassten Wortformen *geschriebener* Sprache in klassifizierten Texten belletristischer, journalistischer, (populär-)wissenschaftlicher und sonstiger Art des Zeitraums von Ende des siebzehnten Jahrhunderts bis heute und ca. 1,5 Millionen Wortformen *gesprochener* Sprache aus Diskussionen, Interviews, Vorträgen, Berichten, Erzählungen, Reportagen der 1960er und 1970er Jahre – kontinuierlich dokumentiert und fortgeschrieben. Dazu stehen den Lexikografen wie auch den Benutzern von WiW diese Korpora und entsprechende Recherche- und Analysewerkzeuge² zur Verfügung.

² Diese Werkzeuge wurden am IDS von Cyril Belica entwickelt, siehe auch <http://www.ids-mannheim.de/kt/corpora.shtml>.

So werden auch Mehrworteinheiten als usuelle Wortverbindungen mit den empirischen Methoden der Korpusanalyse, v. a. der statistischen Kookkurrenzanalyse, identifiziert: „usuelle Wortverbindungen [...] sind über das Einzelwort hinausgehende sprachliche Erscheinungen, die als komplexere Einheiten reproduziert werden können und deren Elemente einen höheren Wahrscheinlichkeitsgrad des Miteinandervorkommens besitzen, als das bei okkasionellen Wortverbindungen der Fall ist.“ (Steyer, 2000, S. 108)

Es wird also unterschieden zwischen einzelwortbezogenen Kookkurrenzen – wie Kollokationen (binäre Relationen zwischen Autosemantika), usuellen synsemantischen Kookkurrenzpartnern oder usuellen phrasalen Mustern – einerseits und Mehrwortverbindungen andererseits – das sind neben Phraseologismen auch schwach- bzw. nicht-idiomatische usuelle Mehrwortverbindungen wie kommunikative Formeln, Zitate, Sprichwörter, Funktionsverbgefüge, synsemantische Fügungen etc.

Ausgehend von der Kookkurrenzanalyse der Einzelwörter lassen sich demnach an Hand der Listen signifikanter Partner Kandidaten für Mehrworteinheiten finden: Bei *aufbinden* finden wir als statistisch signifikantesten Kollokationspartner *Bären*, an zweiter Stelle *Bär*, es folgen – sortiert nach dem Grad lexikalischer Kohäsion – *lassen*, *Schnürsenkel*, *hat*, *erzählte*, *will*, *Schuhe*, ...

Durch intellektuelle Bearbeitung durch den Lexikografen, durch automatisierte kontextsensitive Recherche zur Erhebung phrasaler Muster und durch Validierung der Verbindung durch die Kookkurrenzanalysen der jeweiligen potenziellen Partner können *Bären* und *Bär* als Mehrwortkandidaten eingeordnet werden: *aufbinden* erscheint als signifikanter Partner auch in deren Kookkurrenzliste. *Schnürsenkel* und *Schuhe* weisen *aufbinden* ihrerseits nicht als signifikanten Partner aus.³ Sie eignen sich demnach eher dazu die Selektionsrestriktionen auf den semantischen Rollen der Argumentpositionen von *aufbinden* zu beschreiben, während *lassen*, *wollen* oder *können* etwas über die pragmatische Verwendung des Verbs aussagen.

3.3. Informationen zum Lemma

Nachdem also *Bären aufbinden* als usuelle Wortverbindung durch statistische Signifikanz identifiziert ist und zudem eine idiomatische Qualität festgestellt werden kann, die die Verbindung von *Bär* und *aufbinden* über den Status einer einzelwortbezogenen Kollokation zwischen zwei Autosemantika hinaus begründet, kann es als Mehrworteinheit beschrieben werden.

