

Phraseographie Japanisch–Deutsch: Lexikographische
Perspektiven bei der Entwicklung digitaler Nachschlagewerke

Dissertation

zur

Erlangung des Grades Doktor der Philosophie
in der Philosophischen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von

Iris Vogel

aus

Coburg

2014

Gedruckt mit Genehmigung der Philosophischen Fakultät
der Eberhard Karls Universität Tübingen

Dekan: Prof. Dr. Jürgen Leonhardt

Hauptberichterstatter: Prof. Dr. Robert Horres

Mitberichterstatterin: Prof. Dr. Viktoria Eschbach-Szabo

Tag der mündlichen Prüfung: 04. November 2013

Tübingen, TobiasLib 2014

Inhaltsverzeichnis

1	Einleitung	9
2	Besonderheiten der japanischen Phraseologie	16
2.1	Grundbegriffe und Definitionen	16
2.1.1	Die terminologische Vielfalt in der phraseologischen Forschung	16
2.1.2	Begriffsfelder und Definitionen in der lexikographischen Praxis	21
2.1.3	Terminologische Standards der Sprachtechnologie	24
2.1.4	Zusammenfassung	26
2.2	Phraseologische Merkmale und ihre Auswirkungen auf die Phraseographie . .	27
2.2.1	Polylexikalität vor dem Hintergrund des japanischen Schriftsystems . .	27
2.2.1.1	Segmentierung und Komponentenanalyse	28
2.2.1.2	Orthographische Variation	31
2.2.2	Stabilität und Variabilität	34
2.2.2.1	Lexikalische und morphologische Restriktionen	35
2.2.2.2	Variation innerhalb der Grenzen struktureller Festigkeit . . .	37
2.2.3	Idiomatizität	41
2.2.3.1	Kognitive Festigkeit	41
2.2.3.2	Motiviertheit	42
2.2.3.3	Metaphorische lexikalische Bedeutung der Komponenten . .	43
2.2.3.4	Kognitve Metaphern	44
2.2.3.5	Polysemie	45
2.2.3.6	Idiomatische Äquivalenz	47
2.2.4	Zusammenfassung	50
3	Phraseologie in Theorie und Praxis	51
3.1	Anforderungen an die Phraseographie aus linguistischer Perspektive	51
3.1.1	Terminologie in Titel und Umtexten	52

3.1.2	Materialauswahl	54
3.1.3	Gestaltung der Nennform	55
3.1.3.1	Wahl der morphosyntaktischen Form	56
3.1.3.2	Darstellung der externen Valenz	56
3.1.3.3	Darstellung morphosyntaktischer Restriktionen	58
3.1.4	Kennzeichnung	58
3.1.5	Umgang mit Varianten	59
3.1.6	Makrostruktur	60
3.1.6.1	Die exhaustive Alphabetisierung	61
3.1.6.2	Ordnung nach Leitelementen	62
3.1.6.3	Onomasiologische Ordnung	63
3.1.7	Mikrostrukturelle Einordnung	65
3.1.8	Angabe der Bedeutung	65
3.1.8.1	Pragmatische Aspekte in der Bedeutungsangabe	67
3.1.8.2	Art der Bedeutungserklärung	68
3.1.8.3	Angabe zur Etymologie	69
3.1.8.4	Beispielangaben	70
3.1.9	Grammatische Angabe	72
3.1.10	Mediostruktur	73
3.1.11	Illustrationen	74
3.1.12	Sonstiges	75
3.1.13	Zusammenfassung	76
3.2	Phraseographische Praxis in japanischen Printwörterbüchern	76
3.2.1	Allgemeine Wörterbücher	77
3.2.1.1	Allgemeines	77
3.2.1.2	Phraseologieauffassung, Terminologie und Aufbau	78
3.2.1.3	Gestaltung der Nennform	78
3.2.1.4	Umgang mit Varianten und Synonymen	80
3.2.1.5	Makrostrukturelle Zuordnung	81
3.2.1.6	Mikrostruktur	83
3.2.1.7	Zusammenfassung	84
3.2.2	Phraseologische Wörterbücher	84

3.2.2.1	Phraseologieauffassung, Terminologie, Inhalt und Aufbau . . .	85
3.2.2.2	Gestaltung der Nennform	86
3.2.2.3	Umgang mit Varianten	87
3.2.2.4	Makrostrukturelle Zuordnung	87
3.2.2.5	Mikrostruktur	97
3.2.2.6	Zusammenfassung	100
3.3	Anforderungen an die elektronische Phraseographie aus computerlinguisti- scher Perspektive	101
3.3.1	Automatische Extraktion von Mehrwortlexemen	103
3.3.2	Zugriffsmethoden	105
3.3.3	Onomasiologische Kategorisierung	108
3.3.4	Darstellung lexikalischer Einheiten	109
3.3.5	Organisation lexikalischer Einheiten	111
3.3.6	Vernetzung und Verweise	112
3.3.7	Multimedia	114
3.3.8	Grammatische Angabe	115
3.3.9	Adaptivität hinsichtlich Nutzergruppe und Nutzungssituation	115
3.3.10	Benutzerpartizipation und Qualitätskontrolle	117
3.3.11	Zusammenfassung	118
3.4	Datenformate und Standards in der elektronischen Lexikographie	119
3.4.1	Allgemeine Standards für die Textkodierung	120
3.4.1.1	Unicode	120
3.4.1.2	XML	122
3.4.2	Standardisierungsbestrebungen der Computerlexikographie	123
3.4.2.1	Das Wörterbuchformat der Text Encoding Initiative	124
3.4.2.2	Das Lexical Markup Framework	129
3.4.3	Zusammenfassung	140
3.5	Phraseographische Praxis in Online-Wörterbüchern	141
3.5.1	Ulrich Apels WaDokuJT	141
3.5.1.1	Datenstruktur	142
3.5.1.2	Suchfunktionalität auf WADOKUDE-IW	143
3.5.1.3	Artikelansicht für Phraseme im WADOKUDE-IW	144

3.5.1.4	Benutzeradaptivität und Benutzerführung	146
3.5.2	JapanKnowledge+	147
3.5.2.1	Benutzerführung	148
3.5.2.2	Suchfunktionalität	148
3.5.2.3	Ergebnisanzeige	151
3.5.2.4	Phraseme	152
3.5.2.5	Mikrostruktur	154
3.5.3	»Deutsche Redewendungen« – ETTINGERPHRASEOLOGIEDE-IW . . .	154
3.5.3.1	Terminologie, Phraseologieauffassung und Materialauswahl .	155
3.5.3.2	Makrostruktur	156
3.5.3.3	Nennform und Mikrostruktur	158
3.5.3.4	Sonstiges	160
3.5.4	Zusammenfassung	161

4 Das Online-Wörterbuch für japanische Idiome und ihre deutschen Entsprechungen **163**

4.1	Aufbau und Funktionalität	164
4.1.1	Titel und Terminologie	164
4.1.2	Benutzungshinweise und Orientierungshilfen	164
4.2	Materialauswahl	165
4.2.1	Phrasembestand	166
4.2.2	Referenzkorpora	167
4.3	Kennzeichnung	169
4.4	Nennform	170
4.4.1	Wahl der morphosyntaktischen Form	170
4.4.2	Integration der Phonetik-Angabe	171
4.4.3	Nennform-Varianten	172
4.5	Semantik-Angabe	174
4.5.1	Wörtliche Bedeutung der Komponenten	174
4.5.2	Idiomatische Bedeutung und Äquivalenz	177
4.5.3	Pragmatik	181
4.5.4	Etymologie	183

4.5.5	Synopse-Artikel	184
4.5.6	Multimedia	186
4.5.7	Beispiele und Belege	187
4.5.8	Kategorisierung	188
4.6	Grammatik-Angabe	190
4.6.1	Grammatische Klasse	191
4.6.2	Interne Phrasenstruktur	191
4.6.3	Externe Valenz	192
4.6.4	Belegbare Formen	192
4.7	Polylexikalität und orthographische Variation	193
4.8	Modellierung struktureller Stabilität und Variabilität	193
4.8.1	Morphosyntaktische Restriktionen und Präferenzen	194
4.8.2	Morphosyntaktische, phonetische und lexikalische Variabilität	196
4.8.3	Modifikation	197
4.8.4	Transitiv-intransitiv Varianten im Wörterbuch	198
4.9	Zugriffstrukturen	201
4.9.1	Variabilität und Suchfunktionalität	202
4.9.2	Ergebnisanzeige	203
4.9.3	Mediostruktur	205
4.10	Technische Realisierung	209
4.10.1	Komponenten	209
4.10.2	Datenbasis	209
4.10.3	MeCab-Parser	211
4.10.4	Externe lexikalische Daten	211
4.10.5	Datenstruktur	212
4.10.5.1	Der phraseologische Wörterbucheintrag (<i>Mwe Entry</i>)	213
4.10.5.2	Lemma (<i>Lemma</i>)	214
4.10.5.3	Bedeutungsangabe (<i>Sense</i>)	215
4.10.5.4	Kategorie (<i>Category</i>)	216
4.10.5.5	Synopseartikel (<i>Synopse</i>)	217
4.10.5.6	Verweis (<i>Reference</i>)	218
4.10.5.7	Beispiel (<i>Example</i>)	218

Inhaltsverzeichnis

4.10.5.8	Beleg (<i>Citation</i>)	219
4.10.5.9	Kompatibilität zu Standards und Interoperabilität	220
4.11	Fazit	221
Literatur		225

1 Einleitung

Die zunehmende Digitalisierung von Ressourcen erweckt den Anschein, dass sich der Welt der Wissensmedien unbegrenzte Möglichkeiten eröffnen. Im Zuge dessen wachsen auch die Anforderungen an die Funktionalität, Verfügbarkeit, Umfang und Vernetztheit elektronischer Ressourcen. In »Lexicographer's Dreams in the Digital Age« trägt Schryver (2003) zusammen, was sich Experten von der digitalen Wende der Lexikographie erhoffen. Die Wörterbücher der Zukunft sollen einen größeren Umfang bieten und gleichzeitig den Nutzer gezielter zur gesuchten Information führen. Ihre Entwicklung erfolgt korpusbasiert und sie sind mit diversen externen Wissensressourcen vernetzt. Flexibel und intelligent passen sie sich an die Bedürfnisse individueller Nutzer an und lassen sich nahtlos in dessen Arbeitsumgebung integrieren. Aufgrund der vielfältigen Anforderungen wagt Schryver (2003) die Prognose, dass das Wörterbuch des 21. Jahrhunderts ein digitales Nachschlagewerk sein wird.

Als Nigel Portwood der Oxford University Press (OUP) im August 2010 in einem Interview die Einschätzung äußerte, dass die nächste Auflage des Oxford English Dictionary in einer Dekade vermutlich nicht mehr als gedruckte Version erscheinen wird, sorgte das für großes Aufsehen. Kurz darauf dementierte die Presseabteilung der OUP, dass es eine Entscheidung vor der Fertigstellung des Werkes geben werde (Oxford Times 2010). Zwei Jahre später verkündete der Macmillan Verlag offiziell, dass es ab 2013 keine Druckversionen seiner Wörterbücher mehr geben wird (Rundell 2012). Ob sich dieser Trend verstetigt, bleibt abzuwarten. In jedem Fall lässt sich sagen, dass die Verlage zunehmend auf digitale Medien als Plattform für ihre Wissensressourcen setzen.

Die vorliegende Arbeit befasst sich mit der Frage, wie das digitale Medium für die Neukonzeption eines phraseologischen Wörterbuchs eingesetzt werden kann. Im Zusammenhang mit der regen Aktivität der phraseologischen Forschung in den letzten Jahrzehnten, die sich in einer beträchtlichen Anzahl Publikationen widerspiegelt, kommt wiederholt Unmut über die phraseographische Praxis zum Ausdruck. Es werden Desiderate formuliert und die verhaltenen Reaktionen aus dem Bereich der phraseographischen Praxis moniert. Dobrovols'kij erklärt dieses Phänomen damit, dass die Phraseologieforschung sich nie ernsthaft mit ihrer »wichtigsten

Aufgabe« befasst habe, und zwar der »präzisen und systematischen Beschreibung der Phraseme aller Klassen in allen möglichen Sprachen« (Dobrovol'skij 2002, S. 363). Weiterhin heißt es:

Die Vernachlässigung der »niedrigen Empirie« zugunsten der »hohen Theorie« hat dazu geführt, daß es trotz jahrzehntelangen Aktivitäten vieler Phraseologieforscher in verschiedenen Ländern der Welt kein einziges (mir bekanntes) Wörterbuch gibt, in dem der Benutzer die notwendigen Informationen für die produktive (nicht nur rezeptive) Beherrschung der Phraseme finden kann. (Dobrovol'skij 2002, S. 363)

Verschiedentlich gibt es Bestrebungen, theoretische Überlegungen zur Modernisierung der Phraseographie einer praktischen Umsetzung anzunähern. Einen meta-phraseographischen Ansatz verfolgt Martín (2001), indem er für das Sprachpaar Deutsch-Spanisch existierende Nachschlagewerke im Hinblick auf ihren Umgang mit Phrasemen untersucht und Vorschläge für die Verbesserung der lexikographischen Praxis unterbreitet. Eine detaillierte und strukturierte Analyse deutscher Wörterbücher findet sich bei Stantcheva (2002). Während bei Martín (2001) Fragen der bilingualen Phraseographie im Vordergrund stehen, widmet sich die konstruktiv-kritische Betrachtung von Stantcheva (2002) allen Bereichen der Phraseographie. Arbeiten mit ausgeprägtem Praxis-Anteil finden sich bei Kim-Werner (1996) für die Phraseologie des Deutschen und Koreanischen und Zhu (1998) für das Sprachpaar Deutsch-Chinesisch. Beide Werke beinhalten im Anhang den Prototyp eines bilingualen Wörterbuchs für die somatischen Phraseme der untersuchten Sprachen. Im Vordergrund stehen insbesondere Fragen der Kodierung sprachlicher und kultureller Unterschiede im phraseologischen Lexikon. Arbeiten jüngerer Datums beziehen die Möglichkeiten der elektronischen Lexikographie in ihre Betrachtungen ein. Keil (1998) beschreibt Konzeption und Aufbau eines elektronischen Phrasem-Lexikons, welches die spezifischen Eigenschaften deutscher Redewendungen adäquat abzubilden versucht.¹ Eine korpusbasierte phraseologische Ressource ist die Idiom-Datenbank, die durch das Projekt »Kollokationen im Wörterbuch« der Berlin-Brandenburgischen Akademie der Wissenschaften erstellt wurde.² Mit dem wesentlich weiter gefassten Bereich usuelle

¹ Die resultierende Ressource *Phraseo-Lex* scheint jedoch nicht mehr verfügbar zu sein.

² Verfügbar unter <http://kollokationen.bbaw.de/>.

Wortverbindungen beschäftigt sich ein gleichnamiges Projekt am Institut für deutsche Sprache (IDS) in Mannheim, welches sich der korpusanalytischen Erforschung und Beschreibung rekurrenter Sprachgebrauchsmuster widmet.

Es gibt auch multilinguale Projekte, die sich die Möglichkeit empirischer Datenerhebung zu Nutze machen. So stellt Ješensek (2009) ein phraseologisches Lexikon für Deutsch, Slowenisch, Slowakisch und Ungarisch vor, welches in erster Linie phraseodidaktische Ziele verfolgt. Mit phraseologischen Ähnlichkeiten, die in einer Reihe von Sprachen weltweit existieren, beschäftigt sich »Widespread Idioms in Europe and Beyond« (Pirainen 2012). Theoretische Überlegungen zur Konzeption einer multilingualen Sprichwörter-Datenbank stellt Hrisztova-Gotthardt (2010b) an.

Darüber hinaus werden zunehmend Idiom-Wörterbücher für einzelne Sprachen online zur Verfügung gestellt. Onomasiologische Ansätze finden sich beispielsweise in Espinal (2005) und Ettinger (2009c).

In der japanischen Linguistik wird der Phraseologie und theoretischen Lexikographie bei Weitem nicht die Aufmerksamkeit gewidmet, die sie im deutschen Sprachraum erfährt. Für das Sprachpaar Japanisch-Deutsch stehen eine Reihe kontrastiver Analysen zur Verfügung (Itoh 2004; Itoh 2005; Ueda 2002; Ueda 2003; Ueda 2004). Eine Konnex zu der lexikographischen Praxis wird nicht hergestellt. Der lexikographische Forschungsdiskurs in Japan ist tendenziell deskriptiv (Kokuritsu Kokugo Kenkyūsho 2009; Kurashima 2002; Satake 2000; Kanetake 2000; Takashima 2003) und insbesondere der Themenkomplex Phraseologie wird von Fragen aus dem Bereich »Englische Fremdsprachendidaktik« dominiert (Nesi 2013; Inoue 2011; Yagi 2011; Tono 2001; Nakao 1998). Das relative Desinteresse an phraseologischen Themen in der japanischen Linguistik steht im Kontrast zu einer sehr aktiven lexikographischen und phraseographischen Praxis sowie einer entsprechenden Vielfalt der Wörterbuchlandschaft. Einem dem deutschen Sprachraum vergleichbaren Forschungsdiskurs zwischen den Disziplinen fehlt in Japan vermutlich ein Spannungsfeld, das sich aus einer Diskrepanz zwischen theoretischem Forschungsstand und praktischer Umsetzung speist.

Der Reichtum an phraseologischen Wörterbüchern beschränkt sich auf monolinguale und bilingual japanisch-englische Wörterbücher. Das Sprachpaar Japanisch-Deutsch ist mit nur wenigen Nachschlagewerken vertreten. Zurückverfolgen lässt sich diese Tradition bis in das 19. Jahrhundert. In diversen Beiträgen über »Japanische Sprichwörter« in den »Mittheilungen der deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens« (Knobloch 1874; R. Lange 1875;

R. Lange 1876; R. Lange 1876; R. Lange 1880; Ehmann 1893-1897) werden Sprichwörter mit Umschrift und Übersetzung versehen. Diese Sequenz in Verbindung mit diversen anderen Quellen mündete schließlich 1897 in einem Supplementband zu den Mitteilungen mit insgesamt 3729 Sprichwörtern (Ehmann 1897; Stalph und Suppanschitsch 1999).³ Deziert phraseologische Werke fokussieren Sprichwörter mit unterschiedlicher Schwerpunktsetzung: Arnold-Kanamori (2001) stellt in seinem Werk »Großer Zampano – Japanische Sprichwörter« volkstümliche und literarische Bezüge in den Mittelpunkt.⁴ Ästhetische Aspekte akzentuiert Berndt in seinen Werken, indem er seine Übersetzungen japanischer Sprichwörter von Kalligraphen illustrieren lässt (Berndt 1984; Berndt und Simon 1992). Ein aktuelles Werk zur Phraseologie Japanisch-Deutsch, welches sich neben Sprichwörtern auch Idiomen widmet, sucht man vergeblich.⁵

Die vorliegende Arbeit eruiert Perspektiven, wie diese Lücke in der Erfassung der japanisch-deutschen Idiomatik mit Hilfe digitaler Ressourcen geschlossen werden kann. Als Subklassen spielen insbesondere Idiome und Kollokationen eine Rolle, die sich durch das Merkmal deutlicher Idiomatizität respektive einer phraseologisch gebundenen Bedeutung auszeichnen. Die Analyse und Modellierung beschränkt sich auf Idiome im engeren Sinne. Sprichwörter und Routineformeln sowie spezielle Phänomene (z. B. 4-Zeichen-Phraseologismen) zählen zur Phraseologie im weiteren Sinne und sind somit nicht Gegenstand dieser Arbeit. Die Arbeit ist von ihrem integrativen Ansatz geprägt, der theoretische Anforderungen aus der Phraseologie und Computerlinguistik ebenso berücksichtigt wie die phraseographische Tradition. Durch die Einbindung externer lexikalischer Ressourcen und Werkzeuge geht die Arbeit über die bisherigen überwiegend monolithisch konzipierten phraseologischen Wörterbücher hinaus. Nicht zuletzt

³ Reinirkens »Sprichwörter und Redensarten Deutsch-Japanisch« basiert auf einer Umkehrung der Direktionalität eines Teilbereichs der Ehemann'schen Sammlung (Reinirkens 1955).

⁴ In der Einleitung des »Großer Zampano – Japanische Sprichwörter« bezeichnet Arnold-Kanamori (2001) sein eigenes Werk als »über weite Strecken a) nicht übersetzbar b) hintergründig zu lesen und c) weitschweifig«. Die Bedeutungserklärungen enthalten neben Hinweisen auf die Bedeutungs-genese bzw. -etymologie auch Literaturverweise und eine Einordnung in das phraseologische Umfeld inkl. Varianten.

⁵ Zwar nimmt sich auch das GROSSES JAPANISCH-DEUTSCHES WÖRTERBUCH-PW dem Thema Phraseologismen an, eine Fertigstellung des Werkes ist jedoch bislang nicht in Sicht. Verfügbar ist derzeit nur der erste Band mit der Artikelstrecke A-I (nach der 50-Laute-Tafel).

ermöglicht die Wahl des Objektbereichs Phraseologie Japanisch-Deutsch eine intensive Auseinandersetzung mit verschiedenen Formen der Variabilität und deren Kombinatorik, für welche sich die japanische Sprache aufgrund der ausgeprägten Komplexität des Schriftsystems besonders eignet. Ziel ist die prototypische Entwicklung und theoretische Beschreibung eines Systems, das die Natur des Phänomens möglichst adäquat in der digitalen lexikalischen Ressource abbildet und dabei die zentralen Anforderungen der einschlägigen Forschung erfüllt.

Der erste Teil dieser Arbeit gibt einen Überblick über die Merkmale des Objektbereichs, der lexikographisch repräsentiert werden soll. Dabei geht es in Kapitel 2.1 zunächst um die Terminologie und die Eingrenzung des Sachbereichs in den für diese Aufgabe relevanten Disziplinen und Sprachen. Ziel ist einerseits, eine begriffliche Klärung für die vorliegende Arbeit zu erreichen, die Erkenntnisse dienen jedoch auch als Grundlage für die Erarbeitung der Dokumentation und Hilfetexte im dritten Teil der Arbeit. Kapitel 2.2 analysiert die Charakteristika japanischer Phraseme als lexikalische Einheiten, die von den Merkmalen Polylexikalität, Idiomatizität und Festigkeit geprägt sind. Der Schwerpunkt liegt hier auf den Aspekten, die für die bilinguale Lexikographie allgemein und digitale Lexikonentwicklung im Besonderen relevant sind. So spielt die Komplexität des japanischen Schriftsystems für den Bereich Mehrworteinheiten im digitalen Lexikon eine entscheidende Rolle, da die Berücksichtigung orthographischer Variabilität ein entscheidender Faktor für den Zugriff ist. Darüber hinaus wird ein Überblick über die typischen Ausprägungen von Stabilität und Variabilität im Japanischen gegeben sowie die semantischen Besonderheiten der Phraseme diskutiert.

Vor dem Hintergrund der in Teil 1 thematisierten Besonderheiten der Idiome steht im zweiten Teil der Arbeit die phraseographische Theorie und Praxis im Mittelpunkt. In Kapitel 3.1 werden die Bedürfnisse und Anforderungen der Phraseologie zusammengefasst. Dabei wird besonderer Wert auf aktuellere Arbeiten gelegt und solche, die sich auf digitale Wörterbücher beziehen oder diese mitberücksichtigen. Da es in der japanischen Linguistik keine vergleichbare Diskussion gibt, konzentriert sich dieses Kapitel auf Anforderungen allgemeiner Art, die die spezifisch japanische Problematik nicht im Blick haben. Um den Bezug zur japanischen Phraseographie herzustellen, werden in Kapitel 3.2 repräsentative japanische Lexika mit Blick auf ihren Umgang mit Phrasemen untersucht.

Das Kapitel 3.3 ist den Anforderungen an die Phraseographie aus computerlinguistischer Perspektive gewidmet. Aufgrund der sich rasch wandelnden Paradigmen stehen aktuelle Arbeiten

im Fokus der Anforderungsanalyse. Mit dem Thema Standards der Sprachtechnologie (insbesondere TEI und LMF) und ihre Anwendbarkeit auf den Phänomenbereich Phraseologie befasst sich Kapitel 3.4. Abschließend wird in Kapitel 3.5 anhand einer Auswahl elektronischer Online-Lexika die tatsächliche Diskrepanz zwischen den Desiderata des theoretischen Forschungsdiskurses und dem Status Quo der praktischen elektronischen Phraseographie eruiert.

Im dritten Teil dieser Arbeit (Kapitel 4) wird ein System vorgestellt, das anstrebt die Erfahrungen der lexikographischen Tradition und die Anforderungen aus den verschiedenen Bereichen technisch für das Sprachpaar Japanisch-Deutsch umzusetzen. Dies geschieht mit dem Ziel der Optimierung von Suche, Vernetzung und Präsentation der Daten und orientiert sich sofern möglich an den Modellen und Vorgaben existierender Standards. Schnittstellen zu bestehenden Werkzeugen und Ressourcen sollen zeigen, dass ein digitales Wörterbuch weit mehr sein kann als eine Datenbanklösung für lexikalische Daten. Da es sich um ein für den menschlichen Benutzer optimiertes System handelt, werden auch verschiedene Nutzerperspektiven und Nutzungssituationen bei der Implementierung in Betracht gezogen. Neben den traditionellen Gestaltungsmöglichkeiten spielt hier die Einbindung von multimedialen Inhalten sowie andere Techniken der Visualisierung eine Rolle. Das Kapitel widmet sich überwiegend der Beschreibung der Ressource aus Anwenderperspektive. Technische Aspekte der Systemarchitektur werden in Kapitel 4.10 erläutert.

Um die Lesbarkeit der Arbeit auch einem Personenkreis zu ermöglichen, der der japanischen Sprache nicht mächtig ist, werden alle Begriffe und Beispiele zunächst in Umschrift angegeben und bei Erstnennung mit einer deutschen Übersetzung versehen. Die Umschrift erfolgt nach den Regeln des modifizierten Hepburn-Systems⁶. Längere Zitate werden in Originalskript ohne Umschrift wiedergegeben. Bei englischsprachigen Zitaten wird auf die Übersetzung verzichtet. Die japanischen Beispiele werden mit der in der kontrastiven Linguistik üblichen Glossierung ausgezeichnet. In der Regel werden dabei die Postpositionen an die zugehörigen Elemente angehängt, nur in Kapitel 2.2.1.1 weicht die Darstellung abhängig von dem Ergebnis der Phrasensegmentierung gegebenenfalls ab.

⁶ Die Silbe »n« wird durchgängig als »n« transkribiert, auch vor den Vokalen »b«, »m« und »p«. Die Striche auf »ē«, »ō« und »ū« markieren die Längungen von Vokalen. Die Partikeln für Objekt, Fokus und Richtung werden ihrer Lesung entsprechend als »o«, »wa« und »e« transkribiert.

Die in der vorliegenden Arbeit verwendeten Wörterbuchsiglen wurden in Anlehnung an Engelberg und Lemnitzer (2009, S. 273) aus einer Zusammensetzung aus Herausgeber oder Verlag, Titel und Nummer der Auflage erstellt. Auch hier wurden die japanischen Bezeichnungen nach dem modifizierten Hepburn-System transkribiert. Anhand der Erweiterungen ist zu erkennen, ob es sich um ein gedrucktes Werk (PW), eine CD-Rom-Ausgabe (CW) oder ein Internet-Wörterbuch (IW) handelt.

Für die Darstellung der Datenstrukturen wurde die Unified Modelling Language (UML) gewählt, da man mit deren Hilfe unabhängig von einer konkreten Implementierung Objekte und Abläufe zu visualisieren vermag. XML-Notation kommt in der Modellierung der Schnittstellen zum Einsatz und wird auch in dieser Form präsentiert. Die Implementierung der vorgestellten Strukturen und Mechanismen erfolgte in Ruby on Rails und MySQL.

2 Besonderheiten der japanischen Phraseologie

2.1 Grundbegriffe und Definitionen

Die folgenden Kapitel geben einen Überblick über die in der aktuellen phraseologischen Forschung und phraseographischen Praxis verwendeten Begriffe im Deutschen und im Japanischen. Da eine umfassende Analyse der Usancen den Rahmen der vorliegenden Arbeit sprengen würde, liegt der Fokus auf der Terminologie, die in den aktuellen theoretischen und praktischen Werken vermehrt Verwendung findet. In Anlehnung an die Erfahrung von Burger u. a. (2007, S. 9) wird es nicht als sinnvoll erachtet, phraseologische Terminologie und das klassifikatorische System, die unterschiedlichen Forschungskontexten entstammen, über Sprachgrenzen hinweg zu vereinheitlichen. Statt dessen werden die Begriffe vor ihrem jeweiligen thematischen und sprachlichen Hintergrund vorgestellt und zueinander in Beziehung gesetzt.

2.1.1 Die terminologische Vielfalt in der phraseologischen Forschung

Die rege Aktivität in der phraseologischen Forschung hat im Laufe der letzten Jahrzehnte eine Vielzahl konkurrierender Begriffe hervor gebracht, die über Sprach- und Disziplingrenzen hinweg nur begrenzt kompatibel sind (Burger u. a. 2007, S. 1–10). In der gegenwärtigen Forschung haben sich im deutschsprachigen Raum insbesondere »Phraseologismus«, »Phrasem« sowie »Mehrwortlexem« als übergeordnete Bezeichnungen für die Objekte des Phänomenbereichs durchgesetzt. Die wichtigsten Unterklassen stellen »Idiom« und »Kollokation«, die sich durch das Vorhandensein beziehungsweise Fehlen des Merkmals der Idiomatizität unterscheiden. »Routineformeln« werden entweder als eigene Klasse verankert oder – da ihre Abgrenzung von den bestehenden Klassen vornehmlich auf pragmatischen Gesichtspunkten beruht – in Abhängigkeit ihres Idiomatizitätsgrades mit »Idiomen« respektive »Kollokationen« verortet. Burger definiert die Objektklassen »Phraseologismus« und »Idiom« wie folgt:

(1) Als »Phraseologismus« bezeichnen wir eine Verbindung von zwei oder mehr Wörtern dann, wenn die Wortverbindung in der Sprachgemeinschaft, ähnlich wie ein Lexem, als feste Verbindung gebräuchlich ist (Burger 2002, S. 392)

(2) Diejenige Teilklasse der Phraseologismen, deren Komponenten eine durch die syntaktischen und semantischen Regularitäten der Verknüpfung nicht voll erklärbare Einheit bilden (=Merkmal der »Idiomatizität«), bilden den Bereiche der »Idiome«. Wenn man den Gesamtbereich der Phraseologismen meint (=Kriterium 1), spricht man auch von »Phraseologie im weiteren Sinne«. Die Brauchbarkeit der Definitionen hängt davon ab, ob es gelingt, den Begriff der »Festigkeit« zu operationalisieren, eine Kategorie wie »Gebräuchlichkeit« über die intuitive Einschätzung hinaus intersubjektiv festzumachen und die »Irregularität« der Verknüpfung differenziert zu erfassen. Schließlich ist auch die stillschweigende Voraussetzung, dass es bei einem Phraseologismus um die Verbindung mindestens zweier Wörter (was als »Polylexikalität« bezeichnet wird) geht, nicht unumstritten. (Burger 2002, S. 392)

Entsprechend hat sich als Gegenpart zu dem Begriff »Idiom« in der Phraseologie »Kollokation« als Unterklasse von Phrasemen etabliert, die sich durch das Merkmal schwacher oder gänzlich fehlender semantischer Umdeutung auszeichnen (Burger u. a. 2007, S. 7), wobei meist eine der Komponenten eine phraseologisch gebundene Bedeutung aufweist (Dobrovolskij und Filipenko 2007, S. 714). Dies ist nicht zu verwechseln mit der weiten Definition von Kollokation, unter welchem in der Korpuslinguistik alles gefasst wird, was mit statistischen Methoden als zusammengehörig isoliert werden kann (Burger u. a. 2007, S. 9).

Das Japanische verfügt ebenfalls über eine Fülle¹ von Begriffen für den Bereich der Phraseologie. Geprägt ist das Forschungsgebiet von der durch Miyaji geformten Terminologie (Miyaji 1982) (vgl. Abbildung 2.1), denn sein hierarchisches Modell wird wiederholt zitiert²:

¹ Allein als mögliche Entsprechungen für das englische *idiom* führt das SANSEIDŌ-GENGOGAKU-DAIJITEN-PW sieben japanische Termini an: *kanyōku* (慣用句), *kanyōgo* (慣用語), *kanyōgoku* (慣用語句), *kanyō hyōgen* (慣用表現), *kanyō rengo* (慣用連語), *seiku* (成句) und *jukugo* (熟語), jedoch ist diese Aufzählung in dem linguistischen Speziallexikon wenig hilfreich, da eine semantische oder strukturelle Abgrenzung der Begriffe zueinander nicht erfolgt.

² Zitiert wird die hierarchische Gliederung von Miyaji unter anderem bei Nakajima und Tonoike (1994, S. 238), Otani (2005, S. 537) und Kunihiro (1985, S. 7).

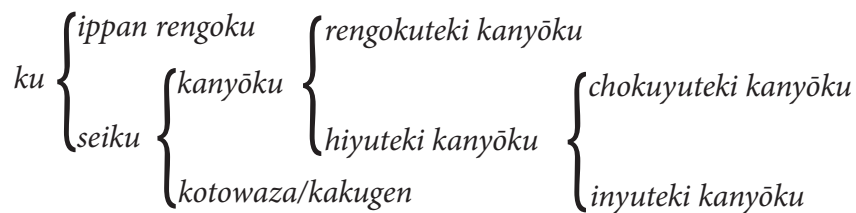


Abbildung 2.1: Klassifikation nach Miyaji (1985, S. 63)

Miyaji beginnt seine terminologische Klassifikation mit dem Wurzelknoten *ku* (句 Phrase), welchem *seiku* (成句) als obersten Begriff phraseologischer Terminologie untergeordnet ist, der alle als feste Verbindung gebräuchlichen Phrasen umfasst. Die weitere Untergliederung erfolgt abhängig von der Stabilität des Ausdrucks. So bezeichnet man eine Reihung von Worten, die als Phrase (*ku*) nur schwach miteinander korrelieren als *ippan rengoku* (一般連語句)³, *kanyōku* (慣用句) zeichnen sich im Gegensatz dazu durch vergleichsweise starke Kohäsionsmerkmale aus. Innerhalb der Gruppe der *kanyōku* differenziert Miyaji zudem zwischen *hiyuteki kanyōku* (比喩的慣用句) mit idiomatischer Bedeutung und *rengokuteki kanyōku* (連語的慣用句)⁴ mit kollokativem Charakter. Idiomatische *kanyōku* werden zusätzlich in metaphorische (*inyuteki kanyōku* 隱喩的慣用句) und metonymische (直喩的慣用句 *chokuyuteki kanyōku*) untergliedert. *Kotowaza* (諺) und *kakugen* (格言) werden als Sprüche oder Redeweisen charakterisiert, die im Volk über die Jahrhunderte hinweg überliefert wurden und allgemeine Regeln und Wahrheiten des Lebens auf satirische beziehungsweise belehrende Art und Weise wiedergeben (Miyaji 1985, S. 65). Obwohl diese sich auch deutlich durch ihre satzartige Struktur und ausgeprägte Stabilität

³ Nach eigenen Angaben verwendet Miyaji mit *ippan rengoku* eine Eigenprägung für etwas, das gemeinhin als *rengo* (連語) bezeichnet wird.

⁴ In früheren Arbeiten verwendet Miyaji auch die Begriffe *rengoteki kanyōku*, *rengo seiku* (連語成句) bzw. *rengoseikuteki kanyōku* (連語成句的慣用句) (Miyaji 1985, S. 63).

von *kanyōku* unterscheiden, spielen morphosyntaktische Charakteristika bei den Definitionen praktisch keine Rolle.⁵

Miyaji illustriert die Unterschiede zwischen den Subklassen mit Hilfe der folgenden Beispiele:

Subklasse	Beispiel	wörtlich	dt. Entsprechung
<i>rengoteki kanyōku</i>	<i>kaze-o hiku</i>	Erkältung-AKK ziehen	<i>sich erkälten</i>
<i>ippan rengoku</i>	<i>kumo-ga nagareru</i>	Wolken-NOM strömen	<i>die Wolken ziehen vorüber</i>
<i>hiyuteki kanyōku</i>	<i>abura-o uru</i>	Öl-AKK verkaufen	<i>Zeit vertrödeln</i>
<i>kotowaza</i>	<i>me-wa kuchi hodo-ni</i>	Augen-FOCUS Mund	<i>die Augen sagen mehr</i>
	<i>mono-o iu</i>	mehr als Dinge-AKK sagen	<i>als der Mund</i>
<i>kakugen</i>	<i>katte kabuto-no o-o</i>	Gewonnen haben	<i>Wenn du gewonnen</i>
	<i>shimeyo</i>	Helm-Poss Riemen-AKK schließen	<i>hast, ziehe den Helmriemen straffer.</i>

Tabelle 2.1: Beispiele japanischer Phraseme nach Subklasse

Aus den Beispielen in Tabelle 2.1 geht hervor, dass die Klassen *ippan rengoku*, *rengoteki kanyōku* und *hiyuteki kanyōku* ein Kontinuum bilden. Insbesondere in Bezug auf die Unterschiede zwischen *rengoteki kanyōku* und *ippan rengoku* wird deutlich, dass strukturelle Festigkeit sich auch in der Semantik widerspiegelt.⁶ So kann man argumentieren, dass *hiku* (引く ziehen) in *kaze-o hiku* eine phraseologisch gebundene Bedeutung aufweist, die an die Definition von Kollokation erinnert. Obwohl gerne zitiert ist diese Klassifikation in der japanischen Phraseologie nicht unumstritten. Momiyama (1997, S. 32) wirft beispielsweise die Frage auf, ob Miyajis Definition von *kanyōku* mit seiner Klassifikation konsistent ist:

⁵ Aufgrund ihres historischen Hintergrunds werden die Phänomene *kotowaza* und *kakugen* auch als *koji seigo* (überlieferte Redewendung, 故事成語) zusammen gefasst.

⁶ Den Zusammenhang zwischen struktureller und semantischer Festigkeit weist Ishida (2003) nach und schlägt eine Methode vor, mit deren Hilfe semantische Festigkeit anhand struktureller Merkmale gemessen werden kann.

Außerdem definiert [Miyaji] in selbigem Werk *kanyōku* als »Ausdruck, der eine Verbindung aus mindestens zwei Wörtern ist, deren Zusammenhalt relativ stark ist und die als Ganzes eine festgelegte Bedeutung haben« (S. 238). Wenn man die Beschreibung »als Ganzes eine festgelegte Bedeutung« als »eine Bedeutung, die sich nicht aus den Komponenten der Einzelteile erschließt« interpretiert, ist es inkonsistent angesichts der Definition von *kanyōku* in selbigem Werk *rengoteki kanyōku* als eine Subklasse von *kanyōku* aufzufassen.⁷

Wie Miyaji terminologisch das Wortfeld strukturiert, ist in Bezug auf die Nomenklatur der *kanyōku* offensichtlich, denn alle Begriffe der Subknoten von *kanyōku* sind Komposita desselben. Warum er jedoch diesen Begriff wählt, wird aus seinen Erläuterungen nicht deutlich, denn er selbst stuft ihn als »instabil« (安定していない *antei shite inai*) ein (Miyaji 1985, S. 62). Im Kontext von *rengo* hingegen begründet er die Entscheidung gegen die Verwendung dieses etablierten Begriffs mit widersprüchlichen Definitionen in den großen Wörterbüchern. Auch positioniert er die Kategorie *rengoteki kanyōku* an einer von *ippan rengo* abgekoppelten Position, obwohl diese Kategorien inhaltlich ein Kontinuum bilden.

Der hauptsächliche Problempunkt ist jedoch, dass Miyajis Terminologie weder mit den Usancen im Forschungsfeld konsistent ist⁸, noch mit dem alltäglichen Gebrauch der Begriffe⁹, die jeweils Idiomatizität als integratives Merkmal der *kanyōku* betrachten. Das wiederum tangiert auch die benachbarten Phänomene, denn seine Definition von *seiku* stimmt nicht mit der in einschlägigen Wörterbüchern überein, die das Merkmal der Idiomatizität impliziert, während Miyaji an dieser Stelle seines hierarchischen Modells noch keine Trennung zwischen idiomatischen und nicht-idiomatischen Ausdrücken vollzieht.

Auch die kontrastive Phraseologieforschung in Japan sieht Idiomatizität durchweg als ein entscheidendes Merkmal von *kanyōku* und verwendet es als japanisches Äquivalent für »Idiom«

⁷ Im Original heißt es: なお、同書は慣用句を「単語を二つ以上の連結体であって、その結びつきが比較的固く、全体で決まった意味を持つ言葉」(p. 238)と定義している。「全体で決まった意味を持つ」という記述を「構成語の意味の総和からは導き出せない意味」と解釈すると、連語成句的慣用句を慣用句の一種に含めることは、同書の慣用句の定義に照らしても適切でないことになる。

⁸ Unter anderem in Kunihiro (1985, S. 5–6), Nakajima und Tonoike (1994, S. 238), Momiyama (1997, S. 29–30) sowie Otani (2005, S. 537).

⁹ So denotiert laut SHŌGAKUKAN-NIHONKOKUGODAIJITEN-PW₂ *kanyōku* eine Einheit von mindestens zwei Wörtern, die in ihrer Gesamtheit eine besondere Bedeutung haben.

(Itoh 2002; Ōtsubo 1985). Im bilingualen Kontext wird der umstrittene Begriff zu Gunsten des Fremdwortes *idiomu* ganz gemieden. Vergleicht man das Wortfeld im Japanischen und Deutschen, ergeben sich folgende Übereinstimmungen: *ku* lässt sich in etwa mit »Phrase« im Sinne einer Verbindung mehrerer Wörter unabhängig von anderen Kriterien gleichsetzen. Eine Phrase, die eine gewisse strukturelle Stabilität aufweist, bezeichnet im Japanischen *seiku*. Der zentrale Begriff *kanyōku* würde im Deutschen am ehesten dem Terminus »Idiom« entsprechen, nach Miyajis Definition »Phrasem« beziehungsweise »Phraseologismus«. Für *rengo* eignet sich als Pendant der Begriff »Kollokation«. Eine Unterscheidung zwischen Miyajis *rengoteki kanyōku* und *ippan rengoku* ist im Deutschen nicht üblich.

Genau genommen gibt es im Deutschen für *kanyōku* als Sammelbegriff für Kollokationen und Idiome, abzüglich der Sprichwörter und Aphorismen (Miyaji), keine einfache Entsprechung, da »Sprichwörter« nicht als ein von Idiomen separierbares Phänomen gelten.

2.1.2 Begriffsfelder und Definitionen in der lexikographischen Praxis

In Titeln und Umtexten phraseologischer Wörterbücher trifft man auf eine Mischung aus alltagssprachlichen Begriffen und wissenschaftlicher Terminologie. Während die Betitelung der Werke Anhaltspunkte liefern kann, wie das Werk marktwirtschaftlich effektiv präsentiert werden soll (Burger u. a. 2007, S. 5), lassen Umtexte durch darin formulierte Ansichten Rückschlüsse auf das »phraseologische Bewusstsein« der Redaktion zu (Stantcheva 2002, S. 6).

Für den deutschen Sprachraum werden Werke, die im Kern nicht-satzwertige undurchsichtige Ausdrücke umfassen, meist durch die Begriffe »Idiomatik«, »Redensarten« oder »Redewendungen« gekennzeichnet.¹⁰ Enthalten Wörterbücher Sprichwörter und Zitate, werden diese meist explizit im Titel genannt, wie im Falle des »Duden, das große Buch der Zitate und Redewendungen« oder das »Wörterbuch der sprichwörtlichen Redensarten«.

Für die Idiomatik des Deutschen ist der DUDEN-REDEWENDUNGEN-PW₃ das Standardwerk für nicht-satzwertige Idiome. Die Erstauflage von 1982 ist ein gutes Beispiel, um zu illustrieren, wie verwirrend der Terminologiegebrauch in der phraseographischen Praxis des Deutschen

¹⁰ Beispiele sind unter anderen der Duden Band 11 »Duden Redewendungen: Wörterbuch der deutschen Idiomatik« (DUDEN-REDEWENDUNGEN-PW₃), »Duden Redensarten: Herkunft und Bedeutung« (DUDEN-REDENSARTEN-HERKUNFTUNDBEDEUTUNG) oder das zweibändige »Deutsche Idiomatik: die deutschen Redewendungen im Kontext« (SCHEMANN-DEUTSCHEIDIDIOMATIK-PW).

ist, denn Titel und Umtexte wurden damals offenbar nicht aufeinander abgestimmt. Zum einen werden die »sprichwörtlichen Redensarten«¹¹ aus dem Titel in den Umtexten nicht weiter erwähnt, des Weiteren wird in der Einleitung zwischen Sprichwörtern und Redensarten unterschieden.¹² Dass es sich hierbei nicht um einen Einzelfall handelt, erläutert Burger (2010, S. 179–180) anhand der Ungenauigkeiten in den einleitenden Texten zum *Duden Gesamtwörterbuch* und *Duden Universalwörterbuch*. Für die zweite Auflage mussten die »sprichwörtlichen Redensarten« im Titel weichen, dafür wurde in der Einleitung die Aufzählung unter anderem um »Redensarten« und »feste Wendungen« erweitert:

Redensarten, Redewendungen, feste Wendungen, idiomatische Wendungen, feste Verbindungen, Idiome, Wortgruppenlexeme, Phraseologismen, Phraseolexeme – mit diesen und anderen Bezeichnungen versucht die Sprachwissenschaft, bestimmte sprachliche Erscheinungen zu fassen, die man als »vorgeformte«, nicht frei gebildete Wortketten, als eine Art »Fertigbauteile« umschreiben kann. (Alsleben 2002, S. 9)

Damit wurde in der zweiten Auflage die Verbindung zwischen Titel und Inhalt hergestellt, jedoch dienen auch hier die meisten der genannten Begriffe allein als thematische Einleitung. Die Ergänzung um eine Allegorie, die den Phänomenbereich als »Fertigbauteile« umschreibt, versucht eine zusätzliche Verständnishilfe zu geben. Auch in dieser erweiterten Einleitung fehlt eine Ausdifferenzierung des Wortfelds. Zur Erklärung der Charakteristika und Eingrenzung des Objektbereichs werden im weiteren Verlauf des Einleitungstexts »feste Wendung« sowie

¹¹ »Sprichwörtliche Redensart« wird laut Burger dann verwendet, wenn ein Phrasem von einem Sprichwort abgeleitet wird oder der Autor eine Affinität zur ethnologischen Forschung ausdrücken möchte. Der Begriff wurde als Lehnwort für das französische *façon de parler* in die deutsche Sprache aufgenommen und wird heutzutage vor allem im Bereich der Sprachdidaktik verwendet (Burger u. a. 2007, S. 7).

¹² Dieser Band beginnt die Einleitung mit einem Versuch einer Heranführung des Benutzers an die linguistische Terminologie: »Feste oder idiomatische Verbindungen, Redewendungen, Idiome, Wortgruppenlexeme, Phraseologismen, Phraseolexeme - mit diesen und anderen Termini versucht man in der Sprachwissenschaft, bestimmte sprachliche Erscheinungen zu fassen, die man als »vorgeformte«, nicht frei gebildete Wortketten umschreiben kann (DUDEN-REDEWENDUNGEN-PW 1998, S. 7).« Überwiegend verwendet die Einleitung den Begriff *festen Wendungen*, der weder in Titel noch in der Aufzählung von (angeblich) in der Sprachwissenschaft gängigen Fachbegriffe vorkommt.

die aufgrund ihrer umgangssprachlichen Prägung in der Sprachwissenschaft gemiedenen Begriffe »Redewendung« und »Redensart« verwendet, wobei erstere als Oberbegriff für den Phänomenbereich dienen und »Redensart« satzförmige Aussprüche bezeichnet, die formelhaften Charakter haben wie beispielsweise *du kriegst die Tür nicht zu* (Alsleben 2002, S. 13).

Die phraseologischen Wörterbücher für das Japanische sind zahlreich, in ihrer Betitelung werden jedoch relativ konsistent die Begriffe eingesetzt, die auch in der linguistischen Forschung bekannt sind: Für das weitere Gebiet der Sprichwörter wird meist *kotowaza* verwendet und *kakugen* als Sammelbegriff für Sinnsprüche mit historisch-moralischem Hintergrund. *Kanyōku* steht insbesondere für verbale Phraseologismen mit zumindest schwacher idiomatischer Bedeutung, während der Anglizismus *idionu* (イディオム engl.: *idiom*) dem bilingualen Kontext vorbehalten ist. Für Kollokationen existieren der von dem englischen *collocation*) abgeleitete Begriff *korokēshon* (コロケーション) und das japanische *nihongo hyōgen* (日本語表現 japanische Ausdrücke). Das aus dem Chinesischen entlehnte *seigo* (成語) ist gleichbedeutend mit *seiku* und findet sich als allgemeine Bezeichnung für Phraseme beispielsweise im Titel des *Seigo-rin* (成語林 Hain der idiomatischen Redewendungen), einem der umfassendsten Phrasologismen-Lexika, es ist auch Bestandteil der Komposita *yoji seigo* (四字成語) und *koji seigo* (故事成語), welche alternative Bezeichnungen für die gebräuchlicheren Begriffe *yoji jukugo* (四字熟語 Vier-Zeichen-Phraseme) und *koji kotowaza* (故事諺 Sprichwörter Sammlungen) darstellen. Die Begriffe *rengo* und *seiku* treten nicht in Erscheinung.

Dass es in der japanischen Phraseographie ähnlich unpräzise Abgrenzungen zwischen den Teilklassen gibt, thematisiert das YONEKAWA-NIHONGO KANYŌKUJITEN-PW in seiner Einleitung:

In der Vergangenheit gab es zwar bereits Werke, die als *Kanyōku jiten* betitelt wurden, viele erfassten jedoch die Bedeutung von *kanyōku* nicht richtig und mischten *kotowaza*, *kakugen* und *kanyōku*.¹³ (Otani 2005, S i)

Auch gibt das YONEKAWA-NIHONGO KANYŌKUJITEN-PW im weiteren Verlauf der Einleitung Auskunft darüber, wie der fragliche Begriff aus Sicht der Redaktion richtig zu verstehen ist:

Als *kanyōku* bezeichnet man Phraseme (*seiku*), die sich aus mindestens zwei Wörtern zusammensetzen, dabei einen relativ festen Zusammenhalt aufweisen und als Gesamtheit

¹³ Im Original heißt es: 過去において、「慣用句辞典」と名のつくものがあつたが、諺・格言・慣用句などが混在したものが多く、慣用句を正しく理解していなかった。

eine bestimmte Bedeutung haben. Und obwohl sie ebenfalls zu den Phrasemen (*seiku*) gehören, unterscheiden sie sich von den Sinnsprüchen (*kakugen*) und Sprichwörtern (*ko-towaza*).¹⁴ (Otani 2005, S. i)

Diese Definition deckt sich fast wörtlich mit der in der Sprachwissenschaft gängigen, jedoch geht daraus keine klare Abgrenzung gegenüber Sprichwörtern und Maximen hervor. Auf sie folgt eine Kurzdefinition der Begriffe sowie ein Verweis auf die im Anhang enthaltene ausführliche Klassifikation und Charakterisierung, die den Stand der einschlägigen Forschung widerspiegelt (Otani 2005, S. 537–563).

Zusammenfassend kann man sagen, dass in beiden Sprachen eine Vielzahl von Begriffen zur Verfügung stehen, um das Spektrum der Phraseologie abzudecken. Jedoch ist der aktive Gebrauch im Japanischen auf wenige Begriffe beschränkt. Auch gibt es deutlichere Übereinstimmungen zwischen der in der linguistischen Forschung verwendeten Terminologie und der lexikographischen Praxis.

2.1.3 Terminologische Standards der Sprachtechnologie

In der Sprachtechnologie findet sich nur ein kleiner Teil der Begrifflichkeiten aus der phraseologischen Forschung und Praxis. Durch die Vorherrschaft statistischer Methoden liegt derzeit das Augenmerk auf der automatischen Extraktion und der semantischen Zuordnung von Phrasemen und weniger auf ihrer detaillierten Analyse. Dementsprechend überwiegen in der Terminologie allgemein deskriptive Begriffe und auf strukturellen Merkmalen basierende Definitionen. Verwendung finden beispielsweise »phraseologische Einheit«, »Mehrworteinheit« beziehungsweise »Mehrwortlexem«, sehr häufig auch »Kollokation« oder im Englischen *multi word unit* und *fixed expression*. Subtile, auf semantischen Kriterien basierende Gruppierungen, die für die Phraseologieforschung prägend sind, spielen derzeit in der Sprachtechnologie eine untergeordnete Rolle.

Um dem Wildwuchs der Terminologie entgegen zu wirken, wurde von der International Standards Organisation (ISO) speziell für die Terminologie von sprachtechnologischen Ressourcen eine Arbeitsgruppe (TC37/SC4) eingerichtet (Ide und Romary 2004, S. 135–137). Die Ergebnisse dieser Expertengruppe werden in einem sogenannten *Data Category Registry* (DCR) in

¹⁴ Im Original heißt es: 「慣用句」とは単語の二つ以上の連結体であって、その結びつきが比較的固く、全体で決まった意味を持つ成句である。そして同じ成句でも格言[...]や諺[...]とは異なる。

Form von Datenkategorien (*data categories*) mit Definitionen und Beispielen gepflegt und über die Nutzeroberfläche ISOCAT¹⁵ der Öffentlichkeit zur Verfügung gestellt (Kemps-Snijders u. a. 2008). Dies ist ein Versuch, die Expertise der Forschungsgemeinschaft zu bündeln und damit Transparenz und einheitliche Verwendung der Begrifflichkeit zu gewährleisten. DCR und ISOCAT sind noch im Aufbau begriffen und weit von einer umfassenden terminologischen Erfassung des Objektbereichs entfernt. Für den Phraseologie-Bereich sind derzeit in ISOCAT nur die Begriffe *idiom* und *phraseological unit* verfügbar:

Idiom: A group of words in a fixed order that have a particular meaning that is different from the meanings of each word understood on its own. (Wright 2009a)

Phraseological unit: Any group of two or more words that form a unit, the meaning of which frequently cannot be deduced based on the combined sense of the words making up the phrase. (Wright 2009b)

Problematisch an der Definition von *idiom* ist, dass von einer festen Wortstellung ausgegangen wird, die keine sprachübergreifende Gültigkeit besitzt. Möglicherweise wurde diese Definition mit Blick auf die Idiomatik des Englischen entwickelt, betrachtet man jedoch beispielsweise die des Deutschen, variiert die Anordnung der Komponenten gegebenenfalls bereits zwischen der Nennform (z. B. *jemanden einen Korb geben*) und flektierten Formen (z. B. *sie gibt ihrem Verehrer einen Korb*). Im Vergleich zu der Definition von *phraseological unit* fällt auch auf, dass die Gesamtbedeutung des Idioms nicht zur *Summe der Einzelworte* in Beziehung gesetzt wird, sondern zu der Bedeutung der darin enthaltenen Worte. Zusammenfassend kann man sagen, dass die Definition von *idiom* nicht den Stand der Forschung entspricht und in der Erklärung von *phraseologicalUnit* eine eindeutige Aussage wünschenswert wäre, ob nicht-idiomatische Ausdrücke auch Teil des Phänomenbereichs sind.

Vor dem Hintergrund der terminologischen Vielfalt auf diesem Gebiet erscheint die Standardisierungsbestrebung überfällig, jedoch ist sie auf die Bereitschaft einer breiten Basis von Experten angewiesen, welche für ihre Sprache und ihre Disziplin Beiträge leisten. Bedauerlicherweise ist die Abdeckung im Bereich Phraseologie noch lückenhaft und den Definitionen fehlt teilweise der Bezug zum aktuellen Stand der Forschung.

¹⁵ Die DCR Datenbasis ISOCAT ist verfügbar unter <http://www.isocat.org>.

2.1.4 Zusammenfassung

Die Gegenüberstellung der Terminologie in den relevanten Sprachen und Disziplinen hat gezeigt, dass sich hinter der terminologischen Vielfalt abhängig vom fachlichen und sprachlichen Kontext unterschiedliche Blickwinkel verbergen. In der vorliegenden Arbeit werden die terminologischen Unterschiede zwischen den Disziplinen und Sprachen insoweit beibehalten, als sie die Gebiete und die Phänomene in den untersuchten Sprachen widerspiegeln.

Definitiv wird auf die gegenwärtigen Arbeiten von Burger aufgesetzt, da die von ihm verwendete Klassifikation die Möglichkeit der differenzierten Betrachtung bietet. Der Begriff Phraseologie steht somit für das Forschungsgebiet, deren Oberbegriff für die sprachlichen Einheiten, »Phraseologismus« wird jedoch durch den prägnanteren Begriff »Phrasem« ersetzt. Aufgrund fehlender semantischer Subklassifikationsmöglichkeiten wird im sprachtechnologischen Kontext statt dessen der strukturell geprägte Begriff »Mehrwortlexem« verwendet.

Als Subklassen spielen insbesondere Idiome und Kollokationen eine Rolle, die sich durch das Merkmal deutlicher Idiomatizität respektive einer phraseologisch gebundenen Bedeutung auszeichnen. Im Zusammenhang mit der Analyse japanischer Phraseme wird aufgrund der Unterschiede zwischen der Terminologie im Japanischen und Deutschen der Begriff *kanyōku* weitgehend beibehalten. Routineformeln werden als gesonderte Form betrachtet und sind somit nicht Gegenstand dieser Arbeit.

2.2 Phraseologische Merkmale und ihre Auswirkungen auf die Phraseographie

Die Charakteristika der *kanyōku* umfassen strukturelle und semantische Merkmale, die sich mit Hilfe der Stichworte Polylexikalität, Stabilität, Idiomatizität und Reproduzierbarkeit resümieren lassen. Jedes dieser Charakteristika hat eine Reihe von Auswirkungen auf die phraseographische Praxis. Das Kapitel skizziert zunächst ganz grundlegende Fragen, die der Komplexität des japanischen Schriftsystems geschuldet sind. Darauf werden verschiedene Aspekte des Kontinuums zwischen Stabilität und Variabilität betrachtet. Schließlich mündet die Diskussion der semantisch-pragmatischen Charakteristika in die Frage nach Äquivalenzbeziehungen zwischen idiomatischen Phrasemen.

2.2.1 Polylexikalität vor dem Hintergrund des japanischen Schriftsystems

Für das Japanische kann in Abhängigkeit von der Leitschrift grundsätzlich zwischen *kanji*-Lexika (漢字 chinesische Schriftzeichen), *kana*-Lexika (仮名 Silbenschriftzeichen) und *rōmaji*-Lexika (ローマ字 lateinische Buchstabenschrift) differenziert werden (Lewin 1991, S. 2617). Die Wahl des Verfahrens ist abhängig von Wörterbuchinhalt und -funktion: *kanji*-Lexika dienen der Recherche von Lesung und Bedeutung chinesischer Schriftzeichen und deren Komposita im Japanischen. Sie bieten so die Möglichkeit, ein Zeichen ohne Kenntnis der Lesung oder Bedeutung nach Typ (Erscheinungsbild) nachzuschlagen. Aufgrund der Komplexität des Schriftsystems¹⁶ konzentriert sich die Information in *kanji*-Nachschlagewerken meist auf die wesentlichen Angaben zu Form, Kombinatorik und Bedeutung der Schriftzeichen.

Mit den Details einzelner Lexeme, zu welchen auch *kanyōku* gehören, befassen sich *kana*- und *rōmaji*-Wörterbücher. Für Nachschlagewerke aller Art mit Japanisch als Ausgangssprache überwiegt in der japanischen Lexikographie das *kana*-Syllabar als Leitschrift in Kombination mit dem auf der 50-Laute-Tafel (*gojū onzu* 五十音図) basierenden Ordnungssystem.¹⁷

¹⁶ Für eine Einführung in die japanische Schrift und Schriftgeschichte siehe auch Müller-Yokota (1989). Eine Übersicht über die der japanischen Schrift inhärenten Probleme gibt Eschbach-Szabo (2001)

¹⁷ Die Behauptung Trippels, es gäbe im Japanischen keine dem Alphabet entsprechende Sortierkonvention bzw. japanische Systeme würden nach Strichzahl des Schriftzeichens sortieren (Trippel 2006, S. 44), geht an der

Die 50-Laute-Tafel gliedert die Silben im Japanischen nach den Vokalen *a-i-u-e-o* in fünf Reihen, die in Kombination mit den Konsonanten insgesamt 50 Laute ergeben. Dies allein ist als Ordnungssystem für die Sortierung von Lexemen für das moderne Japanisch nicht ausreichend. In Benutzungshinweisen werden meist zusätzliche Regeln definiert u. a. für stimmhafte Laute, Längungs- und Wiederholungszeichen sowie für Homophone. Darüber hinaus enthält die Auflistung Silben, die im modernen Japanisch obsolet sind (unter anderem *we* und *wi*), die Partikel *o* sowie der einzige einzelne Konsonant *n*.¹⁸ Einzelne Lernerwörterbücher romanisieren die japanischen Lexeme und ordnen sie nach Alphabet (z. B. KŌDANSHA-KODANSHA'S ROMANIZED JAPANESE ENGLISH-PW und SŌTAKUSHA-SHIBATA'S JAPANESE ENGLISH-ROMAJI-PW).

2.2.1.1 Segmentierung und Komponentenanalyse

Das Thema Segmentierung befasst sich mit der Frage, welche eigenständigen Teile (Token) eine Zeichenkette beinhaltet. Im Zusammenhang mit Phrasemen entsprechen diese Token in der Regel dem in der Phraseologie verbreiteten Konzept der »Komponente«¹⁹. Als Komponenten werden die Einheiten eines Phrasems betrachtet, die als separate Einträge im Lexikon geführt werden und auch für sich alleine stehen können. Für das Japanische sind das im Kern

lexikographischen Realität in Japan vorbei. Strichzahlindizes finden sich tatsächlich als sekundäre Zugriffsmethoden in Lexika, um das Nachschlagen unbekannter Schriftzeichen zu ermöglichen. Aufgrund der Ineffizienz dieses Verfahrens (die Strichzahl ist kein geeignetes Differenzierungsmerkmal und deren Bestimmung vergleichsweise fehleranfällig) ist diese Methode jedoch als primäre Ordnungsstruktur nicht üblich. Gängig ist insbesondere die Anordnung auf Basis struktureller Konstituenten (das sogenannte »Radikalsystem«) (Kerde und Otto 1996, S. 18). Daneben existieren nach Lesung oder nach Zeichenbedeutung geordnete Schriftzeichennachschlagewerke (Lewin 1991, S. 2617). Eine ausführliche Beschreibung der *kanji*-Nachschlagemethoden in elektronischen Wörterbüchern findet sich bei Sharpe (1995), wobei durch die rasante Entwicklung auf dem Gebiet die von ihm beschriebenen Methoden zunehmend an Bedeutung verlieren.

¹⁸ Wenn in dieser Arbeit von dem auf der 50-Laute-Tafel basierenden Ordnungssystem die Rede ist, wird die Berücksichtigung der zusätzlich in Wörterbucheinleitungen definierten Sonderregelungen impliziert.

¹⁹ Parallel zu dem Begriff »Komponente«, den beispielsweise Burger (Burger 2002) bevorzugt, findet sich in anderen Werken auch der Terminus »Konstituente« (Itoh 2005).

Nomina, Verben, Adjektive (auch Verbaladjektive) und Attribute. Partikeln und Verbalsuffixe werden nicht als eigenständige Komponenten betrachtet, sondern der sie modifizierenden Komponente zugerechnet.

Die im europäischen Sprachraum übliche Herangehensweise, zunächst bei der Tokenisierung die Zeichenketten zwischen Leerzeichen (und Satz- bzw. Sonderzeichen) als erste grobe Unterteilung eines Textes zu interpretieren, ist für die japanische Sprache wie auch für andere asiatische Sprachen (u. a. Chinesisch und Koreanisch) nicht geeignet, da in der Regel keine explizit gekennzeichneten Einschnitte unterhalb der Satzebene existieren.²⁰ Auch wenn in anderen Sprachen die Markierung in der Regel nur der erste Anhaltspunkt für eine Segmentierungsroutine ist, sind die Methoden der natürlichen Sprachverarbeitung für das Japanische stärker als die für mitteleuropäische Sprachen von statistischen Verfahren und Wortlisten abhängig (Halpern 2000).

Dies führt dazu, dass es mehr noch als in westlichen Sprachen die Möglichkeit unterschiedlicher Segmentierungsvarianten gibt. Die Wahl der Methode, nach welcher die einzelnen Komponenten bestimmt werden, tangiert nicht nur die Eingrenzung des Phänomenbereichs, sie bestimmt auch, welche Elemente zur weiteren Verarbeitung isoliert werden. Schwierigkeiten bereitet beispielsweise eindeutige Bestimmung der Komponenten eines Mehrwortlexems, wenn eine Kombination mehrerer Komponenten als Kompositum lexikalisiert ist. Dies wiederum hat Auswirkungen auf das Phänomen der (orthographischen) Variabilität und auf die Komponentenstruktur. Um ein Ergebnis reproduzierbar zu machen, ist jedoch notwendig, einen Segmentierungsmechanismus mit einem klar definierten Regelsatz zu beschreiben.

Am Beispiel des Idioms *hara-no mushi-ga osamaranai* wird im Folgenden illustriert, wie sich unterschiedliche Herangehensweisen in der Phraseographie, Computerlinguistik und kontrastiven Linguistik auf die Segmentierung eines Mehrwortausdrucks auswirken können.

In der Phraseographie wird die Zeichenkette so zerlegt, dass jede Komponente einem Lexem entspricht, das als Lemma oder Stichwort eines Wörterbuchartikels beziehungsweise eines Eintrags in einer lexikalischen Ressource dienen kann. Beispiel 1 zeigt die entsprechende Aufteilung der Zeichenkette in ihre Komponenten.

²⁰ Ausnahmen finden sich beispielsweise in Kinderbüchern und Lehrbüchern, die teilweise oder ganz auf die Verwendung von *kanji* verzichten, und Wortgrenzen mit Leerzeichen kenntlich machen, um das Lesen (oder Vorlesen) zu erleichtern.

- (1) 腹 の 虫 が 治まら ない
hara no mushi ga osamara nai
 Bauch GEN Wurm SUBJ beruhigen NEG
sich ärgern; sich nicht beruhigen können

Das Ergebnis besteht aus sechs Einheiten, darunter gleichermaßen Autosemantika (die Nomen *hara* und *mushi* sowie das Verb *osamaru*) und Synsemantika (die Partikeln *no* und *ga*, sowie das Verbalsuffix *nai*).

Die Computerlinguistik bedient sich automatisierter Verfahren. Die morphosyntaktische Analyse derselben Phrase unter Einsatz des MeCab-Parsers (siehe Kapitel 4.10.3) liefert eine etwas andere Aufteilung:

- (2) 腹の虫 が 治まら ない
haranomushi ga osamara nai
 Bauchwurm SUBJ beruhigen NEG
sich ärgern; sich nicht beruhigen können

Der MeCab-Parser segmentiert mit Hilfe eines Lexikons, gewichtet Alternativen jedoch auf Basis statistischer Methoden. Er separiert ebenfalls Verbstamm (*osamara*) und Verbalsuffix (*nai*), erkennt aber *haranomushi* als Kompositum und isoliert somit aus der Zeichenkette nur vier Elemente.

In der kontrastiven Phraseologismenforschung werden Komponenten in der Regel auf Basis semantischer Kriterien bestimmt. In den einleitenden Kapiteln von Itoh (2005) werden die Regeln für die Segmentierung der Beispiele nicht offen gelegt, vornehmlich werden dabei jedoch Autosemantika als Kandidaten für Komponenten berücksichtigt. Anhand strukturell ähnlicher Beispiele 3 und 4 wurde eine analoge Aufteilung in Komponenten für Beispiel 5 durchgeführt.

- (3) 身の毛が よだつ
minoke-ga yodatsu
 Haar-NOM sträuben
entsetzlich; haarsträubend; schrecklich (Itoh 2005, S. 86)

- (4) 眉一 動かさない
mayuhitotsu ugokasanai
 Augenbraue-AKK nicht bewegen
ohne mit der Wimper zu zucken (Itoh 2005, S. 167)

- (5) 腹の虫が 治まらない
haranomushi-ga osamaranai
Bauchwurm-SUBJ nicht beruhigen
sich ärgern; sich nicht beruhigen können

Die Ergebnismenge besteht lediglich aus zwei, statt der vier beziehungsweise sechs Einheiten der vorangegangenen Analysen, da jeweils das Kompositum *haranomushi* mit der Partikel *ga* und der Verbstamm *osamara* mit dem Verbalsuffix *nai* eine Komponente bilden.

Welcher Mechanismus für die lexikographische Arbeit besonders geeignet ist, lässt sich nicht pauschal sagen. Für die Berechnung der orthographischen Varianten wird ein Mechanismus benötigt, der möglichst granulare Ergebnisse liefert (Beispiel 1). In grammatischen Angaben werden in der Regel Verbalsuffixe nicht separiert (Beispiel 5). Reproduzierbar und am engsten an der Realität der lexikalischen Einheiten im Lexikon ist Ausgabe des MeCab-Parsers (Beispiel 2).

2.2.1.2 Orthographische Variation

Im japanischen Printlexikon erfolgt der Zugriff auf die Inhalte in aller Regel über die Lesung, die durch eine (relativ) stabile *kana*-Zeichenkette repräsentiert wird. Orthographische Variation spielt daher eine untergeordnete Rolle. Anders verhält es sich im elektronischen Medium, wo erwartet wird, dass man auf eine beliebige Repräsentation der Zeichenkette zugreifen kann. Problematisch sind insbesondere *kanji-kana*-Mischformen, denn jede kleine Abweichung in der Repräsentation kann den Zugriff auf die Inhalte des Lexikons behindern.

Die folgende Darstellung beschränkt sich auf die quantitativ prominenten Arten orthographischer Variation durch die unterschiedliche Verwendung von *kanji* und *kana*. Dazu gehören die Kombinatorik von *kanji*- und *kana*-Verschriftlichung, alternative *kanji*-Schreibweisen und Differenzen in dem Einsatz von *okurigana* (送り仮名 »begleitende Zeichen«). Die vereinzelt auftretenden Varianten durch den Gebrauch von historischen *kana* oder Verdoppelungszeichen (蝶々 versus 蝶蝶) werden aufgrund ihres verschwindend geringen Anteils im Lexikon der Gegenwartssprache nicht weiter thematisiert. Die einfachste Form der orthographischen Variation betrifft nahezu alle japanischen Phraseme, da jede *kanji*-Zeichenkette ebenso anhand

ihrer phonetischen Merkmale in *kana* repräsentiert werden kann.²¹ Die aus der Printlexikographie übernommene Herangehensweise, sowohl *kanji*- als auch *hiragana*-Repräsentationen in den Index elektronischer Lexika aufzunehmen, ermöglicht üblicherweise den Zugriff über die phonetische Repräsentation in *hiragana*.

Bei Einzelwortlexemen ist diese Herangehensweise auch völlig unproblematisch, da hier das Alles-oder-Nichts-Prinzip die Regel ist. Ein Problem ergibt sich jedoch bei Mehrwortlexemen, da die Wahl der Schriftzeichenart für jede einzelne Komponente variieren kann. In natürlich-sprachlichen Texten ist es beispielsweise nicht unüblich für *hara-no mushi-ga osamaranai* das Verb *osamaranai* in *hiragana* (おさまらない) statt in *kanji* (u. a. 治まらない) anzugeben. Wurde diese *kanji-kana*-Mischform jedoch nicht bei der Indizierung berücksichtigt, läuft die Suche danach mit großer Wahrscheinlichkeit ins Leere. Lediglich wenn für jede durch *kanji* repräsentierbare Komponente eine zusätzliche *hiragana*-Variante berücksichtigt wird, ist der Bereich der Variation durch *kanji-kana*-Mischformen abgedeckt. In den untersuchten elektronischen Wörterbüchern wurde diesem Punkt keine Beachtung geschenkt.

Der zweite Typ basiert auf der Möglichkeit für bestimmte Begriffe unterschiedliche *kanji*-Schreibweisen zu verwenden.

(6) 鼻毛を 延ばす
hanage-o *nobasu*
Nasenhaare-AKK wachsen lassen
eine Schwäche für Frauen haben

(7) 鼻毛を 伸ばす
hanage-o *nobasu*
Nasenhaare-AKK wachsen lassen
eine Schwäche für Frauen haben

Gelegentlich unterscheiden sich die Schriftzeichenvarianten um Bedeutungsnuancen oder pragmatische Aspekte. Häufiger jedoch lassen sich den unterschiedlichen Varianten aufgrund der phraseologischen Umdeutung keine Bedeutungsnuancen mehr zuzuordnen, was auch in den Beispielen 6 und 7 der Fall ist. Unterschiedliche *kanji*-Schreibweisen treten überwiegend in

²¹ In Ausnahmefällen ist auch die Verwendung von *katakana* möglich, beispielsweise wenn ein Begriff besonders betont wird.

Korpusbeispielen in Erscheinung, gelegentlich unterscheiden sich jedoch auch die Lemmaformen verschiedener Wörterbücher. So lassen sich zwar in den Beispielen 6 und 7 keine Bedeutungsunterschiede feststellen, dennoch präsentiert die überwiegende Anzahl der vorliegenden Wörterbücher *hanage-o nobasu* in Form von Beispiel 6 (伸ばす). Eine Ausnahme bildet das SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW durch die Verwendung des selteneren *kanji* (延ばす). Bei dem dritten Typ orthographischer Variation handelt es sich um die relativ unscheinbaren *okurigana* (送り仮名 begleitende *kana*), genauer gesagt den Teil der *okurigana*, die in einem Lexem, welches *kanji* beinhaltet, angefügt werden, um die japanische Lesung des vorangegangenen chinesischen Ideogramms zu stützen.²² Beispiel 8 zeigt die Unterschiede zwischen zwei *okurigana*-Varianten des Verbs *okonau*.

- (8) 行 な う
 行 う
oko na u
 begleiten

Zwar gibt das japanische »Amt für Kunst und Kultur« (文化庁 *bunkachō*) Empfehlungen zum Einsatz der Hilfszeichen, die vom Innenministerium im Jahre 1973 bekannt gegeben wurden, verbindlich sind diese aber nur für den Gebrauch in Gesetzestexten und anderen amtlichen Veröffentlichungen sowie Zeitungen, Zeitschriften und Rundfunk (Bunkachou 1978). In diesem Dokument wird explizit den Künsten und der Wissenschaft die Möglichkeit der individuellen Handhabung eingeräumt. Auch gibt es innerhalb dieses Regelwerkes eine beträchtliche Anzahl von Fällen, in welchen eine Entscheidung über die Verwendung der *okurigana* dem eigenen Ermessen unterliegt.

Das Problembewusstsein für unterschiedliche Handhabung der *okurigana* innerhalb eines Werkes zeigt sich an der veränderten Herangehensweise in der Neuauflage des IWANAMI-KŌJIEN. Beispiele 9 und 10 illustrieren, dass im IWANAMI-KŌJIEN⁵ teilweise die Darstellung desselben Lemmas hinsichtlich des Gebrauchs von *okurigana* variiert.

²² Flexionsformen, die in *hiragana* Silbenschrift an Ideogramme angehängt werden können, werden ebenfalls als *okurigana* bezeichnet. Da diese jedoch nicht den Lesehilfen angehören, sondern bedeutungstragende Suffixe sind, deren explizite Nennung im geschriebenen Japanisch außer Frage steht, wird an dieser Stelle nicht weiter auf sie eingegangen.

- (9) 肩を 落す
kata-o otosu
 Schulter-AKK fallen lassen
entmutigt sein; enttäuscht sein; den Kopf hängen lassen
- (10) 気を 落とす
ki-o otosu
 Geist-AKK fallen lassen
verzweifelt sein, enttäuscht sein

Ab der sechsten Auflage des IWANAMI-KŌJIEN₆ wurde eine Vereinheitlichung der Orthographie einzelner Lemmata durchgeführt, jedoch wurde dies nicht in Form einer allgemeinen Richtlinie auf das gesamte Werk ausgeweitet. So wird in *otosu* das optionale *to* dargestellt, das optionale *na* in *okonau* jedoch weggelassen. Problematischer ist jedoch, dass die jeweils nicht lemmatisierte Variante der *kanji-kana*-Zeichenkette 落す in der digitalisierten Version des IWANAMI-KŌJIEN₆-CW nicht in den Index aufgenommen wurde. Unter den untersuchten Wörterbüchern modelliert nur das SANSEIDŌ-DAIJIRIN-IW alle *okurigana*-Varianten einer Zeichenkette explizit, jedoch nur für Einwortlexeme.

2.2.2 Stabilität und Variabilität

Phraseme zeichnen sich im Gegensatz zu freien Wortverbindungen durch eingeschränkte Modifizierbarkeit aus. Diese »Starrheit«, »Fixiertheit« oder »Festigkeit« ist im Falle von Restriktionen, die zu strukturellen Irregularitäten führen, oder bei *kanyōku* mit einem hohen Idiomatizitätsgrad besonders ausgeprägt.

Die unveränderlichen Phraseme markieren jedoch nur den extremen Pol eines Kontinuums, das sich bis hin zur freien Phrasenbildung erstreckt. Dazwischen gibt es eine Vielzahl von Abstufungen in der Ausprägtheit des Merkmals Festigkeit. In der Phraseographie werden diese meist nur erfasst, wenn Restriktionen die Lemmatisierung eines *kanyōku* tangieren, beispielsweise weil nur eine bestimmte Form des Phrasems gebräuchlich ist, die nicht mit der üblichen Grundform übereinstimmt. Ein noch häufigeres Phänomen ist die Variabilität in der produktiven Ausprägung, die beispielsweise erlaubt, eine lexikalische Komponente durch eine bedeutungsähnliche auszutauschen, ohne dass die idiomatische Bedeutung beeinträchtigt wird.

Nicht zum Gegenstandsbereich der vorliegenden Arbeit gehören okkasionelle Abwandlungen, die spontane Verstöße gegen die Gebrauchsregeln darstellen (z. B. um die wörtliche und übertragene Bedeutung gleichzeitig zu aktivieren) sowie unbeabsichtigte Varianten (sogenannte »Fehler«).

2.2.2.1 Lexikalische und morphologische Restriktionen

Lexikalische Restriktionen liegen dann vor, wenn sich die Komponenten eines Ausdrucks nicht beliebig durch bedeutungsähnliche oder synonyme Ausdrücke ersetzen lassen. Burger (2002, S. 395) erklärt die Unterschiede zur freien Phrasenbildung wie folgt:

D. h. während in einer nicht-phraseologischen Wortverbindung die syntagmatische Verknüpfbarkeit von Autosemantika nicht auf einzelne Wörter beschränkt ist, sondern sich auf Paradigmen (semantische Felder o. ä.) von Wörtern bezieht, handelt es sich bei den Idiomen um Paradigmen mit nur einem Element (als Grenzfall) oder Paradigmen mit einem sehr begrenzten Inventar.

Beispiel 11 zeigt einen typischen Fall eines *kanyōku*, das durch den Austausch einer Komponente seine idiomatische Bedeutung verliert.

- (11) 油を 売る
abura-o uru
Öl-ACC verkaufen
Zeit vergeuden

Ersetzt man beispielsweise das Verb *uru* durch das bedeutungsähnliche sino-japanische *hanbai suru* (販売する), wird die wörtliche Bedeutung *mit Öl handeln* in den Vordergrund gerückt.

Von morphosyntaktischen Restriktionen spricht man, wenn unter anderem die Passivierung, die Negierbarkeit oder eine Änderung des Tempus eines Ausdrucks nicht frei gestellt ist. Innerhalb dieses Bereichs kann überdies zwischen Phrasemen differenziert werden, die in einer bestimmten Form erstarrt sind, und solchen, die gewissen Restriktionen unterworfen, innerhalb dieser Grenzen jedoch frei flektierbar sind (Otani 2005, S. 556).

In der japanischen Phraseologie werden bis zu fünfzehn verschiedene mögliche Transformationen unterschieden, mit Hilfe derer die Stabilität eines Ausdrucks unter Beweis gestellt werden kann (Otani 2005, S. 539). Für die Phraseographie sind insbesondere die Restriktionen interessant, die verhindern, dass verbale *kanyōku* lemmatisiert werden können, da diese potentiell Auswirkungen auf die Gestaltung der Nennform und die Indizierung haben. Die folgende

Auflistung gibt eine Übersicht über die die Grundformbildung verhindernden morphosyntaktischen Restriktionsformen in den exemplarischen Daten.

1. Negation (z. B. *kuchi-ga heranai*, wörtl.: »der Mund ist nicht leer« für *nie um eine Antwort verlegen sein*)
2. Passiv (z. B. *itakumo-nai hara-o sagurareru*, wörtl.: »der Bauch wird abgetastet, obwohl er nicht einmal weh tut« für *zu Unrecht verdächtigt werden*)
3. Kausativ (z. B. *hanagusuri-o kikaseru*, wörtl.: »Nasenmedizin schnuppern lassen« für *bestechen*)
4. Aspekt (z. B. *kao-ni kaite aru*, wörtl.: »es steht im Gesicht geschrieben« für *aus der Mimik deutlich werden*)
5. Höflichkeit (z. B. *o-me-ni kakaru*, wörtl.: »die (werten) Augen anhängen«²³ für *die Ehre haben (jemanden treffen zu dürfen)*)

Häufig sind Anomalien im Bereich Negation. Folgende Phänomene lassen sich beobachten:

1. Idiome, die ausschließlich in negierter Form auftreten und
 - a) in einer bestimmten Flexionsform erstarrt sind (z. B. *te-mo naku*, wörtl.: »auch die Hand nicht« für *einfach*)
 - b) innerhalb dieser Grenzen frei flektierbar sind (z. B. *kuchi-ga heranai*, wörtl.: »der Mund ist nicht leer« für *nie um eine Antwort verlegen sein*)
2. Idiome, die ausschließlich in affirmativer Form auftreten, u. a. weil sie
 - a) einen andauernden Zustand beschreiben (z. B. *kubi-o nagaku shite matsu*, wörtl.: »den Hals lang machen beim Warten« für *ungeduldig warten*)
 - b) etwas Negatives zum Ausdruck bringen (z. B. *atama-ni kuru*, wörtl.: »zum Kopf kommen« für *sich über etwas ärgern*)

²³ Die Höflichkeitsform *o-me-ni kakaru* ist als starre Form zu verstehen, da die nicht höflichkeitsmarkierte Form *me-ni kakaru* eine andere Bedeutung (dt. »sichtbar sein«) hat.

- c) bereits durch Aspekt, Passiv, Voluntativ o. ä. markiert sind (z. B. *kao-ni kaite aru*, wörtl.: »es steht (jmd.) im Gesicht geschrieben« für *aus der Mimik deutlich werden*)
- d) in einer bestimmten Form erstarrt sind (z. B. *kuchi-ga sakete mo*, wörtl.: »auch wenn der Mund reißt« für *verschwiegen sein*)
- e) auf einem Vergleich basieren (z. B. *karite kita neko-no yō*, wörtl.: »wie eine geliebene Katze« für *brav; gut erzogen*)

3. Idiome, die sowohl in negierter als auch in affirmativer Form auftreten, wobei

- a) die negierte Form die ursprüngliche ist und auch heute noch in den meisten Fällen verwendet wird (z. B. *hara-no mushi-ga osamaranai*, wörtl.: »der Bauchwurm lässt sich nicht beruhigen« für *sich ärgern*)
- b) historisch der affirmative Gebrauch überwiegt, während gegenwärtig die negierte Form üblich ist (z. B. *fu-ni ochiru*, wörtl.: »in die Eingeweide fallen« für *einverstanden sein*, das historisch meist affirmativ verwendet wurde, heutzutage jedoch überwiegend in negierter Form auftritt)

Prototypische Phraseme mit quasi absolutem Stabilitätsmerkmal finden sich unter den Punkten 1a und 2d. Den Gegenpol bilden die Formen unter Punkt 3, die zwar keinen morphosyntaktischen Restriktionen unterliegen, jedoch statistisch eine klare Präferenz zu einer der beiden Formen aufweisen. Zwischen diesen beiden Extremen befindet sich eine heterogene Menge von Ausdrücken mit diversen transformationellen Merkmalen, deren Besonderheit meist nur lexikographisch erfasst wird, wenn sie die Bildung der Wörterbuch-Grundform tangiert. So wird beispielsweise die Einschränkung in Bezug auf die Negierbarkeit der *kanyōku* unter den Punkten 2a und 2b in den untersuchten Wörterbüchern nicht kodifiziert, während die unter Punkt 1b angeführten Phänomene in denselben Lexika durch eine negative Lemmaform eindeutig markiert sind.

2.2.2.2 Variation innerhalb der Grenzen struktureller Festigkeit

Stabilität gehört zu den grundlegenden Eigenschaften von Phrasemen in beiden Sprachen, dennoch wurde im vorangegangenen Kapitel am Beispiel der Negation bereits deutlich, dass Restriktionen nur in seltenen Fällen zu unveränderlichen Formen führen. Der Spielraum an Variabilität innerhalb der Grenzen struktureller Festigkeit tangiert nahezu alle Phänomene, die

von Restriktionen betroffen sind. Diese Abweichungen von Standard-Mustern sind ein viel diskutiertes Thema in der phraseologischen Forschung. Für die deutsche Phraseologie wurden diverse Typen struktureller Veränderung erfasst (vergleiche Fleischer (1997), Burger (2010) und Korhonen (1992)):

- Morphosyntaktische Veränderung einzelner Komponenten
- Variation in der Reihenfolge der Komponenten
- Austausch einzelner lexikalischer Komponenten
- Erweiterung oder Reduktion des Komponentenbestandes
- Unterschiede in der externen Valenz

Typische Auswirkungen der Variation sind demnach Abweichungen in Numerus, Rektion, Artikelgebrauch, Negationsart, Lautstruktur sowie dem fakultativen Charakter gewisser Elemente.

Aufgrund des geringen Verwandtschaftsgrads erweist sich diese für die Phraseologie des Deutschen entwickelte Typisierung von Varianten als nur begrenzt auf das Japanische übertragbar, da spezifische Phänomene des Japanischen in dem Schema nicht erfasst sind. Die folgenden typischen Variationsmöglichkeiten, die sich als Ergebnis der Analyse des japanischen Materials heraus kristallisiert haben, beruhen auf einem Vergleich der Lemmatisierungsstrategien diverser Wörterbücher.

1. Morphosyntaktische Veränderung einzelner Komponenten, z. B.:

- a) Variation des Kasus des Präpositionalobjekts (z. B. *heso-de/ga cha-o wakasu*, wörtl.: »mit dem Bauchnabel Wasser kochen/der Bauchnabel kocht Wasser« für *etwas vollkommen Absurdes tun*)
- b) Variation der Transitivität und folglich der internen Valenz (z. B. *kata-no ni-ga oriru/kata-no ni-o orosu*, wörtl.: »eine Last steigt von den Schultern« versus wörtl.: »eine Last von den Schultern abladen« für *von einer Last befreit sein* bzw. *sich von einer Last befreien*)
- c) Variation der Verbform/Verbendung (z. B. *hara-no mushi-ga osamaranai/osamara-zu* oder des Typs *hō-ga ochiru yō/ochisō*, wörtl.: »die Backen scheinen herunter zu fallen« für *etwas außergewöhnlich Köstliches*)

- d) Variation der Flexionsform (z. B. Aspekt *me-ga koeru/koete iru*, wörtl.: »die Augen sind geschult« für *ein gutes Urteilsvermögen haben*)
 - e) Variation in der Polarität (z. B. *shō-ga au*, wörtl.: »die Natur (von Personen) passt« für *zueinander passen; sich verstehen* versus *shō-ga awanai* für *nicht zueinander passen; sich nicht verstehen*)
2. Variation der Reihenfolge der Komponenten (z. B. *he-no kappa*, wörtl.: »Wasserkobold des Furzes« versus *kappa-no he*, wörtl.: »Furz des Wasserkobolds« für *ganz einfach*)
3. Lexikalische Variation durch
 - a) optionale Elemente (z. B. *mushi-mo korosanu [kao-o shite iru]*, wörtl.: »[ein Gesicht machen als könne man] kein Insekt töten« für *harmlos aussehen*)
 - b) Substitution einzelner Komponenten durch (Quasi-)Synonyme (z. B. *hara-o minu-ku/yomu*, wörtl.: »die Eingeweide sind transparent/durchschaubar« für *durchschaubar sein; die Intention durchblicken lassen*)
 - c) Substitution einzelner Komponenten durch (Quasi-)Antonyme (z. B. *kuchi-ga omoi/karui* wörtl.: »der Mund ist schwer/leicht« für *wortkarg/indiskret sein*)
4. Phonetische Variation (Lesung) (z. B. *iyke/iyaki-ga sasu*, wörtl.: »Missfallen sticht« für *überdrüssig werden*)
5. Variation der Wortart durch
 - a) attributive Verwendung von Phrasemen (z. B. *hara-ga kuroi*, wörtl.: »der Bauch ist schwarz« versus *haraguroi*, wörtl.: »schwarzbäuchig« für *bösartig; gemein*)
 - b) Nominalisierung (z. B. *mushi-ga shiraseru*, wörtl.: »die Würmer teilen mit« versus *mushi-no shirase* wörtl.: »die Mitteilung der Würmer« für *eine böse Vorahnung haben*)

Das Raster struktureller Variation weist gewisse Ähnlichkeiten mit dem von Fleischer und Burger für die Klassifikation der deutschen Beispiele entwickelten auf, sprachspezifisch sind jedoch die durch das Schriftsystem bedingte Möglichkeit der Variation der Lautstruktur ohne Veränderung der morphosyntaktischen Merkmale (Punkt 4) und eine relativ große Anzahl von Paradigmen im Bereich der Transitivity von Verben (Punkt 1b). Auf der anderen Seite fehlen

Beispiele für Abweichungen in Numerus, Artikelgebrauch und in der externen Valenz²⁴, erstere, da diese für das Japanische nicht relevant sind, letztere da deren systematische Erfassung in japanischen Nachschlagewerke eine Ausnahme ist.²⁵

Otani (2005, S. 539–540) stellt fest, dass Variation nicht immer arbiträr ist, sondern sich bestimmte Phrasenstrukturen typischerweise ineinander überführen lassen (siehe Tabelle 2.2).

Phrasenstruktur 1	Phrasenstruktur 2	Typ der Variation
<i>NgaV_{negation} nai/naku</i>	<i>NgaV_{negation} nu/zu</i>	Sprachstadium
<i>NoV_{transitiv}</i>	<i>NgaV_{intrans}</i>	Transitivität
<i>NniV_{transitiv}</i>	<i>NniV_{intrans}</i>	Transitivität
<i>NdeNoV</i>	<i>NgaNoV</i>	Kasus

Tabelle 2.2: Variationsmuster und Variationstyp

Um das reale Ausmaß an Variabilität zu eruieren, bedarf es einer umfassenden Korpusanalyse, die aufgrund der derzeit noch sehr begrenzten technischen Möglichkeiten im Rahmen dieser Arbeit nicht realisierbar ist. Dass mit wesentlich mehr Variationsmöglichkeiten zu rechnen ist, zeigt das Ergebnis einer händisch ausgewerteten Korpus-Analyse für *hara-no mushi* in Tabelle 2.3.

Die Mehrzahl dieser Formen lassen sich durch Kenntnis der Lemmatisierungsregeln auf die Formen *hara-no mushi-ga osamaranai* beziehungsweise *hara-no mushi-ga osamaru* zurückführen. Schwerer zu fassen ist die Änderung der Phrasenstruktur in Form von Tilgung oder Substitution von Partikeln beziehungsweise Ergänzung von Elementen.

²⁴ Die Angabe der externen Valenz ist eine formalisierte Darstellung der Einbindung des Phrasems in den Satzkontext. Extern sind alle Elemente, die nicht zu den Komponenten des Phrasems (dem phraseologischen Kern) gehören.

²⁵ Als einziges der vorliegenden Idiomlexika der Gegenwartssprache befasst sich das YONEKAWA-NIHONGO-KANYŌKIJITEN-PW auch mit der Kodifizierung der externen Valenzen als explizite Grammatikangabe in einer standardisierten Form. Dies geschieht jedoch nicht wie im Deutschen üblich in der Nennform, sondern in einer Sektion *yōhō* (Gebrauchsweise) im Artikeltext (vergleiche Kapitel 3.2.2.5).

Korpusbeispiel	Variationstyp
腹の虫治まらぬ <i>hara-no mushi osamaranu</i>	Tilgung Partikeln, Attributivform, Sprachstadium
腹の虫は治まらない <i>hara-no mushi-ha osamaranai</i>	Hervorhebung durch Themenpartikeln
腹の虫がおさまりません <i>hara-no mushi-ga osamarimasen</i>	Höflichkeitsstufe
腹の虫が治まっていない <i>hara-no mushi-ga osamate inai</i>	Aspekt
腹の虫が治まらず <i>hara-no mushi-ga osamarazu</i>	Sprachstadium
腹の虫も多少は収まる <i>hara-no mushi-mo tashō-wa osamaru</i>	einschränkende Ergänzung, Polarität

Tabelle 2.3: Beispiele für Variationsformen von *hara-no mushi-ga osamaranai* in Korpusbeispielen

2.2.3 Idiomatizität

Idiomatizität ist das wesentliche Merkmal, durch welches sich Idiome beziehungsweise *kan-yōku* von dem weiteren Gebiet der Phraseologie unterscheiden. Man spricht auch von fehlender Kompositionalität der Gesamtbedeutung idiomatischer Phraseme, die jedoch wie andere Merkmale auch unterschiedlich stark ausgeprägt ist. Das Kapitel beleuchtet verschiedene Facetten dieses Charakteristikums, begonnen bei der Annahme kognitiver Festigkeit, die das Zentrum der Phraseologie markiert.

2.2.3.1 Kognitive Festigkeit

Mit kognitiver Festigkeit bezeichnet man die Tatsache, dass Phraseme im mentalen Lexikon wie Einzelwortlexeme als Einheit gespeichert, abgerufen und produziert werden. Opaque Idiome gelten in der Phraseographie als »Worte mit Leerzeichen«, die als kognitiv-semantische Einheit wie Vokabeln zu behandeln sind. Am deutlichsten manifestiert sich kognitive Festigkeit in Idiomen mit unikalen Elementen, die aufgrund ihrer fehlenden individuellen lexikalischen Eigenschaften nur in vordefinierten Kontexten Sinn ergeben.

- (12) 身の毛が よだつ
 minoke-ga *yodatsu*
 Körperhaare-SUBJ viele-stehen?
 (aus Furcht) Gänsehaut bekommen

Das Verb *yodatsu* (historisch auch *iyodatsu*) in Beispiel 12 ist ein unikales Element, da es ausschließlich in Zusammenhang mit *minoke-ga* in Erscheinung tritt. Eine eigenständige Bedeutung lässt sich nur durch die Interpretation der einzelnen Schriftzeichen erahnen. In den untersuchten Wörterbüchern ist keine Bedeutungsangabe für das Verb alleine registriert, sondern es wird durchweg im Kontext von *minoke-ga yodatsu* erläutert. Man kann davon ausgehen, dass dieses *kanyōku* auch kognitiv als eine Einheit gespeichert wird.

In der Phraseographie spielt kognitive Festigkeit eine Rolle, wenn es um die Frage der Zuordnung von Komponenten zu Einzelwortlexemen geht. Sind die Komponenten unikale Elemente oder in ihrer freien Bedeutung polysem, kann möglicherweise keine eindeutige Zuordnung erfolgen.

2.2.3.2 Motiviertheit

Obwohl die kognitive Festigkeit als prototypisches Kriterium für *kanyōku* gilt, ist der Großteil des Phänomenbereichs in Bezug auf die semantischen Eigenschaften ambig. Ist es möglich, auf der Basis von Hintergrundwissen die idiomatische Gesamtbedeutung zur wörtlichen Bedeutung in Beziehung zu setzen, spricht man in der Phraseologie von Motiviertheit oder Motivierbarkeit²⁶. Diese verhält sich umgekehrt proportional zu der Idiomatizität eines Ausdrucks und damit auch zu seiner kognitiven Festigkeit. Man muss jedoch bedenken, dass die Frage, wie transparent ein Ausdruck für den Nutzer ist, in Abhängigkeit des jeweiligen kulturellen Hintergrunds variiert und nicht objektiv messbar ist.

Um die Motiviertheit eines Idioms zu verdeutlichen, dient in der Regel die etymologische Angabe.²⁷ In ihr wird durch Erläuterungen der Herkunft ein Bezug zwischen der wörtlichen und übertragenen Bedeutungsebene hergestellt.

²⁶ Zum Unterschied zwischen Motiviertheit und Motivierbarkeit vergleiche Burger (2003, S. 66–67).

²⁷ Dobrovolskij und Piirainen (2010) kommen zu dem Schluss, dass die etymologische Basis nur selten eine Rolle in der Motiviertheit von Idiomen spielt.

Von »Remotivierung« spricht man, wenn beispielsweise in einem Textzusammenhang die freie Bedeutung der Wortverbindung oder die einzelner Komponenten aktiviert wird. Im Zusammenhang mit der Phraseographie ist der Begriff negativ besetzt, denn der Versuch, die idiomatische Bedeutung anhand der wörtlichen Bedeutung zu erklären, läuft Gefahr, Verwirrung zu stiften (vergleiche Kapitel 3.1).

2.2.3.3 Metaphorische lexikalische Bedeutung der Komponenten

Eine mögliche Brücke zwischen wörtlicher und übertragener Bedeutung kann die metaphorische Bedeutung einzelner Komponenten sein. Häufig sind nicht alle Teile des Phrasems in gleicher Weise ihrer wörtlichen Bedeutung entrückt. Diese teil-idiomatischen Phraseme (oder Teil-Idiome) beinhalten meist eine Metapher, die den nicht-wörtlichen Bestandteil des Ausdrucks darstellt.

Momiyama (1997) plädiert dafür, Phraseme, in welchen Metaphern isoliert werden können, nicht zu den *kanyōku* zu rechnen. Basierend auf der metaphorischen Bedeutung von *me* (Auge) im Sinne von *Urteilsvermögen* und *te* (Hand) im Sinne von *Hilfe* spricht er den Beispielen in 13 und 14 die Idiomatizität ab.

(13) 目が^が 高い
me-ga takai
 Auge-NOM hoch
ein scharfes Auge für etwas haben; anspruchsvoll sein

(14) 手が^が 足りない
te-ga tarinai
 Hand-NOM nicht ausreichen
nicht ausreichend Hilfe/Personal haben

Momiyama (1997) argumentiert, dass man die übertragene Bedeutung der Gesamtphrase erschließen kann, wenn man sich der metaphorischen Bedeutung einer der Komponenten bewusst ist. Die Abgrenzung ist schwierig, denn die Kriterien aufgrund derer er die Beispiele 15 und 16 wiederum zu den *kanyōku* zählt, obwohl sie sich bezüglich der Metaphorik eines ähnlichen Mechanismus bedienen, sind nicht transparent.

(15) 目が^が 肥える
me-ga koeru
 Auge-NOM fruchtbar

einen guten Blick für etwas haben

- (16) 目がめ 利きく
me-ga kiku
 Auge-NOM wirken
etwas kritisch beurteilen können; einen Expertenblick haben

Im Lexikon wäre die Folge von Momiyamas Postulat, dass für einen Eintrag als *kanyōku* zunächst geklärt werden muss, ob sich dessen Bedeutung nicht aus der Summe der Komponentenbedeutungen unter Einbindung metaphorischer Bedeutungsangaben erschließen lässt. In diesen Fällen würde ein Ausdruck wie in Beispiel 13 möglicherweise als Kollokation mit Bezug zur metaphorischen Lesart modelliert werden. Auf diese Weise ließe sich der Phänomenbereich eingrenzen, jedoch ist diese Herangehensweise immer abhängig von der Realisierung der jeweiligen Komponentenbedeutungen im Lexikon. Auch fühlt man sich hier an das Henne-Ei-Problem erinnert, denn die Frage nach dem Ursprung der Metapher wird in den meisten Fällen nicht ohne weiteres zu klären sein.

2.2.3.4 Kognitive Metaphern

Wenn eine bestimmte Domäne in Form einer anderen verstanden wird, spricht man in der kognitiven Linguistik auch von kognitiven Metaphern. Diese sind in der Linguistik seit geraumer Zeit stärker ins Interesse gerückt, insbesondere wenn es um die Frage geht, inwieweit unser Denken von »konzeptuellen Metaphern« quasi vorgeformt wird. Lakoff und Johnson weiten diese konzeptuellen Metaphern von der Sprache ausgehend auf den Bereich der Kognition aus, indem sie diese mit Erfahrungen, Denken und Weltverständnis der Menschen in Verbindung bringen (Lakoff und Johnson 1980). In strukturierten mentalen Domänen greifen wir auf Basis-Metaphern zurück, von welchen wir abstrahieren und so unser Wissen über das Vertraute einsetzen, um Abstraktes — wie beispielsweise Emotionen — (sprachlich) in den Griff zu bekommen (Kövecses 1986).

Dabei werden Metaphern als universal angesehen, wenn sie auf grundlegenden Erfahrungen der menschlichen Existenz beruhen. Auch wenn in der sprachlichen Realisierung Unterschiede auftreten, wird davon ausgegangen, dass die metaphorische Basis in unterschiedlichen Sprachen konstant ist. Im Gegensatz hierzu sind Metaphern, die auf kulturellen Faktoren oder Konventionen beruhen, häufig sprachspezifisch und ihre Bedeutung für Außenstehende nicht

transparent. Das Erkennen und Verdeutlichen von konzeptuellen Metaphern ist deshalb in der Phraseographie insbesondere für didaktische Zwecke von Interesse.

Empirische Versuche liefern Evidenz dafür, dass konzeptuelle Metaphern als kognitiv-linguistischer Baustein unseres Denkens fungieren. Durch Lerner-Experimente wird gezeigt, dass Konvention ein entscheidender Faktor in der Abbildung metaphorischer Sprache auf die ihr zu Grunde liegende Bedeutung ist. Diese Verknüpfungen werden beim Erlernen von Idiomen auf Basis der vorliegenden Informationen erstellt und bei Verwendung aktiviert. Ein Vergleich der durchschnittlichen Zeitspanne, welche von L2-Lernern zum Verständnis von Standard-Phrasen und Phrasenneubildungen benötigt wurde, zeigt, dass der Rückgriff auf kreative Neubildungen vergleichsweise ineffektiv ist (Keysar, Shen und Horton 2000). Hieraus folgern die Autoren der Studie, dass im Falle einer bekannten Metapher kein Rückgriff auf das Basis-konzept erfolgt, während bei unbekanntem Formulierungen — auch wenn sie der gleichen Konzept-Domäne angehören — ein Ableitungsprozess das Verständnis verlangsamt.

Andere Untersuchungen zeigen außerdem, dass man die Integration eines unbekanntem Ausdrucks in das mentale Lexikon des Lernenden nahezu beliebig steuern kann.²⁸

Für die bilinguale Phraseographie kann man dies eventuell so deuten, dass man nicht zwingend Basiskonzepte benötigt, um durch konzeptuelle Metaphern das Erlernen von Idiomen zu erleichtern. Geht man beispielsweise davon aus, dass *me* (目 Auge) im Japanischen eine kognitive Metapher für *Urteilsvermögen* ist, könnte man dadurch unter anderem die Motivierung der Beispiele 13, 15, 16 und 17 erklären.

2.2.3.5 Polysemie

Das Thema Polysemie spielt in der phraseologischen Forschung bislang eine untergeordnete Rolle. Statt dessen stehen meist Fragen der Lesarten-Disambiguierung im Vordergrund, wenn sowohl idiomatische als auch kompositionelle Lesart möglich sind.

In Zusammenhang mit Motiviertheit der Bedeutung lassen sich folgende Formen der Polysemie unterscheiden:

²⁸ In einem Versuch mit unbekanntem Idiomen wurden Muttersprachler mit gegensätzlichen Interpretationen desselben Testkorpus konfrontiert. Im Zuge des Erlernens dieser Ausdrücke wurde nicht nur erfolgreich eine »rationale« Verbindung zwischen wörtlicher und übertragener Bedeutung hergestellt, im Nachhinein wurde die neu erlernte Bedeutung häufig als »transparent« eingestuft (Keysar und Bly 1995).

- voll-idiomatische *kanyōku* mit mindestens einer phraseologischen Lesart (z. B. *atama-ni kuru*, wörtl.: »zum Kopf kommen« für 1) *sich ärgern* und 2) *keinen klaren Kopf haben* (z. B. aufgrund einer Erkältung))
- teil-idiomatische *kanyōku* mit einer phraseologischen Lesart
 - welche in Teilen motiviert oder motivierbar sind (z. B. *me-ga takai*, wörtl.: »die Augen sind hoch« für *Expertise besitzen*)
 - bei welchen keine synchron erkennbare Beziehung zwischen der idiomatischen Lesart und den Komponentenbedeutungen besteht (z. B. *hanage-o nobasu*, wörtl.: »die Nasenhaare wachsen lassen« für *eine Schwäche für Frauen haben*)
- Idiome mit disjunktiver idiomatischer und kompositioneller Bedeutung
 - bei welchen eine synchron erkennbare Beziehung zwischen den Lesarten besteht (z. B. *hana-ga takai*, wörtl.: »eine hohe Nase haben« für *arrogant sein*)
 - bei welchen keine synchron erkennbare Beziehung zwischen den Lesarten besteht (z. B. *abura-o uru*, wörtl.: »Öl verkaufen« für *Zeit verschwenden*)
- Idiome, bei welchen beide Lesarten simultan realisiert werden (z. B. *kubi-o furu*, wörtl.: »den Hals/Kopf schütteln« für *etwas verneinen*)

Bei schwach idiomatischen Phrasemen können beide Lesarten gleichzeitig aktiv sein, wie man insbesondere anhand der Verbalisierung von kodifiziertem Verhalten (Kinogrammen) beobachten kann. Hierbei stellt sich ähnlich wie bei der Motivierung durch die metaphorische Bedeutung einer Komponente die Frage, ob man noch von Idiomatizität des Phrasems sprechen kann, oder ob die *übertragene* Bedeutung nicht vielmehr in der nonverbalen Kommunikationssituation angelegt ist. Dies gilt zumindest für die *echten* Kinogramme, welchen eine reale Aktion zu Grunde liegt.

Polysemie tritt nicht nur in Verbindung mit unterschiedlich deutlich ausgeprägter Idiomatizität auf, vereinzelt gibt es auch *kanyōku* mit mehreren disjunktiven idiomatischen Bedeutungen. Hümmer und Stathi (2006) gehen davon aus, dass das Phänomen noch nicht ausreichend linguistisch und lexikographisch erfasst ist, denn ihrer korpuslinguistischen Untersuchung zu Folge verhalten sich Mehrwortlexeme in Bezug auf Polysemie nicht wesentlich anders als Einzelwortlexeme.

2.2.3.6 Idiomatiche Äquivalenz

In der kontrastiven Phraseologie unterscheidet man drei Stufen von Äquivalenz:

1. Vollständige Äquivalenz
2. Partielle Äquivalenz
3. Fehlende Äquivalenz (phraseologische Lücke)

Im Wesentlichen werden dabei folgende Aspekte der Phraseme berücksichtigt:

1. Syntax
2. Lexik
3. Bildhaftigkeit/Domäne
4. Pragmatik/Konnotation/Perspektive

Die Aspekte greifen teilweise in einander, und sind nicht in allen Fällen gleichermaßen relevant. Die deutlichste Abgrenzung lässt sich meist zwischen lexikalischer Besetzung und Bildhaftigkeit vollziehen.

Vollständige Äquivalenz erfordert die Übereinstimmung aller Aspekte. Man kann diese (theoretisch) nachweisen, indem man zeigt, dass die Austauschbarkeit der Äquivalente in beliebigen Kontexten gegeben ist, ohne dass dies zu Veränderungen in der Bedeutung führt. Als Beispiel vollständiger Äquivalenz führt Itoh (2005, S. 48) Beispiel *me-o tsuburu* (17) an.

- (17) 目を つぶる
me-o tsuburu
Auge-Akk zudrücken
ein Auge zudrücken; etwas nachsichtig/wohlwollend übersehen

Itoh erklärt den Zusammenhang zwischen wörtlicher und übertragener Ebene, indem er »Auge« eine metaphorische Bedeutung im Sinne von »Kontrolle« zubilligt, woraus er im Zusammenhang mit »zudrücken« als den absichtlichen Verlust der Kontrolle die Gesamtbedeutung *etwas (wohlwollend) übersehen* ableitet. Da diese Art von Metapher in beiden Sprachen existiert, kann man von *interlingualer Äquivalenz* sprechen²⁹:

1. Syntaktische Struktur: AkkObjekt+Verb
2. Lexikalische Besetzung: *me*=Auge, *o*=AkkPart, *tsuburu*=zudrücken

²⁹ Ehegötz (1990, S. 220) nennt dieses Phänomen »phraseologische Internationalismen«.

3. Bildhaftigkeit: übertragen aus dem »Absichtlichen Verlust der Kontrolle«

4. Drückt ein wohlwollendes Übersehen von Hinderungsgründen aus: Erlauben, Zulassen

Betrachtet man das Beispiel genauer, wird deutlich, dass dennoch feine Unterschiede zwischen den Äquivalenz-Kandidaten auszumachen sind. Zum einen ist eine Graduierung des Ausdrucks *ein Auge zudrücken* im Deutschen durch *beide Augen zudrücken* möglich, die im Japanischen keine Entsprechung hat. Auch muss in diesem Fall berücksichtigt werden, dass sich die Phraseme hinsichtlich ihrer kompositionellen Bedeutung unterscheiden: *me-o tsuburu* kann im Japanischen auch verwendet werden, wenn man die Augen schließt, weil die Lichtverhältnisse es erfordern. Das Äquivalent der kompositionellen Lesart wäre folglich *die Augen zukneifen* oder *blinzeln*. *Ein/beide Auge/n zudrücken* kann zwar auch in seiner wörtlichen Bedeutung verstanden werden, es ist aber sehr ungebräuchlich, da es eine Fremdeinwirkung suggeriert (z. B. die Augen eines Verstorbenen würde man zudrücken, aber für die eigenen Augen würde man das Verb *schließen* verwenden). Folglich ist es besser, in dem genannten Beispiel die idiomatische Bedeutung isoliert von der nicht-idiomatischen zu betrachten und von *funktionaler Äquivalenz* auf der idiomatischen Ebene zu sprechen.