Das bedeutet zunächst die Angabe der Lemmazeichengestalt <LzGA>, der Nennform, des Formativs. Die Auffassungen darüber, was die geeignete Nennform einer Mehrworteinheit ist, gehen weit auseinander. (Dobrovolskij, 1993, S. 59) spricht hier vom „Haupt-Dilemma der Phraseologie [...], u. zw. die lexikographische Universalität der Nennform, die alle möglichen Realisationen abdeckt, dabei aber selbst oft sehr befremdend anmutet, vs. die psychologisch reale und damit benutzerfreundlichere Aussageform, die aber selten imstande ist, alle denkbaren Realisierungsnuancen zu vertreten.“

Zwar wird durch die informationstechnische Speicherung und Verarbeitung des Wortmaterials eine deterministische Zuordnung der Beschreibung der Mehrworteinheit unter ein oder mehrere ihrer Basiselemente mit all ihrer häufig beklagten Inkonsistenz und Redundanz unnötig, da ein und derselbe Wörterbuchartikel nach verschiedensten Kriterien abrufbar und multipel und variabel zuzuordnen ist, doch bleibt die lexikografische Realisierung der Gestalt einer solchen Einheit und die lexikologische Begründung dieser Gestalt ein kritischer Punkt.

³ Interessanterweise ist der signifikanteste Kollokationspartner von *Schuhe* „schieben Schuld“!

So fordert Korhonen (1990) als Nennform das vollständige Phrasenmuster mit obligatorischer und fakultativer Valenz – in unserem Fall z. B. *jmd. bindet jmdm. einen Bären auf*, während Dobrovolskij (1993) in einer Mischform bis zu vier Konstituenten (v. a. Autosemantika) als „Stichwörter“ der Wörterbuch-üblichen Nennform und einer sog. Aussageform voranstellt, wie auch Dormeyer et al. (1998) eine Wörterbuch-übliche Nennform mit einer Liste der Basiselemente indizieren. Chang und Fischer (2000, S. 37) wiederum gehen noch über diesen Ansatz hinaus: Für *einen Korb bekommen/geben/austeilen/erhalten/...* setzten sie eine metaphorische Lesart von *Korb* an, „employing a locally defined metaphor to capture the link between basket-transfer and rejection. It adds one more layer of flexibility since it is not lexically specific with respect to the verb of transfer, depending only on the presence of a transfer event with the basket as the object of transfer.“

WiW wählt in ähnlicher Weise als Nennform die invarianten lexikalischen und eventuelle, die Mehrworteinheit konstituierenden, metalexikalischen Bestandteile der Wortverbindung (wie z. B. das – semantisch begrenzt – variable Verb in *Korb* + VERB⁴). Im Beispiel bleiben demnach die beiden lexikalischen Invarianten *Bären aufbinden*.⁵ Auf den ersten Blick mag eine solche Lösung zu abstrakt erscheinen, doch bietet der hypermediale Wörterbuchartikel – als konsequente technische Fortentwicklung einer Datenbank wie der Deutschen Idiom-Datei Dobrovolskijs (Dobrovolskij, 1993) – die phraseografische Möglichkeit alle Information zu typischen Verwendungs- und Phrasenmustern aufzuführen, aufzufinden und an beliebiger Stelle anzuzeigen.

In unserem Beispiel wird also der Lemmatyp <MehrWortLemma> gewählt, die Lemmazcichengestaltangabe <LzGA> bestimmt und der Typus der Mehrworteinheit <MWKla>/<Phraseologismus> festgehalten.

```
<MehrWortLemma><LzGA>Bären aufbinden</LzGA>
<MWKategorisierung><TXTobj>Bären aufbinden</TXTobj> ist ein
<MWKla><Phraseologismus typ="unmarkiert">verbaler Phraseologismus
</Phraseologismus></MWKla>.</MWKategorisierung>
...
<SchAus><FormVA>Bär aufbinden
<AngabeZusatz><VerwA>Diese Form ist in nur knapp 8 Prozent der
Belege realisiert.</VerwA></AngabeZusatz></FormVA></SchAus>
...
<Kookkurrenz><AutosemKp>
<NominalKp>Wähler</NominalKp>
<NominalKp>Presse</NominalKp>
<NominalKp>Öffentlichkeit</NominalKp>
<NominalKp>Publikum</NominalKp>
<Verbalkp>lassen</Verbalkp>
<Verbalkp>haben</Verbalkp></AutosemKp></Kookkurrenz>
```

Die partielle Ausarbeitung der Informationsdimension **Schreibung & Aussprache** (<SchAus>) zeigt hier die Formvariante <FormVA> *Bär aufbinden*, i. e. lexikalische Variabilität, mit einem entsprechenden Kommentar.

⁴ So variabel das Transfer-Verb lexikalisch auch ist, materialisiert sich doch ohne dieses VERB die metaphorische Bedeutung von *Korb* nicht.

⁵ Dies wird an anderer Stelle im Detail erläutert und begründet werden.

Eine lexikografische Auswertung der Kookkurrenzanalyse des Phraseologismus in <Kookkurrenz> deutet auf die Verwendung des Phraseologismus im öffentlichen (politischen) Diskurs hin⁶ und gibt Hinweise auf die typische Belegung seiner Argumentstellen.

3.4. Zur Lesart: Bedeutung & Verwendung

Polyseme Lesarten einzelner Lemmata und die jeweils zugehörigen Angaben werden in WiW durch die Lesarten-Identifikation <LesIDA> innerhalb der Informationsdimension **Bedeutung & Verwendung** (hier: <MWL-BedVer>) unterschieden. Im Gegensatz zu der in Wörterbüchern üblichen Anordnung der Lesarten in numerisch gegliederten Aufzählungen ermöglicht diese sprachlich fixierte Trennung der Lesarten nicht nur eine unkompliziertere Hinzufügung, Löschung oder Umdefinition von Lesarten, sondern auch Auswahl und Anzeige von bzw. Verweise auf einzelne Lesarten, ohne dass die Verständlichkeit der Teilartikel durch das Fehlen der übrigen (traditionell linear vor- bzw. nachgeordneten) Teile beeinträchtigt würde.

Der knappen Lesarten-Identifikation folgt die ausführlichere semantische Paraphrase <SemParA>. Der qualitative Unterschied aber zwischen *Bären aufbinden* und z. B. *hinters Licht führen* oder *lügen*, der über die neutrale Formulierung in <SemParA> hinausgeht, kann in der Dimension **Pragmatik** ausführlich diskutiert werden.

```
<MWL-BedVer><MWL-Semantik><LesIDA>die Unwahrheit erzählen</LesIDA>
<MWSemEA><SemParA>jemandem einen Sachverhalt als wahr mitteilen,
obwohl man sich über seine Unwahrheit im Klaren ist</SemParA>
```

Die Einordnung der Lesarten in satzsemantische (hier: <Praedikator>) und ontologische Klassen (hier: <EREIGNIS>) folgt einer Klassifikation, die – u. a. ausgehend von Polenz (1988), respektive Sanfilippo (1999) – ebenfalls im Rahmen des WiW-Projekts erarbeitet wurde. Näheres dazu findet sich in Haß-Zumkehr (2000).

```
<SemKla><Praedikator><Handlungspraedikator typ="resultativ">
<SemRolle rolle="Agens">Derjenige, der jemanden anschwandelt: v.a.
<FillerTyp entitaet="BELEBTES:MENSCH">Personen</FillerTyp>.
</SemRolle><SemRolle rolle="Contraagens">Diejenigen, die
angeschwandelt werden: v.a. <FillerTyp entitaet="BELEBTES:MENSCH">
Personen und Personengruppen</FillerTyp>, wie z.B. <Filler>Wähler
</Filler>,<Filler>Presse</Filler>, <Filler>Öffentlichkeit</Filler>
oder <Filler>Publikum</Filler>.</SemRolle>
<ReferenzA><EREIGNIS><HANDELN typ="KOMMUNIZIEREN"/></EREIGNIS>
</ReferenzA></Handlungspraedikator></Praedikator></SemKla>
```

Da es sich in unserem Beispiel um einen verbalen Phraseologismus handelt, der eine Handlung beschreibt, werden zudem die semantischen Rollen der konstruktionsexternen (Wotjak,

⁶ Im Lesart-bezogenen – hier nicht ausgeführten – Abschnitt zur Pragmatik kann ein solcher Befund nach Belieben interpretiert und diskutiert werden.