Im Falle von partieller Äquivalenz sind verschiedene Konstellationen denkbar. Für die Phraseographie steht in der Regel die Bildhaftigkeit und die pragmatisch-funktionale Ebene im Vordergrund. Diese Ebenen müssen jeweils auf Vergleichbarkeit geprüft werden, um eine Aussage über den Äquivalenzgrad treffen zu können. Ein Beispiel hierfür ist *kubi-o tsukkomu*:

- (18) 首を 突っ込む
kubi-o *tsukkomu*
 Hals/Kopf-Акк hineinstecken
sich einmischen

Beispiel 18 weist im Bereich der Bedeutungsbeschreibung »sich einmischen (in etwas, das einen nichts angeht)« Ähnlichkeiten mit dem Idiom *seine Nase in etwas hinein stecken* auf. Die syntaktische Struktur ist vergleichbar und die Bildhaftigkeit stammt aus derselben Domäne. Auf lexikalischer Ebene sind Unterschiede erkennbar (*Hals/Kopf* versus *Nase*), jedoch bedienen sich beide Phraseme semantisch-topologisch ähnlicher Begrifflichkeit. Der Unterschied, der dazu führt, dass *seine Nase in etwas hinein stecken* sich nicht als Äquivalent für *kubi-o tsukkomu* qualifiziert, manifestiert sich auf der pragmatischen Ebene. Aus dem Vergleich der Anwendungskontexte wird deutlich, dass das japanische *kubi-o tsukkomu* eine Einmischung mit

dem Ziel der Veränderung der Vorgänge bezeichnet, während *seine Nase in etwas hinein stecken* eine durch Neugier motivierte Einmischung ohne direkte Folgen ist.

Einschränkungen auf eine bestimmte Domäne können auch Unterschiede in der Bedeutungsweite zur Folge haben. So kann in bestimmten Kontexten *kubi-ni naru* (vergleiche Beispiel 19) mit *es geht jmd./etw. an den Kragen* übersetzt werden. Meist bezieht sich das japanische Idiom jedoch auf eine berufliche oder andere Position, die für die betroffene Person auf dem Spiel steht.

- (19) 首に なる
 kubi-ni *naru*
 Hals/Kopf-DAT werden
 *das Leben verlieren; den Job verlieren; ausgeschlossen werden (aus einer Partei/
 Gesellschaft etc.)*

Im Vergleich hierzu bedeutet *es geht jmd./etw. an den Kragen* eine Bedrohung der Existenz. Der Ursprung der beiden Idiome liegt in der Erfahrung, dass der Hals/Kragen ein neuralgischer Punkt der menschlichen Existenz ist. Und dennoch fokussieren die Idiome unterschiedliche Bereiche des menschlichen Lebens. Obwohl man hier von einer ähnlichen Domäne oder kognitiven Metapher ausgehen kann, unterscheiden sie sich in Bezug auf ihre pragmatischen Merkmale deutlich. Wie sich aus der Valenz-Angabe *es geht jmd./etw. an den Kragen* heraus lesen lässt, kann die Rolle des Gefährdeten auch von einer Sache oder etwas Abstraktem eingenommen werden, während der japanische Ausdruck *kubi-ni naru* aufgrund seiner engeren Bedeutung im Wesentlichen auf Menschen (oder gegebenenfalls andere Lebewesen) beschränkt ist.

Wenn man vollständige Äquivalenz als eine Übereinstimmung von allen wesentlichen Merkmalen der Idiome definiert, würde man möglicherweise zu dem Schluss kommen, dass es diese zwischen distanten Sprachen nicht gibt.³⁰ Häufig ist die Frage auch nur innerhalb eines konkreten Kontextes zu klären. Beispiel 20 ist ein Fall, in welchem es sehr schwer ist, unabhängig

³⁰ Gehweiler (2006, S. 20) hält es für noch nicht bewiesen, dass idiomatische Äquivalenz auf bilingualer Ebene existiert und auch im Beispielparadigma der vorliegenden Arbeit konnte kein derartiger Nachweis erbracht werden. Selbst das zuvor genannte *me-o tsuburu* ist auf idiomatischer Ebene nicht vollständig äquivalent zu dem deutschen *ein Auge zudrücken*, da die »wohlwollende« Motivation für das absichtliche Übersehen nur im Deutschen eine zwingende Voraussetzung für die Verwendung des Idioms ist (vergleiche Kapitel 4.5.2).

von einer bestimmten Situation die Frage zu klären, ob ein Äquivalent für die idiomatische Bedeutung existiert.

- (20) 首を ひねる
kubi-o hineru
Hals-ACC drehen
zweifeln; den Kopf schief legen

Da die idiomatische Bedeutung auf eine Kopfbewegung zurückzuführen ist, die im westeuropäischen Raum nicht zu den konventionellen Gesten gehört, ist die Übertragung der wörtlichen Ebene nur deskriptiv möglich.

2.2.4 Zusammenfassung

Das Kapitel gibt einen Überblick über die für die Phraseographie relevanten Besonderheiten der japanischen Phraseologie. Die Themen Polylexikalität und Variabilität sind aufgrund der Komplexität des japanischen Schriftsystems eng mit einander verwoben. Insbesondere orthographische Variation ist in der Regel mit anderen Variationsformen kombinierbar. Auch zwischen Variabilität und Idiomatizität gibt es Überschneidungen, denn eine entscheidende Frage bei der Kodierung der Phraseme im Wörterbuch wird sein, wie sich die jeweilige Variationsform auf die übertragene Bedeutung auswirkt. Die Herausforderung bei der Implementierung des Wörterbuchs besteht folglich nicht in der Repräsentation einzelner Merkmale, sondern in der dieses Merkmalsgefüges.

3 Phraseologie in Theorie und Praxis

3.1 Anforderungen an die Phraseographie aus linguistischer Perspektive

In der phraseologischen Forschung wird der Status Quo der lexikographischen Erfassung von Phrasemen in der Regel als unbefriedigend bewertet. Stantcheva (2002) untersucht den Stellenwert der Phraseologie in einsprachigen allgemeinen Wörterbüchern des Deutschen seit Ende des 18. Jahrhunderts und schließt ihre Untersuchung mit der Feststellung, dass sich der theoretische Forschungsfortschritt im Bereich der Phraseographie nicht in einer nennenswerten Verbesserung der lexikographischen Praxis widerspiegelt (Stantcheva 2002, S. 175). Pilz (1987) vergleicht allgemeine Wörterbücher und phraseologische Wörterbücher für das Deutsche hinsichtlich ihres Umgangs mit der Phraseologie und stellt fest, dass die allgemeinen Wörterbücher den Bestand besser abdecken als Spezialwörterbücher. Wie Schemann (1991, S. 1022) ist er der Meinung, aus quantitativen Gesichtspunkten könne man auf phraseologische Spezialwörterbücher ganz verzichten (Pilz 1987, S. 135–136), allerdings folgert Schemann daraus: »Eine der vordringlichsten Aufgaben der Lexikographie ist demnach, möglichst vollständige phraseologische (einsprachige und zweisprachige) Wörterbücher zu erstellen«.¹

Auch wenn die von der Phraseologie angeregten Verbesserungen beginnen in die lexikographische Praxis einzufließen, herrscht zwischen dem Stand der Phraseologie-Forschung und der praktischen Phraseographie eine deutliche Diskrepanz (Burger 2010, S. 179–197). Als mögliche Ursachen für das Auseinanderklaffen von phraseologischer Theorie und Praxis wird unter anderen die »schier unübersehbare Fülle« phraseologischer Literatur der vergangenen 20 Jahre

¹ Schemann widmet sich aktiv dieser Aufgabe durch die Herausgabe zweier Werke: dem einsprachigen SCHEMANN-DEUTSCHEIDIOMATIK-PW und dem bilingualen SCHEMANN-IDIOMS-PW.

genannt (Pilz 2002, S. 299–301). Scholze-Stubenrecht² (Scholze-Stubenrecht 1988, S. 298–299) spricht von einem distanzierten Verhältnis zwischen Theorie und Praxis der Phraseographie, welches es in eine »fruchtbare Wechselwirkung« zu verwandeln gälte. Als Gründe für die verzögerte Reaktion der lexikographischen Praxis auf die Vorschläge aus der linguistischen Forschung nennt er vor allem die Vernachlässigung der Vermittlung der Forschungsergebnisse und das fehlende Praxisbewusstsein der einschlägigen Forschung.

Aufgrund der Fülle einschlägiger Literatur konzentriert sich die folgende Zusammenstellung auf die Kernpunkte im linguistischen Diskurs um die Optimierung phraseologischer Nachschlagewerke. Da diesem Thema im japanischen Raum bislang kaum Bedeutung beigemessen wird, werden die Desiderata zunächst in ihrem sprachlichen Kontext belassen.

3.1.1 Terminologie in Titel und Umtexten

Wie bereits in Kapitel 2.1 erläutert, konnte in Bezug auf die Terminologie in der phraseologischen Forschung bestenfalls der terminologische Wildwuchs eingedämmt werden. Von einer Einigung auf eine einheitliche Terminologie kann nicht die Rede sein. Der terminologische Gebrauch in der Phraseographie kann von Seiten der Phraseologie bestenfalls auf seine innere Konsistenz und Übereinstimmung mit aktuellen Forschungsergebnissen geprüft werden. So kritisiert Burger die terminologische Willkür in der Phraseographie und verteidigt sie zugleich mit dem Argument, dass *bis vor kurzem* in der phraseologischen Terminologie der westlichen Forschung Chaos geherrscht habe (Burger 1983, S. 15).³ Die begrifflichen Fragen rund um Beteiligungen und Metatexte können somit aus linguistischer Perspektive nur bedingt kritisiert werden

² Werner Scholze-Stubenrecht ist Chefredakteur in der Dudenredaktion des Bibliographischen Instituts Mannheim und damit maßgeblich beteiligt an der Herausgabe einschlägiger Nachschlagewerke.

³ »Die verwendeten Terminologien sind voller Willkür und häufig ohne Zusammenhang mit dem in der Linguistik üblichen Vokabular. Die terminologische Insuffizienz wird man den Wörterbuchmachern allerdings kaum zum Vorwurf machen können, da im Bereich der Phraseologie *bis vor kurzer Zeit* [vom Verf. kursiv gedr.] in der westlichen Forschung ein Chaos herrschte (vgl. die terminologischen Zusammenstellungen bei (Pilz 1978))« (Burger 1983, S. 15)

und dementsprechend halten sich auch Kritiker meist mit konkreten Vorschlägen zum Terminologiegebrauch in der Phraseographie zurück.⁴

In der Tat zeichnen sich in den jüngeren Werken bereits positive Veränderungen im Bereich des Terminologiegebrauchs ab (Burger 2010, S. 178), dennoch wird auf unklare Ausdrucksweisen allgemein, mangelnde Konsistenz in den Umtexten, zwischen Umtexten und Artikelteil (Wörterbuchverzeichnis) (Cowie 1999, S. 205–206), sowie das Fehlen einer klaren Trennung zwischen Phrasemen und Satzbeispielen im Besonderen hingewiesen. Obwohl sich die meisten Kritiker mit konkreten Vorschlägen zur Differenzierung und Nomenklatur von Phrasentypen zurückhalten, herrscht Einigkeit, dass die Wichtigsten in den Umtexten grundlegend charakterisiert und nach Möglichkeit anhand von Beispielen gegeneinander abgegrenzt werden sollten (Stantcheva 2002, S. 38; Burger 1983, S. 14). Um den Bedürfnissen unterschiedlicher Benutzerkreise (Laien, Fachpublikum) gerecht zu werden, betont Stantcheva (2002, S. 176) die Notwendigkeit einer Zweiteilung und Mehrfachadressierung der Zielgruppen in den Umtexten.

Relevant sind terminologische Fragen auch für die Betitelung, insbesondere im Bereich Spezialwörterbücher. Bei kommerziellen Werken wird meist marktwirtschaftlichen Überlegungen größeres Gewicht beigemessen als den wissenschaftlichen Usancen (Burger u. a. 2007, S. 5; Almind 2005b). So stellt Schemann (1991, S. 2019) fest, dass in der Wörterbuchlandschaft im deutschen und europäischen Kontext der Terminus »Phraseologie« oder »phraseologisch« vergleichsweise selten Verwendung findet. Statt dessen überwiegen Begriffe wie *Idiom*, *Ausdruck* oder *Redewendung*, häufig auch in Kombinationen (z. B. *sprichwörtliche Redewendung*), die sich nur bedingt auf die in der Sprachwissenschaft gängigen Termini abbilden lassen.⁵ In jedem Fall sollten die Begriffe im Titel in Vorwort oder den Benutzungshinweisen eingehend erklärt und zu der dem Inhalt des Werkes und der darin verwendeten terminologischen Systematik in Beziehung gesetzt werden. (Dass dies nicht immer der Fall ist, wurde bereits in Kapitel 2.1.2 an einem Beispiel illustriert).

⁴ Eine Ausnahme macht Dobrovol'skij, der anregt, die vier elementaren Phrasentypen *Idiome*, *Kollokationen*, *grammatische Phraseme* und *Sprichwörter* zu unterscheiden (Dobrovol'skij 2009, S. 151–152).

⁵ Zwar unterbreitet er keine konkreten Vorschläge für eine bessere Betitelung, das zwei Jahre später von ihm publizierte Werk trägt jedoch den Titel *Deutsche Idiomatik: die deutschen Redewendungen im Kontext* (SCHEMANN-DEUTSCHEIDIOMATIK-PW: Schemann, Hans 2011).

3.1.2 Materialauswahl

Das Thema Materialauswahl umfasst mehrere Facetten. Zunächst stellt sich die Frage, wie sich der Phänomenbereich Phraseologie in die Wörterbuchlandschaft eingliedern lässt. Als Elemente natürlicher Sprache sind alle Phrasentypen unumstrittener Bestandteil allgemeiner Wörterbücher. Uneinigkeit herrscht in der Frage, ob spezialisierte Wörterbücher den gesamten Objektbereich oder nur einen Teil davon abdecken sollten. Außerdem steht zur Debatte, wie eng oder weit der Phrasem-Begriff gefasst werden soll und nach welchen Kriterien die inhaltliche Auswahl erfolgen soll (möglichst vollständig, nach Frequenz, nur aktiver oder passiver Gebrauch etc.). Auch vor diesem Hintergrund wäre eine terminologische Vereinheitlichung hilfreich.

Da allgemeingültige Kriterien zur Abgrenzung der Teilbereiche fehlen, bemerkt Schemann (1991, S. 1021), dass die einzig sinnvolle Frage in diesem Zusammenhang die nach dem Kernbereich des Lexikons ist. Für ihn stehen die Merkmale Polylexikalität und Idiomatizität im Vordergrund, jedoch räumt er die Möglichkeit ein, dass in Abhängigkeit von der Zielsetzung eines Werkes variabel ist, wo letztendlich die Grenzziehung zwischen Peripherie der Phraseologie und freier Phrasenbildung erfolgt. Auch Scholze-Stubenrecht (1988, S. 286–287) plädiert für eine Fokussierung auf den Kernbereich der Phraseologie. Er unterbreitet den Vorschlag, sich bezüglich der Abgrenzung zur Peripherie an Fleischers Unterscheidung zwischen Phraseolexem und Phrasem zu orientieren. Konkret bedeutet dies, dass zum Kernbereich der Phraseologie Wortverbindungen gehören, die mindestens ein Autosemantikon enthalten und alle drei Hauptmerkmale (Idiomatizität, Stabilität und Lexikalisierung) aufweisen (\equiv Phraseolexeme⁶). Als einziger spricht sich Dobrovolskij (2009) für die Aufteilung der Phrasentypen in unterschiedliche Werke auf. Die Beschränkung auf nur jeweils einen Phrasentyp⁷ hält er für vorteilhaft, denn »Der Benutzer hat das Recht zu wissen, was er im betreffenden Wörterbuch finden kann.« (Dobrovolskij 2009, S. 151–152). Er setzt beim Benutzer ein grundlegendes Verständnis der Phrasem-Typologie voraus und die Bereitschaft, das Nachschlagewerk in Abhängigkeit der Art der Wortverbindung zu wählen.

⁶ Dieses Verständnis des Begriffs ist nicht zu verwechseln mit der weit gefassten Definition von *Phraseolexem* bei Pilz (1978), der diesen Begriff geprägt hat.

⁷ Gemeint sind hier die oben erwähnten vier Typen: *Idiome, Kollokationen, grammatischen Phraseme* und *Sprichwörter*.

Von einem phraseologischen Wörterbuch, das für allgemeine Benutzer konzipiert ist, erwartet Pilz (1987, S. 137) eine möglichst vollständige Erfassung des »phraseologischen Sprachschatzanteils«, die jedoch Fachtermini und metaphorische Einzelwortlexeme (auch Komposita) ausschließt. Die Erfassung regionaler und dialektaler Varianten hält er hingegen für sinnvoll (Pilz 1987, S. 140). Einsprachige Lernerwörterbücher der Phraseologie sollten sich nach Dobrovol'skij (2002) auf Phraseme konzentrieren, die für den aktiven Gebrauch eines Nichtmuttersprachlers relevant sind. Im bilingualen Wörterbuch können zusätzlich Vergleichskriterien, die auf dem Verhältnis zwischen Phrasemen der Ausgangs- und der Zielsprache basieren, bei der Materialauswahl eine Rolle spielen. So inkludiert Kim-Werner (1996, S. 75–77) nicht nur »intersprachlich idiomatische« Phraseme (für welche in der Zielsprache ein idiomatisches Äquivalent vorhanden ist), sondern auch »weniger idiomatische« Phraseme. Als dritte Gruppe, die in jedem Fall im bilingualen Wörterbuch enthalten sein sollte, nennt sie die Phraseme, die auf der Ebene der Komponentenbedeutungen in der jeweiligen Sprache Ähnlichkeiten aufweisen, deren phraseologische Bedeutung jedoch divergiert (»falsche Freunde«).

3.1.3 Gestaltung der Nennform

Die Diskussion um die Gestaltung der Nennform (auch Grundform oder Wörterbuchform) spielt eine zentrale Rolle in der Phraseographie. Die Nennform trägt wesentlich zur Identifikation des Phrasems bei und ist gegebenenfalls Grundlage für das System, nach der die strukturelle Einordnung vollzogen wird. In ihr können bereits wesentliche Merkmale des Phrasems (morphosyntaktische Restriktionen, externe Valenz) kodiert sein. Der Prozess, durch welchen die kanonische Form eines Ausdrucks ermittelt wird, wird gemeinhin auch als *Lemmatisierung* bezeichnet. Wiegand (2010, S. 164) definiert ihn folgendermaßen:

Lemmatisierung liegt vor genau dann, wenn man bei flektierenden Sprachen von allen belegbaren paradigmazugehörigen Flexionsformen und/oder von allen belegbaren Formvarianten eines sprachlichen Ausdruckes zu einer kanonischen Form übergeht mit dem Zweck, dass diese Nennform als Zugriffstextelement einer äußeren Zugriffsstruktur sicheren externen Zugriff auf dieses Zugriffstextelement und durch Ausführung sich anschließender Benutzungshandlungen auf das gesamte Formparadigma und/oder die Menge der Formvarianten eröffnet.

Diskutiert wird im Zusammenhang der Gestaltung der Nennform von Phrasemen überwiegend die Wahl der morphosyntaktischen Form, die Darstellung der externen Valenz und der

Restriktionen sowie die Möglichkeit, Varianten in die Nennform zu integrieren. Diese Informationen können theoretisch auch in den Artikeltext integriert werden, die westeuropäische Tradition verankert sie jedoch überwiegend in der Nennform. Varianten, die nur bedingt in der Nennform kodifiziert werden können, sind Thema eines separaten Unterkapitels (vgl. Kapitel 3.1.5).

3.1.3.1 Wahl der morphosyntaktischen Form

Welche morphosyntaktische Form als Grundform eines Lexems betrachtet wird ist sprachspezifisch, in der Regel gilt jedoch, dass die Form so allgemein wie möglich und dennoch allgemein verständlich sein soll (Stantcheva 2002, S. 47). Die Vorgehensweise ist ähnlich der Lemmatisierungsmethode bei Einzelwortlexemen, beispielsweise wird für verbale Phraseme im Deutschen der Infinitiv Präsens gewählt und dem entsprechend im Japanischen die Finalform (終止形 *shūshikei*). Die morphosyntaktische Form von Phrasem-Lemmata sollte sich von Satzbeispielen grundsätzlich unterscheiden, da bei der Lemmatisierung versucht wird, die Phrase auf ihren phraseologischen Nukleus zu reduzieren. Gelingt dies, sind nur die Komponenten, die zum unveränderlichen Kern des Phrasems gehören, Teil der Nennform. Kontextelemente werden weggelassen oder gegebenenfalls durch Platzhalter ersetzt (vgl. Kapitel 3.1.3.2).

3.1.3.2 Darstellung der externen Valenz

Die Darstellung der externen Valenz erfolgt in der Regel in sogenannten Strukturformeln, die sowohl in die Nennform als auch in den Artikeltext integriert werden können. Hierbei werden obligate Elemente, auch wenn diese nicht zum unveränderlichen Kern des Phrasems gehören, in Form von Platzhaltern (*etw., jmd., jmdm, etc.*) repräsentiert. Wird die Unterscheidung zwischen statischen und variablen Elementen nicht deutlich gemacht, entstehen Beispielformulierungen, die die Grenze zwischen phraseologischem Kern und Kontext verschwimmen lassen (sogenannte »aktualisierte Beispiele«) (Burger 2010, S. 189).

In diesem Zusammenhang ist die übliche Vorgehensweise, bei der Angabe von Strukturformeln nur die obligaten Objekt-Valenzen anzugeben. Burger hält diese Herangehensweise für unzureichend, da das Subjekt in der Regel aufgrund des Infinitivs getilgt wird (Burger 2010, S. 190). Er plädiert dafür, die Valenzangabe nach dem Muster des KEMPCKE-WÖRTERBUCH-DAF-PW um die semantischen Informationen zum Subjekt zu erweitern. Wie Abbildung 3.1

am Beispiel von *jemandem etwas verübeln* zeigt, wird im KEMPCKE-WÖRTERBUCH-DAF-PW ein durch Schrägstriche markierter Platzhalter der üblichen Valenzangabe voranstellt, der die semantischen Restriktionen des Subjekts verdeutlicht.

verübeln [fɛ|'y:b|n], verübelte, hat verübelt /jmd./
jmdm. etw. ~ SYN 'jmdm. etw. übel nehmen'; ↗
FELD I.6.2: diese abfällige Bemerkung hat er ihm
sehr verübelt; er hat mir sehr verübelt, dass ich ihn
nicht darüber informiert hatte ❖ ↗ **übel**

Abbildung 3.1: Darstellung der externen Valenz im KEMPCKE-WÖRTERBUCH-DAF-PW

Auch Kæjrr betont die Wichtigkeit, die externe und semantische Valenz vollständig anzugeben (Kjaer 1987, S. 166–170). Für das Beispiel *baden gehen* schlägt sie folgende Valenzbeschreibung vor:

baden gehen

1. »scheitern«
 - a) j-d. geht (mit etw.) baden (z. B. mit einem Vorschlag)
 - b) j-d. geht (bei etw.) baden (z. B. bei einem Vorhaben)
 - c) etw. geht (durch etw.) baden (z. B. die Ferien durch den Regen)
2. »in Konkurs geraten«
 - a) j-d. geht (mit etw.) baden (z. B. der Händler mit seinem Geschäft)
 - b) etw. geht baden (z. B. ein Geschäft, eine Firma)

Bemerkenswert an diesem Beispiel ist, dass sich die Polysemie des Ausdrucks auch in seiner Valenzbeschreibung niederschlägt. Die Kodierung paradigmatischer Relationen im Wörterbuch kann folglich die Lesartendisambiguierung unterstützen. Jedoch ist eine solch ausführliche Darstellung der Valenz nur im Artikeltext sinnvoll, zumal sich eine vollständige Darlegung der externen Valenz in der Nennform negativ auf die Verständlichkeit auswirken kann (Koniczna 2005, S. 80; Korhonen 1990, S. 201).

Der relativ ausführlichen Darstellung mit Platzhaltern oder in Beispielsätzen steht die Kodierung der externen Valenz in Strukturformeln gegenüber. Sie birgt den Vorteil, dass sie weniger Platz beanspruchen als explizite Hinweise oder Beispielsätze, die Ersparnis geht jedoch auf Kosten der intuitiven Verständlichkeit und eine Erklärung dieser Vorgehensweise in den Umtexten ist daher unumgänglich (Stantcheva 2002, S. 179).

3.1.3.3 Darstellung morphosyntaktischer Restriktionen

Die Kodierung morphosyntaktischer Restriktionen in der Nennform gibt ersten Aufschluss über die Grammatik eines Phrasems. Burger (1983, S. 34–36) schlägt folgende Regeln vor: Ist die Nennform die konsensuale Grundform des Phrasems, bedeutet dies in der Regel, dass sie keinen morphosyntaktischen Restriktionen unterliegt. Unterliegt der Ausdruck dennoch Beschränkungen (wirken sich die Restriktionen auf andere Formen als die Grundform aus), sollte die entsprechende Information in einem metasprachlichen Kommentar kodiert werden. Weicht die Nennform von der konsensualen Grundform ab, weist dies auf das Vorliegen morphosyntaktischer Restriktionen hin, die die Bildung der konsensualen Grundform verhindern. Diese Regeln gelten analog für semantische Restriktionen.

Burger äußert sich angesichts der unsicheren Ergebnisse der einschlägigen phraseologischen Forschung weitgehend zufrieden mit der Kodierung morphosyntaktischer Restriktionen in der neueren deutschen Phraseographie (Burger 2010, S. 130). Eine Ausnahme bilden die bereits erwähnte Verwendung aktualisierter Beispiele an Stelle der kanonischen Form, die dazu führt, dass die Markierung von Restriktionen in der Nennform an Aussagekraft verliert (vgl. auch (Stantcheva 2002, S. 181)). Scholze-Stubenrecht (1988, S. 291) hält hingegen eine Kodierung von morphosyntaktischen Restriktionen aufgrund mangelnder Stabilität der Phraseme generell für wenig sinnvoll.

3.1.4 Kennzeichnung

Mängel in der systematischen Kennzeichnung von Phrasemen in allgemeinen Wörterbüchern werden häufig moniert (u. a. Burger (2010, S. 183–184) Stantcheva (2002, S. 177–178) Korhonen (1990, S. 198)). Gemein ist den Kritiken, dass sie eine einheitliche und eindeutige Markierungsmethode fordern, die es ermöglicht, Phraseme auf den ersten Blick von Satzbeispielen zu unterscheiden. Die Systematik sollte dabei mit der in den Umtexten verwendeten Terminologie (und der Phraseologieauffassung) überein stimmen und in den Benutzungshinweisen hinreichend erklärt werden. Zur Wahl stehen hier graphische, typographische und verbale Markierungsmethoden. Werden Symbole oder Abkürzungen verwendet, sind diese gesondert in Verzeichnissen zu sammeln und zu erklären.

Präferiert werden graphische Kennzeichnungen durch die vorangestellten Symbole »+« oder »*« sowie verbale Marker, die den aktuellen Forschungsstand reflektieren (Stantcheva 2002,

S. 178–179; Wiegand 2010, S. 169). Wird die metasprachliche Formulierung als Kennzeichnungstyp gewählt, hält Burger es für sinnvoll, phraseologische Subklassen zu distinguieren (Burger 1989, S. 595).

3.1.5 Umgang mit Varianten

Bei der Frage nach Varianten herrscht Konsens, dass diese nach Möglichkeit im Wörterbuch zu verankern sind. Als Varianten gelten nach Barz (1992, S. 36) »usualisierte koexistierende Formalternanten mit invarianter phraseologischer Bedeutung«. Wie eine variante Form sinnvoll im Lexikon verankert werden kann, hängt von ihrer konkreten Realisierung ab. Für die Integration in eine gemeinsame Nennform eignen sich Formen, die nur geringfügig voneinander abweichen. Bei der Zusammenfassung mit Hilfe eines vollständigen Artikels und einem oder mehreren Verweiseinträgen gibt es keine Einschränkungen.

Für die Integration aller synonymen oder quasi-synonymen Einträge in einen Wörterbucharikel, auf welchen nötigenfalls von anderer Stelle im Wörterverzeichnis verwiesen wird, spricht sich Pilz (1987, S. 138) aus.⁸ Daraus kann sich die Notwendigkeit ergeben, die alphabetische Ordnungsstruktur zugunsten der Zusammenstellung von Kovarianten zu durchbrechen, wenn dies für den Benutzer informativer ist (Barz 1992, S. 42–43).

Die Integration von synonymen Varianten in einer gemeinsamen Nennform kann sinnvoll sein, vorausgesetzt eine semantische, stilistische und strukturelle Übereinstimmung kann festgestellt werden (Korhonen 1990, S. 200; Barz 1992, S. 42–43; Petermann 1983, S. 182–183). Dabei kann die Auffindbarkeit durch entsprechende Verweise aus dem Wörterbuchverzeichnis heraus gesichert werden. Dies ist insbesondere für Benutzer mit nicht-muttersprachlichem Hintergrund von Bedeutung, da diese das Phrasem in genau der Form auffinden sollten, in welcher sie es gesucht haben.

Scholze-Stubenrecht plädiert für den Einsatz exemplarischer Varianten statt der gängigen Verwendung von sprachlichen Unversalien (*an einem Ort weht zu einer Zeit ein scharfer/schärferer usw. Wind*) oder Formalisierungen (*loc.Adj. + weht + temp.Adj. + ein scharfer/schärferer usw.*

⁸ Um zu verdeutlichen, wann ein Verweis notwendig ist, schlägt er vor, die Varianten *jmd. jedes Wort (einzeln) aus der Nase ziehen müssen* und *jmd. die Wörter (einzeln) aus der Nase ziehen müssen* in einem Artikel zusammenzufassen, während er *jdm. die Würmer (einzeln) aus der Nase ziehen müssen* durch einen Verweis von einer der beiden Varianten auf den Eintrag der anderen realisieren würde.

Wind), da eine vollständige Erfassung aller Varianten ohnehin nicht möglich ist. Eine begrenzte Anzahl an Varianten (*hier/bei uns usw. weht jetzt/seit neuestem ein scharfer/rauher/anderer o.ä. Wind*) ist übersichtlich und schafft Anhaltspunkte, anhand derer der Benutzer die Verwendung im Einzelfall erschließen kann (Scholze-Stubenrecht 1988, S. 291). Statt einer exhaustiven Nennung aller Restriktionen hält er auch eine kodifizierte Angabe wie beispielsweise »nicht durch Attribute modifizierbar« für denkbar (Scholze-Stubenrecht 1988, S. 295).

Eine weitere Möglichkeit ist die Nennung der Kovarianten im Beschreibungsteil, die Burger (1989, S. 598) vorwiegend für die Angabe lexikalischer Varianten vorsieht, bei welchen ein Austausch durch ein Lexem erfolgt, das in freier Verwendung nicht synonym ist. Diese eignet sich insbesondere für Phraseme, für welche die Bestimmung der Komponenten problematisch ist, da eine nicht klar umrissene Menge von Lexemen als potentielle Kandidaten in Frage kommen (z. B. bei *sich lang und breit...*, das mit einer Reihe von *verba dicendi* verwendet werden kann (Burger 1983, S. 39)).

3.1.6 Makrostruktur

Herbst bezeichnet die Frage, wie Phraseme in eine alphabetische Makrostruktur eingeordnet werden sollen, als »[d]as grundlegende Problem, das sich bei der Erfassung von Idiomen in allgemeinen Wörterbüchern wie in Spezialwörterbüchern stellt« (Herbst 2004, S. 222). In der Phraseographie haben sich verschiedene Ansätze herausgebildet, die jeweils unterschiedliche Probleme mit sich bringen und deren Nutzen durchaus umstritten ist. Um diese Ansätze strukturiert zu diskutieren, muss man diese in Zusammenhang mit der Wahl der Menge der Leitelemente, auf welchen die primäre Ordnungsstruktur (Makrostruktur) des Lexikons definiert ist, diskutieren.

Die einfachste Herangehensweise ist die exhaustive Alphabetisierungsmethode, die die gesamte Sequenz von Komponenten des Mehrwortlexems als Leitelement verwendet, ohne Rücksicht auf die Wortgrenzen oder den semantischen Wert der Elemente zu nehmen. Alternativ kann eine Menge von Stichwörtern (Einzelwortlexemen) nach bestimmten Kriterien als Basis für die Anordnung der Phrasemeinträge dienen. Die Verknüpfung von Lexikoneinträgen zu Stichwörtern erfolgt im semasiologischen Lexikon auf Basis der darin enthaltenen Lexeme (Komponenten), im onomasiologischen Lexikon auf Basis der idiomatischen Gesamtbedeutung, die durch einen Begriff repräsentiert wird. Die onomasiologische Zuordnung bedient sich dabei eines

kontrollierten Vokabulars, während die lexemische Zuordnung auf Basis der Gesamtmenge der (relevanten) Komponenten erfolgt.

In Bezug auf die Anordnung der Leitelemente werden zwei Methoden unterschieden: die sequentielle Anordnung unter Verwendung graphemischer (oder phonetischer) Merkmale (z. B. initialalphabetisch oder nach der Systematik der 50-Laute-Tafel) und die semantisch-hierarchische Anordnung mittels taxonomischer Relationen, wie sie aus Thesauren bekannt ist. Burger (2010, S. 200) bezeichnet die Zuordnung auf Basis der wörtlichen Lesart als eine Gruppierung nach »Sachgruppen«, während er im Kontext onomasiologischer Wörterbücher von einer Gruppierung nach »Begriffen« spricht.

3.1.6.1 Die exhaustive Alphabetisierung

Die exhaustive Alphabetisierungsmethode wird in der deutschen Phraseographie kaum eingesetzt⁹, da die Nennform der Phraseme im Deutschen meist mit Synsemantika (Artikeln, Indefinitpronomen sowie bedeutungsschwachen Verben) eingeleitet wird, wodurch die ersten Elemente dieser Zeichenkette selten spezifisch sind und sich deshalb als Leitelemente disqualifizieren (Klappenbach 1980, S. 127). Die gängige Praxis, bei verbalen Phrasemen Kontextelemente in der Nennform durch die Platzhalter *jmd./etw.* zu ersetzen, um die externe Valenz zu verdeutlichen, sorgt zusätzlich für eine Gleichförmigkeit, da diese meist auch in vorderer Position auftreten. Dementsprechend wird dieser Methode auch vergleichsweise wenig Aufmerksamkeit in der linguistischen Diskussion gewidmet.

Eine modifizierte Form dieser Methode wurde in Verbindung mit dem initialalphabetischen Ordnungssystem für das DUDEN-ZITATEREDEWENDUNGEN-PW gewählt. Möglicherweise basiert die redaktionelle Entscheidung auf der Tatsache, dass dieses aus der Benutzung allgemeiner Wörterbücher bekannte Prinzip dem Benutzer den Zugang zum Spezialwörterbuch erleichtert. In diesem Werk werden Phraseme rein initialalphabetisch sortiert, wobei das jeweils erste Wort als Leitelement hervorgehoben wird. Ausgenommen von der Ordnungsrelation werden hierbei jedoch bestimmte und unbestimmte Artikel, Reflexivpronomen und Platzhalter in der Valenzangabe (z. B. *jemanden* oder *etwas*). Dies führt dazu, dass häufige Synsemantika wie

⁹ Wiegand schlägt diese für Phraseme vor, die kein »Inhaltswort« aufweisen, wobei er die gesamte kanonische Form als Leitelement heranzieht. Diese »Buchstaben-für-Buchstaben-Methode« steht im Gegensatz zur »Wort-für-Wort-Methode«, bei welcher die Anordnung die Wortgrenzen berücksichtigt (Wiegand 2010, S. 165–169).

»mit« oder »ich« lange Artikelstrecken nach sich ziehen. Dennoch findet diese Vorgehensweise auch Befürworter. So beschreibt Hrisztova-Gotthardt (2010b, S. 54) die Vorgehensweise dieses Werkes als »hinsichtlich der Anordnung eine sehr gelungene Lösung« und auch Wiegand (2010, S. 170) meint zum Thema lange Artikelstrecken, dass Benutzer sich nicht daran stören, da sie ohnehin daran gewöhnt seien, lange Strecken mit den Vorsilben »auf-« und »ab-« zu durchforsten.

3.1.6.2 Ordnung nach Leitelementen

Der allgemeine Usus in der Phraseographie im deutschen Sprachraum ist, Mehrwortlexeme Einzelwortlexemen als reduzierten Leitelementen zuzuordnen. Wie bereits in Kapitel 3.1.6.1 erläutert eignet sich das erste Element in der Nennform deutscher Phraseme aufgrund des Fehlens distinktiver Merkmale nicht als Leitelement. Folglich ist die Diskussion alternativer Methoden zur Bestimmung des Leitelements im deutschen Forschungsdiskurs ein wichtiges Thema.

Die denkbar einfachste Möglichkeit ist das exhaustive Prinzip, in welchem das Mehrwortlexem unter allen Komponenten verfügbar gemacht wird. Zwar handelt es sich bei der sogenannten Mehrfachlemmatisierung um die benutzerfreundlichste Variante (Scholze-Stubenrecht 1988, S. 296), jedoch wird im Printbereich diese exhaustive Kodierungsmethode meist aus ökonomischen Gründen nicht in Erwägung gezogen (Konieczna 2005, S. 76). Zudem birgt sie das Risiko von Redundanzen und Inkonsistenzen in der Angabe der Nennform und der Bedeutungserläuterung (Korhonen 1990, S. 203; Stantcheva 2002, S. 41).

Stantcheva (2002, S. 41) plädiert statt dessen für ein morphosyntaktisches Prinzip, bei welchem die Wortklassenzugehörigkeit der Komponenten über die Zuordnung entscheidet. Der Nachteil dieser Vorgehensweise ist, dass sie ein grundlegendes Verständnis der Wortklassen voraussetzt. Das Phrasem nach der Komponente einzuordnen, die den »semantischen Kern« oder das »semantische Zentrum« des Phrasems darstellt, hält Stantcheva hingegen für ungeeignet, da diese Vorgehensweise weitgehende Kenntnisse über das Nachzuschlagende voraussetzt und für vollidiomatische Mehrwortlexeme ungeeignet ist.

Burger (2010, S. 185) spricht sich für die Mischform aus morphosyntaktischem und semantischem Prinzip aus, nach dem das »erste Substantiv bzw. erste sinntragende Wort« über die

Zuordnung entscheidet. Existieren mehrere potentielle Kandidaten kann der Einsatz von Verweisen sinnvoll sein (Korhonen 1990, S. 203; Wiegand 2010, S. 169)¹⁰. Nur Phraseme, die keine Inhaltsworte enthalten, sollen nach der exhaustiven mechanischen Alphabetisierungsmethode behandelt werden.

Obwohl der Ansatz ähnlich ist, sind die »sinntragenden Elemente« bei Burger nicht gleich den Inhaltsworten bei Wiegand. So bezieht Wiegand den Begriff »Inhaltswort« auf »phraseologische Komponenten, deren Form mit der Form der phrasemexternen Benennungen gleich ist« (Wiegand 2010, S. 168) und schließt damit unikale Komponenten aus. Der ebenfalls nach einem semantisch-logischem Prinzip arbeitende Pilz (1987, S. 131–132) plädiert sogar dafür, *nicht für Geld und gute Worte* unter der obligatorischen Komponente »nicht« anzuführen, welches als Synsemantikon weder bei Burger noch bei Wiegand den Kriterien eines Leitelements entspräche. Die jeweiligen Forderungen differieren jedoch auch in ganz grundsätzlicher Art und Weise. So spricht sich Schemann (1991, S. 1027) explizit für eine lockere Handhabung des Stichwortprinzips aus, da es ermöglicht, sehr umfangreiche Idiomgruppen zu bilden, die nicht zwingend bestimmten Kriterien folgen, während Stantcheva (2002, S.176) eine strikte Umsetzung von eindeutigen Regeln fordert, die in den Umtexten detailliert dargelegt werden.

3.1.6.3 Onomasiologische Ordnung

Die Fragen der onomasiologischen Ordnung polarisiert wie kein anderes Thema aus diesem Bereich den Forschungsdiskurs. Laut Pilz handelt es sich dabei um das einzige zufriedenstellende Ordnungssystem (Pilz 1987, S. 131) und auch Hausmann (1983, S. 200) plädiert für die Zuordnung von Phrasemen zu Leitbegriffen in Abhängigkeit von der Gesamtbedeutung des Ausdrucks.

Schemann sieht hingegen die Propagierung des onomasiologischen Verfahrens als zeitweiliges Phänomen, das sich nicht durchsetzen konnte, da es nicht den Bedürfnissen der Benutzer

¹⁰ Bei Wiegand heißt es: »Bei Phrasemen die mehr als ein Inhaltswort aufweisen, wird unter dem zweiten und dritten Inhaltswort sowie gegebenenfalls unter weiteren Inhaltswörtern auf das erste verwiesen, und zwar durch eine verweisvermittelnde Phrasemangabe im Postkommentar des Artikels zum ersten Inhaltswort, der damit ein rechtserweiternder Wörterbuchartikel ist.« Dabei soll auch dieser Postkommentar durch eine Phrasemidentifizierungsangabe eingeleitet werden (Wiegand 2010, S. 169).

entspricht (Schemann 1991, S. 1028). Das Hauptproblem ist seiner Einschätzung nach die unklare Abgrenzung der Sinngruppen gegeneinander und das Fehlen eines extern geschlossenen Ordnungsschemas. Den Nutzen solcher Nachschlagewerke hält Schemann im Gegensatz zum alphabetischen Verfahren für sehr begrenzt. Dies gilt seiner Meinung nach nicht für die idiomatischen synonymischen Wörterbücher.¹¹

Auch Burger sieht das Hauptproblem darin, den Objektbereich in eine intersubjektiv akzeptable Ordnung zu bringen (Burger 2010, S. 201). Dem nach Sachgruppen angeordneten Wörterbuch fehlt in seinen Augen eine passende Benutzerfrage, denn auch die Frage, welche Phraseme für ein bestimmtes Sachgebiet (z. B. mit wörtlicher Bedeutung rund um das Wortfeld »Familie«) zur Verfügung stehen, hält er für nicht zufriedenstellend gelöst. Den Nutzen von nach Gesamtbedeutung (Zielbereich) geordneten onomasiologischen Wörterbüchern sieht er vornehmlich in Bezug auf die Frage, welche Phraseme für einen bestimmten Zielbereich zur Verfügung stehen, wie sich Bedeutungen dieser Phraseme gegeneinander abgrenzen und welchem Zielbereich ein (unbekanntes) Phrasem zuzurechnen ist. Abgesehen von der ersten Fragestellung wird jedoch laut Burger zunächst der alphabetische Index und nicht die onomasiologische Struktur konsultiert werden. Die eigentliche Benutzerfrage onomasiologischer Herangehensweisen, welche Phraseme für einen Zielbereich zur Verfügung stehen, sieht er auch nur dann potentiell beantwortet, wenn der passende »Leitbegriff« gefunden wird. Dabei zieht er die systematische Anordnung der Begriffe aufgrund der Benutzerfreundlichkeit dieser Methode der alphabetischen vor.

Hessky betont die Bedeutung onomasiologischer Gruppierungen für das Sprachlernen und setzt diese in ihrem Werk »Deutsche Redewendungen: ein Wörter- und Übungsbuch für Fortgeschrittene« in die Praxis um. Dabei ist sie sich der Probleme in der Ausarbeitung der Systematik bewusst (Hessky und Ettinger 1997, S. XXIV–XXVI). Ungeachtet des Nutzens einer derartigen Zugriffsmöglichkeit für die Sprachproduktion bezweifelt Scholze-Stubenrecht (1988, S. 296), ob sich eine breitere Benutzergruppe hierfür interessiert und argumentiert, das gleiche Nachschlagebedürfnis werde bereits im Kern von Synonymwörterbüchern erfüllt.

¹¹ Dementsprechend ist das von Schemann publizierte SCHEMANN-IDIOMS-PW onomasiologisch aufgebaut.

3.1.7 Mikrostrukturelle Einordnung

Werden Phraseme nicht als separate Wörterbuchartikel realisiert, sondern in bestehende Strukturen integriert, stellt sich die Frage, an welcher Stelle im Artikel sie verankert werden. Für nicht-idiomatische und teil-idiomatische Ausdrücke besteht die Möglichkeit, diese auf Basis ihrer Bedeutung einer bestimmten Lesart des Lemmas zuzuordnen. Alternativ können sie einen separaten Block (meist am Ende des jeweiligen Wörterbuchartikels) bilden, zusammen mit den opaquen Idiomen, die aufgrund ihrer fehlenden Motivierung keiner Lesart zugeordnet werden können.

Obwohl insbesondere bei metaphorischen Idiomen die Verlockung groß ist, den idiomatischen Zusammenhang durch eine Lesarten-Zuordnung lexikographisch zu manifestieren, ist der einstmalige Vorschlag, Phraseme in Abhängigkeit von ihrer Idiomatizität zu verankern, in jüngeren Publikationen nicht mehr präsent (Burger 1989, S. 596–597; Burger 2010, S. 188–189). Statt dessen herrscht weitgehend Konsens, dass Phraseme ungeachtet ihrer Idiomatizität auf mikrostruktureller Ebene nicht mit einzelnen Lesarten in Verbindung gebracht werden sollen (Stantcheva 2002, S. 43–44).

Die Positionierung der Phraseme in einem separaten Bedeutungspunkt unabhängig von ihrer Idiomatizität ist vergleichsweise intuitiv. Hier gilt jedoch zu beachten, dass eine Integration phraseologischer Angaben im Block unübersichtlich werden kann (Schemann 2009, S. 104). Die Orientierung kann hier ebenso wie bei der Haupt-Lemmareihe durch Ordnungsstrukturen und Markierungen erleichtert werden. Neben der naheliegenden alphabetischen Ordnungsstruktur ist die typologische Anordnung ebenfalls denkbar (Kühn 2003, S. 110). Scholze-Stubenrecht hält selbst eine Anordnung nach morphosyntaktischen Kriterien für eine praktikable Lösung (Scholze-Stubenrecht 1988, S. 293).

3.1.8 Angabe der Bedeutung

Schemann (1991, S. 1022) unterscheidet vier sich ergänzende Arten von Bedeutungsangaben

1. Definitionen
2. Bedeutungsumschreibungen
3. Angabe von synonymischen Lexemen oder Äquivalenten (im bilingualen Kontext)
4. Beispiele

Allerdings stellt Schemann fest, dass Definitionen im strengen Sinne in keinem Werk vorherrschen, meist werden definitorische Elemente und Bedeutungsumschreibungen mit synonymischen Angaben gemischt. Dabei bemängelt er das Fehlen einer Ordnung der unterschiedlich gearteten Angaben. Neben der üblichen Unterscheidung zwischen denotativen, stilistischen (oder konnotativen) Aspekten in der Bedeutungserläuterung unterscheidet er außerdem zwischen Herangehensweisen, die die Gesamtbedeutung in den Vordergrund rücken und von solchen, die eine analytisch-definitorische Herangehensweise wählen (Remotivierung). Im bilingualen Kontext unterscheidet Schemann außerdem zwischen Äquivalenten, die sich auf die Situation der Ausgangssprache beziehen und solchen, die Äquivalente für Situationen in der Zielsprache darstellen. Eine einheitliche Vorgehensweise kann er jedoch nicht feststellen. Außerdem unterscheidet er auf struktureller Ebene Bedeutungsangaben nach ihrer Länge (Einzelwortlexeme, Phrasen, Sätze oder sogar kleine Texte).

Die Anforderungen an die Ausarbeitung der Bedeutungsangabe sind durchweg hoch. Dobrovol'skij (2002, S.363) fordert:

Zur Materialbeschreibung läßt sich zunächst sagen, daß jedes lexikographisch erfaßte Phrasem in semantischer, syntaktischer und pragmatischer Sicht vollständig und explizit beschrieben werden muß.

Damit ist er nicht allein. Auch Pilz verlangt nach einer differenzierten semantischen Erläuterung, sowie der Darlegung aller grammatischen, stilistischen, pragmatischen und soziokulturellen Besonderheiten (Pilz 1987, S. 139).

Darüber, wie dies geschehen soll, herrscht jedoch wenig Einigkeit. Dobrovol'skijs (2002, S. 366) Vorschlag einer »Bedeutungserläuterung, die das Wesen der Semantik von (1) [*den Bock zum Gärtner machen*] mehr oder weniger adäquat treffen würde«, stellt hohe Ansprüche an die Sprachkenntnisse des Benutzers:

»einer Person, die fähig zu sein scheint, in einem bestimmten Tätigkeitsbereich, der mit Verantwortung und/oder Macht verbunden ist, großen Schaden anzurichten, erlauben eben diese [sic] Tätigkeitsbereich zu übernehmen«

Burger hält es für »verfehlt, eine Genauigkeit und Explizitheit in der Paraphrasierung zu verlangen, wo der Sprachgebrauch dem nicht entspricht« (Burger 2009, S. 33). Er plädiert für eine »weniger merkmalthaltige Paraphrase«.

3.1.8.1 Pragmatische Aspekte in der Bedeutungsangabe

Es ist unumstritten, dass pragmatische Aspekte in der Bedeutungserklärung mancher Phraseme eine wichtige Rolle spielen. Der pragmatische Kommentar kann stilistisch-zeitlichen Angaben, Sprechereinstellungen und Angaben zur Illokution beinhalten. In den Punkten, dass die Bedeutungsbeschreibung von Phrasemen von ihrem Idiomatizitätsgrad abhängig gemacht werden sollte, dass eine historische Remotivierung zu vermeiden ist und der semantische Kommentar einheitlich zu gestalten ist, herrscht Einigkeit. In Bezug auf die Kodifizierung gibt es jedoch verschiedene Herangehensweisen.

Traditionell dient eine Auswahl an Stichworten der Attributierung der Phraseme mit stilistischen Merkmalen. Die stilistisch-pragmatische Angabe gilt jedoch als relativ unspezifisch und es fehlen intersubjektiv gültige Bewertungskriterien für die Zuordnung solcher Merkmale zu Phrasemen (Scholze-Stubenrecht 1988, S. 293).

Burger (2010, S. 198–199) sieht die Stilangabe im LANGENSCHIEDT-WÖRTERBUCHDAF-PW als einen Schritt in die richtige Richtung, in welcher statt der umstrittenen traditionellen stilistischen Markierungen (z. B. *umgangssprachlich, salopp, gehoben, vulgär* etc.) kommunikationsorientiert zwischen *gesprochenem* und *geschriebenem* Sprachstil unterschieden wird. In Bezug auf stilistisch-historische Markierung legt er besonderen Wert darauf, *veraltendes* Sprachmaterial zu kennzeichnen. Außerdem sieht er Bedarf, die Informationen zu Sprechereinstellungen in einer Weise auszudifferenzieren, dass aus der Erklärung deutlich wird, auf welchen Aktanden sich die Charakterisierung bezieht. Für die Kodifizierung illokutiver Aspekte regt er eine Integration in die Nennform an. Demnach tritt beim Vorliegen eindeutiger Restriktionen an Stelle der üblichen Form (*vor seiner eigenen Tür kehren*) eine Realisierung des Phrasems, welche die pragmatische Information impliziert (z. B. *du solltest vor deiner eigenen Tür kehren /er sollte vor seiner eigenen Tür kehren*). Alternativ können illokutive Aspekte (z. B. *für jüngere Beine haben*) durch einen entsprechenden Vermerk (z. B. *Aufforderung zum helfen* beziehungsweise *Hilfsbereitschaft*) in der Bedeutungserklärung kodiert werden.

Auch in Bezug auf die stilistische Kennzeichnung ist die Einheitlichkeit und angemessene Erläuterung in den Umtexten unabdingbar. Kühn (1983, S. 422) schlägt allerdings vor, auf konnotative Angaben zur Stilfärbung /Stilschichtung zugunsten von aussagekräftigen pragmatischen Kommentaren zu verzichten. Er stellt die Tradition der denotativen Bedeutungsbeschreibung in Frage, indem er die Auffassung vertritt, »die Bedeutung eines Sprachzeichens [sei] als Formulierung seiner Verwendungsregeln« zu verstehen. Die herkömmlichen Definitionen

sind demnach durch pragmatische (konnotative) Aspekte zu »Gebrauchsanweisungen« zu ergänzen. Restriktionen auf einen bestimmten Personenkreis (z. B. Eltern, Verwandte, Männer, Frauen) sowie potentielle Sprechereinstellungen und Sprachhandlung sollen dabei durch konsequente Textanalyse zum Vorschein gebracht werden (Kühn 2003, S. 115).

Scholze-Stubenrecht (1988, 294 und 300) hält Forderungen, wie die von Kühn formulierten, vor dem Hintergrund technischer und ökonomischer Beschränkungen in der kommerziellen Lexikographie für nicht praktikabel. Ursächlich ist vor allem die Länge der auf diese Weise kodierten Bedeutungsangaben. Darüber hinaus sieht er in dieser Herangehensweise eine potentielle Doppel-Kodierung von Information, da die konnotativen Aspekte der Bedeutung in der Regel bereits in Anwendungsbeispielen enthalten sind.

3.1.8.2 Art der Bedeutungserklärung

Die Paraphrasierung von Phrasemen durch synonymische oder – im Bereich der bilingualen Lexikographie – äquivalente Lexeme ist ein häufig eingesetztes Mittel der Bedeutungsbeschreibung, da diese in der Regel kürzer ausfällt als eine Definition oder Bedeutungsumschreibung. Sowohl die Verwendung von Einwortlexemen als auch die von Phrasemen ist dabei nicht unumstritten. Besonders kontrovers wird der Einsatz synonymischer Phraseme im Kontext von Idiomen diskutiert. Die vorherrschende Meinung, dass synonyme Phraseme die Idiombedeutung nicht adäquat wiedergeben können, wird durch die Tatsache begründet, dass diese meist nicht vollständig semantisch invariant seien (Dobrovol'skij 2002, S. 365–376; Burger 2010, S. 191–194). Darüber hinaus besteht die Gefahr, dass diese aufgrund ihrer komplexen semantischen Merkmale möglicherweise Verständnisprobleme für den Benutzer bergen.

Schemann (2009, S. 109–116) hinterfragt, ob die Regel, eine natürliche Sprache »akzeptiere« keine synonymen Ausdrücke, auf die Idiomatik übertragbar ist. Er widerlegt die Allgemeingültigkeit dieser Regel, indem er auf die große Anzahl synonymier idiomatischer Einheiten verweist, die er in seiner Analyse zu Synonymiegruppen des Deutschen herausgearbeitet hat. Auch geht er davon aus, dass diese Ergebnisse auf andere Sprachen übertragbar sind (Schemann 2009, S. 110).

Im Hinblick auf die Paraphrase von Idiomen mit Einwortkorrelaten herrscht ebenfalls Uneinigkeit. Dobrovol'skij (2002, S. 365–376) begründet seine Skepsis gegenüber einer Bedeutungserklärung durch lexematische Entsprechungen damit, dass die Existenzberechtigung von Phrasemen auf ihrem Mehrwert gegenüber ihren »quasi-synonymen Einzelwortkorrelaten« basiert

(Dobrovol'skij 2009, S. 159). Statt dessen präferiert er umfassende Bedeutungserklärungen in Form von Definitionen, die sich aus extensiven Textanalysen speisen. Burger (1989, S. 596) argumentiert hingegen, dass die pauschalisierende Aussage, Phraseme ließen sich nicht durch Einwortkorrelate erklären, vorwiegend für Routineformeln gelte. Er vertritt den Standpunkt, dass sich insbesondere die Bedeutung von opaquen Idiomen durchaus durch Einwortlexeme vermitteln lässt. Der Ansatz von Kühn (1983, S. 422) sieht eine Mischform vor, in welcher beispielsweise *kalter Kaffee* sowohl durch lexematische Entsprechungen *uninteressant*, *witzlos* als auch in Abgrenzung zu seinen phraseologischen Verwandten (z. B. *alter Hut*) erklärt wird.

3.1.8.3 Angabe zur Etymologie

Obwohl das Interesse seitens der Nutzer an der Etymologie und Motiviertheit der Phraseme durch häufige Anfragen an die Redaktion von Wörterbuchverlagen erkennbar ist (Scholze-Stubenrecht 1988, S. 295–296), sind die Meinungen hinsichtlich dieser Angabe gespalten. Benutzerstudien zeigen, dass etymologische Angaben keinen statistisch signifikanten positiven Effekt auf den Lernerfolg von L2-Lernern haben (Lew und Szczoaniak 2011). Zwar können konzise etymologische Angaben das Verständnis opaquer Idiome in typischen Lernumgebungen unterstützen, da diese aber eine tiefer gehende Auseinandersetzung mit der Materie verlangen, können sie in oberflächlichen Lesesituationen, wie beispielsweise beim Konsultieren eines Wörterbuchs, hinderlich sein (Lew und Szczoaniak 2011, S. 341–342; Boers, Demecheleer und Eyckmans 2004).¹²

Es herrscht Konsens, dass im Wörterbuch die synchrone Bedeutung von Phrasemen im Vordergrund stehen sollte. Eine Erklärung mit Hilfe einer historischen Rückführung auf die Motivationsbasis (Remotivation) ist nach Ansicht der meisten Lexikographen in jedem Fall zu vermeiden (Kühn 2003, S. 111; Burger 2010, S. 191–194; Burger 2009, S. 36–37; Stantcheva 2002, S. 52).

Ob die Angabe der Etymologie dem Nutzer beim Erlernen eines Idioms hilfreich ist, kann nur für den jeweiligen Einzelfall entschieden werden (Burger 2010, S. 194). Insbesondere im Falle von Komponenten, die nicht mehr im aktiven Gebrauch sind, stellt sich dem Lerner die

¹² Die Diskrepanz zu den Ergebnissen von Boers, Eyckmans und Stengers (2007) erklären Lew und Szczoaniak (2011) durch Unterschiede im Versuchsaufbau.

Frage, was die ursprüngliche Bedeutung des Begriffs ist (z. B. *Kerbholz* bei *etwas auf dem Kerbholz haben*). Auch beeinflusst die bildliche Komponente die »lexikalische Kompatibilität« des Idioms wesentlich, weshalb diese dem Verständnis motivierter Idiome zuträglich ist (Dobrovol'skij 2002, S. 367). Beachtenswert in diesem Kontext ist, dass die historisch-etymologischen Angaben in der Bedeutungserklärung deutlich von anderen Teilen der Bedeutungsangabe zu trennen sind (Burger 2010, S. 193). Da in vielen Fällen die Herkunft der Idiome nicht eindeutig geklärt ist, hält Pilz (2002, S. 308) sprachhistorische und soziokulturelle Erläuterungen nur in entsprechend umfangreichen Werken für wirklich sinnvoll.

3.1.8.4 Beispielangaben

Beispielsätze und Belege dienen meist der Illustration des Verhaltens eines Phrasems im Kontext. Sie können ergänzend oder als Ersatz für formalisierte andere Angaben dienen, wie bereits an verschiedenen Stellen in den vorangegangenen Kapiteln deutlich wurde. In der Angabe von exemplarischen Kontexten werden konstruierte Beispielsätze von authentischen (Text-)Belegen (auch Zitate genannt) unterschieden. Insbesondere Belege erbringen traditionell Nachweise über die synchrone und/oder diachrone sowie anderweitig bemerkenswerte Verwendung eines Ausdrucks. Als dritte Form wird eine Mischform diskutiert, die sich an authentischen Textbelegen orientiert, diese aber zum Zwecke der Nutzung im Lexikon vereinfacht.

Da den konstruierten Beispielsätzen das Stigma anhaftet, dass sie die Realität des Sprachgebrauchs nicht ausreichend abbilden, nutzt die Lexikographie zunehmend die Möglichkeiten der Korpuslinguistik um umfangreiche Belegensammlungen anzulegen. Die Frage, die die Lexikographen und Linguisten beschäftigt, ist, wie mit dem Material umzugehen ist und welche Aufgabe Beispielsätze beziehungsweise Belege übernehmen sollen (Ettinger 2009a; Heine 2009).

Beide Formen der Beispielangabe dienen mehrheitlich der Ergänzung weiterer Angaben im Wörterbuchartikel. Im Regelfall ergänzen sie den Wörterbuchartikel um Kontextinformationen hinsichtlich der morphosyntaktischen, pragmatischen oder semantischen Merkmale eines Ausdrucks. Dabei illustrieren (konstruierte) Beispielsätze meist den typischen Gebrauch eines Ausdrucks, während Belege vorwiegend als Nachweis für die synchrone oder eine spezifische Verwendung des Ausdrucks angeführt werden. Da die Art der Angabe unterschiedliche Funktionen erfüllt, weist Burger (2009, S. 35) darauf hin, dass die Wahl zwischen Beispielen und Belegen nicht willkürlich erfolgen sollte und die Systematik des Vorgehens in den Umtexten einer Erläuterung bedarf. Für konstruierte Beispielangaben wird vorgebracht, dass sie besser

die typischen Merkmale eines Phrasems zu vermitteln vermögen als authentische Belege, jedoch unter dem Vorbehalt, dass die Qualität von der Begabung des jeweiligen Lexikographen abhängt (Földes 1994, S. 198; Mellado Blanco 2009, S. 10–11).

Authentische Belege aus textuellen Korpora gelten für den seriösen lexikographischen Prozess als unabdingbar und gehören zu den wesentlichen Anforderungen an die moderne Redewendungen-Lexikographie, da nur mit Hilfe von Beispielen aus einem ausgewogenen Korpus alle Merkmale eines Phrasems adäquat beschrieben werden können (Fellbaum 2004; Hausmann 1983, S. 200; Koller 1987, S. 111; Kühn 2003, S. 114). Kontextuelle Belege illustrieren neben der tatsächlichen Verwendung im Idealfall die Bedeutung, Form, Grammatik und Pragmatik des jeweiligen Phrasems, allerdings können nicht redigierte Korpusbeispiele ein Problem für den (fremdsprachlichen) Benutzer darstellen und auch die sprachliche Korrektheit kann nicht immer sicher gestellt werden (Ješensek 2009, S. 73). Burger sieht den Nutzen der Belege in der Ergänzung der pragmatischen Angaben sowie Abgrenzung der Quasi-Synonyme gegeneinander (Burger 2009, S. 34–35).

Pilz (2002, S. 304–305) begegnet dem »Zwang des Zitat-Belegs«, wie er es nennt, wenig verständnisvoll. Er stellt die definitorische Funktion von Beispielsätzen in den Vordergrund und bemängelt Sorgfalt bei der Auswahl der Belege. Statt der Verwendungsvielfalt sollten seiner Ansicht nach insbesondere Gebrauchsbeschränkungen durch Beispiele illustriert werden.

Schemann (2009, S. 105) misst den Beispielangaben zentrale Bedeutung für die Phraseographie bei und argumentiert, durch die Angabe von gelungenen (Kontext-)Beispielen würde sich die Notwendigkeit einer Definition nicht nur erübrigen, er hält die theoretische Definition sogar als potentiell hinderlich für das Verständnis von Phrasemen (Schemann 2009, S. 107–109). Illustrative Beispiele, wie er die an authentischen Kontexten konstruierten Beispiele nennt, repräsentieren die Bedeutung besser, sofern es gelingt, synonyme oder quasi-synonyme Ausdrücke einzubetten. Für besonders geeignet hält er in diesem Zusammenhang die Beispielbildung anhand von Dialogen, da durch diese Methode pragmatische Determiniertheit in die Angabe integriert werden kann (Schemann 2009, S. 108–109).¹³ Als ideal betrachtet er jedoch eine Kombination aus *ad hoc* gebildeten Beispielen und Kontextbelegen (Schemann 2009, S. 107).

¹³ Der semantische Kommentar in SCHEMANN-IDIOMS-PW und SCHEMANN-DEUTSCHEIDIOMATIK-PW ist durchweg mit Anwendungsbeispielen versehen. Im SCHEMANN-DEUTSCHEIDIOMATIK-PW erfolgt die Bedeutungserklärung sogar ausschließlich durch Belege.

Burger (2009, S. 39–40) attestiert den an realen Kontexten modellierten konstruierten Beispielen in SCHEMANN-DEUTSCHEIDIOMATIK–PW einen gewissen Nutzen, bezweifelt jedoch, dass diese durch ihre implizierte Paraphrasierung eine Bedeutungserläuterung zu ersetzen vermögen.

Scholze-Stubenrecht (1988, S. 296–297) beschäftigt hingegen die Frage der Beispielwahl, genauer gesagt, ob eher typische, oder eher unerwartete individuelle Abweichungen im Wörterbuch hinterlegt werden sollen. Er argumentiert, dass ein deskriptives Wörterbuch einen Überblick über die gesamte Bandbreite der Nutzung eines Phrasems verschaffen soll, während in einem Wörterbuch mit normativem Anspruch den typischen Beispielen Vorrang eingeräumt werden sollte.

3.1.9 Grammatische Angabe

Angaben zur Grammatik gibt es in unterschiedlicher Ausführlichkeit. Die einfachste Version ist die Angabe der grammatischen Klasse, beispielsweise *VP* für Verbalphrasen, *NP* für Nominalphrasen und *PrepP* für Präpositionalphrasen (Dobrovolskij 2009, S. 157). Erweiterte Angaben enthalten zusätzlich Regeln, die das syntaktische Verhalten — insbesondere die morphosyntaktischen Restriktionen — repräsentieren.

Dobrovolskij (2002, S. 369) unterscheidet zwei grundsätzlich verschiedene Herangehensweisen:

- (a) Im ersten Fall gehen wir davon aus, daß die relevanten Regeln in der sogenannten Grammatik der Idiome und nicht in den entsprechenden Wörterbucheinträgen zu fixieren sind.
- (b) Im zweiten Fall wird vorausgesetzt, daß die Besonderheiten des syntaktischen Verhaltens jedes konkreten Idioms letzten Endes nur individuell definierbar sind, denn es handelt sich meistens nicht um prospektiv wirkende Regeln, sondern um retrospektiv wirkende Erklärungen. Daraus folgt, daß die Fähigkeit eines Idioms, sich einer bestimmten syntaktischen Transformation zu unterziehen, im entsprechenden Wörterbucheintrag explizit markiert werden muß.

Eine vollständige Beschreibung hält er nur im maschinenlesbaren Wörterbuch für sinnvoll und zeigt am Beispiel der Passivierbarkeit, wie schwierig es ist, Regeln für das syntaktische Verhalten von Idiomen zu erstellen. Er präferiert eine »doppelte Beschreibung«, die in nicht völlig evidenten Fällen auf die Fixierung der syntaktischen Besonderheiten im Wörterbuchartikel abzielt

und regt gleichzeitig an, dass sich die Phraseologie der Entwicklung einer Idiom-Grammatik widmen sollte. Die Definition eines solchen Regelwerks, die das Zusammenspiel semantischer und syntaktischer Merkmale adäquat beschreibt, könne die Markierung syntaktischer Transformationen im Wörterbuch vereinfachen (Dobrovolskij 2002, S. 371–372). Auch Rothkegel hält aus computerlinguistischer Perspektive eine Untersuchung der Regelmäßigkeiten in den morphosyntaktischen Restriktionen mit dem Ziel der Klassenbildung für wichtig (Rothkegel 2007, S. 1032). Pilz (2002, S. 307) wiederum bezweifelt, dass morphosyntaktische Angaben sinnvoll geschweige denn notwendig sind.

3.1.10 Mediostruktur

Verweise ermöglichen Beziehungen zwischen Einheiten im Wörterbuch, die die sequentielle Anordnung durchbrechen. Sie sind in der Phraseographie ein probates Mittel, um Verbindungen zwischen Phrasemen oder zwischen Phrasemen und Stichwörtern herzustellen. Dabei kommen sie nicht nur innerhalb des Hauptteils des Wörterbuchs vor, auch Register sind Teil der Verweisstruktur.

In erster Linie dienen Verweise der Datenkonsistenz und werden häufig aus ökonomischen Gesichtspunkten der Mehrfachlemmatisierung vorgezogen. Ein ausgefeiltes Vernetzungs- und Verweissystem im Wörterbuch kann aber auch zur Erschließung semantischer phraseologischer Felder genutzt werden (Pilz 1987, S. 140; Schemann 2009, S. 115; Rothkegel 2007, S. 1034; Kjaer 1987, S. 177).

In den vorangegangenen Kapiteln wurden bereits Verweise zwischen folgenden Elementen thematisiert:

- zwischen Einträgen von Varianten
- von Komponenten, die sich nicht als Leitelement qualifizieren, zu Phrasemen
- zwischen semantisch assoziierten Phrasemen (Synonymen, bedeutungsähnlichen Ausdrücken und Antonymen)
- von Phrasemen zu ihrer etymologischen Quelle (meist extern, z. B. Bibelzitate)
- zwischen Phrasemen und ihrer onomasiologischen bzw. semasiologischen Kategorie

Im mehrsprachigen Wörterbuch gibt es überdies die Möglichkeit, Äquivalente durch Verweise zu verknüpfen (Schemann 1991, S. 1028). Für digitale Nachschlagewerke sind auch Querverweise auf Internetressourcen¹⁴ möglich (Ješensek 2009, S. 79) und Schryver (2003, S. 180) prognostiziert gar, dass im elektronischen Wörterbuch alles, was sinnvoll und effektiv verlinkt werden kann, in Zukunft auch verlinkt sein wird. Das betrifft nicht nur die bislang erwähnten internen Verweise, sondern auch externe Verweise auf lexikalische Ressourcen, Korpora, multi-mediale Datenbanken etc..

Verweissysteme sollten vollständig und leicht verständlich sein, sich auf die Gesamtbedeutung des Ausdrucks beziehen und Mehrfachnennungen vermeiden (Pilz 1987, S. 130–137). Im Zusammenhang von Mehrfachlemmatisierung und Verweisen ist auf Einheitlichkeit in der Nennform und innere Konsistenz im Bedeutungsteil zu achten (Korhonen 1990, S. 203; Kühn 2003, S. 112).

3.1.11 Illustrationen

Der Begriff Illustration bezieht sich in dieser Arbeit auf nicht-textliche Elemente, die durch Visualisierung das Verständnis der textuellen Informationen unterstützen oder ergänzen. Pilz (1987, S. 140) befürwortet den Einsatz von Illustrationen (Ikonographie), sofern diese das Verständnis der Bedeutung unterstützen und nicht nur aus ökonomischen Erwägungen eingesetzt werden. Lew und Szczeoaniak (2011, S. 331) unterscheiden für den Bereich Phraseologie in Abhängigkeit davon, welche semantischen Ebenen in der Illustration wiedergegeben werden, vier Arten von Illustrationen:

1. Die Illustration repräsentiert sowohl die wörtliche als auch die übertragene Bedeutung des Idioms.
2. Nur die übertragene Bedeutung ist illustriert.
3. Nur die wörtliche Bedeutung ist illustriert.
4. Die wörtliche Bedeutung wird humoristisch verwertet.

¹⁴ Ješensek (2009, S. 79) nennt als webbasierte Hilfsmittel Wörterbuchsammlungen, Glossare, Textkorpora, fremdsprachendidaktische und andere sprachbezogene Ressourcen.

Eine weitere Möglichkeit der Nutzung von Illustrationen ist die Verbildlichung des Gebrauchs der Phraseme in Situationen, in welchen sie typischerweise Verwendung finden (Földes 1994, S. 176–177).

In einer Benutzerstudie zeigen Lew und Szczeoaniak (2011, S. 341–342), dass die Integration von Illustrationen der wörtlichen Bedeutung von idiomatischen Phrasemen den kurz- und langzeitigen Lernerfolg bei Studenten statistisch signifikant verbessern kann und empfehlen deshalb den Einsatz von bildlichen Komponenten in Wörterbüchern und Lernmaterialien. Aber auch die Visualisierung der metaphorischen Bedeutung findet aktive Befürworter, die sich auf Erkenntnisse der kognitiven Linguistik berufen (Naciscione 2010; Rothkegel 2007, S. 1034).

3.1.12 Sonstiges

Ein weiteres Desideratum ist, Frequenzangaben ins phraseologische Wörterbuch aufzunehmen, um beispielsweise hochfrequente Phraseme zu markieren. So bemängelt Hausmann, dass bereits in FRIEDERICHIDIOMATIK-PW »Geläufiges und Seltenes markierungslos nebeneinander« steht (Hausmann 1983, S. 200). Aus dem Bereich der Übersetzungswissenschaft wird darüber hinaus postuliert, korpusbasiert Häufigkeiten auf Ebene einzelner Merkmale zu bestimmen, um beispielsweise semantische und pragmatische Angaben nach ihrer Vorkommenshäufigkeit zu gewichten oder bestimmten Domänen zuordnen zu können (Koller 1987, S. 113; Bowker 2013, S. 385).

Ilson (1987, S. 199) hält die tabellarische Zusammenfassung von Ausdrücken, die auf derselben konzeptuelle Metapher basieren, für eine sinnvolle illustrative Ergänzung im Wörterbuch. Er schlägt vor, diese Information als nicht-sequentielle textuelle Information am Eintrag des Stichworts der Konzeptbedeutung zu verankern. Moon (2004) beschreibt einen kognitiv-linguistischen Ansatz, bei welchem Idiome im Kontext der dazugehörigen konzeptuellen Metapher (*conceptual metaphor*) kodiert werden. Das Ergebnis wird im MACMILLAN-ENGLISH-DICTIONARY-PW in einer Info-Box präsentiert, wie Abbildung 3.2 illustriert.

Metaphor

An argument is like a **fight** or **war**, with people attacking each other's opinions and defending their own.

*She tried to **defend** herself against his **attacks** on her ideas. ♦ She **shot down** his argument. ♦ That is an **indefensible** point of view. ♦ I decided to pursue another **line of attack**. ♦ We had a big **fight** last night, and I went home early. ♦ There was a lot of **conflict** over what to do next. ♦ They **clashed** over who to appoint. ♦ It was a real **battle of wits**. ♦ We **did battle** with the council about the plans. ♦ I've **crossed swords** with them before. ♦ Tom is always the first to leap to her **defence**.*

→ ANGRY, CRITICIZE, IDEA

Abbildung 3.2: Language awareness: Konzeptuelle Metapher *argument is war* im MACMILLAN-ENGLISH-DICTIONARY-PW

3.1.13 Zusammenfassung

Nur skizzenhaft konnte aufgezeigt werden, welche Postulate die linguistische Forschung an die Phraseographie stellt. So divers die Ansichten sich im Detail präsentieren, kristallisieren sich bestimmte grundlegende Forderungen heraus.

Ein durchgängiges Thema ist die Frage der inneren Kohärenz des Nachschlagewerkes, die sich darin äußert, dass redaktionelle Entscheidungen in den Umtexten dokumentiert sind und im Werk auch konsequent umgesetzt werden. Das beginnt bei der Wahl der Terminologie in Titel und Umtexten und durchdringt alle Teile des Wörterbuchs. Auch wird wiederholt thematisiert, wie wichtig es ist, die Inhalte und Struktur des Wörterbuchs auf die Benutzergruppe abzustimmen. Beispielsweise ist eine Zuordnung von Phrasemen zu Stichworten, die auf semantischen Kriterien beruht, für ein Lernerwörterbuch ungeeignet, da das Wissen um die Bedeutung beim Nutzer nicht als Gegeben vorausgesetzt werden kann. Als drittes allgemeines Prinzip lässt sich aus der Vielfalt der Forderungen extrahieren, dass in jedem Bereich eine exakte Analyse und modulare Realisierung unterschiedlicher Informationssegmente der Kodierung vorangehen müssen, um den unterschiedlichen Anforderungen gerecht werden zu können.

3.2 Phraseographische Praxis in japanischen Printwörterbüchern

In den folgenden Kapiteln werden die für die Phraseographie typischen Vorgehensweise in monolingualen japanischen und bilingualen japanisch-englischen bzw. japanisch-deutschen Wörterbüchern exemplarisch zusammen gefasst. Die Beispiele haben jedoch rein illustrativen

Charakter und erheben keinen Anspruch auf Vollständigkeit. In erster Linie geht es darum, die Methoden der traditionellen Lexikographie im Umgang mit dem Phänomenbereich Phraseologie zu analysieren. Weiterhin sollen die in Konventionen der japanischen Lexikographie herausgearbeitet werden, da diese für die Erwartungshaltung der Wörterbuchbenutzer im Umgang mit *kanyōku* prägend sind. Aufgrund der terminologischen Differenzen wird auch in diesem Kapitel von *kanyōku* die Rede sein, da dies der in der japanischen Lexikographie übliche Begriff ist.

3.2.1 Allgemeine Wörterbücher

Zunächst wird der Frage nachgegangen, wie *kanyōku* in der Regel im allgemeinen einsprachigen japanischen Wörterbuch (*kokugo jiten* 国語辞典) kodifiziert werden. Als allgemein werden Wörterbücher bezeichnet, die sich der Beschreibung des lexikalischen Inventars einer Sprache widmen, in Abgrenzung zu Spezialwörterbüchern, die sich auf eine bestimmte Teilmenge des Wortschatzes konzentrieren, wozu auch phraseologische Wörterbücher gerechnet werden (siehe Kapitel 3.2.2).

Exemplarisch sollen an zwei Werken, dem einbändigen IWANAMI-KŌJIEN⁶ und dem 12-Bände umfassenden SHŌGAKUKAN-NIHONKOKUGODAIJITEN², zwei verschiedene Ansätze im Umgang mit *kanyōku* gezeigt werden. Da die lexikographischen Strategien in Bezug auf die Repräsentation von *kanyōku* nur rudimentär in den Umtexten erklärt werden, basiert die folgende Darstellung überwiegend auf Stichprobenanalysen.

3.2.1.1 Allgemeines

Die Wörterbücher des Iwanami-Verlags genießen traditionell in Japan ein hohes Ansehen und gelten insbesondere in akademischen Kreisen als zuverlässige Quelle. Das allgemeine einsprachige Wörterbuch KŌJIEN des Iwanami-Verlags gilt als »National-Wörterbuch« und erreichte in seiner 6. Auflage (2008) schätzungsweise eine Auflage von 340000 Exemplaren (Asahi Shinbun 2008). Hinsichtlich seines Bekanntheitsgrades ist es in etwa vergleichbar mit dem Duden. Das KŌJIEN Wörterbuch wird auch in diversen elektronischen Formaten angeboten, und steht in Konkurrenz zu SANSEIDŌ-DAIJIRIN-IW und SHŌGAKUKAN-DAIJISEN-PW, auch wenn es um die Ausstattung der extrem populären digitalen Taschenwörterbücher in Japan geht.

Das SHŌGAKUKAN-NIHONKOKUGODAIJITEN² ist das umfassendste einsprachige Wörterbuch zur Gegenwartssprache in Japan. Die insgesamt 13 Bände der zweiten Auflage erschienen zwischen 2000 und 2002. Nach eigenen Angaben umfasst das Werk etwa 88 Millionen Schriftzeichen in 503000 Einträgen. Das Wörterbuch war ursprünglich nur in Printform verfügbar und aufgrund seines Umfangs auf den stationären Gebrauch limitiert. Inzwischen hat eine Online verfügbare elektronische Version die Einsatzmöglichkeiten des Werkes und möglicherweise damit auch seinen Nutzerkreis deutlich erweitert (siehe Kapitel 3.5.2.5).

3.2.1.2 Phraseologieauffassung, Terminologie und Aufbau

Beide Werke halten sich in ihren Umtexten bezüglich ihres Umgangs mit dem Phänomenbereich und der verwendeten Terminologie bedeckt. In den Benutzungshinweisen werden die Begriffe *rengo*, *kanyōku* und *kotowaza* verwendet. Ein Versuch einer Definition, Charakterisierung oder eine Abgrenzung der Begriffe zueinander wird nicht unternommen. Als Hilfestellung für den Benutzer wird lediglich auf die makrostrukturelle Einordnung der *kanyōku* im Wörterbuch eingegangen.

3.2.1.3 Gestaltung der Nennform

Die konventionelle Lemmatisierungsmethode im Japanischen beinhaltet im Wesentlichen die Umwandlung der Flexionsendungen in Verben und Adjektiven in die Finalform (*shūshikei*). Dieses Verfahren wird analog auch für Phraseme angewandt, wobei aufgrund der agglutinierenden Natur der Sprache und der Phrasenstruktur bei verbalen *kanyōku* meist die Endung des letzten Wortes durch die Lemmatisierung berührt wird. Phraseme mit stark ausgeprägter Präferenz zu einer bestimmten Form (z. B. die negierte Form in *hara-no mushi-ga osamaranai*) oder transformationellen Defekten werden von der Lemmatisierung ausgenommen.

Die Markierung der Nennform von *kanyōku* erfolgt im Japanischen meist in ähnlicher Form wie die der Einzelwortlexeme. Allerdings wird bei Einzelwortlexemen im gedruckten Wörterbuch die *kana*-Form (phonetische Form) als Nennform verwendet, auf welche in Klammern die übliche *kanji*- oder *kanji-kana*-Mischform folgt. Bei *kanyōku* wird hingegen in der Regel die *kanji-kana*-Mischform als Lemma-Angabe verwendet und Lesungen werden in Klammern nach den Schriftzeichen (oder als *furigana* 振り仮名 Lesehilfen) angegeben. Das IWANAMI-KŌJIEN⁶ führt nur vereinzelt im Falle von irregulären Lesungen oder schwierigen Schriftzeichen

Lesehilfen an und aus Gründen der Platzersparnis werden die *kana* in halber Größe und in zwei Spalten komprimiert gedruckt. SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ hingegen gibt für alle *kanji* Lesungen in Klammern nach den jeweiligen Schriftzeichen an.

Beide Werke integrieren *kanyōku* als Sublemmata in den Hauptartikel. Bei Betrachtung des IWANAMI-KŌJIEN₆ wird das jedoch deutlicher, da das Lemma des zugehörigen Hauptartikels durch einen Strich ersetzt wird. Dadurch wird die Ausdehnung des Nischenartikels in der Lemmareihe auf den ersten Blick ersichtlich. Nachteilig ist jedoch, dass Sublemmata mental um das Hauptlemma ergänzt werden müssen, da die erste Komponente fehlt.

Das SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ geht einen anderen Weg: Indem es jeweils die erste Komponente in *hiragana* angibt, unabhängig von ihrer gängigen Darstellungsform, versucht es möglicherweise die Orientierung im gesamten Werk zu erleichtern, da durchweg für alle Begriffe, die von der Ordnungsstruktur erfasst sind, eine Umschrift in *kana* erfolgt. In Kombination mit den nachgestellten *hiragana*-Lesehilfen für *kanji* und den Einschüben in eckigen Klammern zur Integration von Varianten in die Nennform (siehe Kapitel 3.2.1.4) entstehen stellenweise Formen, die nicht auf einen Blick erfasst werden können. Positiv fällt auf, dass die Grenzen der variablen Elemente durchweg markiert sind, indem das austauschbare Segment mit einem Gleichheitszeichen (=) eingeleitet wird und die Alternative ebenfalls durch Gleichheitszeichen und Klammerung exakt kenntlich gemacht wird.

はらの虫(むし)が治(おさ)まらない「承知
 (しょうち)しない」 癩(しゃく)にさわって我慢
 ができない。腹立つ心を抑えがたい。*人情本・春の
 若草(1830-4)初・五回「今日の様に言われちゃア何
 程私様な者でも腹の虫が承知為ないヨ」*家族会
 議(1935)〈横光利一〉「計画を狂はしてやらなきア、腹
 の虫が納らないですよ」
 はらの虫(むし)の居所(いどころ)が悪(わる)い
 きげんが悪い。

Abbildung 3.3: Eintrag für *hara-no mushi-ga osamaranai* im SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂

3.2.1.4 Umgang mit Varianten und Synonymen

Auch in Bezug auf den Umgang mit Varianten gibt es grundlegende Unterschiede in den Ansätzen. Im IWANAMI-KŌJIEN₆ werden Varianten am Ende des Phrasemeintrags in der Form *auch [in der Form]* angehängt. Verweise auf synonyme Einträge stehen am Ende des Artikels und werden durch einen Pfeil markiert. Das SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ arbeitet vergleichsweise selten mit Verweisen. Varianten werden statt dessen häufig als eigene Einträge realisiert, wenn sie sich nicht in die Lemma-Angabe integrieren lassen.¹⁵ Sporadisch finden sich Angaben zu antonymen Formen¹⁶ in Bedeutungserklärungen sowie Angaben zur Herkunft,

¹⁵ Beispiele für die Realisierung von Varianten als separate Lemmata sind *hara-ga tatsu* (腹が立つ), *hara-o tateru/tatsu* (腹を = 立てる [= 立つ]) und *hara-ga nietatsu* (腹が煮え立つ).

¹⁶ Beispiele für antonyme Formen sind *hara-ga iru* (腹が居る) für *sich beruhigen* als Gegenstück zu *hara-ga tatsu* im SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂.

insbesondere, wenn es sich um Ableitungen handelt.¹⁷ Eine Markierung des Verweises erfolgt überwiegend bei leeren Einträgen, die rein aus Verweisen bestehen¹⁸ Insgesamt kann man sagen, dass das SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ aufgrund seines Umfangs eine Vielzahl an unterschiedlichen Formen erfasst, welchen es überwiegend eigenständige Artikel widmet. Eine weitere Besonderheit des Werkes ist, dass darin auch regionale *kanyōku* berücksichtigt werden. Darüber hinaus existieren in beiden Wörterbüchern Verweise, im IWANAMI-KŌJIEN₆ sind sie jedoch präsenter und werden häufiger als reine Verweiseinträge ohne eigenen Inhalt realisiert.

3.2.1.5 Makrostrukturelle Zuordnung

In beiden Nachschlagewerken werden die *kanyōku* in einem separaten Block an das Ende des Artikeltextes gestellt. Das bedeutet, dass eine Zuordnung zu einer Lesart bei polysemen Haupteinträgen nicht erfolgt. Die Sublemmata unterscheiden sich durch typographische Merkmale¹⁹ von den Hauptlemmata und Komposita. Die Anordnung der *kanyōku* zueinander richtet sich analog zur Anordnung der Hauptlemmata zuvorderst nach der 50-Laute-Tafel.

Obwohl beide einsprachigen Wörterbücher *kanyōku* als Unterartikel erfassen, die auf den Hauptartikel der ersten Komponente folgen, unterscheidet sich die Makrostruktur bei genauem Hinsehen deutlich, wie am Beispiel von *hara* illustriert werden kann.

¹⁷ Ein Beispiel für Angaben zur Herkunft findet sich unter *hara-ga kitayama* (腹が北山 wörtl.: der Bauch ist Nordberg) für *hungrig sein*, welches von *hara-ga kita* (腹が来た wörtl.: der Bauch ist gekommen) abgeleitet wurde.

¹⁸ Ein Beispiel für einen reinen Verweis findet sich bei *hara-ga itashi* (腹が痛し) ⇒ *hara-ga itai* (はら (腹) が痛い wörtl.: der Bauch schmerzt) für *sehr komisch sein*.

¹⁹ In IWANAMI-KŌJIEN₆ und SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ werden *kanyōku* und *kotowaza* durch kleinere Schriftart und gothischen Schriftstil markiert.

IWANAMI-KŌJIEN-PW:

*hara ~ hara-no mushi**hara* (腹 Bauch)- *ga ieru* (~ 癒える)- *ga itamu* (~ が痛む)- *ga iru* (~ が居る)

[...]

- *o waru* (~ を割る)*hara* (大角)*hara* (爬羅)*bara* (散)

[...]

haraaka (腹赤)*haraashi* (腹悪し)*haraashige* (腹葦毛)

[...]

hara-no fue (大角)*hara-no mushi* (腹の虫)- *ga osamaranai* (~ がおさまらない)*harabai* (腹這い)

SHŌGAKUKAN-NIHONKOKUGODAIJITEN-PW:

*hara ~ hara-no mushi**hara* (腹 Bauch)*haraashi* (腹悪し)*haraashige* (腹悪しげ)*hara-ga ii* (はらがいい)*hara-ga ieru* (腹が癒える)*hara-ga itai* (腹が痛い)*hara-ga iru* (腹が居る)

[...]

hara-no mushi (腹の虫)*hara-no mushi ga-osamaranai* (腹の虫が治まらない)*hara-no mushi-no idokoro-ga warui* (腹の虫の居所が悪い)

[...]

hara-o waru (腹を割る)*hara* (大角)*hara* (爬羅)*bara* (散)

[...]

haraaka (腹赤)*haraashige* (腹葦毛)*haraashige* (腹葦毛)

[...]

hara-no fue (大角)*hara-no mushi* (腹の虫)- *ga osamaranai* (が治まらない)

Abbildung 3.4: Vergleich Makrostruktur des Lemmas *hara* in IWANAMI-KŌJIEN₆ und SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂

Abbildung 3.4 zeigt die Unterschiede anhand einzelner Markierungspunkte: Im IWANAMI-KŌJIEN₆ sind *kanyōku* mit *hara* (腹) als erster Komponente zwar an den Haupteintrag *hara* (腹 Bauch) gekoppelt, *hara-no mushi* und *haraashi* (腹悪し boshaft) werden jedoch nicht als Sublemmata sondern als eigenständige Haupteinträge modelliert und dadurch unterliegen sie der Hauptordnungsstruktur. Dementsprechend ist *hara-no mushi-ga osamaranai* im IWANAMI-KŌJIEN₆ ein Untereintrag von *hara-no mushi* (腹の虫 wörtl.: Bauchwürmer *schlechte Laune*) und nicht wie im SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ von *hara*. Dort werden *haraashi* und *hara-no mushi* mit allen anderen *kanyōku* einer Sublemmareihe *hara* zugeordnet. Obwohl die

Regeln für die makrostrukturelle Ordnung eindeutig und die zu Grunde liegende Phraseologieauffassung gleich zu sein scheinen, führen Inkonsistenzen im Umgang mit Komposita in *kanyōku* zu deutlichen Unterschieden in der lexikographischen Praxis.

3.2.1.6 Mikrostruktur

In keinem der beiden Werke enthalten die als Untereinträge verankerten Phrasembeschreibungen Angaben zur Form. Es wird weder zwischen den in den Benutzungshinweisen erwähnten Phrasemtypen *kanyōku* und *kotowaza* unterschieden, noch beinhalten die Werke explizite Information zu morpho-syntaktischen Besonderheiten des Phrasems (z. B. Valenz und Restriktionen). Der Schwerpunkt liegt in beiden Werken auf der Bedeutungserklärung, innerhalb derer mit Hilfe von Nummerierung unterschiedliche Lesarten markiert werden. Inhaltlich fokussiert IWANAMI-KŌJIEN₆ idiomatische Lesarten, kompositionelle bleiben meist unerwähnt. So ist beispielsweise bei *me-o tsuburu* (目を瞑る) die wörtliche Bedeutung *blinzeln* nicht hinterlegt, die übertragenen Bedeutungen 1. *schlafen* 2. *sterben* 3. *etwas willentlich übersehen* 4. *ertragen, resignieren* durchaus. Im SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ beginnt die Bedeutungserklärung mit der faktischen Bedeutung *me-o tojiru* (目を閉じる die Augen schließen), bevor zu den übertragenen Lesarten übergegangen wird.²⁰

Strukturell besteht die Bedeutungsangabe für *kanyōku* meist aus einer Mischung von Definitionen, synonymen Einzelwortlexemen und paraphrasierenden Erklärungen. Sofern vorhanden werden auch synonyme *kanyōku* eingesetzt. Eine ausschließliche Erklärung durch ähnliche idiomatische Wendungen oder Metaphern (Remotivierung) ist relativ selten²¹. In beiden Werken wird ein hohes sprachliches Niveau voraus gesetzt, wie aus dem sparsamen Umgang mit Lesehilfen und der Wortwahl hervorgeht.

Neben der Bedeutungserklärung finden sich in den Untereinträgen des IWANAMI-KŌJIEN₆ vereinzelt auch Erläuterungen zur typischen Verwendung, die einer Angabe von Kollokationen

²⁰ Ein punktueller Vergleich der Kodierung von *gestural idioms* in Wörterbüchern zeigt, dass die Konkurrenzprodukte SANSEIDŌ-DAIJIRIN-IW und SHŌGAKUKAN-DAIJIEN-PW im Vergleich zu IWANAMI-KŌJIEN₆ mehr Wert auf die Erläuterung der wörtlichen Bedeutungsebene legen (Vogel und Kuchikata im Druck).

²¹ In der Regel ist die Qualität der Wörterbucheinträge relativ hoch, jedoch gibt es auch Ausnahmen. Beispielsweise wird das sehr gängige Idiom *hara-ga tatsu* (腹が立つ *sich ärgern*) im IWANAMI-KŌJIEN₆ mit Hilfe von *shaku-ni sawaru* (しゃくにさわる wörtl.: *Bauchkrämpfe haben*) erklärt.

gleich (siehe Beispiel 21). Flektiert das Phrasem nicht, wird es in der Regel wie in der Sublemma-Angabe durch einen Längenstrich ersetzt.

- (21) 腹を決める 決意する 「引退の一」
hara-o kimeru, ketsui suru »intai-no –«
sich entschließen, einen Entschluss fassen, »Rücktritt –«

Im SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ gibt es keine derartigen Anwendungsbeispiele, statt dessen werden für jede Lesart nach Möglichkeit Erstbelege zitiert. Das IWANAMI-KŌJIEN₆ führt nur für ungewöhnliche beziehungsweise archaische Lesarten Originalbelege an.

3.2.1.7 Zusammenfassung

Das Kapitel zeigt exemplarisch die lexikographische Erfassung von *kanyōku* in zwei renommierten allgemeinen Wörterbüchern des Japanischen. Aufgrund ihrer morphosyntaktischen Struktur ist die makrostrukturelle Einordnung weniger problembehaftet als beispielsweise im Deutschen, dennoch kristallisieren sich bei genauem Hinsehen deutliche Unterschiede in der Anordnung der *kanyōku* heraus. Dies scheint im Wesentlichen durch Unterschiede in der Priorisierung unterschiedlicher Phänomene (Wortbildung und Phraseologie) begründet zu sein. Aufgrund der fehlenden Angaben zu grammatischen und pragmatischen Merkmalen der *kanyōku* eignen sich beide Werke vorwiegend für die Textrezeption, das SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ insbesondere auch für die Rezeption historischer Texte. Auch setzen beide ein hohes sprachliches Niveau voraus und bieten nur spärlich Unterstützung bei der Lesung von *kanji*. Positiv hervorzuheben ist der durchweg systematische Umgang mit *kanyōku*.

3.2.2 Phraseologische Wörterbücher

Im Folgenden werden drei einsprachige phraseologische Spezialwörterbücher des Japanischen vorgestellt, das YONEKAWA-NIHONGO KANYŌKUJITEN-PW, das SŌTAKUSHA-REIKAI KANYŌKUJITEN-PW und das TANNO-KANYŌKUJITEN-PW. Diese Werke wurden ausgewählt, weil sie sich aktiv damit auseinandersetzen, wie man den Besonderheiten der *kanyōku* mit lexikographischen Mitteln gerecht werden kann. Eine Fülle von Nachschlagewerken muss aufgrund des begrenzten Umfangs der vorliegenden Arbeit unberücksichtigt bleiben. Dies gilt insbesondere für die Werke, die sich mit der Phraseologie in ihrer Gesamtheit und nicht mit *kanyōku* im Besonderen befassen.

Von den drei hier skizzierten Werken ist das YONEKAWA-NIHONGO KANYŌKUJITEN-PW das einzige klassisch semasiologisch (nach der 50-Laute-Tafel) konzipierte. Das Werk erschien im Jahre 2005 (3. Auflage 2007) und beinhaltet auf etwas mehr als 500 Seiten nach eigenen Angaben 1.563 Phraseme. Die vergleichsweise geringe Anzahl an Einträgen in Relation zur physischen Größe des Werkes ist auf eine besonders elaborierte Mikrostruktur zurückzuführen, welche in Kapitel 3.2.2.5 eingehender untersucht wird.

Die beiden onomasiologischen Wörterbücher weisen auf die Besonderheit ihres Aufbaus mit Hilfe von Titelzusätzen hin. Das 1992 (11. Auflage im Jahre 2000) im Sōtakusha Verlag erschienene SŌTAKUSHA-REIKAI KANYŌKUJITEN-PW trägt den Untertitel *iitai naiyō kara gyakubiki dekiru* (言いたい内容から逆引きできる mit der Möglichkeit nach dem nachzuschlagen, was man Sagen möchte) und präsentiert auf ebenfalls gut 500 Seiten 3.700 Phrasemeinträge, also mehr als das Doppelte des YONEKAWA-NIHONGO KANYŌKUJITEN-PW. Das *reikai* im Titel weist außerdem darauf hin, dass auf Bedeutungserläuterung anhand von Satzbeispielen Wert gelegt wird. Das zweite onomasiologisch konzipierte Werk ist das 1998 im Jitsugyōsha Verlag erschienene TANNO-KANYŌKUJITEN-PW, das auf knapp 300 Seiten etwa 1100 *kanyōku* behandelt und ebenfalls im Titel damit wirbt, dass man von der Bedeutung ausgehend auf den Inhalt zugreifen kann (*imi kara hikeru kanyōku jiten* 意味から引ける慣用句辞典 Idiomwörterbuch, mit dem man von der Bedeutung ausgehend nachschlagen kann).

3.2.2.1 Phraseologieauffassung, Terminologie, Inhalt und Aufbau

In der Einleitung des YONEKAWA-NIHONGO KANYŌKUJITEN-PW wird dessen Entstehungsgeschichte in direkten Zusammenhang mit der phraseologischen Forschung von Miyaji gestellt, was erklärt, warum der phraseologischen Theorie relativ viel Platz eingeräumt wird. Die verwendete Terminologie und Phraseologieauffassung wurde bereits in Kapitel 2.1.2 ausgiebig erläutert, weshalb an dieser Stelle nicht weiter darauf eingegangen werden soll. Als Zielgruppe werden in der Einleitung Lerner von Japanisch als Fremdsprache sowie japanische Oberschüler genannt. Im Anschluss an den Hauptteil werden auf nahezu siebzig Seiten (S. 537–562) Besonderheiten und verschiedene Typen von *kanyōku* erläutert. Außerdem bietet der Anhang einen Komponentenindex, mittels welchem die *kanyōku* anhand ihrer Bestandteile referenziert werden können.

Das TANNO-KANYŌKUJITEN-PW definiert *kanyōku* in der Einleitung als zwei oder mehr Wörter, die zusammen eine *neue* Bedeutung ergeben. Auf dem Einband ist auch von *jōtōku* (常套句)

und *kimari monku* (決まり文句) die Rede, also zwei verschiedenen umgangssprachlichen Begriffen aus dem Wortfeld Phrasem, die sich in etwa mit »feste Phrase« übersetzen lassen²². Die Nutzung des Wörterbuchs soll laut Vorwort eine »Bereicherung des sprachlichen Ausdrucks« ermöglichen (Vorwort, erste Seite ohne Seitenzahl). Auf die Benutzungshinweise folgt ein Inhaltsverzeichnis, das in fünf Kapiteln, die den übergeordneten Themen entsprechen, die Kategorien, Unterkategorien und Phraseme mit Seitenzahlen vorstellt. Nach dem Hauptteil bietet ein Gesamtindex (nach 50-Laute-Tafel sortiert) die Möglichkeit der semasiologischen Nutzung. Auch das SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW strebt danach, den Ausdruck seiner Nutzer lebendiger und reicher zu machen, indem von der Bedeutung ausgehend ein passender Ausdruck gesucht werden kann (Vorwort S. i). Phraseologieauffassung oder terminologische Fragen finden in der Einleitung keine Erwähnung. Auf das Vorwort folgen Benutzungshinweise und zwei Stichwortverzeichnisse, darunter ein thematisches (vergleichbar mit dem Inhaltsverzeichnis im TANNO-KANYŌKUJITEN-PW, jedoch ohne Aufzählung der Phraseme) und ein nach der 50-Laute-Tafel geordnetes Vollverzeichnis aller Stichworte. Nach dem Hauptteil folgen zwei weitere Verzeichnisse, die die *kanyōku* einmal initialalphabetisch und einmal rückläufig (nach letzter Komponente) auflisten.

3.2.2.2 Gestaltung der Nennform

Die Nennform wird in allen drei Nachschlagewerken in *kanji* angegeben. Kleingedruckte *furigana* rechts der chinesischen Schriftzeichen dienen als Lesehilfe und werden unabhängig vom Schwierigkeitsgrad der *kanji* vorgehalten. Aufgrund ihrer geringen Größe eignen sie sich kaum als Orientierungshilfe.

²² Der Unterschied zwischen *jōtōku* und *kimari monku* konnte anhand der vorliegenden Materialien nicht erfasst werden, da es sich um umgangssprachliche Begriffe handelt, die sich nicht eindeutig auf die linguistische Terminologie abbilden lassen. Im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW wird *jōtōku* mit *kimari monku* gleich gesetzt. Die Definitionen fokussieren den Aspekt der Festigkeit, den man im Deutschen mit »abgedroschen« umschreiben würde (*shinsenmi-no nai hyōgen* 新鮮味のない表現 nicht originelle Phrase).

3.2.2.3 Umgang mit Varianten

Varianten werden nur im YONEKAWA-NIHONGO KANYŌKUJITEN-PW systematisch kodiert. Analog zu Synonymangaben werden variante Formen im Artikeltext verzeichnet und im alphabetischen Index im Anhang indiziert. In den anderen beiden Werken werden sie nur sporadisch erwähnt. Allen drei Werken ist gemein, dass nahezu ausschließlich lexikalische Varianten Eingang finden und diese als Synonyme im Bereich *ruiku* (類句 ähnliche Phrasen) angeführt werden. Die Kennzeichnung mit einem Asterix in diesem Bereich bedeutet, dass das *kanyōku* mit einem separaten Eintrag im Wörterbuch verzeichnet ist. Ein Asterix im Index des SŌTAKUSHA-REIKAI KANYŌKUJITEN-PW weist darauf hin, dass es sich um eine Variante handelt, welche unter einem anderen Stichwort zu finden ist (vgl. Abbildung 3.5).

□を酸っぱくする	* 183
□を滑らす	* 208
□を揃 <small>そろ</small> える	209
□を出す	226
□を叩 <small>たた</small> く	193
□を衝 <small>つ</small> いて出る	209
□を噤 <small>つぶ</small> む	* 167
□を尖 <small>とが</small> らす	77

Abbildung 3.5: Kennzeichnung der Varianten im Index des SŌTAKUSHA-REIKAI KANYŌKUJITEN-PW

3.2.2.4 Makrostrukturelle Zuordnung

Das semasiologische YONEKAWA-NIHONGO KANYŌKUJITEN-PW nimmt die Gesamtzeichenkette des *kanyōku* in seiner Grundform und ordnet diese nach der 50-Laute-Tafel. Da im Japanischen Synsemantika und Platzhalter in den Lemmata eine untergeordnete Rolle spielen (vergleiche Kapitel 4.9.2), ist gegen die exhaustive Methode nichts einzuwenden. Ausnahmen werden in den Umtexten nicht thematisiert und in Bezug auf die Beispielmenge wurde konsequent nach der Systematik verfahren.

Im Unterschied zu semasiologischen Wörterbüchern, in welchen von der Form zum Inhalt beziehungsweise der Bedeutung geführt wird, wird in onomasiologischen Wörterbüchern versucht, dem Benutzer die Möglichkeit zu geben, von der Bedeutung ausgehend ein *kanyōku* zu finden.

Onomasiologische Nachschlagewerke sind meist für die Textproduktion konzipiert. Um den Zugriff über die Bedeutung zu ermöglichen, muss diese abstrahiert und in eine nachvollziehbare Ordnung gebracht werden. Mit unterschiedlichen Herangehensweisen versuchen Wörterbücher die Realität auf der Grundlage ihrer Bedeutung in Konzepte zu unterteilen, um diese in Beziehung zueinander zu setzen. Taxonomien verwenden dabei ausschließlich Teil-Ganzes-Relationen während Ontologien eine Vielfalt an Relationstypen erlauben. Ein Beispiel für ein ontologisches Vorgehen ist Rogets Systematisierung²³, in welchem Schweiß mit (menschlicher) Anstrengung oder Arbeit in Verbindung gebracht wird und nicht wie im Zusammenhang mit Einordnungen, die auf abstrakten beziehungsweise logischen Klassenkonzepten basieren, mit anderen Körperflüssigkeiten.

Die Hauptschwierigkeit für den Benutzer des Nachschlagewerks ist, die vom Lexikographen zu Grunde gelegten Ordnungskriterien nachzuvollziehen. Wünschenswert wäre, dass sie offensichtlich und intuitiv stimmig sind, denn die wenigsten Nutzer wollen zunächst die Benutzungshinweise studieren, wenn sie ein Nachschlagewerk konsultieren. Die grundlegende Frage, wie sich Wörter zueinander verhalten, welche Gemeinsamkeiten und Unterschiede sie ausweisen, oder genauer gesagt, wie man die semantischen Relationen zwischen ihnen exakt definieren kann, lässt sich bis in die Antike zurückverfolgen.²⁴ Bis heute gibt es hierfür keine einheitliche Vorgehensweise, da sich das Sprachgefühl der individuellen Sprecher einer Sprachgemeinschaft in den Nuancen immer unterscheidet.

Eine Möglichkeit, dieses Problem zumindest teilweise zu umgehen, ist die Anordnung der Begriffe in einer Liste, zum Beispiel einer alphabetisch geordneten, ohne hierarchische Struktur. Das bedeutet jedoch, dass dem Nachschlagenden der Begriff, unter welchem ein Ausdruck verzeichnet ist, kennen muss. Eine stufenweise Annäherung an den gesuchten Begriff ist hier nicht möglich. Häufig trifft man auf Mischformen. So wird in den vorliegenden Werken auf oberster Ebene die Welt in eine begrenzte Anzahl von abstrakten Konzepten eingeteilt, die unterschiedlich viele Unterebenen umfassen, deren Kategorien intern alphabetisch (beziehungsweise nach der 50-Laute-Tafel) geordnet sind.

²³ Im Englischen hat sich hingegen der Begriff *Thesaurus* eingebürgert, der auf Mark Rogets (1779-1868) Synonymwörterbuch mit dem Titel *Thesaurus of English Words* (1852) Bezug nimmt.

²⁴ Für eine Geschichte der onomasiologischen Lexikographie siehe Sterkenburg (2003).

Die zwei phraseologischen Wörterbücher mit onomasiologischer Zugriffsstruktur untergliedern den Objektbereich in fünf Hauptkategorien (Themen).

SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW	TANNO-KANYŌKUJITEN-PW
1. <i>kankaku</i> (感覚 Empfindung), <i>kanjō</i> (感情 Gefühl)	1. <i>kōi</i> (行為 Handlung), <i>taido</i> (態度 innere Einstellung)
2. <i>karada</i> (からだ Körper), <i>seikaku</i> (性格 Charakter), <i>taido</i> (態度 innere Einstellung)	2. <i>kankaku</i> (感覚 Empfindung), <i>kanjō</i> (感情 Gefühl)
3. <i>kōi</i> (行為 Handlung), <i>dōsa</i> (動作 Bewegung), <i>kōdō</i> (行動 Tun)	3. <i>jinsei</i> (人生 Leben), <i>ningen kankei</i> (人間関係 zwischenmenschliche Beziehungen), <i>seikatsu</i> (生活 Alltag)
4. <i>jōtai</i> (状態 Zustand), <i>teido</i> (程度 Ausmaß), <i>kachi</i> (価値 Wert)	4. <i>shigoto</i> (仕事 Arbeit), <i>shokuba</i> (職場 Arbeitsplatz)
5. <i>shakai</i> (社会 Gesellschaft), <i>bunka</i> (文化 Kultur), <i>seikatsu</i> (生活 Alltag)	5. <i>seiji</i> (政治 Politik), <i>keizai</i> (経済 Wirtschaft), <i>shakai</i> (社会 Gesellschaft), <i>bunka</i> (文化 Kultur), <i>shizen</i> (自然 Natur)

Abbildung 3.6: Gegenüberstellung der Themen im SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW und TANNO-KANYŌKUJITEN-PW

Abbildung 3.6 zeigt, dass hinsichtlich der thematischen Gliederung nur in einem Punkt Übereinstimmung herrscht. Das Thema KANKAKU, KANJŌ (感覚 · 感情 *Empfindung*²⁵, Gefühl) ist sowohl im SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW als auch TANNO-KANYŌKUJITEN-PW auf oberster Ebene angesiedelt. Die erste Kategorie im TANNO-KANYŌKUJITEN-PW mit dem Titel KŌI, TAIDO (行為 · 態度 *Handeln, Einstellung*) ist ein Konglomerat der Kategorien 2 KARADA, SEIKAKU, TAIDO (からだ · 性格 · 態度 *Körper, Charakter, Einstellung*) und 3 KŌI, DŌSA,

²⁵ Von der Übersetzung von KANKAKU mit *Sinneswahrnehmung* wurde abgesehen, da deren typische Vertreter *sehen* (見る *miru*) und *hören* (聞く *kiku*) in den Kategorien um das Wortfeld TAT/HANDLUNG/VERHALTEN angesiedelt sind.

kŌDŌ (行為・動作・行動 *Handeln, Bewegung, Tun*) des SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW. Auch in den Kategorien 3 JINSEI, NINGEN KANKEI, SEIKATSU (人生・人間関係・生活 *Leben, zwischenmenschliche Beziehungen, Alltag*) und 5 SEIJI, BUNKA, SHIZEN (政治・文化・自然 *Politik, Kultur, Natur*) des TANNO-KANYŌKUJITEN-PW gibt es Teilüberschneidungen mit Kategorie 5 SHAKAI, BUNKA, SEIKATSU (社会・文化・生活 *Gesellschaft, Kultur, Alltag*) des SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW. Die weiteren Kategorien weisen nur punktuell Ähnlichkeiten miteinander auf und sind deshalb kaum vergleichbar.

SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW			TANNO-KANYŌKUJITEN-PW		
Kategorie	Lesung	Deutsch	Kategorie	Lesung	Deutsch
愛情	<i>aijō</i>	Liebe	愛する、恋しい	<i>aisuru, koishii</i>	lieben, lieb
あきらめる、 断念	<i>akirameru, dannan</i>	Aufgeben, Verzicht	あきらめる	<i>akirameru</i>	Aufgeben
あきれる	<i>akireru</i>	staunen	あきれる	<i>akireru</i>	staunen
焦る、焦慮	<i>aseru, shōryo</i>	eilen, Hast	あせる	<i>aseru</i>	eilen
安心	<i>anshin</i>	Beruhigung	安心する	<i>anshin suru</i>	beruhigt sein
怒る、怒り	<i>okoru, ikari</i>	sich ärgern, Zorn	怒る	<i>okoru</i>	sich ärgern
疑う	<i>utagau</i>	Zweifel	疑う、疑われる	<i>utagau, utagawareru</i>	zweifeln, angezweifelt werden
恐れる、恐ろしい	<i>osoreru, osoroshii</i>	Fürchten, fürchterregend	恐ろしい	<i>osoroshii</i>	furchterregend
驚く、驚かす	<i>odoroku, odorokasu</i>	sich erschrecken, jemanden erschrecken	驚く、驚かす	<i>orodoku, odorokasu</i>	sich erschrecken, jemanden erschrecken
記憶	<i>kioku</i>	Erinnerung	覚える	<i>oboeru</i>	sich erinnern
覚悟、決心	<i>kakugo, kesshin</i>	Resignation, Entschlossen- heit	覚悟、決心する	<i>kakugo, kesshin suru</i>	Resignation, entschlossen sein
感謝	<i>kansha</i>	Dankbarkeit	感謝	<i>kansha</i>	Dankbarkeit

3 Phraseologie in Theorie und Praxis

Kategorie	Lesung	Deutsch	Kategorie	Lesung	Deutsch
感動	<i>kandō</i>	Rührung, Ergriffenheit	感動する	<i>kandō suru</i>	gerührt sein
気配り	<i>kikubari</i>	Fürsorge	気配り、配慮	<i>kikubari,</i> <i>hairyo</i>	Fürsorge, Sorge
希望	<i>kibō</i>	Wunsch	希望がある	<i>kibō ga aru</i>	einen Wunsch hegen
嫌う、いやがる	<i>kirau, iyagaru</i>	hassen, verachten	嫌い	<i>kirai</i>	hassen
安楽・気楽	<i>anraku, kiraku</i>	Gemütlichkeit, Behaglichkeit	気楽・気軽	<i>kiraku, kigaru</i>	Gemütlichkeit, Unbeschwert- heit
緊張	<i>kinchō</i>	Nervosität	緊張する	<i>kinchō suru</i>	nervös sein
後悔、悔やむ、悔しい	<i>kōkai, kuyamu,</i> <i>kuyashii</i>	Reue, Bedauern, bedauerlich	くやしい	<i>kuyashii</i>	bedauerlich
恋う、恋しい	<i>kou, koishii</i>	lieben, lieb	恋する	<i>koisuru</i>	lieben
興奮	<i>kōfun</i>	Aufregung	興奮する	<i>kōfun</i>	aufgeregt sein
親しい	<i>shitashii</i>	vertraut	親しい	<i>shitashii</i>	vertraut
愉快	<i>yukai</i>	Spaß	すっきりする	<i>sukkiri suru</i>	sich erfrischt fühlen
尊敬	<i>sonkei</i>	Respekt	尊敬する	<i>sonkei suru</i>	respektieren
楽しい	<i>tanoshii</i>	lustig	楽しい	<i>tanoshii</i>	lustig
同情	<i>dōjō</i>	Mitgefühl	同情する	<i>dōjō suru</i>	mitfühlen
泣く	<i>naku</i>	weinen	泣く	<i>naku</i>	weinen
納得	<i>nattoku</i>	Einverständnis	納得	<i>nattoku</i>	Einverständnis
憎む、憎い	<i>nikumu, nikui</i>	hassen, verhasst	憎む、憎い	<i>nikumu, nikui</i>	hassen, verhasst
恥ずかしい	<i>hazukashii</i>	peinlich	恥ずかしい	<i>hazukashii</i>	peinlich
欲望、欲求、貧欲	<i>yokubō,</i> <i>yokkyū,</i> <i>donyoku</i>	Begierde, Verlangen, Bedürfnis	ほしい	<i>hoshii</i>	Verlangen
満足	<i>manzoku</i>	Zufriedenheit	満足	<i>manzoku</i>	Zufriedenheit

Kategorie	Lesung	Deutsch	Kategorie	Lesung	Deutsch
やけ、無分別	<i>yake,</i> <i>mufunbetsu</i>	Verzweiflung, Unbesonnen- heit	やけになる	<i>yake ni naru</i>	verzweifelt sein
憂鬱	<i>yūutsu</i>	Depression	憂鬱	<i>yūutsu</i>	Depression
記憶	<i>kioku</i>	Erinnerung	忘れない	<i>wasurenai</i>	nicht erinnern
笑う	<i>warau</i>	lachen	笑う	<i>warau</i>	lachen

Tabelle 3.1: Vergleichbare Kategorien des Begriffsfelds EMOTION/GEFÜHL in SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW und TANNO-KANYŌKUJITEN-PW

Vergleicht man exemplarisch die Kategorisierung des Begriffsfeldes EMPFINDUNG/GEFÜHL (*kan-kaku, kanjō*) in den beiden Wörterbüchern (siehe Tabelle 3.1), wird unmittelbar deutlich, dass auf Basis der verwendeten Bezeichnungen für Kategorien nicht in jedem Fall eine eindeutige Zuordnung möglich ist. Konnte auf Begriffsebene keine Übereinstimmung hergestellt werden, wurde anhand des Inhalts auf die ähnlichste Kategorie geschlossen.

Zunächst fällt auf, dass das SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW insgesamt den Teilbereich in nur 55 Kategorien untergliedert, während das TANNO-KANYŌKUJITEN-PW 75 Kategorien vorsieht. Anhand ihrer Stichworte oder ihres Inhaltes konnten 37 Kategorien als (quasi) vergleichbar eingestuft werden), wobei das SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW in der Benennung der Kategorien den Nominalstil bevorzugt, das TANNO-KANYŌKUJITEN-PW indes zum Verbalstil tendiert. Nicht in allen Fällen sind die durch den Vergleich der Inhalte der Kategorien erzielten Äquivalenzen auf der Oberfläche nachvollziehbar.

Dass das TANNO-KANYŌKUJITEN-PW insgesamt mehr Kategorien führt, liegt darin begründet, dass zwölf Einträge Querverweise sind. Dabei werden nicht nur bedeutungsähnliche Begriffe zusammengefasst, die Verweise ermöglichen den direkten Zugriff auf Begriffe, die nicht an erster Stelle genannt werden und solche, die anderen Themen zugeordnet sind. Im SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW wird diese Funktionalität durch einen separaten Stichwortindex zur Verfügung gestellt.

- *unubore* (自惚れ Eitelkeit) ⇒ *jiman*
- *urayamashii* (うらやましい beneidenswert) ⇒ *hoshii*
- *ureshii* (うれしい) froh ⇒ *tanoshii*
- *omoshiroi* (面白い interessant) ⇒ *tanoshii*

- *oboeru* (覚える) sich erinnern ⇒ *wasurenai*
- *kawaii* (可愛い) liebenswert ⇒ *kodomo-ga kawaii*²⁶
- *koisuru* (恋する) lieben ⇒ *aisuru, koishii*
- *keifuku suru* (敬服する) aufsehen ⇒ *sonkei*
- *shinpai suru* (心配する) sich sorgen ⇒ *nayamu, shinpai suru*
- *tsurai* (つらい) hart ⇒ *kurushii, tsurai*
- *nikumu* (憎む) hassen ⇒ *uramu, nikumu*
- *rakutan suru* (落胆する) entmutigt sein ⇒ *gakkari suru*

Folgende Kategorien führt das SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW zusätzlich an:

- *konomu, konomi* (好む, 好み) mögen, Vorliebe)
- *yōkyū, yōbō* (要求, 要望) Wunsch, Verlangen)
- *magokoro* (真心) Aufrichtigkeit)
- *niou, nioi* (におう, 匂い) verdächtig sein)
- *hiiki* (ひいき) Vorliebe)

Teilweise sind die Begriffsfelder unterschiedlich gegeneinander abgegrenzt. Tendenziell wird hierbei im SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW in einer Kategorie zusammengefasst, was im TANNO-KANYŌKUJITEN-PW in mehrere Unterkategorien aufgeteilt wird (vgl. Tabelle 3.2).

YONEKAWA-NIHONGO KANYŌKUJITEN-PW			TANNO-KANYŌKUJITEN-PW		
Kategorie	Lesung	Deutsch	Kategorie	Lesung	Deutsch
落胆・ がっかり	<i>rakutan,</i>	Entmutigung,	落胆	<i>rakutan</i>	Entmutigung
	<i>gakkari</i>	Enttäuschung	がっかりする	<i>gakkari</i>	Enttäuschung
つらい、悲し い、	<i>tsurai,</i>	hart, traurig,	悲しい	<i>kanashii</i>	traurig
	<i>kanashii,</i>				
苦しい	<i>kurushii</i>	bitter	苦しい、辛い	<i>kurushii, tsurai</i>	hart, bitter
心配、悩み、 不安	<i>shinpai,</i>	Sorge,	心配する	<i>shinpai suru</i>	sich sorgen
	<i>nayami,</i>	Kummer,			
	<i>fuan</i>	Unsicherheit	悩む、心配す る	<i>nayamu,</i> <i>shinpai suru</i>	sich sorgen

²⁶ »Niedlichkeit von Kindern« in der Kategorie »Leben«

3 Phraseologie in Theorie und Praxis

Kategorie	Lesung	Deutsch	Kategorie	Lesung	Deutsch
不快、不愉快	<i>fukai, fuyukai</i>	Missfallen, Unannehm- lichkeit	不機嫌、不愉快	<i>fukigen, fuyukai</i>	Missfallen, Unannehm- lichkeit
不機嫌	<i>fukigen</i>	Verdruss			

Tabelle 3.2: Unterschiede in der Kategorisierung des Begriffsfelds EMOTION/GEFÜHL in SÖTAKUSHA-REIKAIKAN-YÖKUJITEN-PW und TANNO-KANYÖKUJITEN-PW

Eine beachtliche Anzahl von Kategorien, die im SÖTAKUSHA-REIKAIKAN-YÖKUJITEN-PW zu dem Themenbereich KANJŌ/KANKAKU gerechnet werden, finden sich im TANNO-KANYÖKUJITEN-PW in anderen Themenkomplexen wieder. Die Nähe zu dem Bereich KÖRPER/CHARAKTER/EINSTELLUNG, welchem die meisten Kategorien zugerechnet werden, ist nachvollziehbar, aber die Zuordnung erstreckt sich über die gesamte Breite der Hauptkategorisierung (vgl. Tabelle 3.3).

SÖTAKUSHA-REIKAIKAN-YÖKUJITEN-PW				TANNO-KANYÖKUJITEN-PW			
Kategorie	Lesung	Deutsch	Hauptkat.	Kategorie	Lesung	Deutsch	Hauptkat.
厚かましい	<i>atsukamashii</i>	unverschämt	KÖRPER/ CHARAKTER/ EINSTELLUNG	厚かましい い、ずう ずうしい	<i>atsukamashii, zūzūshii</i>	unverschämt, verschlagen	EMPFINDUNG/ GEFÜHL
あわてる	<i>awateru</i>	Fassung verlieren	HANDELN/ BEWEGUNG/ TUN	あわて る、落ち 着かない	<i>awateru, ochitsukanai</i>	die Fassung verlieren, sich nicht beruhigen	EMPFINDUNG/ GEFÜHL
大変・は なはだし い	<i>taihen, hanahadashii</i>	schmerzlich, tragisch	KÖRPER/ CHARAKTER/ EINSTELLUNG	痛ましい	<i>itamashii</i>	schmerzlich	EMPFINDUNG/ GEFÜHL
罪・罰	<i>tsumi, batsu</i>	Verbrechen, Strafe	GESELLSCH./ KULTUR/ ALLTAG	後ろめた い	<i>ushirometai</i>	schuldbewusst sein	EMPFINDUNG/ GEFÜHL
美しい	<i>utsukushii</i>	schön	ZUSTAND/ AUSMASS/ WERT	美しい	<i>utsukushii</i>	schön	EMPFINDUNG/ GEFÜHL
体裁 /ば れる	<i>teisai / bareru</i>	Aussehen / ans Licht kommen	KÖRPER/ CHARAKTER/ EINSTELLUNG	うらおも て	<i>uraomote</i>	undurchsichtig sein	EMPFINDUNG/ GEFÜHL

3 Phraseologie in Theorie und Praxis

Kategorie	Lesung	Deutsch	Hauptkat.	Kategorie	Lesung	Deutsch	Hauptkat.
臆病、小心	<i>okubyō, shoshin</i>	Feigheit	KÖRPER/ CHARAKTER/ EINSTELLUNG	臆病	<i>okubyō</i>	Feigheit	EMPFINDUNG/ GEFÜHL
注目	<i>chūmoku</i>	Aufmerksamkeit	GESELLSCH./ KULTUR/ ALLTAG	気を引く	<i>ki-o hiku</i>	die Aufmerksamkeit auf sich lenken	EMPFINDUNG/ GEFÜHL
困る	<i>komaru</i>	Verlegenheit, in Schwierigkeiten sein	ZUSTAND/ AUSMASS/ WERT	困る	<i>komaru</i>	Verlegenheit, in Schwierigkeiten sein	EMPFINDUNG/ GEFÜHL
寂しい	<i>sabishii</i>	einsam	ZUSTAND/ AUSMASS/ WERT	さびしい	<i>sabishii</i>	einsam	EMPFINDUNG/ GEFÜHL
自信	<i>jishin</i>	Selbstbewusstsein	KÖRPER/ CHARAKTER/ EINSTELLUNG	自信がある	<i>jishin-ga aru</i>	selbstbewusst sein	EMPFINDUNG/ GEFÜHL
嫉妬	<i>shitto</i>	Eifersucht	EMPFINDUNG/ GEFÜHL	嫉妬	<i>shitto</i>	Eifersucht	LEBEN/ EHEPAAR
悪がしい・ずるがしい	<i>warugashikoi, zurugashikoi</i>	bösartig, hinterlistig	KÖRPER/ CHARAKTER/ EINSTELLUNG	ずるい	<i>zurui</i>	hinterlistig	EMPFINDUNG/ GEFÜHL
大切・大事	<i>taisetsu, daiji</i>	kostbar, wichtig	ZUSTAND/ AUSMASS/ WERT	大切	<i>taisetsu</i>	kostbar	EMPFINDUNG/ GEFÜHL
反応	<i>hannō</i>	Reaktion, Wirkung	KÖRPER/ CHARAKTER/ EINSTELLUNG	同意する	<i>dōi suru</i>	zustimmen	EMPFINDUNG/ GEFÜHL
眠る	<i>nemuru</i>	schlafen	HANDELN/ BEWEGUNG/ TUN	眠い、眠る	<i>nemui, nemuru</i>	müde, schlafen	EMPFINDUNG/ GEFÜHL
無愛想・冷淡・意地悪	<i>buaisō, reitan, ijiwaru</i>	Ungeselligkeit, Gefühlslosigkeit, Gefühlskälte, Gemeinheit	KÖRPER/ CHARAKTER/ EINSTELLUNG	無愛想	<i>buaisō</i>	Ungeselligkeit, Gefühlslosigkeit	EMPFINDUNG/ GEFÜHL
平然・平気	<i>heizen, heiki</i>	kaltblütig, gleichgültig	KÖRPER/ CHARAKTER/ EINSTELLUNG	平気、平然	<i>heiki, heizen</i>	gleichgültig, gelassen	EMPFINDUNG/ GEFÜHL

3 Phraseologie in Theorie und Praxis

Kategorie	Lesung	Deutsch	Hauptkat.	Kategorie	Lesung	Deutsch	Hauptkat.
油断	<i>yudan</i>	Unvorsichtig- tigkeit	EMPFINDUNG/ GEFÜHL	油断	<i>yudan</i>	Unvorsichtigkeit	HANDELN/ EINSTELLUNG
酔う	<i>you</i>	Trunkenheit	EMPFINDUNG/ GEFÜHL	酒	<i>sake</i>	Alkohol (<i>sake-ni you</i>)	LEBEN/ ZWISCHENM. BEZ./ALLTAG
非情	<i>hijō</i>	Gefühllos	Körper/ Charakter/ Einstellung	冷酷	<i>reikoku</i>	kaltblütig	EMPFINDUNG/ GEFÜHL

Tabelle 3.3: Abweichung Hauptkategorie und Lücken des Begriffsfelds EMOTION/GEFÜHL in SŌTAKUSHA-REIKAI-KANYŌKUJITEN-PW und TANNO-KANYŌKUJITEN-PW

Im TANNO-KANYŌKUJITEN-PW ist zusätzlich eine weitere Untergliederungsebene mit 58 Unterkategorien vorhanden, in welcher unterschieden wird zwischen

- affirmativer und negierter Verwendung (z. B. umfasst *akirameru* die Unterkategorien *akirameru* (あきらめる aufgeben) und *akiramenai* (あきらめない nicht aufgeben))
- aktiver und passiver Verwendung (z. B. beinhaltet *kiraku* (気楽 Erleichterung) die Unterkategorien *ki-o raku-ni suru* (気を楽にする sich entspannen) und *ki-ga raku-ni naru* (気が楽になる Erleichterung empfinden)).
- graduellen Unterschieden (z. B. beinhaltet *warai* (笑い Lachen) die Unterkategorien *hoemu* (微笑む lächeln) und *ōwarai* (大笑い lautes Gelächter))
- verschiedenen Stadien in einem Prozess (z. B. beinhaltet die Kategorie *kibō* (希望 Wunsch) die Unterkategorie *kibō-ga kieru* (希望が消える der Wunsch lässt nach))
- faktitiver und kausativer Verwendung (z. B. umfasst *odoroku* (驚く erschrecken) und *odorokasu* (驚かす jemanden erschrecken))
- Kontextbezügen (z. B. beinhaltet *okoru* (怒る sich ärgern) eine Unterkategorie *josei-ga okoru* (女性が怒る eine Frau erzürnt)).

In einigen wenigen Fällen ist es so, dass eine Kategorie im SŌTAKUSHA-REIKAI-KANYŌKUJITEN-PW einer Subkategorie im TANNO-KANYŌKUJITEN-PW entspricht (siehe *akirameru* oder *manzoku*).

Mit Hilfe dieses Top-Down-Kategorienvergleichs kann nur in begrenztem Rahmen eine Aussage über die Ähnlichkeit der Kategorisierung gemacht werden. Einerseits führt die Zuordnung anhand der Inhalte nicht immer Kategorien zusammen, die offensichtliche Affinitäten aufweisen, auch unterscheiden sich die Inhalte der durch gleiche Bezeichnungen und thematische Zuordnung als äquivalent eingestuft Kategorien teilweise erheblich. So enthält beispielsweise *awateru* im TANNO-KANYŌKUJITEN-PW vier Einträge, von welchen drei im YONEKAWA-NIHONGO-KANYŌKUJITEN-PW jeweils anderen Kategorien (teilweise sogar anderen Themenbereichen) zugeordnet werden.

3.2.2.5 Mikrostruktur

Die Mikrostruktur aller drei phraseologischer Wörterbücher ist durch Markierungen gegliedert. Im YONEKAWA-NIHONGO-KANYŌKUJITEN-PW ist dies am augenfälligsten (siehe Abbildung 3.7). Bei den Einträgen sind hier in der Regel folgende fünf Sektionen gekennzeichnet: Bedeutung (*imi* 意味), Gebrauch (*yōhō* 用法), Anwendungsbeispiele (*yōrei* 用例), Synonyme (*ruiku* 類句) und fremdsprachliche Entsprechungen (*gaikokugo* 外国語). Die Strukturierung der Einträge in den anderen Wörterbüchern ist weniger ausgeprägt, wobei die Bedeutungserklärung (*imi*) als Kern des Artikel unmarkiert direkt auf die Lemma-Angabe folgt. Weitere Kategorien sind im SŌTAKUSHA-REIKAI-KANYŌKUJITEN-PW Ergänzungen, die mit einem umrandeten Quadrat gekennzeichnet sind. Diese enthalten in der Regel Angaben zu Etymologie, Lesarten, Schreibweisen oder ähnlichen Ausdrücken sowie Literatur- beziehungsweise Quellenangaben (*shutten* 出典)²⁷. Das TANNO-KANYŌKUJITEN-PW markiert nur die Angabe von Synonymen und Antonymen (mit *han* 反 für *hantaigo* 反対語). Obwohl auch die anderen Wörterbücher die Angabe von Antonymen vorsehen, wird in der Praxis kaum davon Gebrauch gemacht.

Der Schwerpunkt liegt in allen drei Werken auf der Bedeutungsangabe. Aufgrund der onomasiologischen Strukturierung ist in SŌTAKUSHA-REIKAI-KANYŌKUJITEN-PW und TANNO-KANYŌKUJITEN-PW eine Untergliederung in unterschiedliche Lesarten innerhalb der Artikelkontexte nicht vorgesehen. Alle untersuchten polysemen *kanyōku* wurden nur mit jeweils einer

²⁷ Bei Quellenangabe *shutten* handelt es sich um literarische Quellen, während mit *yōrei* konstruierte Beispielan-gaben oder Belege aus nicht-literarischen (meist Zeitungstexten entnommenen) Quellen gemeint sind.

手を上げる てをあげる (意味) (1) 候補として名乗り出る。(2) 暴力をふるう。(3) 閉口して途中で投げ出す。降参する。(4) 技量を上達させる。(用法) 文型 (1) 「ダレダレが手を上げる」。命令・意志表現は可能。(2) 「ダレダレがダレダレに手を上げる」。(3) 「ダレダレが手を上げる」。(4) 「ダレダレがナニナニの手を上げる」

(用例) (1) ①『朝日新聞』(三〇〇四・九・四朝)「ライブドアが早い段階で手をあげ、それがきっかけになって球界の縮小が避けられたことを忘れてはならない」(2) ②向田邦子『字のない葉書』(一九五三)「癪癪を起こして母や子供達に手を上げる父の姿はどこにもなく」(3) 深谷忠記『秋・津和野殺人ライン』(一九五三)「さしもの彼も、今度ばかりは亜紀子のあまりの身勝手さに怒りを爆発させた。初めて手を上げた」(3) ④「こんな難しい問題は手を上げるしかない」(4) ⑤「酒の方も手を上げる」

(類句) (1) 「名乗りを上げる」(2) 「手を出す」(3) 「お手上げ」「かぶとを脱ぐ」「軍門に降る」「シヤツポを脱ぐ」「白旗を上げる」「旗を巻く」(4) 「腕を上げる」(外国語) 韓国語では「손을(手)을(을)다(上げる)」

Abbildung 3.7: Eintrag für *te-o ageru* im YONEKAWA-NIHONGO KANYŌKUJITEN-PW

Lesart hinterlegt. Das YONEKAWA-NIHONGO KANYŌKUJITEN-PW hingegen listet die unterschiedlichen Lesarten mit laufender Nummerierung, auf welche in den anderen Sektionen referenziert wird. So stehen zum Beispiel der monolithischen Bedeutungsangabe »aufgeben« von *te-o ageru* (手を上げる wörtl.: *die Hand/Hände heben*) im SŌTAKUSHA-REIKAI KANYŌKUJITEN-PW vier Lesarten im YONEKAWA-NIHONGO KANYŌKUJITEN-PW gegenüber (1. sich zur freiwilligen Teilnahme melden, 2. gewalttätig werden, 3. aufgeben 4. seine Fähigkeiten verbessern).

Eine weitere Besonderheit des YONEKAWA-NIHONGO KANYŌKUJITEN-PW ist die grammatische Angabe in der Sektion *yōhō*. Als einziges der untersuchten Wörterbücher integriert es systematisch Satzmuster in seine grammatische Erklärung. Dabei werden Platzhalter für Personen (*daredare*), Dinge (*naninani*) und Orte (*dokodoko*) verwendet. Die Darstellung in *katakana* an Stelle der üblichen Schriftform (*hiragana* oder *kanji*) dient der Abgrenzung der Platzhalter gegen die Phrasem-immanenten Komponenten. Ein Beispiel für die Darstellung der externen Valenz ist *darenani-wa mushi-ga sukanai* (jemanden /etwas-TOP Insekt-SUB mögen-NEG) was in etwa bedeutet »jemanden /etwas (intuitiv) nicht mögen«. Dadurch wird verdeutlicht, dass die Aversion sowohl gegen Personen als auch gegen Dinge gerichtet sein kann und die Anbindung durch Topikalisierung erfolgt. Überdies leisten Erklärungen wie *meistens sagt man das über Personen* oder *in seltenen Fällen auch affirmativ gebraucht* zusätzlich Hilfestellung beim

aktiven Gebrauch des *kanyōku*. Allerdings geht aus den Beispielen hervor, dass die in den Satzmustern exemplifizierten kontextuellen Einbindungsmöglichkeiten nur einen Ausschnitt des realen Gebrauchs darstellen. Auch werden bei polysemen *kanyōku* sowohl Satzmuster als auch Beispielangaben in Beziehung zur Lesart erläutert:

- (22) 「ダレダレガ手を上げる」命令、意志表現は可能。
 »*daredare-ga te-o ageru*« *meirei, ishihyōgen-wa kanō*.
 »jemand-SUBJ Hand-AKK heben« Befehl oder Absichtserklärung-TOP möglich.

Das SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW verwendet überwiegend konstruierte Beispielsätze, Lesehilfen werden nur für die sporadischen Belege verwendet. Den pragmatischen Ansatz betonen die Beispielangaben im TANNO-KANYŌKUJITEN-PW, da diese aus ausgiebigen konstruierten Dialogen bestehen, die die Bedeutung im Kontext verdeutlichen sollen. In dem Eintrag für *hara no mushi ga osamaranai* wird beispielsweise durch eine imaginäre Unterhaltung zwischen zwei Männern der Gebrauch des *kanyōku* illustriert:

»Es wurmt mich, dass der Imawaga die Yōko von der Rezeption gefragt hat, ob sie ihn heiraten möchte. Ich hau ihm eine rein.« »Das sagst du so, aber Yōko-san soll sich doch sehr gefreut haben einen Heiratsantrag von Imagawa zu bekommen.« »Deshalb werde ich mich auch nicht beruhigen, bevor ich ihm (Imawaga) nicht zwei-drei Schläge verpasst habe.«²⁸

Die zwei Vorkommen von *hara-no mushi-ga osamaranai*, die hier mit »es wurmt mich« und »werde ich mich nicht beruhigen« übersetzt wurden, werden auch im Wörterbuchartikel durch Fettdruck hervor gehoben. Trotzdem setzt diese Herangehensweise voraus, dass der Benutzer den Wörterbuchartikel vollständig liest, da die exemplarische Situation die einzige im Wörterbuch kodierte Form der Bedeutungserklärung ist.

Die Angabe von zusätzlichen Informationen, wie synonymen und antonymen *kanyōku*, folgt in allen Werken auf die Beispielangabe. Verweise innerhalb des Wörterbuchs werden jeweils mit einem Stern gekennzeichnet. Die Angabe der fremdsprachlichen Entsprechungen umfasst englische, chinesische und koreanische Übersetzungen. Für das Englische lässt sich sagen, dass die Angabe nicht auf phraseologische Äquivalente beschränkt ist.

²⁸ Im Original heißt es: 「今川が受付の耀子さんにプロポーズしたなんて、腹の虫が治まらない。なぐってやる。」 「そんなこと言っても、耀子さんは今川のプロポーズを大喜びしているそうじゃないか。」 「だから二、三発なぐってならないと、腹の虫が治まらない」.

3.2.2.6 Zusammenfassung

Aus dem Vergleich der Kodierung von *kanyōku* in den unterschiedlichen Wörterbüchern lassen sich nur wenige etablierte Konventionen ableiten. Hierzu gehören beispielsweise die Verwendung der Finalform in der Nennform und die Anordnung der Leitelemente nach der 50-Laute-Tafel. Es handelt sich dabei durchweg um allgemeine Konventionen der japanischen Lexikographie. Bei den untersuchten phraseologischen Wörterbüchern sind die Gestaltung der Nennform und die Strukturierung weitgehend ähnlich. In inhaltlichen Fragen lassen sich indes deutliche Unterschiede erkennen.

Die drei phraseologischen Nachschlagewerke verfolgen verschiedene Ansätze, um die aktive Verwendung der *kanyōku* zu unterstützen. Bei den Werken SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW und TANNO-KANYŌKUJITEN-PW liegt der Fokus auf der Möglichkeit, *kanyōku* nach ihrer Bedeutung nachzuschlagen. Da diese Werke für die Textproduktion konzipiert sind, legen sie keinen Wert auf eine umfassende Erklärung aller Bedeutungsnuancen, sondern konzentrieren sich auf pragmatische Gesichtspunkte. Im SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW nimmt die klassische Bedeutungserklärung nicht wesentlich weniger Raum ein als die Beispiele. Im Gegensatz hierzu liegt der Schwerpunkt beim TANNO-KANYŌKUJITEN-PW strukturell und quantitativ auf den Dialogen, die den Wörterbucheintrag prägen.

Das YONEKAWA-NIHONGOKANYŌKUJITEN-PW ordnet seine Einträge nach traditioneller Methode. Innovativ ist hier die Formangabe, die systematisch die grammatischen Merkmale beschreibt. Die Kodierung der grammatischen Struktur im Wörterbuch ist nicht neu. Im Ansatz ähnliche Beschreibungsstrukturen finden sich auch in anderen phraseologischen Nachschlagewerken (z. B. dem ŌBUNSHA-SEIGORIN-PW). Das YONEKAWA-NIHONGOKANYŌKUJITEN-PW geht jedoch systematischer vor, indem es die Kodierung von externen Valenzen formalisiert und nicht nur für alle Einträge, sondern auch für alle Lesarten umsetzt.

Man kann sagen, dass insbesondere im YONEKAWA-NIHONGOKANYŌKUJITEN-PW bereits einige der Anforderungen der Linguistik umgesetzt wurden. In Bezug auf inhaltliche Fragen zeigt das Werk innovative Lösungsansätze auf. Die Betrachtung der onmasiologisch strukturierten Werke bestätigt hingegen die in Kapitel 3.1.6.3 thematisierten Vorbehalte gegenüber dieser Vorgehensweise bezüglich der Ordnung des Objektbereichs. Zudem wird der Umgang mit polysemen Phrasemen erschwert.

3.3 Anforderungen an die elektronische Phraseographie aus computerlinguistischer Perspektive

Nach computerlinguistischer Theorie ist ein Wörterbuch definiert als ein Referenzwerk mit Informationen zu lexikalischen Einheiten,²⁹ während der Begriff Lexikon im Kern eine geordnete Liste von Worten beschreibt, die mit bestimmten phonologischen, grammatischen und semantischen Merkmalen angereichert sein können.³⁰ In der Praxis werden beide Begriffe mit einer großen Bandbreite von lexikalischen Ressourcen verbunden, angefangen von einfachen lexikalischen Wortlisten (beispielsweise für die Rechtschreibprüfung) über diverse lexikalische Komponenten für natürliche Sprachverarbeitung (beispielsweise Thesauren und Wortnetze) bis hin zu hoch komplexen Datenstrukturen, die multimediale Inhalte und multi-modale Zugriffsmöglichkeiten bieten.

Während aus lexikographischer Perspektive die Genese und Zielsetzung eines Wörterbuchs eine wichtige Rolle bei der Einordnung der Ressource spielt (Müller-Spitzer 2003), steht in der Computerlinguistik das Einsatzgebiet im Vordergrund. Allgemein unterscheidet man zwischen maschinenlesbaren Wörterbüchern (MRD als Abkürzung für *machine readable dictionaries*), die für den menschlichen Benutzer konzipiert sind, und lexikalischen Ressourcen, die für die natürliche Sprachverarbeitung (NLP) geschaffen wurden.

Obwohl gelegentlich versucht wird diese beiden Gebiete zu fixieren (vergleiche beispielsweise Wiegand (1998, S. 58)), sind die Grenzen zwischen maschinenlesbaren Wörterbüchern und NLP-Komponenten fließend. In vielen Fällen werden NLP-Komponenten in elektronische Wörterbücher eingebunden, die in erster Linie für den menschlichen Benutzer konzipiert sind oder in komplexen Anwendungen den menschlichen Nutzer in der natürlichen Sprachverarbeitung unterstützen. Für die Visionäre der Computerlinguistik ist das elektronische Wörterbuch ein

²⁹ Die Definition in Isocat lautet: »A reference work in which you can find information on several or all of the following features: word form, spelling, phonetics, meaning, usage and etymology of (parts of) words, collocations, expressions and proverbs.« (Eerten 2012).

³⁰ Die allgemeine Definition von *lexicon* in Isocat lautet: »ordered collection of different words (or morphemes) of a language, often annotated with phonologic, syntactic and semantic information« (Eerten 2010).

Auslaufmodell, das früher oder später durch intelligente Methoden natürlicher Sprachverarbeitung abgelöst wird beziehungsweise nur noch im Hintergrund der sprachverarbeitenden Applikationen lexikalische Information beisteuert. Im Ergebnis würde der Benutzer nicht mehr direkt auf das Wörterbuch zugreifen sondern nur noch mit komplexen Werkzeugen arbeiten, die ihn in seiner Aufgabe unterstützen, ohne dass er eines weiteren lexikalischen Hilfsmittels bedarf.

Aufgrund der Komplexität virtueller Ressourcen stellt sich die Frage, wodurch sich die Einheit Wörterbuch auszeichnet. H. Bergenholtz (2011) analysiert dänische Internetwörterbücher unter der Vorgabe, dass die Einheit »Wörterbuch« durch seine Funktionalität definiert ist. Demnach handelt es sich um verschiedene Wörterbücher, wenn sich die Funktionalität unterscheidet, auch wenn sie auf einen gemeinsamen Datenbestand zugreifen. Innerhalb der lexikographischen Funktionslehre ergibt dies Sinn, da unterschiedliche Funktionalität und damit verbundene Profile eine zentrale Rolle in der Wörterbuchkonzeption einnehmen. Es ist deshalb durchaus denkbar, dass aus der Perspektive eines Benutzers der Eindruck verschiedener Wörterbücher entsteht.

Vor diesem Hintergrund beschreibt Tarp, was Lexikographen auf dem Weg zur Entwicklung eines »lexikographischen Rolls Royce« beachten sollten:

1. Lexicographers should make full use of the interactive possibilities of electronic media in order to create lexicographic tools which, in an easy and quick way, can help the user to determine the type of lexicographic work to be consulted, the type of extra-lexicographic situation and the type of relevant user characteristics.
2. The access routes should not only be based upon full words, combinations of words, or even orthographic mistakes, but also on other search criteria, e.g. associations.
3. The traditional dictionary articles with static data and fixed structures should be replaced by articles containing dynamic data which are, so to say, unique for each search related to a specific type of user in a specific type of user situation. (Tarp 2009, S. 25–26)

Tarps Anforderungen stehen ganz im Zeichen der lexikographischen Funktionslehre. Allgemeiner formulieren Fuertes-Olivera und Niño-Amo (2011, S. 172–173) Anforderungen an Internet-Wörterbücher:

- An internet dictionary must be easy to find
- It must have smart searches

- It must strengthen interactivity
- It can be part of a reference portal
- It must use an elegant and pleasant layout
- It must use a familiar and reassuring virtual environment
- It must allow the retrieval of data in different formats³¹
- It must be provided with extra software³²
- It must have readable articles³³
- It must be regularly updated and corrected
- It must offer instant, simple, and reduced numbers of results
- It must use advanced searches
- It must display the results logically

Diese beiden Beispiele allgemein gehaltener Anforderungen zeigen bereits die Vielfalt an diversen Wünschen und kontroversen Forderungen in Bezug auf elektronische Wörterbücher.

Der Fokus dieses Kapitels liegt auf den Desiderata, die in der wissenschaftlichen Diskussion im Kontext der Entwicklung von Wörterbüchern stehen, die für den menschlichen Benutzer konzipiert wurden. Aufgrund der raschen Veränderungen auf diesem Gebiet handelt es sich zwangsläufig um eine Momentaufnahme, die keinen Anspruch auf Vollständigkeit erhebt. Die folgenden Kapitel berücksichtigen insbesondere aktuelle computerlinguistische Arbeiten, die für die Phraseographie im Besonderen relevant sind. Eine ausführliche Aufstellung der Anforderungen, die Lexikographen an das digitale Medium stellen, findet sich bei Schryver (2003).

3.3.1 Automatische Extraktion von Mehrwortlexemen

Ein beachtlicher Teil der computerlinguistischen Forschung beschäftigt sich mit der Frage der automatischen Extraktion von Mehrwortlexemen aus Textkorpora. Verschiedentlich wird versucht, das Phänomen Mehrwortlexem rein auf der Basis statistischer Methoden zu fassen. Das

³¹ Mit verschiedenen Datenformaten ist hier beispielsweise die Integration von multimedialen Inhalten gemeint.

³² Zusätzliche Software bedeutet in diesem Zusammenhang zusätzliche Funktionalität, wie beispielsweise das Kopieren und Einfügen von Inhalten.

³³ Lesbar bezieht sich auf Anforderungen an die Darstellung. Beispielsweise soll auf Kontraste geachtet und mit Farben sparsam umgegangen werden.

bedeutet in der Regel, dass der Objektbereich Phraseologie mit der Menge der statistisch signifikanten Kookkurrenzen³⁴ gleichgesetzt wird.

Die Befürworter solcher Verfahren argumentieren, dass es sich um eine neutrale Methode handelt, den Objektbereich einzugrenzen. Auch ist es für den Benutzer in der konkreten Situation unter Umständen nicht ersichtlich, ob er es mit einem Phrasem zu tun hat oder mit einer bloßen Kookkurrenz (Wible 2008). Mittels statistischer Erfassungsmethoden lassen sich überdies unbekannte Phraseme aus Korpora extrahieren und überprüfen (Petrova 2010, S. 147–149).

Vor dem Hintergrund der Unterschiede in der Phraseologie-Auffassung zwischen den Disziplinen ist der Einsatz computerlinguistischer Methoden für die Phraseologie nur bedingt möglich. Für die von einer engen Definition ausgehenden Phraseographie gelten solche Methoden als ungeeignet (Konecny 2010, S. 80), zumal man nicht davon ausgehen kann, dass die derzeitigen Verfahren der statistischen Korpusanalyse zu einer belastbaren Aussage führen, ob ein Phrasem (oder eine bestimmte Form eines Phrasems) existiert (Herold 2007). Ausschlaggebend für die Qualität der Ergebnisse ist unter anderem die Verfügbarkeit eines ausgewogenen Korpus. Hier gilt zu bedenken, dass der Großteil linguistischer Korpora überwiegend auf geschriebener Sprache beruhen, ein nicht unwesentlicher Anteil der Phraseme jedoch nahezu ausschließlich in gesprochener Sprache verwendet wird (Quasthoff, F. Schmidt und Hallsteinsdóttir 2010, S. 37–38). Auch Variation und Modifikation sind bislang nicht annähernd vollständig erfasst, was sich ebenfalls negativ auf die Qualität der Korpusanalyseergebnisse auswirken kann (Hrisztova-Gotthardt 2010a; Copestake u. a. 2002). Vielversprechend sind semi-automatische Methoden, die grammatische und semantische Information in die statistische Analyse integrieren (Fjeld, Nygaard und Bick 2010). Noch liegen für diese Methoden jedoch keine Evaluationsergebnisse vor, so bleibt abzuwarten, wie sich diese in der Praxis bewähren.

³⁴ Als Kookkurrenz bezeichnet man in der Linguistik das gemeinsame Auftreten zweier lexikalischer Einheiten in einer textuellen Einheit. In der Praxis wirkt sich das so aus, dass beispielsweise nach Colsons automatischer Korpusanalyse *there are no* dem Kernbereich idiomatischer Trigramme zugerechnet wird (Colson 2010, S. 32–33).

Ein weiteres Einsatzgebiet für korpusbasierte Methoden ist die Frequenzbestimmung. Dabei wird die Häufigkeit einer vordefinierten Menge von Phrasemen entweder absolut oder in Abhängigkeit einer Textsorte bestimmt (Quasthoff, F. Schmidt und Hallsteinsdóttir 2010). Frequenz-Daten sind ein probates Mittel in der Eingrenzung des Lexembestandes eines Wörterbuchs oder für Häufigkeitsangaben (meist durch Markierung der häufigen Lexeme), in der Phraseographie wird jedoch auch diskutiert, ob sie nicht bei der Wahl der Nennform und Angabe von Varianz eingesetzt werden können (Heid 2007; Fellbaum 2004; Parizoska 2010). Unabhängig von den Problemen derartiger Verfahren muss hinterfragt werden, wie sich diese Vorgehensweise mit der traditionellen Funktion der Nennform (vergleiche Kapitel 4.4.3) vereinbaren lässt. Gegebenenfalls müssen Informationen, die normalerweise in der Nennform kodiert sind, anderweitig zur Verfügung gestellt werden. Darüber hinaus stellt sich die Frage, wie die Auffindbarkeit gewährleistet wird, da für den Benutzer nicht vorhersehbar ist, welche Form eines Phrasems statistisch am häufigsten auftritt.

In diesem Sinne ist nicht unumstritten, ob Flexibilität in der Gestaltung der Nennform überhaupt eine sinnvolle Vorgehensweise ist, denn auch aus computerlinguistischer Perspektive stellt das Fehlen eines Standards für die Nennform von Phrasemen in Wörterbüchern ein ernstzunehmendes Problem dar (Oppentocht und Schutz 2003, S. 219).

3.3.2 Zugriffsmethoden

Direktheit und Schnelligkeit des Zugriffs auf die Inhalte sind wesentliche Qualitätskriterien, an welchen sich elektronische Nachschlagewerke messen lassen müssen.

No matter how outstanding the contents of a dictionary, if the contents one needs at a particular point in time cannot be accessed in a quick and straightforward way, the dictionary de facto fails to be a good dictionary. (Schryver 2003, S. 173)

Das Zusammenspiel von Datenstruktur und Suchfunktionalität entscheidet darüber, wie unmittelbar und rasch der Benutzer von seiner Suchanfrage zum Ergebnis gelangt.³⁵ Die Je-Schneller-Desto-Besser These findet jedoch nicht nur Befürworter. Forschungsergebnisse im Bereich

³⁵ Ein weiterer wichtiger Faktor ist — insbesondere im Zusammenhang mit komplexen Schriftsystemen wie dem Japanischen — die Möglichkeiten der Eingabe. Obwohl sie in der Forschung eine wichtige Rolle spielen, sind sie für die Konzeption des Nachschlagewerkes selbst nicht relevant und werden deshalb an dieser Stelle bewusst nicht in die Diskussion eingebracht.

Didaktik belegen, dass ein zu effektiver Nachschlagevorgang negative Auswirkungen auf die Speicherung im Langzeitgedächtnis hat (Hessky und Ettinger 1997, S. XXXIII). Neben der Suche wird auch der Möglichkeit des Stöberns in Wörterbüchern auch eine gewisse Notwendigkeit beigemessen (Schryver 2003, S. 184–185).

Phraseme stellen höhere Anforderungen an die Zugriffsstrukturen als Einzelwortlexeme. Zwar ist die Indizierung diverser Formen möglich und die Angabe der Nennform relativ unabhängig davon, dennoch können Restriktionen und Variation den Zugriff erschweren. Die Möglichkeiten der orthographischen Variation potenzieren sich mit der Anzahl der Komponenten, die ein Mehrwortausdruck enthält. Und Variation und kreative Abwandlungen können ebenso wie in der Korpusanalyse nur in gewissem Rahmen antizipiert und abgedeckt werden.

In jedem Fall müssen die Daten in einer Weise aufbereitet werden, die eine erfolgreiche Suche erleichtert. Dazu gehört in der Regel vorhandene Platzhalter aus den indizierten Zeichenketten zu tilgen, damit diese die Suchergebnisse nicht negativ beeinflussen (Almind, H. Bergenholtz und Vrang 2006, S. 171). Um dem Problem der Variabilität zu begegnen, kann die Indizierung aller (bekannten) Varianten der Nennform Abhilfe schaffen (Oppentocht und Schutz 2003, S. 217). Methoden der Benutzerpartizipation können eingesetzt werden, um die Liste der möglichen Varianten zu erweitern, denn die Praxis hat gezeigt, dass es sich bei *nicht gefundenen Einträgen* nicht selten um Varianten existenter Einträge handelt. Ergänzend kann eine Ähnlichkeitssuche den Benutzer unterstützen, falls bei der Eingabe des Suchbegriffs Fehler passieren, indem beispielsweise im Fall einer leeren Ergebnismenge Teile der eingegebenen Zeichenkette mit der Datenbasis abgeglichen und eventuelle partielle Treffer dem Benutzer präsentiert werden. (Almind, H. Bergenholtz und Vrang 2006, S. 171–172; Oppentocht und Schutz 2003, S. 223). Damit könnte L2-Lernern geholfen werden, die sich an etwas Gehörtes nicht exakt erinnern oder es nicht in das Schriftsystem umsetzen können. Gouws (2009) unterscheidet zwischen deskriptiver und präskriptiver Funktion in Abhängigkeit von der Nutzungssituation. Er vertritt den Standpunkt, dass auch in Zeiten des raschen Wandels von Ressourcen Wörterbücher eine Standardisierungsfunktion erfüllen und deshalb die der Norm entsprechenden lexikalischen Einheiten oder Formen im Vordergrund stehen sollten. Der Zugriff auf abweichende Elemente soll beispielsweise für die Textrezeption zwar ermöglicht werden, jedoch indirekt (mittels Verweis) zu präskriptiven Einheiten führen.

Odiijk (2004) regt an, Variabilität in Phrasemen systematisch zu beschreiben. Der von ihm vorgeschlagene Standard soll nach Möglichkeit unabhängig von einer bestimmten Theorie und

technisch realisierbar sein. Dabei wird morphosyntaktische Variabilität in Idiom-Mustern kodifiziert, die deren Struktur eindeutig identifizieren. Auch eine Verlinkung zwischen den Phrasemen und den Einträgen ihrer Komponenten soll den Zugriff erleichtern.

Eine grundsätzliche Frage der Systemarchitektur ist auch, welche Bereiche durchsuchbar sein sollen. Es kann sehr hilfreich sein, durchaus auch gewünscht, dass die unterschiedlichen Informationstypen (Definitionen, Synonyme, Antonyme, Etymologie etc.), die im Artikeltext des Wörterbuchs enthalten sind, gezielt abgefragt und angeordnet werden können (Oppentocht und Schutz 2003; Hrisztova-Gotthardt 2010b). H. Bergenholtz (2011, S. 42–49) beschreibt am Beispiel dänischer Internetwörterbücher, wie die Suche in Abhängigkeit von der Funktion des Wörterbuchs³⁶ gestaltet werden soll. Vereinfacht³⁷ dargestellt werden für die Textrezeption (*Meaning of Fixed Expressions*) nur die Felder mit Lemma-Angaben indiziert, für die Textproduktion (*Use of Fixed Expressions*) hingegen auch Informationen zum kontextuellen Gebrauch mit in die Suche einbezogen. Wird nach einem Ausdruck mit einer bestimmten Bedeutung (*Fixed Expression with a certain meaning*) gesucht, werden zusätzlich zu den Lemma-Angaben alle Arten von Bedeutungsangaben bei der Suche berücksichtigt und für den Wissensaufbau (*Knowledge about Fixed Expressions*) wird jedes Vorkommen des Suchbegriffs in den Lemma-Angaben, Bedeutungsangaben, Kommentare (z. B. etymologische Angaben) und Assoziationen erfasst (vgl. Tabelle 3.4)

Art der Angabe	Lemma	Bedeutung	Kollokation	Beispiel	Synonym	Kommentar
<i>Textrezeption</i>	+	-	-	-	-	-
<i>Textproduktion</i>	+	-	+	+	-	-
<i>Best. Bedeutung</i>	+	+	-	-	-	+
<i>Wissensaufbau</i>	+	+	-	-	+	+

Tabelle 3.4: Zugriff in Abhängigkeit von der Funktion des Lexikons (Vereinfachte Darstellung)

Kritisch bemerkt H. Bergenholtz (2011), dass die Ergebnisliste der Suche nach einem Phrasem mit einer bestimmten Bedeutung möglicherweise ausufern könnte. Lange Trefferlisten führen

³⁶ In diesem Fall wird zwischen vier Wörterbüchern unterschieden, die sich eine Datenbasis teilen.

³⁷ Die von H. Bergenholtz (2011) vorgestellte Datenstruktur ist wesentlich elaborierter. Die Angaben sind granularer, umfassender und werden gegeneinander gewertet.

zu einer Informationsflut, aus welcher das Wesentliche mühsam extrahiert werden muss. Das widerspricht der Vorstellung eines guten lexikographischen Werkzeugs, das nach Ansicht von H. Bergenholtz und I. Bergenholtz (2011) einer bestimmten Nutzergruppe schnellen und zuverlässigen Zugang zu den benötigten Informationen bietet.

Lew (2011) sieht die Entwicklung von spezialisierten Werkzeugen durch die Übermacht der Popularität von *Google*-Suchen gefährdet. Er bemerkt, dass nur noch eine kleine Gruppe von Experten über die Fähigkeit verfügt, gezielt spezielle Werkzeuge einzusetzen. Benutzerstudien stützen diese These, indem sie zeigen, dass trotz des Vorhandenseins entsprechender Werkzeuge in Artikel eingebettete Mehrwortlexeme von den meisten Nutzern nicht gefunden werden, weil sie nach der gesamten Zeichenkette suchen (Heid 2011, 295 u. 302–303).

Neben der Verfeinerung der Suchmethodik ist auch eine Ausweitung des Datenbestands eine mögliche Herangehensweise um die Trefferergebnisse zu verbessern. Tono (2009, S. 57–60) hebt den Nutzen der parallelen Suche von Begriffen in mehreren Wörterbüchern³⁸ hervor. Der Mehrwert liegt dabei gerade in der Tatsache, dass sich die Informationen der einzelnen Wörterbücher gegenseitig ergänzen. Auch der Funktionalität einiger japanischer elektronischer Taschenwörterbücher, die Übersetzungsäquivalente über Wörterbuchgrenzen hinweg erfassen können, misst er eine wichtige Bedeutung bei.

3.3.3 Onomasiologische Kategorisierung

Die onomasiologische Suche im elektronischen Wörterbuch soll durch Kategorisierung differenzierten Zugriff auf semantische Informationen ermöglichen (Oppentocht und Schutz 2003, S. 222). Die Computerlinguistik begegnet der Möglichkeit des onomasiologischen Zugriffs mit weniger Skepsis als die traditionelle Lexikographie. L'Homme, Robichaud und Leroyer (2013, S. 216–219) zeigen in einer Pilotstudie, dass sich die Möglichkeit des onomasiologischen Zugriffs überwiegend positiv auf die Performanz bei der Nutzung eines elektronischen Kollokationenwörterbuchs auswirkt. Das mag daran liegen, dass das elektronische Medium andere Möglichkeiten bietet als die traditionelle Lexikographie und die Notwendigkeit, sich auf eine Makrostruktur im Werk festzulegen, entfällt. Ebenso wie in der Printlexikographie müssen

³⁸ In Tono (2009) ist die Einheit Wörterbuch dadurch spezifiziert, dass sie einen distinkten Datenbestand hat.

jedoch Daten zur Verschlagwortung oder Bildung von Kategorien erhoben und durch entsprechende Zugriffsstrukturen zugänglich gemacht werden. Dies kann sowohl durch hierarchische Navigationsstrukturen geschehen als auch durch die Integration in die Suchfunktionalität. Durch eine Verschlagwortung können die Probleme der Zuweisung einzelner Kategorien teilweise vermieden werden, indem ein Phrasem über Begriffsfelder, die verschiedene Aspekte seiner Bedeutung repräsentieren, indiziert wird. Um den relativ hohen redaktionellen Aufwand einzudämmen, kann nach der Basiskategorisierung der Nutzer in die Erweiterung und Ausdifferenzierung der Verschlagwortung mit einbezogen werden (Almind, H. Bergenholtz und Vrang 2006, S. 23–24).³⁹ Darüber hinaus kann die onomasiologische Kategorisierung auch genutzt werden, um die Darstellung von Phrasemen zu optimieren (Heid 2007, S. 1042).

3.3.4 Darstellung lexikalischer Einheiten

Auf die konkrete Darstellung von Online-Inhalten kann nur begrenzt Einfluss genommen werden, da Hard- und Software sehr unterschiedliche Ergebnisse produzieren (Almind 2005a). Darüberhinaus sorgen die rasche Weiterentwicklung im Bereich Informationstechnologie sowie wechselnde Moden und »Best Practices« in Applikationsentwicklung und Webdesign dafür, dass Vorgaben in diesem Bereich relativ schnell als überholt gelten. So galt bis vor kurzem als Grundregel für die Digitalisierung von lexikalischen Ressourcen, dass auf Verdichtungen verzichtet werden sollte, da Platzbeschränkungen für das elektronische Medium nicht greifen. Aufgrund der zunehmenden mobilen Nutzung des Internets durch Kleingeräte wird diese These erneut hinterfragt werden müssen.⁴⁰

Von konkreten Formatierungsvorschlägen muss deshalb Abstand genommen werden, jedoch gelten die Grundprinzipien des barrierefreien Webdesigns auch für digitale Wörterbücher. Im

³⁹ Nachteil der vorgestellten Methode ist, dass zunächst die Verschlagwortung auf die Begriffe beschränkt ist, die dem Lexikographen innerhalb von 30 Sekunden einfallen. Aus diesem Grund regt H. Bergenholtz (2011, S. 53) an, dass für jeden Eintrag die assoziativen Stichwörter von mehreren Lexikographen gesammelt werden sollten.

⁴⁰ Eine Umfrage im Umfeld wissenschaftlicher lexikalischer Ressourcen zu Folge ist die Nutzung von elektronischen Wörterbüchern auf kleinen Bildschirmen in Deutschland noch wenig verbreitet (Müller-Spitzer, Kopleinig und Töpel 2013). Dies widerspricht jedoch kommerziellen Nutzungsstatistiken, die von einem generellen Trend hin zu mobilen Endgeräten ausgehen.

Kern geht es darum, Informationen gut lesbar und übersichtlich darzubieten, indem wichtige Informationen durch Positionierung oder Größe hervorgehoben werden, während mit graphischen Elementen, Farben und Informationsfülle zurückhaltend umgegangen wird.

Eine für digitale Wörterbücher zentrale Frage betrifft die Darstellung der Suchergebnisse eines Eintrags. Lew (2011, S. 424) nennt drei Möglichkeiten:

1. a menu of target items is presented
2. a menu is presented, but the most likely choice opens by default
3. partial entries are listed

Lew plädiert für die dritte Methode, da eine reine Auflistung (1.) wenig benutzerfreundlich ist und die Variante, bei der der wahrscheinlichste Treffer in voller Länge präsentiert wird (2.), die Gefahr birgt, dass der Benutzer die anderen Möglichkeiten übersieht. Eine Kurzübersicht aller möglichen Treffer (3.) ist nach seiner Einschätzung deshalb die beste Lösung.