1989) Argumente und deren Belegung, also die mehr oder weniger harten Selektionsrestriktionen, diskutiert⁷. Über das Attribut `rolle` sind diese Angaben direkt mit den grammatischen Angaben zur Valenz, i. e. der Komplementklasse und den Realisierungsmöglichkeiten der jeweiligen Rolle, verlinkt. Die vorgegebene Wertemenge des Attributs orientiert sich an Polenz (1988, S. 170–172), die des `<FillerTyp>` an Sanfilippo (1999).

Während obige Angaben für alle Lemmatypen definierbar sind, werden die folgenden Angaben zur semantischen Binnenstruktur ausschließlich für Mehrworteinheiten angegeben. Dazu zählen Angaben

- zum Grad der Idiomaticität bzw. Kompositionalität⁸ `<IdiomA>` (da das Konzept der usuellen Wortverbindungen, wie oben erwähnt, auch nicht-idiomatische, lexikalisch stark kohäsive Wendungen umfasst),
- zu den Basiselementen (deren entsprechende Einzelwortlemma-Artikel mit den zugehörigen lexikalisierten Mehrwortverbindungen bidirektional verlinkt sind),
- zu typischen Verwendungsmustern `<TypVerwMstA>` (die einer Aussageform im Sinne von Dobrovolskij, 1993, am nächsten kommen),
- zu Belegbeispielen möglicher Modifikationen `<KorpModA>` (i. e. strukturelle Variabilität).

```
<IdiomA idiom="vollidiomatisch"></IdiomA>
<SemBinnenstruktur><BasisElement unikal="nein"
idiom="idiomatisch"><TXTObj verwlemma="Baer">Bären</TXTObj>
</BasisElement><BasisElement unikal="nein" idiom="idiomatisch">
<TXTObj>aufbinden</TXTObj></BasisElement></SemBinnenstruktur>
</MWSemEA>
...
<TypVerwMstA QuA-ID="T94DEZ">Wir werden die Regierung fragen,
warum sie uns einen Bären aufbinden will.</TypVerwMstA>
...
<KorpModA><BBeiA QuA-ID="T90NOV">Soviel Bären läßt sich Lummer
natürlich nicht aufbinden.</BBeiA>
<BBeiA QuA-ID="T00FEB">Wahrscheinlich war es auch ein Leichtes
für ihn, der guten, alten Zitty ordentlich einzuheizen und ihr
ein paar Berliner Bären aufzubinden:</BBeiA>
<BBeiA QuA-ID="U98SEP">Diese Schizophrenie ist als 'politisches
Interesse' bekannt und bedeutet, daß dem Wähler jeder beliebige
```

⁷ Es ist nicht ausgeschlossen, dass an dieser Stelle auch konstruktionsinterne Argumente beschrieben werden. Der Aussage von Sternkopf (1992, S. 224) „Wenn Valenz im Sinne von Kreativität/Fähigkeit aufgefaßt wird, was bleibt dann von ihr übrig, wenn die Anzahl der Mitspieler, deren syntaktische Markierung und lexikalische Auffüllung schon vorgegeben, sozusagen kodifiziert sind?“ kann man entgegenhalten, dass die Anzahl der Mitspieler immer vorgegeben ist, während man die kodifizierte lexikalische Auffüllung derselben auch als ein Ende der Skala in der Rigidität der Selektionsrestriktionen begreifen könnte. Da die konstruktionsinterne Valenz in der **Grammatik**-Dimension von WiW durchaus beschrieben wird – das Phraseolexem, wie die anderen Klassen usueller Wortverbindungen auch, nicht als syntaktisch opaquere „black box“ verstanden wird – liegt eine Beschreibung der internen Rollen im Prinzip nicht fern.

⁸ Eine mögliche figurative Kompositionalität wie in *Öl ins Feuer gießen*, i. e. die Motiviertheit und Bildhaftigkeit einer Wendung, kann in einer hier nicht ausgeführten Angabe `<MotiviertA>` beschrieben und diskutiert werden.

Bär aufgebunden werden kann, sofern der Bär beim Umkippen nur freundlich brummt...</BBeiA></KorpModA>

Die Angaben zur **Paradigmatik** sowie die Angaben der Teildimension **Pragmatik** werden im Rahmen dieses Beitrags nicht ausgeführt. Sie sind für alle Lemmatypen gleichermaßen spezifizierbar. Erwähnenswert im Zusammenhang mit Mehrworteinheiten ist allerdings die Möglichkeit, so ausführlich wie erwünscht das illokutive Potenzial der Einheit zu diskutieren.

3.5. Zur Lesart: Grammatik

Die Dimension **Grammatik** (Müller-Landmann, 1999) informiert bei Mehrworteinheiten v.a.

- über die syntaktische Funktion der Lesart (hier: funktion="Verb"),
- die wendungsinterne syntaktische Struktur (<SynBinnenstruktur>),
- die Stabilität derselben bzw. die transformationellen Restriktionen auf ihren Bestandteilen (TransRest..., v. a. nach Fleischer, 1997),
- mögliche syntaktische Anomalien (<SyntAnomal>, v. a. nach Fleischer, 1997, im Beispiel: keine) der Wendung als Ganzes
- und, angelehnt an Zifonun et al. (1997), über die externe Valenz (hier Verbalenz: <Val-Vb>), über welche die Brücke zu den Angaben zu den semantische Rollen in **Bedeutung & Verwendung**, den Selektionsrestriktionen, geschlagen wird.

```
<MWL-Gramm><MWL-Verb funktion="Verb">
...
<SynBinnenstruktur><Unterordnung>
<Regierendes><Wort wart="Verb">aufbinden</Wort>
<TransRestVb><Negation-nicht belegung="belegt"></Negation-nicht>
<Passivbildung belegung="belegt"></Passivbildung> ...
</Regierendes>
<Regiertes>
<Realisierung rolle="Patiens"><NPakk></NPakk></Realisierung>
<RegWort><Wort wart="Nomen" verwlemma="Baer">Bären</Wort>
<TransRestNm><Negation-kein belegung="belegt">
<BBeiA QuA-ID="T87SEP">Pfeiffer wertete Behnkes Äußerungen als
Beleg dafür, daß er der Presse "keinen Bären aufgebunden"
habe.</BBeiA>
</Negation-kein>
<AdjAttribExp belegung="belegt">
<BBeiA QuA-ID="T93MAR">Da haben sich die Berliner Olympioniken in
spe mal wieder einen gewaltigen Bären aufbinden lassen, diesmal
sogar einen spuckhäßlichen.</BBeiA></AdjAttribExp> ...
</RegWort></Regiertes></Unterordnung></SynBinnenstruktur>
...
```

```

<ValVb>
<KompVb><Ksub rolle="Agens"><NPnom></NPnom></Ksub>
<Kobj obligatorisch="ja" rolle="Contraagens"><NPdat></NPdat>
</Kobj><PhrMstA>er/sie/es bindet jemandem einen Bären
auf</PhrMstA></KompVb></ValVb></MWL-Verb></MWL-Gramm>

```

Die syntaktische Binnenstruktur kann nebenordnend sein, wie in *ganz und gar; rank und schlank; Friede, Freude, Eierkuchen*, oder unterordnend, wie im Beispiel *Bären aufbinden* oder in *sich lieb Kind machen; wie die Pest*, etc.

Transformationelle Restriktionen umfassen z. B. bei Verben die (Un-)Möglichkeit der Passivbildung (**Fraktur ist von uns nicht geredet worden* vs. *Der Vogel ist von Otto abgeschossen worden*), bei Adjektiven/Adverbien die (Un-)Möglichkeit der Expansion durch Adverbial (**sehr klein begeben* vs. *+den Dingen einen völlig freien Lauf lassen*) oder bei Nomen die (Un-)Möglichkeit der Numerusvariation (**auf Drähten sein* vs. *ins Auge/in die Augen fallen*).