Vorschläge zur Darstellung einzelner Einträge unterbreiten Modelle aus dem Bereich der lexicographischen Funktionslehre (H. Bergenholtz und Tarp 2002; Almind, H. Bergenholtz und Vrang 2006; H. Bergenholtz 2011). In der von H. Bergenholtz (2011) beschriebenen Implementierung von Phrasemwörterbüchern werden in Abhängigkeit von der Suche für die Darstellung vier Typen unterschieden. Für die Textrezeption werden dem Benutzer nur die Angaben zur Form (inklusive Lemmavarianten) und Bedeutung präsentiert, während für die Textproduktion zusätzlich die Informationen zur Grammatik, Beispielangaben und Kollokationen angezeigt werden. Die assoziative Suche zeigt zunächst nur eine Liste der Basislemmata und Bedeutung, während die Suche aus Interesse zusätzlich den Bereich der assoziierten Kategorien mit einbezieht (siehe Tabelle 3.5)⁴¹.

Art der Ang	Lem.	Bed.	Gram.	Koll.	Beisp.	Syn./Ant.	Anm.	Assoz.
<i>Textrezeption</i>	+	+	-	-	-	-	-	-
<i>Textproduktion</i>	+	+	+	+	+	+	-	-
<i>Best. Bedeutung</i>	+	-	+	-	-	-	-	-
<i>Wissensaufbau</i>	+	+	+	+	+	+	+	+

Tabelle 3.5: Präsentation in Abhängigkeit von der Funktion des Lexikons (Vereinfachte Darstellung)

⁴¹ Analog zu Tabelle 3.4 wurde auch hier eine vereinfachte Darstellung gewählt, aus welcher die Wertung der Felder und die Reihenfolge der Präsentation nicht hervorgeht.

3.3.5 Organisation lexikalischer Einheiten

Die Organisation lexikalischer Daten bildet die entscheidende Grundlage für weitgehend alle Strukturen im elektronischen Nachschlagewerk. Dabei steht die Frage im Vordergrund, welchen Status die verschiedenen lexikalischen Phänomene in der hierarchischen Struktur des Wörterbuchs haben und wie sie sich in Relation zueinander verhalten. Die Thematik dieses Kapitels beschränkt sich auf der Organisation von Einheiten, die von dem Benutzer auch als modular empfunden werden. Fragen der technischen Umsetzung in konkrete Datenstrukturen werden in Kapitel 3.4 vor dem Hintergrund von Standardisierungsbestrebungen diskutiert.

Bezüglich des Status der Phraseme in dem elektronischen Nachschlagewerk gibt es wie im gedruckten Wörterbuch zwei Möglichkeiten: Sie werden strukturell in den Inhalt anderer lexikalischer Einheiten eingebettet, wobei anders als im gedruckten Werk über die Indexierung der direkter Zugriff auf sie möglich ist. Oder sie bilden autonome Einheiten im Lexikon, die über Relationen mit anderen Einheiten im Lexikon verknüpft sind.

Beide Methoden haben ihre Vor- und Nachteile. Eine Einbettung in einen existierenden Wörterbuchartikel ist aus lexikographischer Perspektive die einfachere Variante und entspricht möglicherweise eher dem, was der Benutzer aus traditionellen Nachschlagewerken gewohnt ist. Insbesondere ist der Zugriff über die Komponenten eines Phrasems weniger anfällig für Probleme, die durch Variabilität und Modifikation hervorgerufen werden. Und je länger und modularer die Zeichenkette, desto größer die Gefahr einer partiellen Abwandlung, wie die folgenden Kapitel zeigen.

Dennoch herrscht weitgehend Einigkeit darüber, dass die Einbettung von Phrasemen in die Artikel der Komponenten im elektronischen Medium nicht sinnvoll ist. Statt dessen sollen Phraseme, die einen lexemartigen Status genießen, auch in entsprechender Weise im Wörterbuch verankert werden (Oppentocht und Schutz 2003, S. 220; Heid 2007, S. 1041; Jónsson 2009, S. 168).

Diese Präferenz wird durch Benutzerstudien bestätigt, die zeigen, dass der Zugriff auf Mehrworteinheiten über den (Um)Weg des Wörterbuchartikels der darin enthaltenen Komponenten in der Praxis dazu führt, dass häufig die Suche in die Leere läuft (Heid 2011, S. 295). Wie dem Problem der Variabilität begegnet werden soll, bleibt in diesem Zusammenhang ohne Erwähnung. So beginnt Jónsson sein Plädoyer für die Herauslösung der Phraseme aus der Mikrostruktur von Wörterbuchartikeln mit:

The neutral character of single-word lemmas, which can cover very different and inconsistent properties of the actual word, especially with regard to the analysis of the semantic and grammatical structure (whereas multi-word units are characterised by their clear grammatical and semantic properties). (Jónsson 2009, S. 168)

Dass Jónsson hier von »klaren grammatischen und semantischen Strukturen« spricht, deutet darauf hin, dass er sich tendenziell mit dem von Stabilität geprägtem Zentrum der Phraseologie befasst und die Themen Variabilität und Polysemie der Phraseme ausblendet.

3.3.6 Vernetzung und Verweise

Die Möglichkeit Informationen nicht nur sequentiell anzuordnen, sondern ihnen durch Verlinkung und Verweise eine netzartige Struktur zu geben, löste in der Anfangsphase der elektronischen Lexikographie eine gewisse Euphorie aus. Dabei wird gerne übersehen, dass es sich keineswegs um eine konzeptionelle Neuerung handelt, denn auch traditionelle Nachschlagewerke verfügen über teilweise sehr elaborierte Methoden der internen und externen Verknüpfung von Informationen. Die eigentliche Revolution besteht darin, dass die relativ mühsame Verfolgung von Verweisen im gedruckten Werk im digitalen Medium durch komfortablere Zugriffsmethoden oder dynamische Inhaltsgenerierung ersetzt wurden. In den meisten Fällen ist das, was oberflächlich den Anschein eines elektronischen Wörterbuchs erweckt, bereits ein Konglomerat aus verschiedenen Informationsquellen, die dynamisch eine virtuelle Einheit generieren.

Unterscheiden kann man im Bereich e-Lexikographie grundsätzlich, ob es sich um interne oder Ressourcen-übergreifende Relationen handelt. Bei internen Relationen sind weniger die von Interesse, die sich durch die Datentechnik ergeben, als die Verweise, die auf Relationen lexikalischer Einheiten zueinander beruhen.

Die einfachen Realisierungsmöglichkeiten führen auch dazu, dass die Bandbreite der Verweistypen erweitert wird. So bildet Yihua (2009) in der Konzeption eines bilingualen Wörterbuchs neben den üblichen Relationen (Antonymie, Synonymie) auch Hyponymie, Teil-Ganzes-Beziehungen, kollokative Beziehungen, assoziative Beziehungen, derivative Beziehungen, komparative Beziehungen im Wörterbuch ab. Für den konkreten Fall Phraseographie formuliert Heid (2007, S. 1041) im Zusammenhang mit der Frage, welche Kriterien ein gut ausgearbeitetes internes Repräsentationsformat erfüllen muss, hinsichtlich der Verlinkung den Anspruch, dass Mehrwortlexeme mit den Bedeutungsangaben ihrer Komponenten vernetzt sein müssen. Auch

plädiert er für die Verlinkung von synonymen und antonymen lexikalischen Einheiten, die auch Beziehungen zwischen Mehrwort- und Einzelwortlexemen zulässt. Zur Unterstützung des Verständnisses von Beispielangaben regt Heid (2007, S. 1041) an, lexikalische Einheiten in Beispielsätzen mit ihrem Eintrag im Wörterbuch zu verlinken, um dem Nutzer das einfache Nachschlagen unbekannter Begriffe zu ermöglichen.

Elektronische Wörterbücher lassen sich nahezu beliebig mit anderen lexikalischen und sonstigen Referenzwerken vernetzen und in nicht-lexikalische Anwendungen integrieren (Oppentocht und Schutz 2003, S. 223–226). Bothma (2011) spricht von »elektronischen Informationswerkzeugen« (*e-information tool*), die dem Benutzer durch Verlinkung den Zugang zu externen Ressourcen ermöglichen. Ein gewisser Wandel hat sich auch in der Wertung von Hyperlinks als e-lexikographische Verweismethode vollzogen. Während die Möglichkeit von Hyperlinks anfangs als präferierte Verknüpfungsmethode galt (Storrer 2001), wird inzwischen hinterfragt, ob es nicht besser ist, relevante Information direkt oder in einer separaten Schicht (Popup) anzuzeigen (Oppentocht und Schutz 2003). Fuertes-Olivera und Niño-Amo (2011, S. 181–182) konkretisieren, dass nicht nur auf lexikalische, sondern auch auf textuelle (enzyklopädische) Ressourcen verwiesen werden sollte, um beispielsweise kulturelle Besonderheiten zu verdeutlichen. Neben Hyperlinks werden auch Popup-Texte mit Erläuterungen als Funktionalität in Erwägung gezogen.

Wie viel Information tatsächlich für den Benutzer sinnvoll ist, wird in erster Linie von der lexikographischen Funktionslehre hinterfragt (H. Bergenholtz 2011; Tarp 2011), aber auch Lew (2011, S. 241–242) äußert sich kritisch gegenüber der exzessiven Nutzung von lexikographischen Daten aus verschiedenen Quellen durch Wörterbuchportale, die maschinell Inhalte zu Wörterbuchartikeln verbinden. Er begründet seine Einschätzung damit, dass diese Einträge häufig unübersichtlich und redundant sind. Derartige Internetwörterbücher weisen untereinander deutliche inhaltliche Überschneidungen auf, da sie teilweise auf dieselben Quellen zugreifen. Verlinde (2011, S. 275) stellt fest, dass Portallösungen noch nicht kritisch evaluiert wurden und auch Heid (2011, S. 301) bezweifelt, dass die Funktionalität von Wörterbuchportalen dem Nutzer einen realen Mehrwert bringt. Einerseits bergen diese Portale die Gefahr, dass zu viel Information auf zu engem Raum präsentiert wird, andererseits stellt Heid (2011, S. 302) in seinen Benutzerstudien fest, dass Informationsfülle zwar vom Benutzer als lexikographisches Qualitätsmerkmal gewertet wird, obwohl sie in der Praxis der effektiven Nutzung der lexikalischen Ressource abträglich ist.

3.3.7 Multimedia

Der Einsatz von multimedialen Daten ist im elektronischen Medium in ganz anderer Weise möglich als im gedruckten Werk. Neben dem erweiterten Einsatz der aus der Printlexikographie bekannten Grafiken zur Visualisierung von schwer beschreibbaren Phänomenen wurden bislang nur Audio-Dateien zur Sprachausgabe in größerem Stil genutzt, um dem L2-Lerner den Zugang zu phonetischer Information ohne Kenntnis der linguistischen Kodierung (z. B. IPA) zu ermöglichen (Oppentocht und Schutz 2003). In einem weiteren Schritt kann neben der reinen Sprachausgabe die Visualisierung der auditiven Eigenschaften verwendet werden, um Vergleiche zwischen L2-Lernern und Muttersprachlern anzustellen (Schryver 2003, S. 166). Auch die Einbindung multimedialer Korpusbeispiele ist möglich und wird teilweise bereits praktiziert (Schryver 2003, S. 169).

Grafiken haben sich bewährt, um komplexe Zusammenhänge darzustellen, die nur schwer in Worten fassbar sind. Animationen können Prozesse illustrieren, wie beispielsweise die Strichfolge beim Schreiben chinesischer Schriftzeichen (*kanji*) (Apel 2011, S. 41–43).

Lew (2011, S. 245–246) beobachtet einen Aufwärtstrend bei der Einbindung von Grafiken und Audio-Daten im elektronischen Lexikon. Studien zeigen, dass sich illustrative graphische Elemente positiv auf den Lerneffekt von Phrasemen auswirken (Lew und Szczeaniak 2011; Miura 1996). Im Unterschied hierzu zeigen die Ergebnisse der Untersuchungen von Boers et al. (Boers, Piquer Píriz u. a. 2009; Boers, Eyckmans und Stengers 2007) keine signifikante Verbesserung der Lernerfolge durch grafische Elemente. Dies lässt sich möglicherweise durch die Kopplung von grafischen Elementen mit etymologischen Angaben in dem Versuchsaufbau erklären.

Obwohl aus Perspektive der Übersetzer Bedarf an multimedialen Inhalten in elektronischen Ressourcen besteht (Bowker 2013, S. 398–390), haben diese laut Umfrageergebnissen einen geringen Stellenwert bei den Nutzern (Müller-Spitzer, Kopleinig und Töpel 2013, S. 446–448). Auch Lew (2011, S. 246) äußert Bedenken zum Thema Verwendung von animierten Grafiken und Video-Aufzeichnungen in Wörterbüchern und Lernmedien. Er stützt sich dabei auf eine eigene Studie, die zeigt, wie sich die Konsultation von Videosequenzen zu Gesichtsausdrücken im Nachschlagewerk negativ auf den Lernerfolg auswirkt (Lew und Doroszevska 2009). Möglicherweise spielt hier ähnlich wie im Bereich der Illustrationen die Art und Relevanz der Daten eine wichtige Rolle.

3.3.8 Grammatische Angabe

Die grammatische Angabe wird zur Abstraktion phraseminterner Strukturen und kontextueller Gegebenheiten genutzt. In ihrer einfachsten Form entspricht sie der Angabe des Phrasemtyps, die an Stelle der Wortart-Angabe für Einzelwortlexeme tritt. Sie kann überdies zur Formalisierung phrasemimmanenter oder kontextueller Information dienen. Eingesetzt werden diese Angaben für im Hintergrund ablaufende Prozesse der natürlichen Sprachverarbeitung oder als Orientierungshilfe für den Nutzer.

Zu den computerlinguistischen Einsatzbereichen gehört die Berechnung von Variabilität und die Optimierung der Suche nach unbekanntem Phrasemen. So extrahieren Quasthoff, F. Schmidt und Hallsteinsdóttir (2010) aus einer Liste häufiger Phraseme eine Menge typischer POS-Tag-Muster, um die Suche nach Neologismen zu stützen. Jónsson (2009) beschreibt überdies die Möglichkeit, durch die Angabe syntaktischer Informationen Relationen zwischen lexikalischen Einheiten zu erschließen.

Eine kodifizierte Form der internen Phrasenstruktur und externen Valenz kann dem Nutzer die Möglichkeit bieten, die Einbettung eines Phrasems auf einen Blick zu erfassen, ohne sich mit Beispielsätzen oder Belegen auseinandersetzen zu müssen. Heid (2007, S. 1039) schlägt vor, die Kodifizierung von Phrasemen analog zu der von Einzelwortlexemen vorzunehmen.

- remind <SUBJ₁ OBJ₂ of-OBJ₃>
- in Erinnerung bringen <SUBJ₁ IND-OBJ₃ OBJ₂>

Diese Art der Darstellung setzt ein relativ hohes sprachliches Niveau und grammatisches Verständnis voraus, da aus der formalisierten Darstellung nicht hervorgeht, an welcher Stelle in der Struktur das Phrasem verankert ist.

3.3.9 Adaptivität hinsichtlich Nutzergruppe und Nutzungssituation

Benutzeradaptivität ist ein Punkt, in welchem die elektronische Lexikographie nur sehr wenig aus der lexikographischen Tradition schöpfen kann. Zwar kennt auch die Printlexikographie unterschiedliche Druck-Ausgaben der Wörterbücher für verschiedene Nutzergruppen, aber die Anforderungen der Computerlinguistik gehen inzwischen deutlich über die bisherigen Möglichkeiten hinaus. Die Besonderheit ist, dass im elektronischen Medium mehrere parallele

Darstellungen möglich sind und Anpassungen der Benutzungsoberfläche auch an Unterschiede in der Funktionsweise gekoppelt sein können.

Im einfachsten Fall wird dem Nutzer die Möglichkeit gegeben, durch Anpassung bestimmter Umgebungsvariablen die Ressource auf seine Bedürfnisse auszurichten. Tarp (2011, S. 63–64) betont, dass man bei Anpassungsfragen die Anforderungen individueller Nutzer vor Augen haben sollte. Obwohl er die Abstraktion nach Typen von Nutzern für durchaus sinnvoll und notwendig erachtet, gibt er zu bedenken, dass in der Praxis immer der einzelne Nutzer mit individuellen Bedürfnissen das Wörterbuch benutzt.

Die Benutzeradaptation kann auf verschiedenen Wegen erfolgen. Meist kann der Benutzer aktiv die Präsentation der Inhalte auf der Benutzeroberfläche in elektronischen Medien verändern, es ist jedoch ebenfalls möglich, unterschiedliche vorgeformte Profile für verschiedene Benutzergruppen zu hinterlegen, die diese als Ausgangspunkt für ihre eigenen Anpassungen verwenden können.

Eine äußerst ambitioniertes Ziel ist die Stratifizierung, die verschiedene Benutzergruppen durch Bereitstellung unterschiedlicher Sprachen oder Sprachniveaus beziehungsweise Schwierigkeitsgrade in Metatexten und Hauptteil berücksichtigt (Schryver 2003, S. 184–185).

Tarp (2011, S. 63–64) unterscheidet drei Individualisierungsmethoden: interaktive (durch Ausfüllen eines Benutzerprofils), aktive (Erstellung eines individuellen Master-Eintrags) und passive (durch das Sammeln von Nutzerdaten). Er schlägt eine Kombination aus allen Individualisierungsmethoden vor, die sich an den Nutzungssituationen orientieren.

Lew (2011, S. 235–236) hält die Entwicklung des *Louvain English for Academic Purposes Dictionary* (LEAP)⁴², welches Inhalte in Abhängigkeit der Muttersprache des jeweiligen Benutzers und der gewählten Domäne anpasst, für besonders vielversprechend. Er plädiert dafür, fortgeschrittene Benutzer mittels Kontrollfunktionen selbst entscheiden zu lassen, welche Informationen sie benötigen, indem man ein Ein- und Ausblenden mikrostruktureller Einheiten ermöglicht. Tarp (2011) sieht in diesem Zusammenhang auch mögliche Probleme, denn er geht davon aus, dass in Zukunft die Frage an Gewicht gewinnen wird, wie man im Zusammenhang mit Nutzeradaptivität mit Information umgeht, die der Benutzer irrtümlicherweise nicht als wichtig einstuft.

⁴² Das LEAP-Wörterbuch ist in einer Beta-Version einem begrenzten Nutzerkreis zugänglich (vgl. <http://www.uclouvain.be/en-322619.html>).

3.3.10 Benutzerpartizipation und Qualitätskontrolle

Eine Besonderheit von internetbasierten Wörterbüchern ist die Möglichkeit, durch Online-Datenerhebung das Wissen der Nutzergemeinde als Basis für die Weiterentwicklung zu nutzen. Dräger und Juska-Bacher (2010) stellen eine Methode für den Einsatz von Online-Werkzeugen vor, die insbesondere auf die Erhebung von Daten zur arealen Verteilung, Bekanntheitsgrad und aktuellen Verwendung von Phrasemen abzielt. Die Erfassung erfolgt durch Umfragen und durch die Einbindung interaktiver Komponenten in das phraseologische Nachschlagewerk. Mit Hilfe der Online-Umfragen können gezielt Informationen von Testpersonen erfragt werden, jedoch ist die sorgfältige Erarbeitung der Fragebögen und die Wahl der Probanden ausschlaggebend für die Qualität der Ergebnisse. Die zweite Möglichkeit, in der eine existente Ressource um eine interaktive Komponente erweitert wird, ermöglicht dem Benutzer die direkte Beteiligung am lexikographischen Prozess. Der Unterschied dieses Entwurfs zu den vielen kollaborativ erstellten Referenzwerken ist, dass der Kernbereich des Lexikons von Experten (Phraseographen) erarbeitet wird und die Daten, die von Nicht-Experten zur Verfügung gestellt wurden, entsprechend kenntlich gemacht werden. Ein weiteres bemerkenswertes Anliegen der Entwickler dieser Methode ist, auf die Zitierfähigkeit der sich ständig verändernden Datenbestände zu achten (Dräger und Juska-Bacher 2010, S. 175–176). Auch Almind, H. Bergenholtz und Vrang (2006) räumen der Benutzerpartizipation einen wichtigen Stellenwert ein. Neben dem Faktor der Datenerhebung sehen sie auch den Vorteil einer gewissen Nutzerbindung, denn die Herausbildung eines loyalen Benutzerstamms kann gegebenenfalls auch für die Verbreitung »ihres« Nachschlagewerkes förderlich sein. Nesi (2013, S. 376) äußert Skepsis gegenüber kollaborativ erstellten Wörterbucheinträgen von Idiomen, da diese aufgrund ihrer komplexen Bedeutung und häufigen humoristischen Verwendung in den Medien meist problematisch seien.

Die Frage nach der Sicherung gewisser qualitativer Standards im Lexikon ist gerade im Zusammenhang mit Benutzerpartizipation besonders wichtig. Die Computerlexikographie bietet Methoden der (semi-)automatischen Qualitätskontrolle, wie beispielsweise Möglichkeiten der exhaustiven Kontrolle der Verweisstruktur, der Kontrolle über die Kohärenz der Sprache, die Verifizierbarkeit des Fehlens versteckter Einträge, die Möglichkeit, die definitorische Sprache zu vereinheitlichen, die Überprüfung der Konsistenz der Informationen sowie das Vermeiden von zirkularen Definitionen (Calzolari 1989; Oppentocht und Schutz 2003).

3.3.11 Zusammenfassung

Die Zusammenstellung der Anforderungen an die Phraseographie aus computerlinguistischer Perspektive zeigt vor allem, dass keine Einigkeit herrscht, wie die Blaupause des *lexikographischen Rolls-Royce* der Zukunft aussehen sollte. Die wenigen vorhandenen Nutzerstudien zeigen die Unterschiede zwischen lexikographischem Ideal und praktischer Nutzung von Wörterbüchern sehr deutlich auf. Ein Beispiel hierfür ist die Suchfunktionalität: Während Lexikographen sich ausgeklügelte Suchmasken erträumen, in welchen gezielt Information abgefragt werden kann, möchte die Masse der Nutzer ein Wörterbuch benutzen wie populäre Internet-Suchmaschinen.

Eine zentrale Frage ist, welche Zielgruppe und auch welches Nutzerbild bei der Entwicklung eines Wörterbuchs zu Grunde gelegt wird. Dies zeigt sich auch bei der Thematik der Individualisierungsmethoden. Unklar ist, in welchen Bereichen man dem Benutzer Gestaltungsfreiheit gewähren soll und wie man verhindern kann, dass der Nutzer Einstellungen tätigt, die für ihn nicht vorteilhaft sind.

Ein Punkt, in welchem ein Umdenken stattzufinden scheint, ist die Frage, wieviel Information dem Benutzer zugemutet werden kann. Hier zeichnet sich ein Wandel ab von der Einstellung, dass das, was möglich ist, auch realisiert werden sollte (Portale, Vernetzung), hin zur Beschränkung auf wesentliche Information (Tarp 2013). Zwar lassen sich Benutzer intuitiv durch Masse beeindruckend, aber die bisherigen Studien sind Momentaufnahmen. Ob sich die Erfolgsquote nicht doch auf die längerfristige Nutzung auswirkt, ist noch nicht erforscht.

Die Möglichkeiten für Innovation und deren Evaluation dieser sind noch lange nicht ausgeschöpft. Der rasche Wandel der Paradigmen zeigt die Flexibilität der einschlägigen Forschung, jedoch bleibt als Wermutstropfen die Erfahrung, dass sich Benutzer unter Umständen schwer tun im Umgang mit Ressourcen, die sich signifikant von den gängigen Werkzeugen unterscheiden (Verlinde 2011, S. 275). Will man sich von den bekannten Paradigmen lösen, muss man sich nicht nur mit der Funktionalität selbst auseinandersetzen, sondern auch mit der Frage beschäftigen, wie man Benutzer an neue Funktionalität heranführen kann, und zwar vor dem Hintergrund, dass nur ein geringer Prozentsatz der Benutzer die Hilfetexte konsultiert (Heid 2011).

Auch wenn sich die automatische Extraktion von Beispielen aus Textkorpora in der Praxis bislang nicht bewährt hat (Kilgarriff und Kosem 2013, S. 46–48), ist zu erwarten, dass zukünftig korpuslinguistische Methoden eine größere Rolle in der Phraseographie spielen werden. Als

besonders hilfreich im lexikographischen Prozess gelten Werkzeuge zum Erstellen sogenannter Word Sketches, die typische Verwendungsweisen aus Korpusdaten extrahieren (Rundell 2013, S. 22–23). Derzeit führen Probleme wie Divergenz in der Phraseologie-Auffassung, Probleme mit Variabilität und Modifikation, Polysemie sowie das Fehlen angemessener Korpora dazu, dass die verfügbaren Methoden kein vollständiges Abbild der phraseologischen Realität erstellen können. Auf diesem Gebiet wird sich aller Voraussicht nach schon in der nahen Zukunft viel verändern. So bleibt zu hoffen, dass auch die phraseologische Forschung, und insbesondere auch die Phraseographie, von künftigen Entwicklungen profitieren wird.

3.4 Datenformate und Standards in der elektronischen Lexikographie

Standards gelten in der Computerlinguistik als die Grundvoraussetzung um Interoperabilität und Nachhaltigkeit linguistischer Daten und Werkzeuge zu gewährleisten. Sie können als Blaupause bei der Entwicklung neuer Sprachressourcen dienen, aber nicht selten werden sie unabhängig von der internen Kodierung als Austauschformate mit dem Ziel der Interoperabilität definiert. Obwohl sie einen hohen Stellenwert einnehmen, ist die Anzahl etablierter Standards der Texttechnologie überschaubar.

Unumstritten sind Unicode und XML als Standards zur Zeichenkodierung und zur Textauszeichnung. Abgesehen davon gibt es für die meisten Gebiete der Computerlinguistik Standardisierungsbestrebungen. Etablierte Standards sind jedoch eine Seltenheit. Auch in der elektronischen Lexikographie konkurrieren unterschiedliche Formate um den Status als Standard. Zwei ambitionierte Versuche lexikalischen Ressourcen eine einheitliche Struktur zu geben, die Text Encoding Initiative (TEI) P5 Richtlinien und das Lexical Markup Framework (LMF) der International Standardization Organization (ISO), werden in den folgenden Kapiteln vorgestellt.

Die Richtlinien der Text Encoding Initiative sind im Bereich narrative Texte weit verbreitet und widmen sich seit der Version P5 auch der Kodierung elektronischer Wörterbücher. Das Lexical Markup Framework ist der Versuch der International Standardization Organization, durch ein internationales Experten-Gremium einen ISO-Standard für lexikalische Ressourcen zu definieren.

Die Vorstellung dieser Standards ist aufgrund der Beschränktheit der verfügbaren Dokumentation und fehlender Anwendungsbeispiele auf eine exemplarische Zusammenfassung der wesentlichen Merkmale für den Bereich Phraseographie reduziert.⁴³

3.4.1 Allgemeine Standards für die Textkodierung

Bei Unicode und XML handelt es sich nicht um konkurrierende, sondern um sich ergänzende Standards. Während Unicode die Grundlage für die Kodierung von Zeichen schafft, ermöglicht XML die Kodierung semantischer Meta-Information und hierarchischer Strukturen auf textueller Ebene. Ergänzt werden diese durch die ISO-Normen für Kurzbezeichnungen der Sprachen (ISO 639-2) und Schriftsysteme (ISO 15924). Im Folgenden wird nur auf die zentralen Charaktermerkmale der Standards eingegangen, die für die darauffolgenden Kapitel von Bedeutung sind.

3.4.1.1 Unicode

Vor der Einführung von Unicode waren nur die 128 Zeichen des ASCII-Standards herstellerbeziehungsweise plattformunabhängig. Die für die Erweiterung des lateinischen Alphabets definierten ISO-8859-x Zeichenkodierungen, die den für die Repräsentation des Englischen notwendigen Zeichensatz um diakritische Zeichen erweitern, wurden nach Sprachen beziehungsweise den darin jeweils verwendeten Zeichensatz in 15 Untergruppen untergliedert. Da in den Untergruppen dieselben Kodierungspunkte mit unterschiedlichen Zeichen belegt sind, sind diese ISO-8859-x Zeichenkodierungen nicht untereinander kombinierbar.

⁴³ Kommerzielle Anbieter versuchen, ihre Datenformate zu schützen, indem sie die Spezifikation nicht offen legen. Bedauerndwert ist, dass in diesem Zusammenhang beispielsweise das EPWING-Format nicht in die Diskussion eingebracht werden kann. Hierbei handelt es sich um einen Standard in der japanischen Lexikographie, der von mehreren Verlagen und Firmen unterstützt wird. Dieses Format wurde zwar für die Open Source Gemeinde mit Hilfe von *reverse engineering* erschlossen, eine hinreichende Dokumentation dieses Standards ist jedoch nach meinem Erkenntnisstand leider nicht verfügbar.

Die Mission des Unicode Konsortiums⁴⁴ war deshalb, jedem Zeichen eine eindeutige Nummer zuzuweisen, die unabhängig von der Plattform, dem Programm oder der Sprache genutzt werden kann. Durch die Entwicklung des Unicode-Standards ist es so gelungen, für die konsistente Repräsentation von Text eines Großteils der existierenden Schriftsysteme zu sorgen. Der Unicode 5.2.0 Standard umfasst ca. 100.000 Zeichen, die von der ISO 10646 Arbeitsgruppe und dem Unicode Konsortium ausgewählt wurden, um die Schriftzeichen der heutzutage wichtigen Sprachen (*major languages* Unicode 2010) abzubilden. Darunter fallen neben den europäischen alphabetischen Schriften auch die komplexen Schriftsysteme des Nahen Ostens und des asiatischen Raums. Die Kodierung des Unicode-Zeichensatzes kann wahlweise in UTF-8, UTF-16 und UTF 32 erfolgen, wobei sich UTF-8 als Unicode-Kodierung im Internet zu etablieren scheint (möglicherweise aufgrund der Kompatibilität zu der ASCII-Zeichenkodierung).

Wenn Unicode als etablierter Standard bezeichnet wird, geschieht dies aus westlicher Perspektive, denn dort ist Unicode als einer der wenigen universalen Standards für Datenaustausch und Datenverarbeitung unverzichtbar geworden. Dabei wird oft vergessen, dass in den asiatischen Ländern – allem voran in Japan – dieser Standard nicht unumstritten ist. Der Hauptkritikpunkt der asiatischen Länder an Unicode ist die aus ihrer Sicht unnötige Han-Vereinheitlichung, das heißt die Zusammenlegung von historischen und regionalen Varianten von Schriftzeichen (Glyphen) zu einer orthographischen Einheit oder einem Graphem. So gibt es insbesondere in Japan alternative multilinguale Zeichenkodierungssysteme, die darauf abzielen, die Unterschiede zwischen den Sprachen durch eindeutige Kodierung auf Glyphen-Ebene vorzunehmen. Das bedeutet natürlich eine signifikante Vergrößerung des Zeichensatzes. So enthält die kommerzielle Implementierung eines Konkurrenzprodukts von TRON »Cho Kanji 3« über 170.000 Schriftzeichen.

Inwieweit es sich hierbei um ein politisches Problem handelt, ist schwer einzuschätzen. Tatsache ist, dass aufgrund der Han-Vereinheitlichung die Entwicklung einer einheitlichen Schriftart für den Unicode-Zeichensatz, der auch historische Unterschiede zwischen den Schriftzeichen der asiatischen Sprachen berücksichtigt, nicht möglich ist. Trotz dieser Vorbehalte und Einschränkungen ist auch in Japan bei den online verfügbaren Wörterbüchern Unicode (UTF-8) zumindest im Bereich multilingualer Wörterbücher üblich.

⁴⁴ Das Unicode Konsortium ist eine Non-Profit-Organization die sich ausschließlich durch Mitgliedsbeiträge finanziert. Zu den Mitgliedern zählen sowohl einzelne Personen als auch Unternehmen und Organisationen.

Obwohl Unicode sich als Zeichenkodierung im WWW für multilinguale Ressourcen weitgehend durchgesetzt hat, werden auch im europäischen Raum in der Textkodierung weiterhin die sprachspezifischen ISO-8859-x Zeichenkodierungen verwendet (Kilgarriff und Kosem 2013, S. 32–33). Und auch in Japan sind die in ähnlicher Weise einschränkenden Shift-JIS, ISO-2022 und EUC-JP Kodierungsstandards noch weit verbreitet.

3.4.1.2 XML

XML steht für *Extensible Markup Language* und bezeichnet eine Auszeichnungssprache, die es auf rein textueller Basis ermöglicht, Daten hierarchisch zu strukturieren. 1998 wurde sie als vereinfachte (eingeschränkte) Version der SGML (*Standard Generalized Markup Language*) vom W3C-Konsortium eingeführt und hat sich insbesondere in Verbindung mit Unicode zu einem Standard der Textkodierung entwickelt.

Der Erfolg der Auszeichnungssprache XML liegt in ihrer Einfachheit. Da es sich auch bei den Metadaten um reine Textdaten handelt, reicht ein Texteditor aus, um XML-Datei zu bearbeiten. Die Gliederung der Daten erfolgt durch textuelle Auszeichnung, die eine semantische und hierarchische Auszeichnung der Daten ermöglicht. Hierfür verwendet XML Auszeichnungsregeln, mit welchen der Inhalt mit Meta-Information angereichert werden kann. XML ist dabei sehr flexibel. Die wichtigste Regel ist, dass für die Elemente eine eindeutige hierarchische Struktur definiert werden muss. Dadurch entsteht eine Baumstruktur, in welcher ausgehend von einem Wurzelement Elemente einander gleich-, über- oder untergeordnet sind.

Die Kennzeichnung der Elemente im XML-Dokument erfolgt durch spitze Klammern (< und >), wobei innerhalb der Klammern beliebig definierbare Bezeichner stehen. Für die Einheit *Wörterbuchartikel* definiert beispielsweise TEI ein Element `entry`, der Inhalt wird dabei von Anfangs-Tag `<entry>` und End-Tag `</entry>` eingegrenzt. Leere Elemente werden in der Form `<inhaltsleer/>` realisiert. Mit Hilfe von Attributen innerhalb der Klammerung können Elemente genauer spezifiziert werden (`<entry id='1'>`).

In der XML-Spezifikation sind nur Regeln für die Wohlgeformtheit definiert, beispielsweise, dass es genau ein Wurzelement geben muss und die Elemente terminiert und eine sorgfältig hierarchisch gegliedert sein müssen. Um ein spezifisches XML-Format zu definieren, in welchem vor allem fest gelegt ist, welche Elemente erlaubt sind und an welcher Stelle sie auftreten dürfen, bedarf es einer Spezifikation mit Hilfe eines benutzerdefinierten Schemas. Hierfür stehen verschiedene Schema-Sprachen zur Verfügung (Document Type Definition (DTD)),

Relax-NG, XML-Schema). Diese Schemata können anderen Benutzern oder der Öffentlichkeit zugänglich gemacht werden, um Schnittstellen oder Standardformate zu definieren. Die bekanntesten Auszeichnungssprachen auf Basis von XML sind XHTML als Alternative für HTML im Internet und das Open Document Format (ODF) für diverse Textverarbeitungssoftware. Der Zugriff auf XML-Daten gilt als wenig performant, insbesondere wenn es sich um hochgradig strukturierte Daten handelt, da er entweder über eine Pfad-Abfrage (XPath) oder eine Suche (XQuery) erfolgt. Zwar gibt es Datenbanklösungen für XML, tendenziell wird jedoch XML als Standard für narrativ-textuelle Daten und für Webservices mit geringen Datenvolumen präferiert. Für granulare Daten, insbesondere für performanz-abhängige Datenbank- oder Applikationsschnittstellen sowie Webservices mit hohem Datenaufkommen, werden alternative Datenaustauschformate eingesetzt (u. a. JSON, JSONP, YAML).

3.4.2 Standardisierungsbestrebungen der Computerlexikographie

Für die lexikographische Modellierung lexikalischer Daten in elektronischen Wörterbüchern konnte sich bislang noch kein Standard durchsetzen. Hinsichtlich der Verfügbarkeit wäre nahelegend, sich an den Richtlinien des Standards der Text Encoding Initiative zu orientieren, da es sich um einen offenen Standard handelt. Das Lexical Markup Framework definiert ein Format, das sich zum Ziel setzt, die Modellierung lexikalischer Daten für alle Medien zu ermöglichen. Der Nachteil dieses Standards ist jedoch, dass es sich um ein kommerzielles Produkt handelt, das als Spezifikation von der International Standardization Organization unter der ISO Nummer 24613:2008 angeboten wird.⁴⁵

In den folgenden Kapiteln werden die Standards TEI-Dictionaries und LMF anhand der verfügbaren Dokumentation kurz skizziert und mit Hilfe von Beispielen mit Blick auf ihre Spezifikation für die Phraseographie untersucht.

⁴⁵ In dieser Arbeit wird auf die offen erhältliche Version des Standards in Form des finalen Entwurfs (FDIS = Final Draft International Standard) zurückgegriffen.

3.4.2.1 Das Wörterbuchformat der Text Encoding Initiative

Die Text Encoding Initiative (TEI) ist eine gemeinnützige Organisation bestehend aus Mitgliedern der wissenschaftlichen Gemeinschaft, die für sozial- und geisteswissenschaftliche Disziplinen Richtlinien zur Kodierung von digitalen Texten entwickeln und pflegen. Die TEI Richtlinien (derzeit in der Version P5 1.3.0 vorliegend) stoßen seit Mitte der 1990er Jahre auf breite Akzeptanz im Bibliotheks- und Verlagswesen, bei den Museen und nicht zuletzt in der Sprachwissenschaft. Die Version P5 (1.3.0) der TEI Richtlinien verwendet XML 1.0 als Markup und stellt Strukturvorgaben in Form von Dokumenttypdefinitionen (DTD) und diversen Schemata (RelaxNG, Schematron) zur Verfügung.

Die Spezifikation besteht aus nahezu 500 Elementen, die nach Möglichkeit die Kodierungsanforderungen aller Textsorten in allen Sprachen abdecken sollen. In der Praxis kommt bei der Kodierung einer Textressource immer nur eine kleine Auswahl des Bestandes zum Einsatz, die sich der Benutzer seinen Bedürfnissen entsprechend zusammenstellt. Die TEI inhärente Flexibilität basiert auf dem Versuch, alle möglichen Darstellungsweisen des gedruckten Wörterbuchs modellieren zu können (I. Schmidt und Müller 2001, S. 40). In dieser Flexibilität mag der Grund für die weite Akzeptanz der TEI-Richtlinien liegen.

Der Anwender kann auf verschiedene Sektionen der TEI-DTD zugreifen und sich seine DTDs an seine individuellen Bedürfnisse anpassen. Auf oberster Ebene unterscheidet man Elemente, die Metainformationen zu den Texten enthalten (TEI Header) und solche, die die Textstruktur und -inhalte widerspiegeln.

Der TEI-Header entspricht in etwa einer Titelseite in einem gedruckten Werk. Auffällig ist, dass der Header keine Informationen zur Kodierung des Dokuments enthält, da diese in dem Schema des Dokuments spezifiziert werden. Wie Listing 1 zeigt, umfasst er überwiegend Meta-Daten zum Dokument, die im bibliothekarischen Kontext für die Einordnung des Werkes von Bedeutung sind.

```
<TEI xmlns:xi="http://www.w3.org/2001/XInclude" xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Ulrich Apel's WaDokuJT</title>
        <respStmt>
          <resp>Originally by</resp>
          <name>Ulrich Apel</name>
        </respStmt>
      </titleStmt>
    </fileDesc>
  </teiHeader>
</TEI>
```

```

<extend>about 230.000 headwords</extend>
<publicationStmt>
  <availability status="cc">
    <p>Licence: Creative Commons</p>
    <p>Commercial licence available</p>
  </availability>
  <pubPlace>
    <ref target="http://www.wadoku.eu">http://www.wadoku.eu</ref>
  </pubPlace>
</publicationStmt>
<sourceDesc>
  <p>Originally based on local encoding.</p>
  <p>Content was successfully converted to Unicode.</p>
</sourceDesc>
<encodingDesc>
  <projectDesc>
    <p>This dictionary is financially supported by the WaDoku e.V.!</p>
  </projectDesc>
</encodingDesc>
</fileDesc>
</teiHeader>
[...]
```

Listing 1: Beispielmodellierung eines TEI-Headers für das APEL-WADOKUJT-IW

Für die Kodierung eines Dokuments nach den TEI-Richtlinien stehen zahlreiche Elemente und Attribute zur Verfügung. Um die Auswahl zu erleichtern, sind diese in Module gegliedert. Neben den generischen Modulen (*tei header*, *core component* and *text structure*) gibt es auch eine Auswahl von für Dokument-Typen spezifischen Elementen. Für Wörterbücher gibt es dementsprechend ein Modul *Dictionaries*.

Das Modul war bis zu der vierten Version der TEI-Richtlinien mit »Print Dictionaries« betitelt. In der 2008 publizierten Version P5 wurde das Kapitel in »Dictionaries« umbenannt und der Anspruch erhoben, dass man die Kodierung eines breiten Spektrums mono- und multilingualer Wörterbücher sowie lexikalischer Ressourcen sowohl für den Print- als auch für den Online-Bereich unterstützt. Gleichzeitig wird in der Kapiteleinleitung auf die Schwierigkeit hingewiesen, die durch die sehr unterschiedlichen Nutzer-Anforderungen entstehen und man behält sich vor, zukünftig Erweiterungen und Änderungen an der Spezifikation für Wörterbücher vorzunehmen (TEI Consortium 2008, S.251).

Die Ausweitung der Richtlinien auf die elektronische Lexikographie erfolgt ohne Typisierung in digitale und gedruckte Medien. Es handelt sich nicht um ein neues Format, sondern um

die Kodierung von elektronischen Wörterbüchern entlang der Linien, die die lexikographische Tradition vorgibt. So stehen beispielsweise für Umtexte eine Reihe von Elementen zur Verfügung, mit Hilfe derer das Wörterbuch in vorangestellte Umtexte (<front>), Hauptteil (<body>) und nachgestellte Umtexte (<back>) aufgeteilt wird. Im elektronischen Wörterbuch spielt jedoch die Anordnung der Wörterbuchteile keine Rolle.

Innerhalb des Hauptteils (<body>) kann zwischen drei Artikeltypen gewählt werden. Einzelne Wörterbuchartikel werden mit dem <entry>-Element modelliert, das Element <superEntry> bietet die Möglichkeit, mehrere Wörterbuchartikel in einem Artikel zu inkludieren. Da die TEI Spezifikationen den Selbstanspruch haben, die Heterogenität des gesamten Gebietes abbilden zu können, stellt das Konsortium als Alternative zu <entry> das Element <entryFree> zur Verfügung, das weitgehend freie Gestaltungsmöglichkeiten einräumt.

Das Element »entry« gibt eine an der klassischen Lexikographie orientierte Struktur vor. Ein Beispiel für den Wörterbucheintrag für *kubi* (*Kopf, Hals*) aus dem WaDokuJT modelliert auf Basis des <entry>-Elements zeigt Listing 2.

```
<entry version="1.1" id="620155" xmlns="http://www.wadoku-idioms.de/xml/entry">
  <form type="lemma">
    <orth xml:lang="kanji">首</orth>
    <orth xml:lang="kanji" type="antiq">頸</orth>
    <orth xml:lang="hiragana">くび</orth>
    <orth xml:lang="katakana">クビ</orth>
    <pron notation="latin">kubi</pron>
    <stress>0</stress>
  </form>
  <gramGrp>
    <pos>n</pos>
  </gramGrp>
  <sense n="1">
    <cit type="translation" xml:lang="de">
      <quote>Kopf</quote>
      <gen>m</gen>
    </cit>
    <cit type="translation" xml:lang="de">
      <quote>Haupt</quote>
      <gen>n</gen>
    </cit>
  </sense>
  <sense n="2">
    <cit type="translation" xml:lang="de">
      <quote>Hals</quote>
      <gen>m</gen>
    </cit>
    <cit type="translation" xml:lang="de">
      <quote>Nacken</quote>
      <gen>n</gen>
    </cit>
  </sense>
</entry>
```

```

</sense>
<sense n="3">
  <cit type="translation" xml:lang="de">
    <usg type="style">metaphor</usg>
    <def>Entlassung</quote>
    <gen>f</gen>
  </cit>
</sense>
<entry>

```

Listing 2: Beispielmodellierung des Eintrags für *kubi* mit den Daten des APEL-WADOKUJT-IW nach den TEI P5 Richtlinien

Das Beispiel ist nur eine Möglichkeit, den bilingualen Wörterbucheintrag in TEI zu kodieren. Dennoch illustriert es den klassischen Aufbau eines Wörterbuchartikels mit Lemma-Angabe, Formangabe, Grammatikangabe und semantischem Kommentar. Der Unterschied zur Vorgängerversion (P4) der TEI-Richtlinien manifestiert sich vor allem im semantischen Kommentar, genauer gesagt in der Angabe von Äquivalenten. Die Elemente für Übersetzungsäquivalent (<tr> und <trans>) wurden durch Referenz- und Zitatangaben (<cit> und <quote>) ersetzt, wobei das <cit>-Element durch das Attribut <translation> und die Angabe der Sprache markiert wird. Die Markierung stammt aus dem allgemeinen Teil der TEI Richtlinien. Darin erschöpfen sich auch schon die strukturellen Änderungen zwischen den Versionen P4 und P5 hinsichtlich der Kodierung von Wörterbüchern. Das bedeutet, dass sich die nominelle Ausweitung der Richtlinien auf elektronische Nachschlagewerke nicht wie zu erwarten auch in inhaltlicher Veränderung widerspiegelt.

Für die Kodierung von Idiomen gibt es in den Richtlinien keine Vorgaben oder Beispiele. Kollokationen können mit dem Element <colloc> in der grammatischen Angabe (<gramGrp>) eingebunden werden, aber andere Phrasem-Formen sind nicht vorgesehen. Eine Möglichkeit wäre, sie mit Hilfe von <superEntry> in die Artikelstruktur zu integrieren. In den Richtlinien werden jedoch nur Homographen als Anwendungsbeispiele für dieses Element genannt. Eine andere Möglichkeit wäre, Idiome als separate Einträge zu realisieren und mit dem Haupteintrag oder allen Komponenten zu verlinken.

Abbildung 3 zeigt eine Möglichkeit, den Eintrag für *kubi-ni naru* (*den Job verlieren*) mit Hilfe der TEI P5 Richtlinien zu modellieren. Dabei wurden die Inhalte des entsprechenden Eintrags

im APEL-WADOKUJT-IW in einem vollwertigen Eintrag modelliert und die Form als *phrase* ausgewiesen.⁴⁶

```
<entry version="1.1" id="1612206" xmlns="http://www.wadoku-idioms.de/xml/entry">
  <form type="phrase">
    <orth xml:lang="Jpan">首になる</orth>
    <orth xml:lang="hiragana"><small>くびになる</small></orth>
    <phon notation="latin">kubi ni naru</phon>
  </form>
  <sense>
    <cit type="translation" xml:lang="de">
      <quote>entlassen werden</quote>
    </cit>
    <cit type="translation" xml:lang="de">
      <quote>gefeuert werden</quote>
    </cit>
    <cit type="translation" xml:lang="de">
      <quote>auf die Straße gesetzt werden</quote>
    </cit>
    <cit type="translation" xml:lang="de">
      <quote>hinausfliegen</quote>
    </cit>
  </sense>
  <ref target="620155" type="component"><small>首</small></ref>
</entry>
```

Listing 3: Beispielmmodellierung des Eintrags für *kubi-ni naru* mit den Daten des APEL-WADOKUJT-IW nach den TEI P5 Richtlinien

TEI bietet zusätzlich die Möglichkeit typographische Auszeichnung zu definieren, um genaue Angaben über die exakte Darstellung eines gedruckten Wörterbucheintrags abzubilden. Auch ohne deren Einsatz erinnert die Beispiel-Modellierung an die Mikrostruktur gedruckter Wörterbücher, die linear-sequentiell bestimmte Informationstypen abarbeitet.

Zusammenfassend lässt sich sagen, dass der TEI-Standard versucht, sowohl das oberflächliche Erscheinungsbild als auch die inhaltlichen Zusammenhänge in einer Struktur abzubilden. Dies geschieht unter anderem durch den Einsatz von Paradigmen aus der Textlinguistik, wie beispielsweise *translation* zur Auszeichnung von Äquivalenten. Insbesondere die digitale Lexikographie wird nicht durch dezidierte Strukturen oder Elemente repräsentiert.

⁴⁶ In der Original-Daten (XML-Version vom 8.1.2012 heruntergeladen von www.wadoku.de) wird *kubi-ni naru* als Untereintrag modelliert und als *Verwendungsbeispiel* markiert.

Im Bereich der elektronischen Wörterbuchmodellierung wird der Standard bislang in einigen nicht-kommerziellen Online-Wörterbüchern verwendet. Große lexikographische Projekte, die TEI P5 für elektronische Wörterbücher einsetzen, sind FreeDICT⁴⁷ und das DWDS-Wörterbuch⁴⁸ der Berlin-Brandenburgischen Akademie der Wissenschaften.

Gründe, die gegen den Einsatz sprechen, sind die Kopplung an das XML-Format, das sich bei hochgradig granularen Datenstrukturen ungünstig auf das Datenvolumen auswirkt und von streng hierarchischen Daten ausgeht. Auch wird negativ bewertet, dass nach den TEI-Richtlinien die Struktur unterspezifiziert, potentiell redundant und unübersichtlich ist (Fraser 2011). Und nicht zuletzt definiert TEI nicht ein Wörterbuchformat, sondern das Werkzeug zur Definition eines solchen. So kann eine Basierung auf den TEI-Richtlinien alleine nicht die Interoperabilität mit anderen Ressourcen garantieren.

3.4.2.2 Das Lexical Markup Framework

Das Lexical Markup Framework ist insofern vielversprechend, als es sich unter anderem zum Ziel setzt, Sprachen mit multiplen Orthographien – wie auch das Japanische – und starre, semi-starre und flexible Mehrwortausdrücke zu repräsentieren (Francopoulo u. a. 2009, S. 59). Die folgende Zusammenfassung gibt einen kurzen Überblick über die Gestaltungsprinzipien des Rahmenwerks, konzentriert sich im Detail jedoch auf die Realisierung der Aspekte, die für die Repräsentation von Mehrwortausdrücken in einem japanisch-deutschen Lexikon notwendig sind.

Das Lexical Markup Framework (LMF) definiert einen Standard in Form eines Meta-Modells, um die Erstellung, Verwaltung und Interoperabilität beziehungsweise Weiterverwendbarkeit von lexikalischen Ressourcen für die natürliche Sprachverarbeitung zu verbessern. An der Entwicklung des Standards ISO 24613:2008 durch das Technical Committee 37 Subgroup 4 (TC37/SC4) der International Standards Organization (ISO) waren mehr als 60 Wissenschaftler aus unterschiedlichen Ländern beteiligt (u. a. Gil Francopoulo, Monte George und Nicoletta Calzolari). Erklärtes Ziel des Komitees war es, basierend auf bestehenden Ressourcen die besten Gestaltungskonzepte und Methoden herauszuarbeiten und in ein Modell zu überführen, das

⁴⁷ Verfügbar unter www.freedict.org.

⁴⁸ Verfügbar unter www.dwds.de.

hochgradig modular ist und an die unterschiedlichsten Arten von elektronischen lexikalischen Ressourcen angepasst werden kann.

Wie andere vom TC37 entwickelte Standards gliedert sich die Spezifikation des Rahmenwerks in zwei Ebenen: LMF selbst definiert Strukturen als Klassen, genauer gesagt deren Bezeichnung, Verwendung und Beziehungen zueinander (high level specification). Für die inhaltliche Ebene werden in LMF die Kategorien des ISO Data Category Registry (DCR, Revision von ISO 12620) eingebunden (low level specification). Hierfür wird ein Data Category Subset (DCS) aus vorhandenen Kategorien selektiert, das um benutzerdefinierte Kategorien erweitert werden kann. Die Registrierung eigener Kategorien im DCR wird empfohlen, um die Austauschbarkeit der Daten sicherzustellen. Zwar werden in den Beispielen der LMF-Spezifikation Attribute präsentiert, LMF definiert jedoch keine Eigenschaften, sondern legt nur fest, an welcher Stelle Attribut-Wert-Paare aus bereits vorhandenen Vokabularen eingebunden werden können.

Die Struktur des LMF ist mittels (statischen) UML-Diagrammen spezifiziert, die Klassennamen, Relationen zwischen Klassen und Zugehörigkeit zu Paketen definieren. Eine Document Type Definition (DTD) ermöglicht die Validierung von XML-Datenstrukturen. Auffällig ist darin das *feat*-Element, welches durch generische Attributierung die Verwendung beliebiger Kategorie-Werte-Paare aus dem DCR erlaubt. Modellierungsbeispiele finden sich in den Spezifikationen selbst sowie in Francopoulo u. a. (2009) und Francopoulo u. a. (2006).

Das LMF gliedert sich in ein Kernpaket (Core Package), das die generischen Klassen lexikalischer Informationen enthält, und Erweiterungen für die Modellierung spezifischer lexikalischer Datenbestände. In der Kernkomponente sind die für alle lexikalischen Ressourcen relevanten Klassen (z. B. *Lexical entry*, *sense* und *form*) und deren Relationen zueinander definiert. Die Erweiterungen können nur in Zusammenhang mit dem Kernpaket verwendet werden. Von dem TC37/SC4 entwickelte Erweiterungen machen Vorgaben, die anwendungsorientiert (Machine Readable Dictionaries), inhaltsorientiert (Morphologie, Constraint Expressions) oder eine Kombination aus beidem sind (NLP Semantic, NLP Syntax, NLP Multilingual Notation, etc.). Im Folgenden sollen die Kernkomponente und anhand von Beispielen die Erweiterungen vorgestellt werden, insofern sie für ein bilinguales phraseologisches Wörterbuch von Bedeutung sind. Aufgrund der sehr eingeschränkten Modellierungsmöglichkeiten der konkret für maschinenlesbare Wörterbücher konzipierten Komponenten, werden dabei auch NLP-Komponenten in Betracht gezogen.