Syntaktische Anomalien der Mehrwortverbindung können u. a. sein: unflektierter Adjektivgebrauch (*sich lieb Kind machen*), eine obligatorische Negationskomponente (*nicht hinter dem Ofen/hinter keinem Ofen hervorlocken*) oder Valenzabweichungen (*Lügen strafen*).

Die Valenz der Mehrworteinheiten wird entsprechend ihrer grammatischen Funktion analog zu der einer Einzelwort-Wortart beschrieben, die diese Funktion ebenfalls erfüllt. I. e., einem verbalen Phraseologismus wird dasselbe Komplementierungspotenzial zugestanden wie einem Einwort-Verblemma⁹.

Mit der Phrasenmusterangabe <PhrMstA> wird schließlich auch folgendem Desiderat entsprochen: „Kodifizierte nun das Wörterbuch anstelle der Nennform verbaler Phraseologismen Satzmodelle, so entspräche dies mehr dem Benutzerbedürfnis, indem damit Verwendungsbedingungen mitgeliefert würden.“ (Sternkopf, 1992, S. 223)

3.6. Ausblick

Dieser Beitrag geht v. a. auf die Spezifika von Mehrwortlemmaeinträgen in der WiW-Mikrostruktur ein. Da Einträge zu Mehrwortlemmata gegenüber Einträgen zu Einwortlemmata zusätzlich v. a. die Informationen zum Typ der Mehrworteinheit und zur syntaktischen und semantischen Binnenstruktur einfordern, werden hier große Teile der DTD überhaupt nicht erwähnt.

Es muss als großer Vorteil der DTD-Konzeption in WiW betrachtet werden, dass tatsächlich nahezu alle Angabearten (i. e. Teilbäume der DTD) für alle drei Lemmatypen gleichermaßen gelten. So ist es fast immer möglich, Wortschatzeinheiten nach linguistisch-lexikologischen Eigenschaften und eben unabhängig vom Aussehen des Formativs zu kompilieren: Durch die umfassende, feinkörnige und linguistisch-lexikologisch fundierte Mikrostruktur der Wörterbuchartikel werden die XML-getaggten Informationen zu den Stichwörtern entlang der Mikrostruktur nach unterschiedlichen, expliziten Kriterien indiziert. Im Ergebnis ist jedem Lemma eine scharf strukturierte Menge möglichst theortoleranter Merkmale zugeordnet. Das erlaubt, Anfragen an das WiW-Wörterbuch in quasi jeder beliebigen Detailliertheit zu beantworten.

⁹ Eine Unterscheidung in Idiome mit variablem Subjekt (VPL1) und feste prädikative Phrasen (VPL2), wie sie von Dormeyer et al. (1998, S. 2) vorgenommen wird, scheint daher nicht nötig: Entweder die externe Valenz beinhaltet das <Ksub>-Subjekt – dann handelt es sich um VPL1 – oder aber das Subjekt wird als <NPnom>-Realisierung in der syntaktischen Binnenstruktur als <Regiertes> ausgewiesen – damit ist es VPL2.

Wo ein Benutzer mit einem einfachen lemma-orientierten Nachschlagebedürfnis mit einer derartig kompakten Darstellung der Grundinformation schon zufrieden sein könnte:

Bären aufbinden – jemandem etwas vorlügen oder weismachen – er/sie/es bindet jemandem einen Bären auf – *Wir werden die Regierung fragen, warum sie uns einen Bären aufbinden will.*

wird es einem wissenschaftlich interessierten Benutzer ermöglicht, – jeweils in Abhängigkeit von individuellen wissenschaftlichen Fragestellungen – linguistisch-lexikologische Klassifizierungen empirisch zu bestätigen durch die Zusammenstellung von Wortschatzeinheiten, die bestimmte Merkmale mit bestimmten Ausprägungen tragen (z. B. „Gib mir die Etymologie aller verbalen Phraseologismen, die eine Nominalphrase im Akkusativ inkorporieren und eine kommunikative Handlung beschreiben“).

Ergebnisse solcher externer wissenschaftlicher Wortschatzforschung mittels WiW werden – gebündelt und in Beziehung zueinander gesetzt – zurückgeführt und ein integraler, mit der eigentlichen Lemmastrecke durch Verlinkung verwobener Bestandteil des WiW-Informationssystems sein und damit weiteren beträchtlichen linguistischen Mehrwert schaffen.