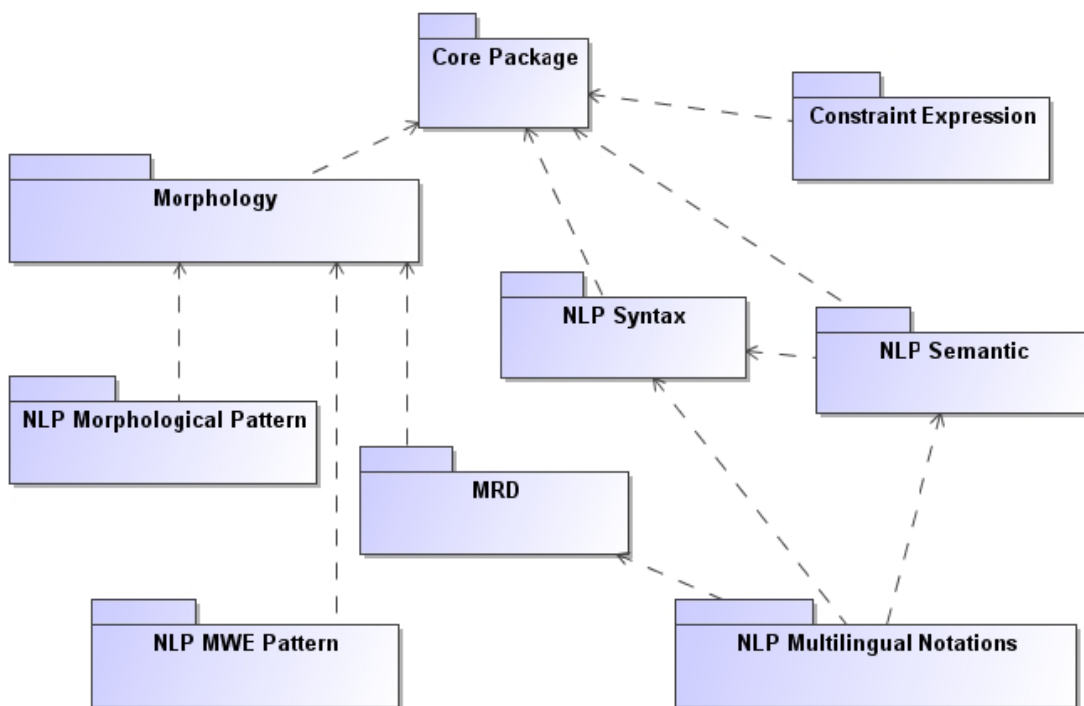


Abbildung 3.8: Komponentenstruktur des LMF (LMFWorking Group 2008, S. 18).

In der Kernkomponente werden die wesentlichen Klassen des Lexikons definiert: Im Zentrum steht die lexikalische Ressource (*Lexical Resource*), die ein oder mehrere Lexika (*Lexicon*) enthalten kann. Metainformation über die verwendeten Standards in der lexikalischen Ressource werden in der Klasse *Global Information* festgehalten. Die einzige weitere obligatorische Komponente ist die abstrakte *Form*-Klasse, mit welcher Lexeme, morphologische und orthographische Varianten, Morphe und deren Eigenschaften (Aussprache, Silbentrennung, Lemmata) modelliert werden.

Ebenfalls zur Kernkomponente gehört der Komplex rund um die Semantik, in welchem multiple und potentiell rekursive Lesarten (*Sense*) angelegt und ihren korrelierenden Definitionen (*Definition* und *Text Representation*) zugeordnet werden. Die *Statement*-Klasse steht ergänzend zur Definition für weitere Spezifikationen zur Verfügung.

Die Modellierung konkreter Inhalte ist an den Einsatz von Attributen geknüpft, die im LMF bewusst unterspezifiziert sind. Da die Modellierung (insbesondere die LMF-DTD) an Stelle der Attribut-Spezifikationen Platzhalter einsetzt, ist die inhaltliche Gestaltung kaum Restriktionen unterworfen. Eine deutliche Vorstellung von dem, was sich die Entwickler des LMF vorgestellt haben, erhält man insbesondere durch die Beispielm modellierungen im Annex der LMF-Spezifikation. Modelliert man einen Eintrag aus einem einsprachig japanischen Wörterbuch anhand der Spezifikation und der Beispiele, erhält man eine hierarchische Struktur aus Meta-Information, Wörterbucheintrag, Form und Bedeutungsangabe, die der TEI-Struktur nicht unähnlich ist (vgl. Abbildung 3.9). Auffällig ist, dass in der Meta-Information keine Angaben zur Genese des Werkes vorgesehen sind. Statt dessen fokussiert *Global Information* Details der Kodierung der gesamten Ressource, indem auf den Ebenen Sprache, Skript und Zeichen auf entsprechende Standards verwiesen wird.

Auf die besonderen Bedürfnisse nicht-europäischer Sprachen geht die LMF-Spezifikation verschiedentlich in Beispielinstanzen ein. Durch den Einsatz der ISO-Normen für Sprachkodierung gibt es die Möglichkeit, zwischen Repräsentationen in unterschiedlichen Schriftsystemen zu unterscheiden. Standardmäßig wird dort als Wert des Attributs *script* mit *Hani* angegeben (LMFWorking Group 2008, S. 26–26), was laut ISO 15924 Schriftzeichen chinesischen Ursprungs umfasst. Eine Ausdifferenzierung der in Japan gebräuchlichen Schriftzeichen (die auch eine geringe Menge rein japanischer Schriftzeichen (*kokuji*) umfasst) ist nicht vorgesehen. Für die *kanji-kana*-Mischformen existiert außerdem *Jpan* als Alias für *Hani*, *hiragana* und *katakana*.

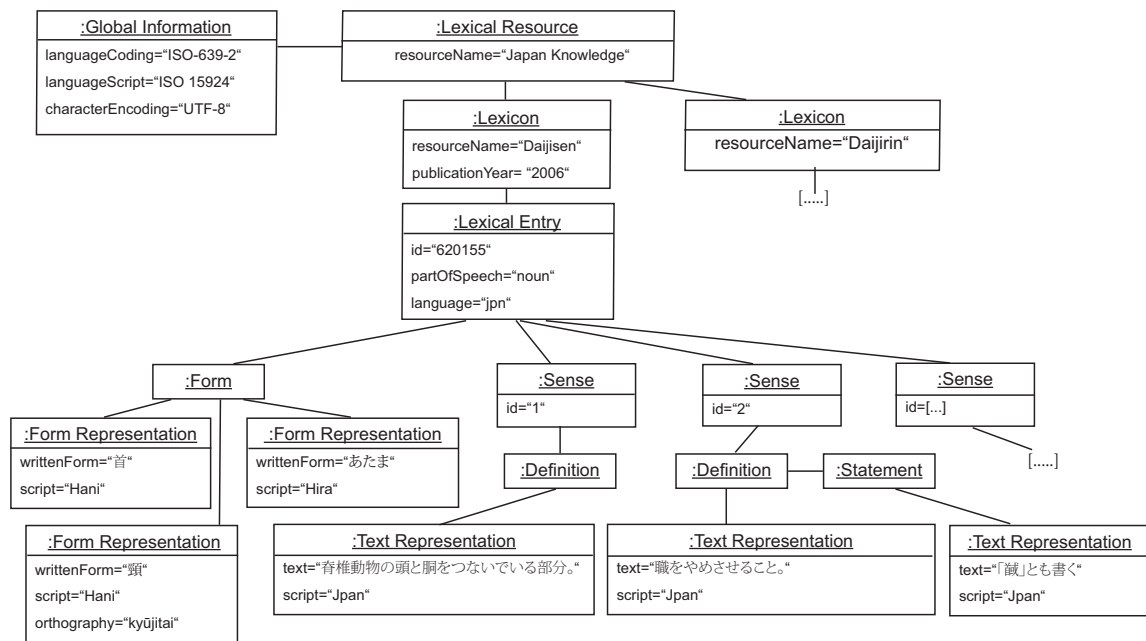


Abbildung 3.9: Beispielmodellierung des Eintrags *kubi* (Hals) aus dem SHOGAKUKAN-DAIJISEN-IW mit Hilfe der Kernkomponente.

Um Schriftzeichen zu markieren, die nicht zum Bestand der *jōyō-kanji*⁴⁹ gehören, wurde in dem Anwendungsbeispiel das Attribut *orthographyName* ergänzt. Im SHŌGAKUKAN-DAIISEN-PW wird diese Information durch ein hochgestelltes Kreuz neben dem Schriftzeichen 𪛗 kodifiziert. Eine komplexere Herausforderung bei der Modellierung des Wörterbuchartikels stellt die Kodierung der von einer Lesart (*Sense*) abhängigen, abweichenden Formrepräsentation »*kubi*«-*to-mo kaku* (「𪛗」とも書く »wird auch 𪛗 geschrieben«) mittels der Klassen und Relationen der Kernkomponente dar. Eine Möglichkeit ist die variante Form in der *Statement*-Klasse zu verankern (siehe Abbildung 3.9). Die Klasse ist für heterogene Informationen vorgesehen, die die Bedeutungserklärung ergänzen (LMFWorking Group 2008, S. 43). Um die Information wie üblich auf Ebene des Formkommentars zu verankern, müsste diese Lesart aufgrund ihrer abweichenden Orthographie in einem separaten Eintrag kodiert werden.

Das Morphologie-Paket gehört zu den nicht anwendungsorientierten Erweiterungen, die weder speziell für den Einsatz in der natürlichen Sprachverarbeitung noch mit Blick auf maschinenlesbare Wörterbücher konzipiert wurden. Das Modell ergänzt die Kernkomponente insbesondere im Bereich der abstrakten Form-Klasse. So wird zwischen der im Wörterbuch verzeichneten Grundform *kubi-ni naru* (*Lemma*), dem Stamm *kubi-ni na* (*Stem*) und in verschiedenen Kontexten vorkommenden abweichenden Formen wie beispielsweise die Vergangenheitsform *kubi-ni natta*, in Wortform-Klassen (*WordForm*) unterschieden. Neben den genannten grammatischen Formen können durch die Klasse *Related Form* auch typisierte Relationen zwischen lexikalischen Einträgen modelliert werden. Das kann im Fall von Derivation genutzt werden, um zum Beispiel eine Verbindung zwischen einem Mehrwortlexem wie *hana-no shita-o nagaku suru* (鼻の下を長くする *unter der Nase lang machen* als Sinnbild für eine lüstern wirkende Mimik) und seiner nominalisierten Form *bikachō* (鼻下長) herzustellen.

Ebenfalls Bestandteil der morphologischen Erweiterung sind die Klassen *ListOfComponents* und *Component* zur Realisierung des Aggregatzustands von lexikalischen Einträgen der Mehrworteinheiten (auch *MWU* für *multi word units*). Die Reihenfolge der Komponenten einer Mehrworteinheit wird durch die sequenzielle Anordnung der Komponenten (*Component*) in der übergeordneten Klasse Komponentenliste (*ListOfComponents*) impliziert. Die einzelnen

⁴⁹ Als *jōyō-kanji* werden die Schriftzeichen bezeichnet, die offiziell in Gebrauch sind. Verpflichtend ist diese Beschränkung auf derzeit 2.136 *kanji* jedoch nur für Schulen, Medien und Behörden.

Komponenten verweisen auf Instanzen lexikalischer Einträge (*Lexical Entry*). Die direkte Referenzierung der Form- oder Semantik-Subklassen ist nicht vorgesehen (siehe Abbildung 3.10).

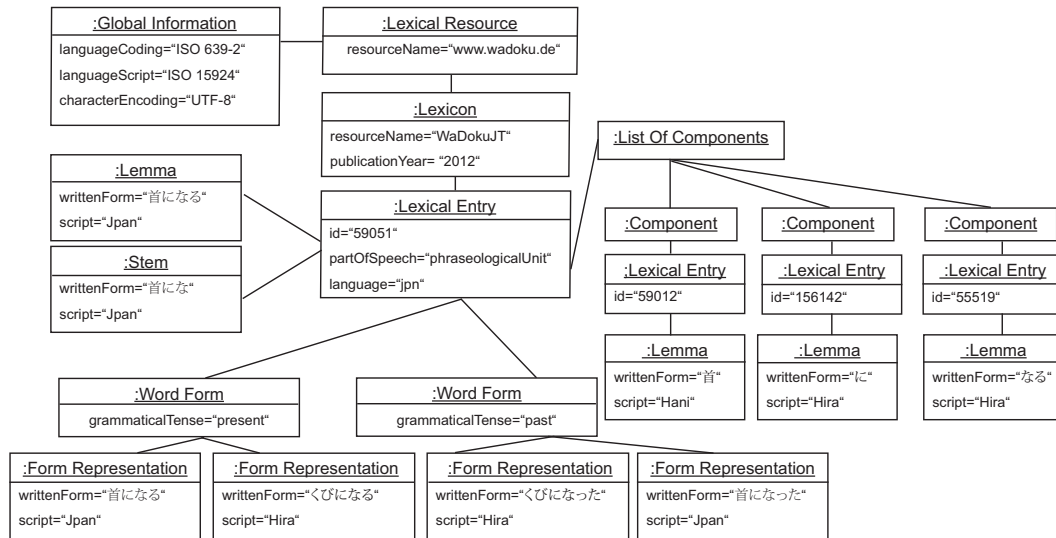


Abbildung 3.10: Beispielmodellierung des Eintrags *kubi-ni naru* (entlassen werden) aus dem WADOKUDE-IW mit Hilfe der morphologischen Erweiterung

Durch die Komponentenliste wird für den gesamten Eintrag (*LexicalEntry*) die Abfolge fest gelegt. Bedeutungsgleiche Varianten mit optionalen Elementen wie beispielsweise *kubi-o yoko-ni furu* (首を横に振る den Kopf seitwärts schütteln) und *kubi-o furu* (首を振る den Kopf schütteln) müssen in separaten Einträgen kodiert werden, da *yoko-ni* nicht als optional markiert werden kann. Dies gilt auch für lexikalische Varianten, in welchen eine Komponente durch eine bedeutungsähnliche oder ein ganzes Paradigma an ähnlichen Begriffen ersetzt werden kann.

Folglich muss für jede variante Form eine eigene Instanz *Lexical Entry* angelegt werden. Die bereits erwähnte Klasse für verwandte Formen (*Related Forms*) ist nicht geeignet, da diese keine direkte Beziehung zu der Bedeutungsangabe des Eintrags impliziert (LMFWorking Group 2008, S. 22).

Ebenfalls nicht vorgesehen ist die Referenzierung einzelner Lesarten polysemer Komponenten. Da in diesem Entwurf die einzige Verbindung zwischen Mehrwortlexem und Komponente auf

Ebene der *Lexical Entry*-Klassen erfolgt, können beispielsweise metaphorische Lesarten von Komponenten nicht zu dem Mehrwortlexem in Beziehung gesetzt werden.

Die MRD (*Machine Readable Dictionaries*) Erweiterung fokussiert die Kodierung von Strukturen, die in für den menschlichen Benutzer konzipierten mono- und bilingualen elektronischen Wörterbüchern eine Rolle spielen. Der Nutzen für die maschinelle Sprachverarbeitung ist hier sekundär. Im Zentrum steht dabei die Modellierung semantischer Information, die durch Kontexte, Domänen und Äquivalente angereichert werden kann.

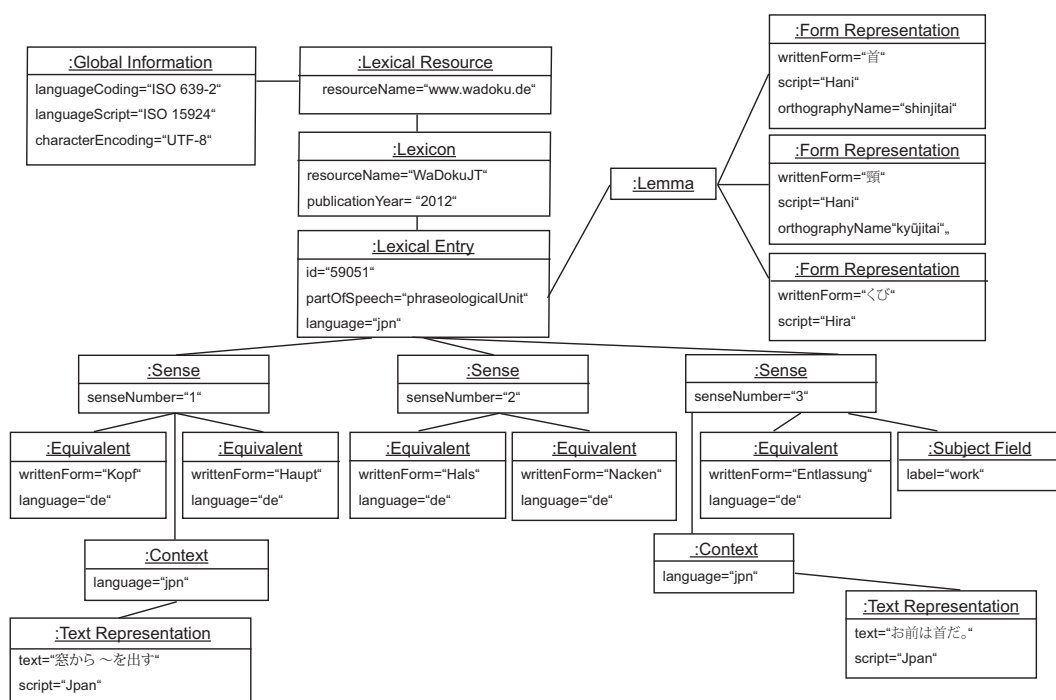


Abbildung 3.11: Beispielkodierung des Eintrags *kubi* aus WADOKUDE-IW mit Elementen der MRD Erweiterung des LMF.

Die Kontext-Klasse (*Context*) bietet die Möglichkeit, Beispielangaben in verschiedenen Sprachen und Skripten zu realisieren. Eine Kontext-Instanz ist jeweils einer Lesart (*Sense*) zugeordnet und kann mit mehreren Textrepräsentationen assoziiert sein, die verschiedenen orthographischen Varianten entsprechen. Für das Japanische bedeutet dies, dass die *kanji-kana*-Mischschrift beispielsweise durch phonetische Repräsentationen in *kana* oder lateinischer Umschrift ergänzt werden kann. In Beispiel 3.11 entspräche dies den Textrepräsentation in *kanji-kana*-Mischform 窓から首を出す, *hiragana*-Umschrift まどからくびをだす und lateinischer Schrift *mado-kara kubi-o dasu*. Nicht vorgesehen ist die Angabe einer Übersetzung (den Kopf zum Fenster hinausstecken), da die Klasse *Context* über das Sprachattribut an eine Sprache gebunden ist. Statt dessen können Relationen zwischen Kontexten über die Klassen *Context Axis* und *Context Axis Relation* kodiert werden (siehe die Erweiterung *NLP multilingual notations extension*).

Für bi- oder multilinguale Wörterbücher können einer Lesart (*Sense*) Instanzen der Äquivalenz-Klassen (*Equivalent*) zugeordnet werden. Die Null-Äquivalenz ist laut Spezifikation für monolinguale Wörterbücher vorgesehen (LMF Working Group 2008, S. 30). In der Praxis ist diese aber aufgrund fehlender entsprechender Begrifflichkeiten in der Zielsprache (lexikalischer Lücken) eine auch in bilingualen Kontexten relevante Konstellation. Abhängigkeiten zwischen Äquivalenz und Kontext werden ebenfalls indirekt über *Context*-Relationen hergestellt.

Die Erweiterung *NLP multilinguale Notation* (*NLP multilingual notations extension*) spezifiziert Zwei- und Mehrsprachigkeit, indem sie Relationen auf syntaktischer und semantischer Ebene auf Achsen abbildet. Die drei wesentlichen Klassen sind *Sense Axis* zur Verbindung zweier verwandter Bedeutungen, *Transfer Axis* zur Kodierung syntaktischer Argumente und *Context Axis* zur Verbindung von lexikalischen Einheiten mit typischen Beispielkontexten. Um divergierende Bedeutungsnuancen in den verschiedenen Sprachen zum Ausdruck zu bringen, gibt es die Möglichkeit, diese in den Attributen der zugehörigen *Relations*-Klassen *Sense Axis Relation* und *Transfer Axis Relation* zu notieren. Darüber hinaus können externe Interlingua-Komponenten (*Interlingual External Ref*) referenziert und Konditionen für Ausgangs- und Zielsprache angegeben werden (*Source Test* und *Target Test*). Einen Überblick über die Konzepte und Einsatzmöglichkeiten der Erweiterung *NLP multilinguale Notationen* findet sich in Francopoulo u. a. (2006).

Besonders vielversprechend für die Kodierung von Phrasemen im Wörterbuch ist die Erweiterung NLP Notationen für Mehrwortausdrücke (*NLP multiword expression patterns*) zur Abbildung interner grammatischer Strukturen. Das LMF bedient sich hierbei der Phrasenstrukturgrammatik um diverse Formen von Wortkombinationen, angefangen bei Komposita über Phrasen bis hin zu ganzen Sätzen, zu beschreiben. Bedingung für den Einsatz der Erweiterung ist die Verknüpfung des lexikalischen Eintrags (*Lexical Entry*) mit einer Komponentenliste (*List Of Components*). Die mit *MWE Pattern* assoziierten lexikalischen Komponenten (*MWE Lex*) werden auf Basis ihrer Anordnung in *List Of Components* referenziert. Die Muster selbst repräsentieren in der Klasse *MWE Pattern* einen bestimmten Typ oder ein bestimmtes Phänomen lexikalischer Kombinatorik, in welcher syntaktische Konstituenten in Knoten (*MWE Node*) zusammengefasst und mittels Achsen (*MWE Edge*) mit den Komponenten (*MWE Lex*) verbunden werden. Den Einsatz der *NLP multiword expression patterns* illustriert Abbildung 3.12. Das Beispiel zeigt, dass die Zuordnung der lexikalischen Elemente des Musters (*MWE Lex*) zu

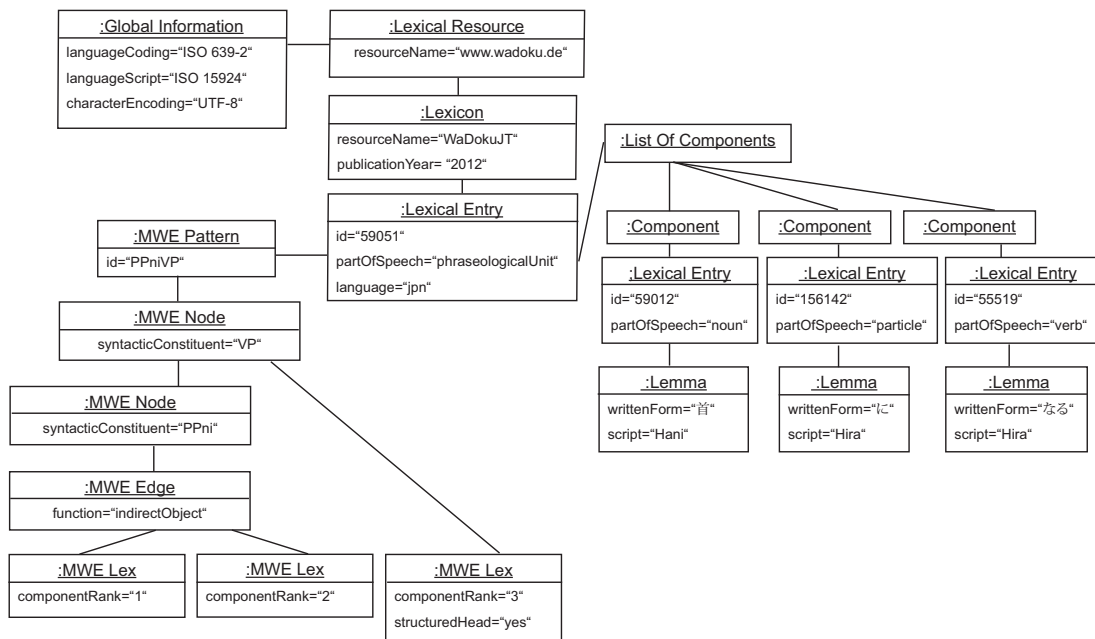


Abbildung 3.12: Mögliche Instanziierung des Mehrwortlexems *kubi-ni naru* mit Hilfe der Erweiterung *NLP multiword expression patterns*.

Komponenten (*Component*) über ein Attribut definiert ist, das auf den Rang der Komponenten (*Component Rank*) in der Komponentenliste (*List Of Components*) referenziert. Dabei ist zu beachten, dass hier der Usus der theoretischen Phraseologie, nur Autosemantika zu den Komponenten zu rechnen, nicht gilt, da für grammatische Strukturen gerade auch Synsemantika eine zentrale Rolle spielen.

Das Lexical Markup Framework bietet einen umfassenden Bausatz von Komponenten für verschiedene Einsatzgebiete, die die Bedürfnisse an die Modellierung von Wörterbüchern sowohl für die natürliche Sprachverarbeitung als auch für den menschlichen Benutzer umfasst. Will man die Charakteristika von Phrasemen anhand des LMF abbilden, muss eine Auswahl diverser Erweiterungen des LMF eingesetzt werden, um einen passenden Rahmen zu generieren. Da Mehrwortlexeme wie Einzelwortlexeme im Modell eigenständige lexikalische Einheiten darstellen, steht das gesamte Repertoire an Erweiterungen zur Verfügung.

Wie praktikabel sich die Arbeit mit diesem Rahmen in der Praxis erweist, lässt sich aufgrund fehlender Umsetzungsanalysen nicht definitiv sagen. Als nachteilig kann sich jedoch erweisen, dass die meisten Relationen auf der Ebene lexikalischen Einträge erfolgen. Dies gilt auch für die Verknüpfung mit den Komponenteneinträgen (*List Of Components*). Die Anordnung der Komponenten in *List Of Components* impliziert dabei ihre Anordnung im Mehrwortlexem. Auf diese implizite Position wird auch durch die Querverweise referenziert, was als Indiz gesehen werden, dass das LMF mit Blick auf eine feste Abfolge der Komponenten entworfen wurde. Variabilität wird nur in Form regelhafter Transformationen berücksichtigt. Auch stellt sich die Frage, wie bei einer vollständigen Modellierung von Einzel- und Mehrworteinträgen Redundanzen vermieden werden können.

Die Möglichkeit, im LMF multiple Orthographien nebeneinander zu kodieren, erweist sich für das Japanische mit seinem komplexen Schriftsystem als sehr nützlich. Durch die Möglichkeit zusätzliche Attribute (*feat*) in den Textrepräsentationen festzulegen, ist diese Lösung nahezu beliebig erweiterbar, wie am Beispiel der Annotation diachroner Varianz in der Orthographie gezeigt wurde.

3.4.3 Zusammenfassung

Die Richtlinien der Text Encoding Initiative und das Lexical Markup Framework sind ähnlich aufgebaut. Es gibt jeweils eine Kernkomponente und einzelne Module, die nach Bedarf eingebunden werden. Bei genauerer Betrachtung kristallisieren sich grundsätzliche Unterschiede heraus. Das beginnt bereits bei der Wahl der Meta-Information. Geprägt von narrativen Ressourcen stellt der TEI-Header Informationen in den Vordergrund, die bei der Einordnung textueller Ressourcen in einen größeren Verbund relevant sind (z. B. Genese oder Mitautorenschaft). Das LMF fokussiert in den Beispielinstanzen der Spezifikation technische Aspekte, die für die maschinelle Verarbeitung der Ressourcen von Bedeutung sind (z. B. Zeichenkodierung, Skript oder Sprache).

Beide Standards haben das Ziel, die Strukturen möglichst aller lexikalischer Ressourcen in einem Regelsatz abzubilden. Hierfür definieren sie sowohl Elemente als auch Attribute, lassen aber gerade bei den Eigenschaften Gestaltungsspielraum.

Die Gründe hierfür sind jedoch sehr verschieden. Aus den Richtlinien der TEI geht hervor, dass bestimmte Bereiche bewusst unterspezifiziert sind, da die Erweiterungsmöglichkeiten den unterschiedlichen Anforderungen der Experten aus verschiedenen Disziplinen ein Werkzeug an die Hand geben soll, mit welchen Ressourcen nach eigenen Vorstellungen modelliert werden können.

Das LMF hingegen zielt darauf ab, ein hochgradig standardisiertes Rahmenwerk zur Verfügung zu stellen, in welchem der Benutzer überwiegend durch Selektion seine lexikalische Ressource konzeptioniert. Hierfür wird das LMF durch die terminologischen ISO-Standards komplettiert. Die Tatsache, dass die Vorgaben des LMF unterspezifiziert sind, liegt insbesondere daran, dass die Entwicklung der terminologischen Ressourcen nicht abgeschlossen ist. So ist das DCR nur mit einem Bruchteil der Datenkategorien bestückt, die für die lexikographische Auszeichnung nötig sind. Eine Kodierung ausschließlich mit Datenkategorien, die bereits im DCR enthalten sind, war bereits für die wenigen Illustrationsbeispiele dieses Kapitels nicht möglich. Dadurch ergibt sich zwar Gestaltungsspielraum, diese Flexibilität geht jedoch auf Kosten der Interoperabilität.

Ein weiterer wesentlicher Unterschied zwischen den Standards manifestiert sich in der Intensität, mit welcher sie sich der Kodierung von Phrasemen im Lexikon widmen. Während die TEI-Richtlinien nur die Einbindung von Kollokationen berücksichtigen, macht das LMF detaillierte Vorgaben für den gesamten Phänomenbereich.

Nicht zuletzt unterscheiden sich die Standards hinsichtlich ihrer Verfügbarkeit. Während TEI in der akademischen Welt und der Open-Source-Gemeinde weithin bekannt ist, bleibt in Bezug auf den ISO-Standard LMF abzuwarten, ob dessen Nutzung nicht aufgrund seines kommerziellen Charakters eine Grenze gesetzt ist.

3.5 Phraseographische Praxis in Online-Wörterbüchern

Um die phraseographische Praxis im Bereich digitale Lexikographie zu eruieren, werden exemplarisch drei Ressourcen vorgestellt, die jeweils einen Wörterbuchtyp repräsentieren. Als bilinguale Ressource für das Sprachpaar Japanisch-Deutsch wird zunächst die Online-Version des APEL-WADOKUJT-IW beschrieben. Als zweites Beispiel wurde das Wörterbuchportal JapanKnowledge und insbesondere das darin enthaltene SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW als digitale Version des SHŌGAKUKAN-NIHONKOKUGODAIJITEN-PW₂ gewählt. Deziert phraseologische Online-Wörterbücher sind bislang selten. Aufgrund seines onomasiologischen Aufbaus interessant ist das dritte Werk: die digitalisierte Version des Lehr- und Übungsbuchs »Deutsche Redewendungen« (ETTINGERPHRASEOLOGIEDE-IW).

3.5.1 Ulrich Apels WaDokuJT

Das APEL-WADOKUJT-IW ist das derzeit umfangreichste und aktuellste japanisch-deutsche Wörterbuch im digitalen Format. Aufgrund seiner nun zehnjährigen Webpräsenz kann man es wohl zurecht als das meist genutzte Nachschlagewerk auf seinem Gebiet bezeichnen. Es entstammt der Initiative von Ulrich Apel, der aufgrund der unbefriedigenden lexikographischen Erfassung seines Forschungsgebiets (japanische Zukunftsforschung und Zukunftsplanung) begann, einen Fachwortschatz zu erstellen. Im Jahr 1999 stellte er diesen als Dateidownload im proprietären Datenbankformat (Filemaker) im Internet zur freien Nutzung und Erweiterung zur Verfügung. Durch die Unterstützung von Thomas Latka wurde auf WADOKUDE-IW auch eine Online-Nutzung der Daten möglich. Apel stieß mit seinem kollaborativen Ansatz auf ein positives Echo und die Internet-Gemeinde nutzte und erweiterte das Wörterbuch bereitwillig. Aufgrund seiner direkten Partizipationsmöglichkeiten entwickelte sich insbesondere WADOKUDE-IW zu einem beliebten Community-Wörterbuch.

Das ursprüngliche Fachwortschatzprojekt wurde zu einem vollwertigen Wörterbuch weiterentwickelt, das unter anderem morphologisches, semantisches, phonetisches, pragmatisches,

etymologisches und in Teilen auch enzyklopädisches Wissen umfasst. Der derzeitige Datenbestand von ca. 100.000 Lemmata beziehungsweise 270.000 Datensätzen entspricht in etwa dem Umfang des sechsbändigen gedruckten GROSSESJAPANISCHDEUTSCHESWÖRTERBUCH-PW⁵⁰, welches am Deutschen Institut für Japanforschung (DIJ) kompiliert wird.

Seit 2007 wird die Weiterentwicklung und Bereitstellung der Daten des APEL-WADOKUJT-IW von einem gemeinnützigen Verein (Wadoku e.V.) unterstützt, der ebenfalls ein Webinterface (WADOKUEU-IW) zur Verfügung stellt. Neben den zwei bereits erwähnten Online-Wörterbüchern sind die Daten im Internet in diversen Versionen im Umlauf und wurden bereits verschiedentlich in Wörterbuchportale oder lexikalische Anwendungen integriert.

Beide Versionen werden regelmäßig aktualisiert, die im folgenden beschriebenen Grundprinzipien sind jedoch über die letzten Versionen weitgehend konstant geblieben. Da sich die Plattform für WADOKUEU-IW noch im Aufbau befindet, stammen die folgenden Beispiele für die Beschreibung des APEL-WADOKUJT-IW überwiegend aus Thomas Latkas Version WADOKU-DE-IW.

3.5.1.1 Datenstruktur

Obwohl sich Datenstruktur und Datenbestände der verschiedenen Versionen des APEL-WADOKUJT-IW auseinander entwickelt haben, gleichen sich auf struktureller Ebene die Wörterbücher hinsichtlich ihres Umgangs mit Idiomen. Die Datenstruktur unterscheidet grundsätzlich zwischen Haupteinträgen und Untereinträgen. Jeder Untereintrag ist wie im gedruckten Wörterbuch einem Haupteintrag zugeordnet. Typische Haupteinträge sind Einzelwortlexeme. Mehrwortlexeme und Anwendungsbeispiele werden als Untereinträge kodiert.

Der gesamte Datenbestand wird in einer einzigen Tabelle gepflegt. Zur Unterscheidung werden die Typen entsprechend markiert. Zu jedem Eintrag gehören phonetische, orthographische, morphologische und semantische Informationen. Die Inhalte sind zusätzlich durch internes Markup ausgezeichnet, um beispielsweise die Grenzen zwischen orthographischen Varianten oder einzelnen Lesarten zu markieren. Die Zuordnung von Untereinträgen zu Haupteinträgen

⁵⁰ Zum gegenwärtigen Zeitpunkt ist nur der erste Band des GROSSESJAPANISCHDEUTSCHESWÖRTERBUCH-PW erschienen, der die Lemmata mit den Anfangsbuchstaben A-I umfasst. Eine Online-Version GROSSESJAPANISCHDEUTSCHESWÖRTERBUCH-IW ist seit Anfang 2013 verfügbar, diese beschränkt sich jedoch auf die Indizierung der Lemmata des gedruckten Werks. Die Inhalte werden in Form von Grafiken präsentiert.

erfolgt durch Referenzierung der eindeutigen Identifikationsnummer. Als Untereinträge gelten alle Mehrwort-Einträge, von Komposita bis zu Beispielsätzen.

Ein entscheidendes Merkmal für die Phraseographie ist der *Worttyp*, gerade dieser ist jedoch bei Untereinträgen meist unterspezifiziert. In der Dokumentation wird kein separater Worttyp für *kanyōku* definiert. Die im Wiki befindlichen redaktionellen Hinweise hierzu sind teilweise redundant und widersprüchlich. Die begrifflich *kanyōku* am nächsten stehende Wortklasse Redensarten wird im »POS« betitelten Kapitel den *mei(shi)* (名(詞) Abkürzung für *Nomen*) und im Kapitel »Worttypen« *monku* (文句 Redensart) zugeordnet. So überrascht es, dass die meisten Einträge als *Beispiele* (例 *rei*) an Stelle von *mei* oder *monku* markiert sind. Im gesamten Datenbestand⁵¹ finden sich nur zwei Einträge mit der Kennzeichnung *kanyōku*, die etwa 30 mit *monku* gekennzeichneten Einträge sind eine Mischung aus unterschiedlichen Typen von Mehrwortlexemen (darunter auch Routineformeln und Satzbeispiele).⁵²

Darüber hinaus gibt es die Möglichkeit, deutsche Entsprechungen mit einem Worttyp in geschweiften Klammern zu annotieren. In vier Fällen wird dieses Annotationsformat genutzt, um diese als »Redew.« zu kennzeichnen.

3.5.1.2 Suchfunktionalität auf WaDokuDe-IW

Für die Suche gibt es ein zentrales Eingabefeld, mit welchem die japanischen Inhalte und die deutschen Entsprechungen durchsucht werden können. Durch die Suchoptionen können das Feld (*Japanisch, Lesung, Deutsch*) und die Art der Übereinstimmung (*beginnt mit, endet mit, enthält, stimmt exakt*) spezifiziert werden. Darüber hinaus sind die Auto-Vervollständigung und Unterscheidung von Groß- und Kleinschreibung an- beziehungsweise abwählbar.

⁵¹ Da der Worttyp nicht zu den durchsuchbaren Elementen der Online-Nutzungsschnittstellen gehört, bezieht sich diese Analyse auf die Download-Version des Wörterbuchs.

⁵² Sprichwörter werden wesentlich häufiger als solche erkannt und markiert. Im gesamten Datenbestand befinden sich 129 Einträge, die als *kotowaza* (諺 Sprichwort) markiert sind. Es ist allerdings zu vermuten, dass die Anzahl der verzeichneten Sprichwörter diese Ziffer wesentlich übersteigt.



Abbildung 3.13: Suchergebnis für *hara-no mushi* auf WADOKUDE-IW (Stand: 17.1.2013).

Die Ergebnisse werden tabellarisch nach Treffergenauigkeit präsentiert. Untereinträge sind den Haupteinträgen nebengeordnet. Ihre Zugehörigkeit wird erst ersichtlich, wenn man durch Wahl eines Eintrags in die Artikelansicht wechselt.

Die Auffindbarkeit der *kanyōku* wird dadurch gewährleistet, dass alle Einträge durchsuchbar sind, unabhängig davon, ob es sich um Haupt- oder Untereinträge handelt. Da sie überwiegend nicht als solche gekennzeichnet sind, sondern zu den Beispielen zählen, kann auf ihre spezifischen Anforderungen bei der Suche nicht reagiert werden. Das bedeutet, dass sie nur in exakt der Form gefunden werden können, in welcher sie indiziert wurden. Wie bei Einzelwortlexemen ist die vollständige Zeichenkette auch in der *hiragana*-Form erfasst, nur in wenigen Fällen sind zusätzliche Varianten kodiert. Das Ergebnis einer Suche nach einer nicht indizierten Variante liefert als Ergebnis die Bedeutung der lexikalischen Einheiten, die durch eine morphologische Analyse bestimmt wurden.

3.5.1.3 Artikelansicht für Phraseme im WaDokuDe-IW

In den wenigen Fällen, in welchen *kanyōku* intern mit der Kennzeichnung *monku* versehen sind, wird dies an der Oberfläche durch »Redensart« markiert (siehe Abbildung 3.14).



Abbildung 3.14: Artikelansicht für das mit *monku* als Worttyp gekennzeichnete *hebi-no namagoroshi* auf WADOKUDE-IW (Stand: 17.1.2013)

Wenn die deutsche Entsprechung ebenfalls markiert ist, tritt die Kennzeichnung als REDEW. an Stelle des Worttyps.



Abbildung 3.15: Artikelansicht für *hara-no mushi-ga osamaranai* auf WADOKUDE-IW (Stand: 17.1.2013)

Generell gilt, dass in der Artikelansicht Haupt- und Untereinträge miteinander verlinkt werden. Beispiel 3.15 zeigt einen Link in der Artikelansicht von *hara-no mushi-ga osamaranai*, der über dem Lemma platziert ist und auf die Komponente *hara-no mushi* verweist.

Umgekehrt erscheint *hara-no mushi-ga osamaranai* als »Beispiel« in der Detailansicht von *hara-no mushi* (auch bei den wenigen Einträgen, die mit *monku* als Worttyp gekennzeichnet sind). Zusätzlich zur Angabe der deutschen Übersetzung werden jeweils Links auf andere Sprachressourcen (*Kanjilexikon*⁵³, *Google*, *Tatoeba*⁵⁴) sowie die Inhalte der kommerziellen einsprachigen Werke SANSEIDŌ-DAIJIRIN-IW und SHŌGAKUKAN-DAIISEN-PW angezeigt. Besonders prominent ist in der Detailansicht die Kommentarfunktion. In der Regel sind *kanyōku* mit dem Haupteintrag verknüpft, der der ersten Komponente entspricht. In manchen Fällen sind bei Untereinträgen mehrere Komponenten verlinkt.⁵⁵

Die Qualität der Bedeutungserklärung von *kanyōku* im WADOKUDE-IW ist sehr unterschiedlich. Regeln für ihre Erstellung fehlen in den redaktionellen Hinweisen. Wie das Beispiel *hara-no mushi-ga osamaranai* in 3.15 zeigt, entsprechen in diesem Fall die Bedeutungserklärungen *aktualisierten Beispielen*. Auch wird der Variabilität der *kanyōku* kaum Rechnung getragen. Nur drei der vielen Varianten von *hara-no mushi-ga osamaranai* sind indiziert, und zwar neben der reinen *hiragana*-Form die *kanji-kana*-Mischformen mit *osamaranai* in *kanji* (治まらない) und in *hiragana*.

3.5.1.4 Benutzeradaptivität und Benutzerführung

Wie bereits erwähnt nimmt die Kommentarfunktion auf WADOKUDE-IW eine relativ prominente Stellung in der Artikelansicht ein. Für registrierte Benutzer ermöglicht die Plattform außerdem das Erstellen, Verwalten und Exportieren von Merklisten und Übungslisten. Eine Anleitung zur Nutzung dieser Funktion wurde unter »Hilfe« nicht gefunden. Auch können registrierte Nutzer Inhalte zum Wiki beisteuern und sich im Forum an Diskussionen beteiligen.

⁵³ Hans-Jörg Bibikos Japanisch-Deutsches Kanji-Lexikon, verfügbar unter <http://lingweb.eva.mpg.de/kanji/>

⁵⁴ Bei *Tatoeba* (www.tatoeba.com) handelt es sich um ein kollaboratives Projekt, welches Parallelkorpora für Satzbeispiele aufbaut.

⁵⁵ Ein Beispiel für die Verknüpfung mehrerer Komponenten ist der Eintrag für *hara-ga itai* (腹が痛い、Bauchschmerzen haben; sich kaputt lachen). Die idiomatische Bedeutung (*sich kaputt lachen*) ist jedoch nicht hinterlegt.

3.5.2 JapanKnowledge+

JapanKnowledge+ (ジャパンナレッジ・プラス) ist ein Beispiel für ein kommerzielles Wissens-Portal. Neben mono- und bilingualen Wörterbüchern stellt JapanKnowledge+ auch historische, biographische, geographische, enzyklopädische Nachschlagwerke sowie einzelne Datenbanken mit historischen Texten und multimedialen Inhalten zur Verfügung. Da es sich um einen kommerziellen Anbieter⁵⁶ handelt, ist der Zugriff auf die Inhalte von JapanKnowledge+ auf einen Kreis registrierter Nutzer beschränkt.⁵⁷

Die derzeit 50 in JapanKnowledge+ verfügbaren Wissensressourcen werden nach erfolgter Anmeldung in thematischen Gruppen auf der rechten Seite des Portals angezeigt. Es ist möglich, einzelne Ressourcen oder Ressourcen-Typen (Gruppen) auszuwählen. Die generische Suche über alle Ressourcen ist voreingestellt. Beim Nachschlagen werden die Ergebnisse in einem neuen Fenster angezeigt, folglich bleibt die Einstellung in dem Ausgangsfenster erhalten. Selektionen gehen jedoch verloren, sobald die Seite neu geladen wird. Wechselt man aber beispielsweise die Suchmethode (von der *One-Look* Gesamtsuche zur Expertensuche oder umgekehrt), wird die Standard-Einstellung wieder hergestellt.

Im Zentrum der folgenden Kapitel steht das SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW Wörterbuch als die digitale Version des SHŌGAKUKAN-NIHONKOKUGODAIJITEN-PW₂. Dieses verfügt als einziges Wörterbuch im Wissensportal über dezidierte Zugriffsmethoden. Für Vergleichszwecke wird das ebenfalls in JapanKnowledge+ enthaltene SHŌGAKUKAN-DAIISEN-IW als digitale Version des SHŌGAKUKAN-DAIISEN-PW herangezogen, das als einsprachig japanisches Wörterbuch mit etwa 265.000 Einträgen mit dem traditionelleren IWANAMI-KŌJIEN⁶ konkurriert. Das zweisprachige deutsch-japanisch/japanisch-deutsche Taschenwörterbuch des

⁵⁶ Das JapanKnowledge-Portal wird von der in Tōkyō ansässigen Firma NetAdvance (ネットアドバンス) angeboten, die sich auf die Bereitstellung von Wissensressourcen spezialisiert. Das Repertoire umfasst ausschließlich Nachschlagewerke und Korpora des *Shōgakukan*-Verlags.

⁵⁷ Dankenswerterweise ermöglicht die Virtuelle Fachbibliothek Crossasia der Staatsbibliothek über die japanologischen Bibliotheken einen zentralen Zugang zu JapanKnowledge+ und anderen wertvollen digitalen Ressourcen.

Portals SHŌGAKUKAN-POKETTOPUROGRESSHIBUDOKUWAWADOKUJITEN-IW eignet sich aufgrund seiner sehr begrenzten Erfassung von *kanyōku* nicht für eine Analyse.⁵⁸ Ein phraseologisches Wörterbuch ist in JapanKnowledge+ nicht enthalten.

3.5.2.1 Benutzerführung

Die Benutzerführung ist in verschiedene Bereiche untergliedert. Unter *FAQ* (*yoku aru shitsumon* よくある質問 Häufige Fragen) findet man Anleitungen für die Lösung technischer Probleme. Über der Wörterbuchauswahl verbergen sich unter den Titeln *Guide* und *tsukaikata* (使い方) die englischen und japanischen Benutzungshinweise. Sie erläutern im Wesentlichen die Suchfunktionalität des Portals. In der japanischen Version wird jedes Kapitel durch ein Video (*screencast*) ergänzt. Informationen zu den einzelnen Ressourcen und deren jeweilige Benutzungshinweise sind unter *kontentsu annai-o miru* (コンテンツ案内を見る Hinweise zum Inhalt betrachten) verlinkt.

3.5.2.2 Suchfunktionalität

JapanKnowledge+ bietet verschiedene Suchfunktionalitäten. Für die Suche über das gesamte Wissensportal steht die sogenannte *OneLook*-Suche zur Verfügung. Es handelt sich um eine einfache Suche, die auf die als Lemmata indizierten Begriffe beschränkt ist.

Zusätzlich gibt es eine Detailsuche über das Wissensportal, die ermöglicht, die Suche mehrerer Begriffe aus maximal drei Eingabefeldern mit Hilfe Boolescher Operatoren (AND/OR/NOT) zu verknüpfen. Für jedes der Eingabefelder kann der Skopus und ein Suchmuster festgelegt werden. Für die Eingrenzung des Suchbereichs stehen zur Auswahl:

1. Lemma (見出し語 *midashigo*)
2. Namensvariante (別名 *betsumei*)
3. Titel (タイトル *taiteru*)
4. Derivationen (派生語 *haseigo*)

⁵⁸ Von den 88.000 Einträgen im SHŌGAKUKAN-POKETTOPUROGRESSHIBUDOKUWAWADOKUJITEN-IW entfallen drei Viertel auf den deutsch-japanischen Teil.

5. Index (索引 *sakuin*)6. Volltext (全文 *zenbun*)

Die Bereiche *Lemma* und *Volltext* beinhalten auch die Namensvarianten und Titel der Ressourcenartikel. Für die Volltext-Suche wird immer eine partielle Suche durchgeführt, für alle anderen Bereiche kann zusätzlich gewählt werden, ob der Anfang, das Ende oder die Gesamtzeichenkette auf Übereinstimmung mit dem Suchbegriff geprüft werden soll. Die Sortierung der Ergebnisse kann durch Parameter justiert werden (vgl. Kapitel 3.5.2.3). Zusätzlich gibt es noch eine Phrasensuche (フレーズ検索 *furēzu kensaku*), die allerdings nicht mit Blick auf ostasiatische Sprachen konzipiert ist.⁵⁹

Abbildung 3.16: OneLook Expertensuche in JapanKnowledge+

Durch die Angabe von Kriterien in der Detailsuche kann gezielt auf die Inhalte der Ressource zugegriffen werden. Den Ergebnissen nach zu urteilen, entspricht die *OneLook*-Suche dem

⁵⁹ Die Phrasensuche führt dazu, dass mehrere durch Leerzeichen getrennte Worte als eine zusammenhängende Phrase gesucht werden.

kombinierten Resultat einer partiellen Suche in den Sektionen *Lemma* und *Titel* in der Detailsuche.

Für das SHÖGAKUKAN-NIHONKOKUGODAIJITEN-IW steht zusätzlich zur *OneLook*-Suche auch eine dezidierte Suchfunktionalität mit dazugehöriger Ergebnisanzeige zur Verfügung. Diese separate Ansicht wird durch einen Link mit dem Titel *kobetsu* (個別 *einzel*n) neben dem Eintrag für das SHÖGAKUKAN-NIHONKOKUGODAIJITEN-IW in der Wörterbuchliste aufgerufen.

Abbildung 3.17: Suchmaske für das SHÖGAKUKAN-NIHONKOKUGODAIJITEN-IW

Folgende Unterschiede zu der Suchfunktionalität des Gesamtportals sind bei dem Aufruf augenfällig: Zunächst gibt es als Standard nur ein Eingabefeld, auf Knopfdruck lassen sich jedoch zwei weitere (mit entsprechender Möglichkeit der logischen Verknüpfung) einblenden. Für jedes Eingabefeld kann wie in der *OneLook*-Suche Skopus und Art der Übereinstimmung gewählt werden. Die Sektionen des Skopus unterscheiden sich jedoch deutlich von der der Gesamtsuche, da sie der Mikrostruktur des Wörterbuchs entsprechen:

1. Lemma (見出し *midashi*)
2. Volltext (全文 *zenbun*)
3. Beispiele (Gesamt) (用例 (全体) *yōrei (zentai)*)
4. Beispiele (Beispieltext) (用例 (本文) *yōrei (honbun)*)
5. Beispiele (Quelle) (用例 (出典情報) *yōrei (shutten jōhō)*)
6. Dialekt (Gesamt) (方言 (全体) *hōgen (zentai)*)
7. Dialekt (variante Formen in Lemma-Angaben) (方言 (異形含む見出し) *hōgen (ikei fukumu midashi)*)

8. Wortgeschichte (語誌 *goshi*)9. Theorien zur Etymologie (語源説 *gogensetsu*)

Zusätzlich ist es möglich, in der Suche Haupteinträge (親見出し *oyamidashi*), Untereinträge (子見出し *komidashi*) und Kanji-Lesungen (字音語素 *jion goso*) ein- beziehungsweise auszuschließen.

3.5.2.3 Ergebnisanzeige

In der Ergebnisanzeige der *OneLook*-Suche werden Treffer in Nachschlagewerken und in Textkorpora separiert. Die Trefferliste ist durchnummeriert und umfasst einen Ausschnitt des Wörterbucheintrags bestehend aus Lemma/Titel, Bezeichnung des Nachschlagewerks, der Sektion, in der der Treffer erzielt wurde (z. B. *midashi* für Lemma) sowie einen Ausschnitt des Artikeltextes (abhängig vom jeweiligen Nachschlagewerk umfasst dieser ein bis zwei Zeilen). Lemma/Titel und ein auf das Artikelfragment folgender Link *tsuzuki-o yomu* (続きを読む *weiterlesen*) führen zu dem jeweiligen Eintrag.

辞典・事典系: 「腹の虫」の検索結果 6件見つかりました。

全文検索

1. はらの虫(むし) [= 中(なか)の虫(むし)] (日本国語大辞典|見出し自体)

①人体の消化器に寄生する虫。回虫(かいちゅう)の類。* 和玉篇〔150後〕「蝘 人腹中長虫 ハラノムシ」* 浮世草子・傾城色三味線〔1701〕大坂・五「腹(ハラ)の虫(ムシ)を残らずたいらげ手なみを見せん」* … (続きを読む)

2. はら-の-むし【腹の虫】(デジタル大辞泉|見出し自体)

1 腹の中にいる寄生虫。回虫の類。2 空腹時の腹鳴りを腹の中に虫がいて鳴くものとした、その虫。3 腹立ちなどの感情を腹の中にいる虫によるものとした、その虫。… (続きを読む)

3. はらの虫(むし)が=治(おさ)まらない [= 承知(しょうち)しない] (日本国語大辞典|見出し自体)

癪(しゃく)にさわって我慢ができない。腹立つ心を抑えがたい。* 人情本・春の若草〔1830～44〕初・五回「今日の様に言われちゃア何程私の様な者でも腹の虫が承知為ないヨ」* 家族会議〔1935〕〈横光利一〉… (続きを読む)

4. 腹(はら)の虫(むし)が治まら・ない(デジタル大辞泉|見出し自体)

腹が立ってがまんできない。「ここで引き下がっては一・ない」… (続きを読む)

5. はらの虫(むし)の居所(じょす)が熏(わる)い(日本国語大辞典|見出し自体)

Abbildung 3.18: Anzeige der Ergebnisse aus drei Wörterbüchern in der OneLook-Suche des JapanKnowledge+Portals

Die Ergebnisanzeige kann durch die Hierarchisierung folgender Sortierkriterien verändert werden:

- Stichworttyp (キ一種別 *kii-shubetsu*)
- Lesung (よみ *yomi*)
- Inhalt (コンテンツ *kontentsu*)
- Informationsmenge (absteigend) (情報量 (多い順) *jōhō (ōi jun)*)
- Informationsmenge (aufsteigend) (情報量 (少ない順) *jōhō (sukunai jun)*)

Die Abfolge der ersten drei Sortierkriterien dieser Auflistung entspricht der Voreinstellung. Auch auf die Anzahl der angezeigten Treffer kann in der Suchmaske Einfluss genommen werden (in den Einheiten 20, 50, 100 Treffer).

In der dezidierten Suche des SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW werden die Ergebnisse ebenfalls in Form einer nummerierten Liste ausgegeben, die jeweils das Lemma (inkl. Lesung) und einen Ausschnitt des Artikeltextes enthält, in welchem zu Anfang auch die Wortart erscheint. Eine Möglichkeit, Einfluss auf die Sortierung zu nehmen, gibt es nicht. Diese erfolgt auf Basis der Treffergenauigkeit (vollständig, initial, partiell).

3.5.2.4 Phraseme

Die Nennform der Phraseme im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW unterscheidet sich deutlich von der anderer Wörterbücher in JapanKnowledge+. Analog zu der Printversion SHŌGAKUKAN-NIHONKOKUGODAIJITEN₂ kodiert auch die Online-Version *kanyōku* in Unter-einträgen, die dem Haupteintrag der ersten Komponente zugeordnet sind. Der Haupteintrag wird nur durch einen Link kenntlich gemacht. Auch in der Online-Version werden orthographische und lexikalische Varianten in der Nennform kodiert, sofern sie die Sortierung nicht beeinträchtigen.

Beispiel 3.19 zeigt, dass die erste Komponente *hara* nicht in der orthographischen Form des Haupteintrags mit dem *kanji* (腹) dargestellt wird, sondern durch die *hiragana*-Umschrift (はら). Hierbei handelt es sich um eine Besonderheit der Online-Version, die in den Benutzungshinweisen mit der Sortierung nach der 50-Laute-Tafel begründet wird. Diese künstliche Mischung der orthographischen Varianten in der Nennform ist ebenso ungewöhnlich wie die Tatsache, dass die Komponenten durch Leerzeichen getrennt werden.