Literaturverzeichnis

- CHANG, N. UND FISCHER, I. (2000): “Understanding Idioms”. In: *Konvens 2000 Sprachkommunikation*, herausgegeben von ITG. S. 33–38. ITG-Fachbericht 161.
- DOBROVOL'SKIJ, D. (1993): “Datenbank deutscher Idiome. Aufbauprinzipien und Einsatzmöglichkeiten”. In: *Germanistik und Deutschlehrausbildung: Festschrift zum hundertsten Jahrestag der Gründung des Lehrstuhls für deutsche Sprache und Literatur Pädagogischen Hochschule Szeged*, herausgegeben von Földes, Csaba, Szeged usw.: Ed. Praesens, S. 51–67.
- DORMEYER, R.; FISCHER, I. UND KEIL, M. (1998): “A Database for Verbal Idioms”. In: *Euralex '98 Proceedings*, herausgegeben von Fontenelle, T.; Hilismann, Ph.; Michiels, A.; Moulin, A. und Theissen, S. Liège, S. 99–109.
- FLEISCHER, W. (1997): *Phraseologie der deutschen Gegenwartssprache*. 2., durchgesehene und ergänzte Auflage.
- FRAAS, C. UND HASS-ZUMKEHR, U. (1998): “Vom Wörterbuch zum lexikalischen Informationssystem. LEKSIS – ein neues Projekt des Instituts für Deutsche Sprache”. *Deutsche Sprache* 98 (4): S. 289–303.
- HASS-ZUMKEHR, U. (2000): “Begründung und Erläuterung einer primär satzsematisch, sekundär ontologisch ansetzenden Lexemklassifikation”. Internes Arbeitspapier. Institut für Deutsche Sprache, Mannheim.
- HASS-ZUMKEHR, U. (2001): “Zur Mikrostruktur im Hypertext-Wörterbuch”. In: *Chancen und Perspektiven computerunterstützter Lexikographie*, herausgegeben von Lemberg, I.; Schröder, B. und Storrer, A. Im Erscheinen.
- KORHONEN, J. (1990): “Zur (Un-)Verständlichkeit der lexikographischen Darstellung von Phraseologismen”. In: *BudaLex '88 Proceedings*, herausgegeben von Magay, T. Budapest: Akademiai Kiado, S. 197–206.
- MÜLLER-LANDMANN, S. (1999): “WiW-Grammatik. Manual für das WiW-Lexikon”. Internes Papier. Institut für Deutsche Sprache, Mannheim.
- MÜLLER-LANDMANN, S. (2000): “Design eines Internet-Lexikons zwischen Recherche und Rezeption”. In: *Euralex 2000 Proceedings*, herausgegeben von Heid, U.; Evert, S.; Lehmann, E. und Storrer, Ch. Stuttgart, S. 97–105.

- POLENZ, P. (1988): *Deutsche Satzsemantik: Grundbegriffe des Zwischen-den-Zeilen-Lesens*. Berlin usw.: de Gruyter. 2., durchgesehene Auflage.
- SANFILIPPO, A. (1999): "Preliminary Recommendations on Lexical Semantic Encoding. Final Report. EAGLES LE34244. The EAGLES Lexicon Interest Group". <http://www.ilc.pi.cnr.it/EAGLES96/home.html>.
- STERNKOPF, J. (1992): "Valenz in der Phraseologie? Ein Diskussionsbeitrag". *Deutsch als Fremdsprache* (4): S. 221–224. München/Berlin.
- STEYER, K. (2000): "Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten". *Deutsche Sprache* 2000 (2): S. 101–125.
- WOTJAK, B. (1989): *Zu Inhalts- und Ausdrucksstrukturen verbaler Phraseolexeme in System und Text*. Leipzig.
- ZIFONUN, G.; HOFFMANN, L. UND STRECKER, B. (1997): *Grammatik der deutschen Sprache*. Berlin usw.: de Gruyter.