JapanKnowledge+ | 閉じる |

日本国語大辞典 | 凡例 |

親見出し: はら【腹・肚】

はらの = 虫(むし) [= 中(なか)の虫(むし)] [他の子見出しへ](#)

Knowledge Searcher を使用する

①人体の消化器に寄生する虫。回虫(かいちゅう)の類。

- *和玉篇 [150後] 「蛭 人腹中長虫 ハラノムシ」
- *浮世草子・傾城色三味線 [1701] 大坂・五「腹(ハラ)の虫(ムシ)を残らずたいらげ手なみを見せん」
- *読本・権説弓張月 [1807~11] 拾遺・五〇回「これを腹(ふく)せば心を清し、(略) 蛭蟻虫(ハラノナカノムシ)を殺す」

②空腹どきの腹鳴りを、腹中に虫がいて、それが鳴くものとしていう、その虫。

- *俳諧・伊勢山田俳諧集 [1650] 寄合「うまさうにたうふをくみは浦山し あはれにもただ鳴はらの虫」
- *俳諧・文政句帖 - 七年 [1824] 一〇月「腹の虫なるぞよ雪は翌あたり」
- *滑稽本・和合人 [1823~44] 四・上「どうも酒の顔を見ちゃア、腹のむしが承知為ねへ」

③腹の中を掃除する、腹の中を掃除する、腹の中を掃除する、腹の中を掃除する、腹の中を掃除する

| 発音 | 辞書 | 表記 |

他の子見出し: 116項目

- はら悪(あ)し
- はら悪(あ)しげ
- はら痛(いた)し
- はらがいい
- はらが癒(い)える
- はらが痛(いた)い
- はらが痛(いた)む
- はらが居(い)る
- はらが太(おお)きい
- はらが堅(かた)い
- はらがかびる
- はらが返(か)やる
- はらが来(き)た

Abbildung 3.19: Eintrag für *hara-no mushi* im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW

In der Nennform kodierte lexikalische und morpho-syntaktische Varianten sind in der Regel auch indiziert. Dies gilt nicht für alle Varianten und ausgerechnet auf die Form, die für die Darstellung des Lemmas gewählt wird, kann über die Suche nicht immer zugegriffen werden. Beispielsweise liefert die Suche nach der Zeichenkette (はらの中の虫) eine leere Ergebnismenge, die nicht in der Nennform kodierte *kanji*-Variante (腹の中の虫) hingegen liefert einen Treffer. Wenn es sich also bei der Nennform um keine gängige Form des *kanyōku* handelt, ist fraglich, ob diese an prominenter Stelle zu präsentieren sinnvoll ist.

Variante Formen, die nicht in der Nennform kodiert werden, sind auch nicht hinterlegt. So bleibt die Suche nach *hara-no mushi-ga osamaranu* (腹の虫が治まらぬ) oder *aita kuchi-ga fusagaranai* (あいた口がふさがらない) erfolglos, während die jeweils andere Negationsform *hara-no mushi-ga osamaranai* (腹の虫が治まらない) oder *aita kuchi-ga fusagaranu* (あいた口がふさがらぬ) das gewünschte Ergebnis liefert.

Bei der *OneLook*-Suche zeigt sich, dass sich die Wörterbücher hinsichtlich der Kodierung von Varianten ergänzen können. Obwohl im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW deutlich

mehr Varianten verzeichnet sind als im SHŌGAKUKAN-DAIISEN-IW, kommt es dennoch vor, dieses vergleichsweise kleine Nachschlagewerk gerade die Variante kodiert, die im Suchindex des wesentlich umfangreicheren fehlt. Beispielsweise enthält das SHŌGAKUKAN-DAIISEN-IW *heso-de cha-o wakasu* mit der ersten Komponente in *hiragana*-Schreibweise (へそで茶を沸かす), die aufgrund der oben erwähnten Problematik im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW nicht auffindbar ist. Ebenso ist die vergleichsweise häufige Negationsform *fu-ni ochinai* (腑に落ちない) des *kanyōku fu-ni ochiru* (腑に落ちる) nur im SHŌGAKUKAN-DAIISEN-IW indiziert.

3.5.2.5 Mikrostruktur

Die Einträge der *kanyōku* werden im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW auch virtuell als Unterartikel präsentiert. Zwar wird der Hauptartikel nicht angezeigt, ein Link über dem Untereintrag mit dem Titel *oya-midashi* (親見出し) verweist jedoch auf ihn.

Die Mikrostruktur des SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW ist analog zu der Printausgabe in Lesarten gegliedert, die jeweils Bedeutungsangaben und Belege enthalten. Wie Abbildung 3.19 zeigt, wurden Textverdichtungen durch die Einführung von Zeilenumbrüchen und Einzügen zwischen Bedeutungsangaben und Belegen aufgelöst.

Rechts neben der Artikelanzeige wird eine Linkliste der *kanyōku* angezeigt, die dem Hauptartikel zugeordnet sind. Auf diese Liste verweist ein Link neben dem Lemma mit dem Titel »zu weiteren Untereinträgen« (他の子見出しへ *hoka-no komidashi-e*). Die Links im oberen Bereich dieser Auflistung führen zu den Sektionen *Aussprache* (発音 *hatsuon*), Wörterbücher (辞書 *jisho*) und Anzeige (表記 *hyōki*) im unteren Bereich der Artikelanzeige.

3.5.3 »Deutsche Redewendungen« -- EttingerPhraseologieDe-IW

Die Webseite ETTINGERPHRASEOLOGIEDE-IW versteht sich als Online-Ausgabe des im Jahre 1997 erschienenen Lernerwörterbuchs für deutsche Phraseologismen mit dem Titel »Deutsche Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene« (Hessky und Ettinger 1997). Als Grund für die Veröffentlichung des gedruckten Werkes im Internet werden im Vorwort zunächst Verzögerungen bei der Neuauflage des vergriffenen Werkes erwähnt, aber dann auch auf die Vorteile des digitalen Medium verwiesen (Ettinger 2009f). Insbesondere werden dabei die Möglichkeit ausgiebiger Belegangaben und die Verfügbarmachung der Information für eine größere Nutzergemeinde (im Zuge der »Demokratisierung des Wissens«)

angeführt. Als Adressatengruppe werden primär Lerner des Deutschen genannt, deren Muttersprache nicht Deutsch ist (Deutsch als Fremdsprache), die jedoch über solide Grundkenntnisse verfügen. Generell wird auf anderem Niveau der Einsatz im Unterricht für muttersprachliche Sprecher des Deutschen nicht ausgeschlossen (Hessky und Ettinger 1997, S. 8–9).

Die Internetausgabe ist eng an das gedruckte Werk angelehnt und umfasst ebenfalls etwa 1.400 Phraseme. Wie die gedruckte Vorlage enthält das Lernerwörterbuch neben den ausführlichen Umtexten und dem Hauptteil auch einen Übungsteil, der Druck-Vorlagen für Übungen zum Erlernen und Memorisieren der Phraseme zur Verfügung stellt.

Vorwort und Benutzungshinweise aus dem gedruckten Werk werden als PDF-Dateien angeboten. Auf eine ausführliche Darstellung der Genese des Werkes folgt ein Überblick über die Geschichte der Phraseographie und die phraseologische Forschung. Auf insgesamt 40 Seiten werden Zielsetzung, Systematik und Benutzung des Werkes sowie die ihm zu Grund liegende phraseologische Theorie eingehend erläutert. Hinzu kommt in der Online-Ausgabe ein Internet-Vorwort in sechs Kapiteln, welches im Wesentlichen Unterschiede und Gemeinsamkeiten der Print- und Online-Ausgabe thematisiert (Ettinger 2009e), sowie eine Bibliographie.

Die umfangreichen Umtexte berücksichtigen alle zentralen Aspekte der phraseographischen Genese. Ausführlich erläutern sie ihren eigenen Standpunkt und ihre Vorgehensweise beim Erstellen des Lern- und Übungsbuches vor dem Hintergrund anderer phraseologischer Werke. Bedauerlich ist lediglich, dass es keine Kurzfassung der Benutzungshinweise gibt, die dem Leser, der nicht an dem vollständigen Hintergrund der Entscheidungsfindung teilhaben möchte, als Orientierungshilfe dient.

3.5.3.1 Terminologie, Phraseologieauffassung und Materialauswahl

Hessky und Ettinger (1997) entscheiden sich aufgrund der in dem Fachbereich vorherrschenden terminologischen Vielfalt bewusst für den allgemeinen Begriff »Redewendung«. Inhaltlich beschränkt sich das Werk auf Phraseme, die neben den Merkmalen Polylexikalität und Idiomatizität auch einen gewissen *semantischen Mehrwert* enthalten. Ersichtlich wird dieser durch semantisch-pragmatische Merkmale, die verhindern, dass man die *Redewendungen* in beliebigen Kontexten austauschen kann (Hessky und Ettinger 1997, S. 9). Auffällig ist, dass satzwertige Phraseme (insbesondere Sprichwörter) nicht berücksichtigt wurden, obwohl die Definition sie nicht explizit ausschließt.

Als Basis für die Auswahl der etwa 1.350 Phraseme wurde der Inhalt des 1993 erschienen Arbeitsbuchs von Hessky (1993) mit dem Phrasembestand gängiger phraseologischer und allgemeiner Wörterbücher des Deutschen verglichen und zusätzlich die »ideolektale Intuition des Bearbeiters« zu Rate gezogen (Hessky und Ettinger 1997, S. 5). Dadurch wird versucht, einen Bestand zu bieten, der einer Auswahl hochfrequenter Phraseme ebenbürtig sein soll. Für die Online-Ausgabe wurde dieser übernommen.

3.5.3.2 Makrostruktur

Der Artikelteil ist analog zu der Systematik des gedruckten Werkes von Hessky und Ettinger (1997) nach Schlüsselbegriffen gegliedert⁶⁰. Dabei wird der Objektbereich in acht Kapitel untergliedert, die durch Buchstaben gekennzeichnet werden:

- A Aussehen des Menschen
- B Zustand des Menschen
- C Eigenschaften des Menschen
- D Einstellung, Beziehung zu den Mitmenschen, zur Umwelt
- E Menschliches Handeln
- F Einschätzung einer Lage, eines Sachverhalts
- G Umwelt – Außenwelt
- H Situationsgebundene Phraseologismen

Die Kapitel beinhalten bis zu 49 Kategorien, die durch eine Buchstaben-Zahlen-Kombination und einen oder mehrere Schlüsselbegriffe (Schlüsselbegriff-Cluster) gekennzeichnet sind. Der Buchstabe, der das Kapitel bezeichnet, wird dabei in die Kennzeichnung übernommen. Schlüsselbegriff-Cluster bestehen aus synonymen oder sich semantisch ergänzenden Begriffen (siehe B.1.7 MORD – TÖTUNG – SELBSTMORD – STERBEN – TOD) sowie antonymen Konzepten (siehe B.3.1 ETWAS SEHEN BZW. NICHT SEHEN (WOLLEN)). Antonyme Konzepte werden aber nicht zwangsläufig in eine Kategorie zusammen gefasst, wie man an den Kategorien B.4.1 WOHLSTAND und B.4.2 VERARMUNG erkennen kann. Die Kategorien sind nicht alphabetisch angeordnet, sondern werden nach inhaltlichen Gesichtspunkten gruppiert. Kapitel B (ZUSTAND DES MENSCHEN) fällt aus der Reihe, da dort zwischen Kapiteln und Kategorien Zwischenebenen eingefügt wurden.

⁶⁰ Die Systematik stimmt weitgehend mit der in Hessky (1993) überein.

B Zustand des Menschen

B.1 PHYSISCHES BEFINDEN

- B.1.1** Gesundheit – Krankheit -Genesung
- B.1.2** Müdigkeit
- B.1.3** Schlaf – Schläfrigkeit – Schlaflosigkeit
- B.1.4** Hunger – Essen
- B.1.5** Durst – Trinken
- B.1.6** Betrunkenheit
- B.1.7** MORD – TÖTUNG – SELBSTMORD – Sterben – Tod⁶¹

B.2 GEFÜHLSMÄSSIGE LAGE⁶²

- B.2.1** Angst – Furcht
- B.2.2** gute Laune – Lachen – Freude – Wohlbefinden – IN Stimmung sein
- B.2.3** Traurigkeit – Weinen
- B.2.4** Überraschung
- B.2.5** Mißmut – Ärger – Wut
- B.2.6** Ungeduld – Neugier
- B.2.7** Enttäuschung
- B.2.8** Sehnsucht – Verlangen – Gier
- B.2.9** Gelassenheit – Gleichgültigkeit – Teilnahmslosigkeit
- B.2.10** Empörung
- B.2.11** Sorge – Kummer – Leid
- B.2.12** Undankbarkeit – Schadenfreude
- B.2.13** Übermut – Leichtsin

B.3 SINNESWAHRNEHMUNG

- B.3.1** Etwas Sehen bzw. NICHT Sehen (wollen)
- B.3.2** Hören und Gehörtes (nicht) verstehen

B.4 MATERIELLE LAGE

- B.4.1** Wohlstand – Reichtum – Luxus

⁶¹ Dass die Begriffe dieser Kategorie teilweise auf der Webseite in Großbuchstaben angegeben sind, ist möglicherweise ein Versehen. Zumindest konnte keine Erklärung für diese typographische Kennzeichnung gefunden werden.

⁶² Die Verwendung von »ss« und »ß« in dieser Auflistung entspricht der des Originals.

B.4.2 Verarmung – Verschuldung – Armut

B.5 GESELLSCHAFTLICHE, MORALISCHE LAGE

B.5.1 Ansehen – Einfluß

B.5.2 Unehrenhaftigkeit

B.5.3 Verführung – Verkommenheit

Diese onomasiologische oder — wie die Autoren präferieren — ideographische Gliederung soll das Lernen erleichtern und das Verstehen synonymmer Phraseme fördern (Hessky und Ettinger 1997, S. 4). Kategorien, die in der gedruckten Ausgabe Unterkapiteln entsprechen, werden in der Online-Version mit jeweils einer Webseite wiedergegeben. Innerhalb der Kategorien sind die Phraseme alphabetisch nach erstem Substantiv angeordnet. Falls das Phrasem kein Substantiv enthält, werden auch Adjektive und nachfolgend Verben berücksichtigt.

Wie im gedruckten Vorbild gibt es im Menü der Webseite drei Zugriffsmöglichkeiten. Zunächst bedient der alphabetische Index die gängigste Zugriffsmethode von der Form auf den Inhalt. Hier werden die Phraseme analog zu der Sortierung innerhalb der Kapitel angeordnet. Als zweite Möglichkeit wird der Zugriff über eine alphabetisch geordnete Liste der Schlüsselbegriffe angeboten. Das Inhaltsverzeichnis, das der Buchgliederung entspricht, wird als letzte Option gelistet. Als Alternative kann durch den Menüpunkt »Buchkapitel« auch direkt auf einzelne Kapitel zugegriffen werden, sofern man deren Kapitel-Kennzeichen kennt, möglicherweise um die Arbeit mit den Übungen zu unterstützen.

Bemerkenswert an der Makrostruktur der Online-Ausgabe ist, dass sie sich strikt an der gedruckten Ausgabe orientiert, also unabhängig von der Zugriffsart die Phraseme in ihrem onomasiologischen Kontext anzeigt. Auch das Fehlen jeglicher integrierter Suchfunktionalität, sowohl für Phraseme als auch für Schlüsselbegriffe, ist auffällig.

3.5.3.3 Nennform und Mikrostruktur

Die Nennform und auch die Paraphrasierung entspricht weitgehend der im DUDEN-REDEWENDUNGEN-PW. Abgesehen von der Komponente, die hervorgehoben wird, weil sie für die Anordnung auf der Seite entscheidend ist, erscheint die Nennform nicht typographisch markiert. Das Beispiel 3.5.3.3 zeigt, dass das Lemma deutlicher in den Korpusbeispielen hervorgehoben wird als in der Lemma-Angabe selbst. An die Nennform direkt angeschlossen sind die durch spitze Klammern markierten eigenen Bedeutungserklärungen. Diese wurden speziell für die

Online-Ausgabe ergänzt und enthalten Hinweise auf grammatische, syntaktische und klassenmatische Restriktionen sowie zur Pragmatik.

15. jmdm. ist eine **Laus** über die Leber gelaufen/gekrochen (ugs.) < Das Verb „gekrochen“ ist äußerst selten im Internet belegt. Mit diesem häufig in der Frageformverwendeten Phrasem erkundigt man sich danach, warum jmd. schlecht gelaunt ist. Ziemlich ungebräuchlich ist die Verwendung in der ersten Person, wie z. B. *"Mir ist eine Laus über die Leber gelaufen".

jmd. ist schlecht gelaunt, ärgert sich scheinbar grundlos über alles und jedes; jmd. ist über etwas (zumeist eine Kleinigkeit) verärgert
(Duden - Redewendungen, 32008)

(<http://www.natura-naturans.de/artikel/leber.htm>): - < Angaben zur Etymologie des Phrasems. > - Das Organsystem Leber-Galle spielt bei der Entstehung der verschiedenen Gemütsnaturen eine wichtige Rolle; dieses Organsystem ist der Hauptregulator unseres emotionalen Befindens. Störungen dieses Systems bringen uns seelisch aus dem Gleichgewicht. Diese Anschauung stammt aus sehr alter Zeit. In einem babylonischen Text findet sich der Satz "*möge sich deine Leber glätten*" als Ausdruck für den Wunsch, der Angesprochene möge sein seelisches Gleichgewicht wiederfinden. Desgleichen lehrt die chinesische Medizin, die Leber sei Sitz der Seele. Im alten China zeigte der Henker nach Hinrichtung eines Verbrechers dessen Leber dem Publikum, um pathologische Größe und Aussehen dieses Organs als Ursache seines Fehlverhaltens zu demonstrieren. In unserem Kulturkreis spricht man bei emotionaler Instabilität davon, daß dem Betroffenen **eine Laus über die Leber gelaufen ist** oder *seine Galle überläuft*.

< Horoskope enthalten immer zahlreiche Phraseme. > - Schütze: **Ihnen scheint heute eine Laus über die Leber gelaufen zu sein**. Doch anstatt Ihren Frust an den Kollegen oder am Partner auszulassen, *sollten Sie ihm auf den Grund gehen*. Im Job *würden Sie sich am liebsten auf Ihren Lorbeeren ausruhen*. Doch dazu ist dieser Tag nicht geeignet. Was zählt, ist gründliche Arbeit und hohe Konzentration.

< Ratschläge zur Stressbewältigung. > - Mobbing ist in diesem Zusammenhang ein bedeutender Belastungsfaktor, der zusätzlich neben den anderen täglichen Belastungen das berühmte "*Fass zum Überlaufen*" bringt. Für die dann häufig auftretenden Symptome gibt es in der Umgangssprache eine ganze Reihe von Umschreibungen. So wird jemand, dem man seinen Stress deutlich ansieht beispielsweise gefragt: "**Ist dir eine Laus über die Leber gelaufen?**"

- Was machst du denn heute für ein Gesicht? **Dir ist wohl eine Laus über die Leber gelaufen?**

[Zurück](#)

[nach oben](#)

Abbildung 3.20: Beispiel für *jemandem ist eine Laus über die Leber gelaufen* aus ETTINGERPHRASEOLOGIEDE-IW (Ettinger 2013)

In Ermangelung eines kohärenten Systems wurde auf die Kodierung von Restriktionen sowie pragmatische Angaben weitgehend verzichtet. Ausführliche Angaben zur Etymologie fehlen in der Druckausgabe laut Vorwort (Hessky und Ettinger 1997, S. 4). In der Online-Version wurden nur Erläuterungen zur Herkunft von Phrasemen aufgenommen, sofern sie als wissenschaftlich gesichert gelten (Ettinger 2009d).

Der augenfälligste Unterschied zwischen gedruckter und virtueller Version sind jedoch die Beispiele. Während in Hessky und Ettinger (1997) die Korpusbelege aus gedruckten Werken stammen, wird in der Online-Version ausschließlich auf das Internet als Korpus zurückgegriffen. Im Vorwort wird diese Wahl vorwiegend damit begründet, dass das Material aus dem WWW einem Korpus der gesprochenen Sprache ähnelt (Ettinger 2009d). Bewusst wurden ausführliche Belege gewählt und mit der Original-Quelle verlinkt, um die im Kontext impliziten Hinweise auf die Verwendung dem Lernenden zur Verfügung zu stellen sowie ihm die Häufigkeit (oder besser: relative Rarheit) des Auftretens von Phrasemen aufzuzeigen (Ettinger 2009d). Hierzu dienen auch vereinzelte Häufigkeitsangaben zu den Phrasemen (oder Phrasemvarianten), die

darüber Auskunft geben, wie viele Treffer eine Suchanfrage bei der Suchmaschine Google zu einem bestimmten Stichtag (2009) gemeldet hat.⁶³

Nicht immer sind die Inhaltseinheiten zueinander und mit den Umtexten konsistent. So erklärt Ettinger (2009e), die Angabe zur Paraphrase sei aus den Standardwerken übernommen und entsprechend gekennzeichnet, tatsächlich unterscheiden sich zitierte Quelle und Original bei Stichproben wesentlich voneinander.⁶⁴ Von noch größerer Tragweite ist die Problematik, dass die Wahl der Lemma-Form und die Bedeutungserklärung häufig keinen Zusammenhang zu den Korpusbeispielen aufweist.

3.5.3.4 Sonstiges

Besondere Beachtung verdienen die Übungen, die auch in der Phraseodidaktik positiv aufgenommen wurden (Heinecke 1999). Sie bauen überwiegend auf der Kategorisierung auf, zusätzlich bereichern Illustrationen und Paraphrasen den Lehrteil des Wörterbuchs. In der Online-Version werden drei der vier Aufgaben als PDF angeboten. Bei der fehlenden Übung handelt es sich um die Arbeit mit Korpusbeispielen, die aufgrund der Umstrukturierung der Einträge womöglich nicht übernommen werden konnte. Der Bogen zur Erfassung der Merkmale von Phrasemen zum Zwecke des Aufbaus einer eigenen Sammlung besteht aus einer statischen HTML-Seite (Ettinger 2009b), die als Druckvorlage konzipiert ist.

Kapitel 6 der Einleitung beinhaltet eine äußerst interessante Typologie der Redewendungen, die sowohl inhaltliche als auch strukturelle Aspekte berücksichtigt. Leider ist eine Verbindung zum Hauptteil des Wörterbuchs nicht erkennbar. Weder sind die in der Typologie genannten Beispiele vollständig im Hauptteil vertreten, noch sind diese der Typologie entsprechend markiert. Der Sinn der Typologie beschränkt sich darauf, auf die jeweiligen Besonderheiten hinzuweisen und darauf, wie sich dies auf die phraseographische Praxis auswirkt. Entsprechend wird von dem Kapitel Typologie auf den Übungsteil verwiesen.

⁶³ Die Frage nach der Aussagekraft einer derartigen Angabe beziehungsweise wie sich die Inhalte des Internets zu dem Anteil der durch Google indizierten Webseiten verhält, wird nicht thematisiert.

⁶⁴ Beispielsweise ist die Bedeutung von *[nicht] jemandens Kragenweite sein* im DUDEN-REDEWENDUNGEN-PW mit *ganz nach meinem Geschmack sein* angegeben, in ETTINGERPHRASEOLOGIEDE-IW wird die negative Form mit *jmdm. nicht zusagen, nicht gefallen* (Duden - Redewendungen, ³2008) zitiert. Darüber hinaus ist der Hinweis auf häufigen Gebrauch in erotischen Texten nicht nachvollziehbar.

3.5.4 Zusammenfassung

Das Kapitel hat gezeigt, dass existierende lexikalische Ressourcen zumindest teilweise versuchen den Anforderungen der phraseologischen und computerlinguistischen Forschung zu entsprechen. Dies gilt nicht nur für digitale lexikalische Ressourcen, die ohnehin einem ständigen Wandel unterworfen sind, auch die gedruckten japanischen Lexika haben in Teilen bereits Anforderungen aus den verschiedenen Bereichen linguistischer Forschung umgesetzt. Insbesondere umfangreiche Ressourcen reagieren jedoch aus nachvollziehbaren Gründen eher träge auf veränderte Anforderungen, wenn diese Änderungen an dem existierenden Datenbestand oder bekannter Funktionalität voraussetzen.

Die auf dem APEL-WADOKUJT-IW basierenden Wörterbücher sind hinsichtlich ihres Umfangs und ihrer Aktualität einzigartig. Die Funktionalität, welche WADOKUDE-IW bietet, ist an die Bedürfnisse der allgemeinen Nutzung angepasst und weitgehend selbsterklärend. Auch aus den Inhalten des Wikis wird deutlich, dass die Ressource als allgemeines Wörterbuch konzipiert ist und darin liegt auch seine Stärke. Die speziellen Anforderungen der Phraseographie werden hingegen nicht systematisch berücksichtigt.

Der Überblick über die Funktionalität von WADOKUDE-IW und JapanKnowledge+ zeigt die Bemühung über spezielle Suchfunktionalität den Zugriff auf die Inhalte der Wörterbücher zu optimieren. Für *kanyōku* werden jedoch keine dezidierten Lösungen vorgehalten, wenn man von der Möglichkeit, die Sublemmata separat zu durchsuchen, absieht. Orthographische und morphosyntaktische Varianten werden von der Suche nicht erfasst.

Abgesehen davon liegen die Stärken des SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW eher in der Natur des Nachschlagewerkes begründet. Die Digitalisierung hat dazu wenig beigetragen. Rätselhaft ist insbesondere die Entscheidung, in der Lemma-Angabe die jeweils erste Komponente in *hiragana* darzustellen, denn im digitalen Kontext trägt diese Kunstform wenig zur Auffindbarkeit bei. Auch für die in die Nennform integrierte Variationsangabe sind im elektronischen Wörterbuch alternative Methoden denkbar.

Ähnliches gilt für ETTINGERPHRASEOLOGIEDE-IW, dessen erklärtes Ziel ist, die Inhalte aus Hessky und Ettinger (1997) einem breiten Nutzerkreis zugänglich zu machen. Dies geschieht in Form einer Online-Publikation, die die Vorteile des Mediums nur in sehr begrenztem Umfang einsetzt. Wesentliche Strukturen des gedruckten Werkes bleiben erhalten, wie beispielsweise die onomasiologische (ideographische) Gliederung und die drei unterschiedlichen Zugriffsstrukturen. Da das Werk in erster Linie als Lehrbuch dient, kann es nicht anhand der Kriterien

eines Nachschlagewerkes gemessen werden. Die Effektivität des gezielten Zugriffs, die im digitalen Nachschlagewerk einen hohen Stellenwert hat, widerspricht der didaktischen Absicht »den Benutzer sogar beim Suchen zum Lernen zu zwingen« (Hessky und Ettinger 1997, S. 20). Dies mag ein Grund sein, warum referentielle Mittel, die das Auffinden erleichtern, nicht ausgeschöpft (Hessky und Ettinger 1997, S. 21) beziehungsweise die statische Form des Lehrbuchs weitgehend beibehalten wurde. Wünschenswert wäre dennoch eine bessere Vernetzung der Informationen innerhalb des Lernerwörterbuchs. Beispielsweise werden in ETTINGERPHRASEOLOGIEDE-IW im Inhaltsverzeichnis nur die Gliederungspunkte des Wörterbuchteils mit den Inhalten verlinkt, Vorworte, Indizes und Bibliographien sind ausschließlich über das Hauptmenü erreichbar.

Auch wenn viele Details der Werke unerwähnt bleiben müssen, zeigt die Darstellung sehr unterschiedliche Herangehensweise im Umgang mit Phrasemen im Online-Wörterbuch. Zusammenfassend kann man sagen, dass keines der untersuchten Wörterbücher die Möglichkeiten des digitalen Mediums ausschöpft. Die Gründe hierfür lassen sich teilweise im Digitalisierungsprozess zurückführen. Im Falle des *nativ digitalen* APEL-WADOKUJT-IW ist es wohl eher ein Problem des fehlenden Problembewusstseins in der Autorengemeinde, die dazu geführt hat, dass Phraseme nicht deutlich von anderen Phrasen unterschieden werden.

4 Das Online-Wörterbuch für japanische Idiome und ihre deutschen Entsprechungen

Die ersten Teile dieser Arbeit waren der lexikographischen Theorie und Praxis gewidmet. Diese zeigen die Möglichkeiten und Perspektiven der Phrasem-Lexikographie auf, die sich abzeichnenden Konturen eines lexikographischen »Rollce Royce« sind noch undeutlich, da unterschiedliche Vorstellungen sich überlagern und teilweise miteinander kollidieren. Der dritte Teil dieser Arbeit versucht nun, aus den diversen Vorlagen eine in sich konsistente Blaupause zu skizzieren.

Ziel ist die Konzeption eines Online-Wörterbuchs für japanische Idiome und deren deutsche Entsprechungen. Folgende Anforderungen stehen im Zentrum der Entwicklungsarbeit:

- direkter Zugriff auf Varianten
- inhaltliche Konsistenz bzw. Vermeidung von Redundanz
- adäquate semantische und grammatische Beschreibung
- onomasiologischer Zugriff
- effektives Arbeiten mit Belegen und Beispielen

Das Kapitel gliedert sich in zwei Teile. Im ersten Teil werden Aufbau und Funktionalität der Ressource aus der Anwenderperspektive beschrieben. Den Ausgangspunkt bilden überwiegend phraseologische Themen, wobei in der Praxis die Aspekte der theoretischen und der digitalen Phraseographie ineinander fließen. Der zweite Teil des Kapitels ist technischen Aspekten der Realisierung gewidmet. Neben grundlegenden Fragen der Systemarchitektur werden auch Aspekte der Interoperabilität mit Standards und Einbindung externer Ressourcen thematisiert.

4.1 Aufbau und Funktionalität

4.1.1 Titel und Terminologie

Traditionell dienen Titel und Einleitung eines Wörterbuchs der Orientierung des interessierten Benutzers bei der Wahl des Referenzwerks. Der Titel soll in beiden Sprachen vermitteln, welchen Phänomenbereich das Wörterbuch fokussiert. Da es als Online-Nachschlagewerk konzipiert ist, soll der Titel zugleich den Anforderungen eines URI entsprechen.

Aufgrund der terminologischen Vielfalt in beiden Sprachen, ist ein Titel zwangsläufig ein Kompromiss. Im bilingualen Kontext sollte zudem darauf geachtet werden, dass er für die Sprecher beider Sprachen verständlich ist. Der Titel *wadoku-idioms* ist eine Synthese aus den Kurzbezeichnungen der Ausgangs- und Zielsprache *wa-doku*¹ und dem englischen Begriff *idioms*, welcher auf die Spezialisierung des Nachschlagewerks hinweist. Der Anglizismus *idioms* wurde als Kompromiss gewählt, da der Begriff im bilingualen Kontext sowohl im Japanischen als auch im Deutschen üblich ist. Dadurch entspricht der Titel weitgehend dem Usus der Bezeichnung japanischer Wörterbücher. Als Paralleltitel wird die deutsche Übersetzung »Japanisch-Deutsches Idiom-Wörterbuch« verwendet.

Diese terminologische Entscheidung wird entsprechend in den Benutzungshinweisen erläutert. Dabei wird der Phrasentyp *Idiom* bzw. *kanyōku* grundlegend charakterisiert und anhand von Beispielen gegen verwandte Phänomene (insbesondere Kollokationen und Sprichwörter) sowie Satzbeispiele abgegrenzt.

4.1.2 Benutzungshinweise und Orientierungshilfen

Die Erläuterung konzeptioneller Aspekte in den Benutzungshinweisen dient primär der Orientierung im Nachschlagewerk. Sie gliedert sich in folgende Themen:

- Terminologie
- Materialauswahl
- Bedeutungsangaben
- Grammatik

¹ *Wa* (和) steht für das Japanische als Ausgangssprache und *doku* (独) für Deutsch als Zielsprache.

- Suchfunktionalität

Zusätzlich gibt es ein Symbol- und Abkürzungsverzeichnis. Die Hinweise richten sich an ein sprachinteressiertes Publikum und sind allgemein verständlich gehalten. Auf eine Zweiteilung nach Nutzergruppen in den Benutzungshinweisen wird verzichtet, da für Detailfragen auf die vorliegende Arbeit verwiesen wird.

Als zusätzliche Orientierungshilfe lassen sich kontextsensitiv Kurztexte als Hilfestellungen einblenden. Diese dienen insbesondere der Erklärung von Symbolen oder Begriffen und enthalten wo nötig Verweise auf weiterführende Information in der Nutzerdokumentation.

4.2 Materialauswahl

Die erste Einschränkung des untersuchten Materials ergibt sich aus dem Fokus auf das Gebiet der *kanyōku*. Die Einschränkung auf ein Phänomen aus dem Bereich der Phraseologie stellt keine konzeptionelle Entscheidung dar. Vielmehr ergibt sie sich zwangsläufig aus der Bemühung, im Rahmen dieser Arbeit eine ausreichende Analysetiefe erreichen zu können. Eine Ergänzung des Wörterbuchs um andere Phrasentypen ist wünschenswert und sollte möglichst nahtlos erfolgen, das heißt in einer Form, die dem Benutzer das Gefühl vermittelt, ein einheitliches Wörterbuch zu benutzen.

Da bei einem approximativen Phrasenbestand von 5000 Elementen² nur exemplarisch gearbeitet werden kann, musste der Phrasen-Korpus weiter eingeschränkt werden. Die im Folgenden erläuterten Kriterien beziehen sich auf den Kernbereich des exemplarischen Materials. Um bestimmte Fälle, die nicht im Kerndatenbestand enthalten waren, ebenfalls abdecken zu können, wurde dieser punktuell erweitert. In Einzelfällen wurden auch Phrasemeinträge modelliert, die anderen Phrasentypen zugehörig sind. Diese Beispiele haben jedoch rein ergänzenden Charakter und dienen ausschließlich der Illustration von Beziehungen innerhalb des Gesamtwörterbuchs. Sie erheben keinerlei Anspruch, jeweiligen phänotypischen Merkmalen des Phrasems gerecht zu werden. Aufgrund unzureichender Materialbasis mussten dialektale Varianten sowie domain-spezifische Phraseme (Fachtermini) unberücksichtigt bleiben.

² Hausmann (1983, S. 200) beziffert die Anzahl von Phrasemen, die seiner Ansicht nach in einem phraseologischen Wörterbuch verzeichnet sein sollten, auf etwa 5.000 Einheiten.

4.2.1 Phrasembestand

Bei der Wahl des Phrasembestandes wurde darauf geachtet, einen Ausschnitt aus der Menge der *kanyōku* zu modellieren, die sich für die Illustration der spezifischen Charakteristika besonders gut eignet. Bei der Selektion wurden mehrere Kriterien kombiniert. Zunächst wurde der Fokus auf verbale Idiome (*dōshi kanyōku*) gelegt, da dieser Typ als zentraler Vertreter der idiomatischen Phraseme gilt und strukturell ein breites Spektrum von absoluter Festigkeit über diverse Formen der Variabilität hinweg abdeckt. Primär wurden dabei *kanyōku* berücksichtigt, die als Komponenten Körperteilbezeichnungen enthalten (Somatismen) und eine Emotion oder einen Charakterzug zum Ausdruck bringen. Diese Gruppe erwies sich als besonders geeignet, da sie bevorzugtes Objekt der kontrastiven Phraseologieforschung ist und dementsprechend einschlägige Veröffentlichungen als theoretische Basis für die Modellierung zur Verfügung stehen (für Japanisch-Deutsch siehe Itoh 2004; Itoh 2005; Ueda 2002; Ueda 2003; Ueda 2004). Der Vergleich der onomasiologischen Kategorisierung in verschiedenen Wörterbüchern in Kapitel 3.2.2.4 zeigt überdies im Teilbereich Emotionen die deutlichste Kongruenz. Dies kommt der exemplarischen Entwicklung einer Kategorien-Komponente zu Gute.

Weitere Schwerpunkte wurden anhand lexikalischer, semantischer, struktureller und phänomenologischer Gesichtspunkte gesetzt, wenn die Datenanalyse Regularitäten vermuten ließ. So enthält die Beispielmenge eine große Anzahl von *kanyōku*, die mit der Komponente *hara* oder *mushi* gebildet werden, aber auch reichlich Beispiele zur Transitivität sowie Idiome, die auf der sprachlichen Kodierung non-verbaler Verhaltens basieren.

Um einen gewissen Verbreitungsgrad zu gewährleisten, wurde aufgrund des Fehlens belastbarer Frequenz-Statistiken darauf geachtet, dass die Idiome zum Lemma-Bestand des einsprachig japanischen Wörterbuchs IWANAMI-KŌJIEN⁶ gehören. Variante Formen entstammen jedoch dem weiteren Kreis der bereits genannten gedruckten und elektronischen Wörterbücher. Die Selektion resultiert in einem Beispielkorpus von etwa 300 Idiomen³, deren Analyse die empirische Grundlage für die Konzeption des Wörterbuches darstellt.

³ Als ein Idiom gilt in diesem Fall eine Menge von Phänomenen, die sich aufgrund ihrer semantischen und strukturellen Ähnlichkeiten in dem Datenmodell zu einem lexikalischen Eintrag zusammen fassen lassen. Die Zahl der Datensätze in der Datenbank, die Lemmata im Wörterbuch entsprechen würden, oder die der indizierten Varianten und indizierten Zeichenketten ist deutlich höher.

4.2.2 Referenzkorpora

Die in Kapitel 3.3.1 thematisierten Vorbehalte in Bezug auf derzeitige Verfahren der phraseologischen Korpusanalyse gelten auch für das Japanische.⁴ Das betrifft sowohl die Verfügbarkeit ausgewogener Korpora für Schrift- und Umgangssprache als auch die Verfügbarkeit verlässlicher Werkzeuge oder Verfahren zur Extraktion und Analyse der *kanyōku*.

Für die vorliegende Arbeit ist die wichtigste Quelle für Belege das Balanced Corpus of Contemporary Written Japanese (BCCWJ), welches seit August 2011 durch das National Institute for Japanese Language and Linguistics⁵ der Öffentlichkeit zur Verfügung gestellt wird. Das BCCWJ beinhaltet eine Mischung von Texten aus Zeitungen, Zeitschriften und dem Internet, sowie aus der japanischen Gegenwartsliteratur und Sachtexten. Es zeichnet sich durch gleichzeitige Aktualität der Inhalte und Heterogenität in den Textsorten aus.

Bis zur Veröffentlichung des BCCWJ wurde für die vorliegende Arbeit das Zeitungs- und Zeitschriftenarchiv Nikkei Telecom Archiv⁶ und das Literatur-Archiv Aozora Bunko⁷ referenziert. Nikkei Telekon ist ein kommerzieller Datenbank-Service, der sich in erster Linie als Rechercheinstrument für Wirtschaftsfragen versteht (Nikkei Digital Media 2012). Das darin enthaltene Zeitschriftenarchiv enthält neben den seit 1986 erschienen Ausgaben der Nikkei Shinbun Inhalte weiterer Zeitungen und Zeitschriften aus dem Wirtschaftsbereich. Der inhaltliche Fokus auf den Themenbereich Wirtschaft und Politik schränkt den Nutzen des Korpus für die Suche nach Belegen für den Gebrauch von *kanyōku* stark ein.

Das dem Projekt Gutenberg ähnliche Aozora Bunko beinhaltet derzeit ca. 11.500 literarische Texte, deren Urheberrecht erloschen ist (Tomita 1997). Das japanische Urheberrecht gilt derzeit bis 50 Jahre nach dem Tod des Autors, nur wenige Autoren stellen ihre Texte freiwillig früher

⁴ C. Hashimoto, Satō und Utsuro (2006-05-19) beschreiben eine erfolgversprechende Methode zur Erkennung und Disambiguierung japanischer Idiome in Korpora, die jedoch detaillierte lexikalische Beschreibungen der Idiome voraussetzt.

⁵ Das National Institute for Japanese Language and Linguistics (NINJAL beziehungsweise *Kokuritsu Kokugo Kenkyūjo* 国立国語研究所) ist eine staatliche Einrichtung, deren Aufgabe die Unterstützung der empirischen und theoretischen Forschung zu allen Bereichen der japanische Sprache ist.

⁶ Das Zeitschriften-Archiv ist für den wissenschaftlichen Gebrauch über das crossasia-Portal der Ostasien-Abteilung der Staatsbibliothek Berlin verfügbar.

⁷ Verfügbar unter <http://www.aozora.gr.jp>.

zur Verfügung (ca. 200 Schriftstücke). Konkret bedeutet das, dass der Fokus auf Texten der Meiji- (1868-1912) und ersten Hälfte der Shōwa-Zeit (1926-1989) liegt, deren Verständnis ein hohes sprachliches Niveau voraussetzt.⁸ Auch sieht der kollaborative Ansatz von Aozora Bunko keine systematischen Mechanismen der Qualitätskontrolle vor.

Die Unterschiede zwischen den Korpora sind nicht nur inhaltlicher Natur, auch in Bezug auf die Analyse-Werkzeuge gibt es wenig Gemeinsamkeiten. Zwar verfügen alle über die Möglichkeit, nach Gesamtausdrücken zu suchen, hinsichtlich ihrer Ausrichtung, Funktionalität und ihres Ergebnisformats sind jedoch deutliche Unterschiede erkennbar. Auch in dieser Hinsicht ist das BCCWJ die einzige Ressource, die die Ansprüche einer linguistischen Analyse reflektiert. Es bietet eine breite Palette an Werkzeugen für die Korpus-Abfrage, die die Formulierung komplexer Suchanfragen mit diversen Platzhaltern, logischen Verknüpfungen und der Kombination von Zeichenketten mit morphosyntaktischen Kriterien ermöglicht. Die Ausgabe ist konfigurierbar und eine Exportfunktionalität erleichtert die Arbeit erheblich. Nikkei Telekom und Aozora Bunko bieten nur rudimentäre Möglichkeiten die Suche einzuschränken und liefern als Ergebnis eine Verweisliste zu Volltexten, aus welchen in einem weiteren Schritt die jeweiligen Textstellen extrahiert werden müssen. Die Suchfunktion der Plattform Nikkei Telekom orientiert sich an den Bedürfnissen journalistischer Recherchen. Sie ermöglicht die Einschränkung auf bestimmte Quellen, Zeiträume und die Erstellung kombinierter und/oder-Suchanfragen. Aozora Bunko wurde mit Blick auf literarische Interessen konzipiert und erschließt die Inhalte mit einem Autoren- und TitelindeX sowie einer integrierten Google-Suche.

Alle verfügbaren Korpora fokussieren Japanisch in der geschriebenen Form. Ein Pendant für die gesprochene Sprache, welche gerade im Bereich Phraseologie einen wichtigen Stellenwert einnimmt, ist für das Japanische derzeit nicht verfügbar. Obwohl die Funktionalität des BCCWJ die Arbeit wesentlich erleichtert, konnte nicht in allen Fällen Belegmaterial extrahiert werden. Denkbare Gründe sind der unvollständig erfasste Variantenreichtum des Objektbereichs sowie das Fehlen von Material für die gesprochene Umgangssprache.

⁸ Die Mehrzahl der in Aozora Bunko enthaltenen Texte sind in klassischem Japanisch unter Verwendung historischer Schriftzeichen verfasst, da sie aus einer Zeit vor der Vereinheitlichung von Umgangs- und Schriftsprache (言文一致 *genbun itchi*) beziehungsweise vor der Schriftreform (国語改革 *kokugo kaikaku*) nach Ende des Zweiten Weltkriegs stammen.

4.3 Kennzeichnung

In der Phraseographie dient die Kennzeichnung der Phrasentypen in erster Linie der Abgrenzung gegenüber Satzbeispielen. Nur sekundär wird sie zur Differenzierung der Subklassen genutzt. Intern kann eine Untergliederung in unterschiedliche Typen der Optimierung von Datenstrukturen und Abläufen zuträglich sein und redaktionelle Hilfestellung bieten.

Da sich diese Arbeit auf das Kerngebiet der *kanyōku* konzentriert, scheint sich eine explizite Markierung zunächst zu erübrigen. Es gibt jedoch mehrere Gründe, weshalb dennoch eine Markierung im System vorgenommen wurde. Wie bereits erwähnt, erfolgt die Beschränkung auf *kanyōku* aus praktischen Erwägungen. Eine grundsätzliche Typisierung dient der Erweiterbarkeit des Systems um zusätzliche Phrasentypen. Dies gilt ebenso für die Möglichkeit, mit externen Datenquellen zu interagieren. Auch zeigt sich in der praktischen Arbeit, dass eine Trennung der Typen *kanyōku* und *rengo* nicht trivial ist, denn in einigen Fällen lassen sich kompositionelle und idiomatische Lesarten nicht klar voneinander trennen (siehe Kapitel 4.5). Die sinnvollste Lösung war, kompositionelle und idiomatische Lesarten in einem Wörterbuchartikel zusammenzufassen und mit Hilfe der Markierung auf Lesarten-Ebene zwischen den Phrasem-Typen zu differenzieren.

Die Typisierung, auf welcher die Markierung basiert, entspricht den Usancen der japanischen Lexikographie:

慣 (**kan**) *kanyōku* – Idiom

連 (**ren**) *rengo* – Kollokation

諺 (**kotowaza**) *kotowaza* – Sprichwort

格 (**kaku**) *kakugen* – Sprüche, Maximen

Die graphische Form der Markierung (deutsche oder japanische Kennzeichnung, explizit oder implizit etc.) gestaltet sich flexibel. Sie kann an den jeweiligen Nutzungskontext beziehungsweise die jeweilige Nutzergruppe angepasst werden. Entscheidend ist die Übereinstimmung mit der in den Umtexten erläuterten Terminologie und systematische Verankerung der Abkürzungen in Symbolverzeichnis und Hilfe-Funktion.

4.4 Nennform

Aufgabe der Nennform ist, ein komplexes Paradigma von Wortformen in einer repräsentativen Form zusammenzufassen, die den Wörterbucheintrag betitelt und sich in eine Ordnungsstruktur fügt. Sie bietet die Möglichkeit der impliziten Kodierung von Informationen nach den Konventionen der lexikographischen Tradition.

Anders als in Printphraseographie ist die Lemma-Angabe im digitalen Wörterbuch nicht primär für den Zugriff auf den Wörterbuchartikel ausschlaggebend und muss deshalb nicht zwingend alle für den Eintrag relevanten Formen repräsentieren, da weitere Formen im Index hinterlegt sein können. Zwar wird anders als im SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW die Nennform zwingend in den Suchindex aufgenommen, dennoch wird im Folgenden als Nennform die Form bezeichnet, die für den Nutzer sichtbar den Wörterbucheintrag betitelt. Für den Zugriff relevant sind darüber hinaus alle indizierten Formen.

4.4.1 Wahl der morphosyntaktischen Form

In Bezug auf die Wahl der morphosyntaktischen Form für die Nennform werden die in der japanischen Phraseographie üblichen Konventionen beibehalten, soweit sie auch im digitalen Wörterbuch sinnvoll sind. Als Lemma wird folglich die Form gewählt, die traditionell der Wörterbuchgrundform von *kanyōku* entspricht. Enden *kanyōku* mit einem flektierbaren Verb oder Adjektiv, werden diese auf die konventionelle Grundform (Finalform *shūshikei* 終止形) reduziert, es sei denn, morphosyntaktische Restriktionen verhindern deren Bildung oder es gibt eine deutlich ausgeprägte Präferenz zu einer Flexionsform. In derartigen Fällen tritt der Konvention gemäß die flektierte Form an Stelle der Grundform als Lemma in Erscheinung.

Die Phraseme werden nach Möglichkeit auf ihre phraseologischen Grenzen reduziert. Treten optionale Elemente regelhaft auf, werden diese als Nennform-Varianten (siehe Kapitel 4.4.3) hinterlegt. Dies gilt auch für Erweiterungen oder Abwandlungen, die kollokativen Charakter haben. Nicht registriert werden okkasionelle Varianten, da diese nicht systematisch erfasst werden können.

4.4.2 Integration der Phonetik-Angabe

Die Phonetik der Phraseme wird nach dem Vorbild der japanischen Spezialwörterbücher in *kana*-Umschrift kodiert. Akzent-Angaben konnten aus Ermangelung verfügbarer Daten nicht berücksichtigt werden.

Für die Art und Weise, wie die phonetische Angabe in Relation zum Lemma positioniert beziehungsweise in dieses integriert wird, gibt es keine etablierten Konventionen. Häufig werden bei Phrasemen *furigana* oder geklammerte Angaben verwendet oder Kompromiss-Lösungen, wie die Darstellung der ersten Komponente in *hiragana* im SHÖGAKUKAN-NIHONKOKUGODAI-JITEN-IW (vergleiche Kapitel 3.5.2.4).

Für alle Methoden lassen sich Pro- und Contra-Argumente finden. Dies hängt teilweise auch mit der traditionellen Multifunktion der Nennform zusammen. Eine phonetische Darstellung der Gesamtzeichenkette ist insbesondere für die Auflistung bei sortierten Ergebnisanzeigen sinnvoll, sie eignet sich jedoch weniger für das schnelle optische Erfassen und Wiedererkennen einer längeren Zeichenkette.

Als Betitelung eines Eintrags ist die Anzeige der Lesehilfe für einzelne Schriftzeichen über dem Text (*furigana*) wesentlich übersichtlicher. Da sie relativ unauffällig sind, besteht hier die Gefahr, dass sie nicht wirklich wahrgenommen werden. Sprachdidakten hingegen fürchten, dass umgekehrt durch die *furigana* die *kanji* weitgehend ignoriert werden könnten, weshalb sie für eine phonetische Angabe unterhalb der *kanji* plädieren (Engel 2012).

Für andere Nutzungssituationen und Nutzergruppen sowie die Anzeige der Recherche-Ergebnisse sind diverse alternative Präsentationsformen denkbar:

1. Konfigurierbare Positionierung (z. B. integriert/oberhalb/unterhalb/links/rechts der *kanji-kana*-Mischform)
2. Einblenden auf Anfrage (z. B. durch Positionierung eines Zeigers)
3. Ausblenden nach einer benutzerdefinierten Anzeige-Zeit
4. Anzeige nur ab einem benutzerdefinierten *kanji*-Schwierigkeitsgrad
5. Abspielen einer Audio-Datei (automatisch beim Laden/auf Anfrage)

Als Standard-Form wird die Angabe von Lesehilfen über dem eigentlichen Lemma gewählt, die als barrierefreies Informationsformat einem großen Nutzerkreis dienen soll. Um nicht unnötig von der *kanji-kana*-Mischform abzulenken, werden dabei nur Lesehilfen für die *kanji* ausgegeben (siehe Beispiel 23).

はら むし おさ
(23) 腹の虫が治まらない

In der Ergebnisauflistung wird die Gesamtzeichenkette als *furigana* über der *kanji-kana*-Mischform angezeigt (siehe Beispiel 24), da diese für die Sortierung der Treffer untereinander ausschlaggebend ist.

はらのむしがおさまらない
(24) 腹の虫が治まらない

Die Optionen Lesehilfen auf Anfrage und zeitweiliges Einblenden stehen als Zusatzoption für entsprechend konfigurierte Benutzungsumgebungen zur Verfügung. Aus Ermangelung geeigneter Daten konnten die Anzeige in Abhängigkeit des Schwierigkeitsgrades und Audio-Lesehilfen nicht im Rahmen dieser Arbeit realisiert werden.

Die Frage der Gewichtung der phonetischen Angabe in Relation zu der repräsentativen Schriftform ist umstritten. Aus Sicht der Didaktik wäre eine denkbare Herangehensweise, dass man die Angabe in den Vordergrund rückt, die Gefahr läuft, ignoriert zu werden. In Folge würde beispielsweise für die Text-Rezeption die phonetische Angabe prominent präsentiert, während in einer Lernumgebung die *kanji-kana*-Mischform im Mittelpunkt stünde. Dies bedeutet jedoch, dass man in der konkreten Nutzungssituation durch das Beiläufige von der wesentlichen Information ablenkt. Eine angemessene Lösung hierfür kann nur in Verbindung mit Benutzerstudien erarbeitet werden und ist somit nicht Teil der vorliegenden Arbeit.

4.4.3 Nennform-Varianten

Als Nennform-Varianten werden normative Varianten eines *kanyōku* bezeichnet, die als Kandidaten für die repräsentative Lemma-Angabe in Frage kommen. Hierzu gehören insbesondere die Formen, die in einer verdichteten Darstellung in eine Lemma-Angabe integriert würden oder zur Repräsentation ein und desselben Phänomens in unterschiedlichen Wörterbüchern Verwendung finden. Nicht als Nennform in Erscheinung treten indizierte Varianten. Sie werden einzelnen Nennformvarianten zugeordnet, die statt der indizierten Variante angezeigt werden. Typische Beispiele für die indizierten Varianten sind Grundformen von Phrasemen, die

aufgrund von Restriktionen nicht in dieser Form in Erscheinung treten (siehe Kapitel 4.8 und 4.10.5.2).

Nennform-Varianten dienen dazu, dem Nutzer die Wiedererkennung des Gesuchten zu erleichtern, indem bei einem Treffer die der Suchanfrage ähnlichste Form statt einer repräsentativen Lexikon-Grundform präsentiert wird. Die alternative Anzeige von Nennform-Varianten ersetzt die verdichtete Darstellung der Varianten in einer Lemma-Angabe, die sich negativ auf den Wiedererkennungswert auswirkt.

Diese Architektur hat den Vorteil, dass gängige Formen eines *kanyōku* gleichberechtigt in der Datenbasis angelegt und verwaltet werden können. Existieren mehrere Nennform-Varianten, was bei etwa einem Drittel der Einträge der Fall ist, wird nur jeweils eine Form in der Trefferliste angezeigt, wenn das Suchergebnis mehrere einem Eintrag zugehörige Formen zulässt. Dies gilt ebenso für das Lemma in der Artikelansicht. Nennform-Varianten werden als »Formvarianten« in einem optionalen Segment der Mikrostruktur angezeigt.

Da in der Trefferliste einer Suchanfrage mehrere Nennform-Varianten desselben Wörterbuchartikels enthalten sein können, jedoch nur eine als Lemma des Wörterbuchartikels angezeigt werden soll, ist eine Gewichtung der Nennform-Varianten zueinander notwendig. Aufgrund fehlender Frequenz-Statistiken orientiert sich diese an dem Usus in der japanischen Lexikographie. Ist dieser nicht eindeutig⁹, wird die im IWANAMI-KŌJIEN₆ verwendete Variante als primäre Form registriert.

⁹ Beispielsweise sind für *osamaranai* in *hara-no mushi-ga osamaranai* unterschiedliche Formen als Lemma-Form in den Lexika hinterlegt. Die häufigsten Wörterbuchformen sind die *kanji*-Version 治まらない, die sich in SHŌGAKUKAN-NIHONKOKUGODAIJITEN-PW₂, TANNO-KANYŌKUJITEN-PW, KARADAKOTOBAJITEN sowie YONEKAWA-NIHONGO-KANYŌKUJITEN-PW findet und die Verwendung der *hiragana* in IWANAMI-KŌJIEN₆, KENKYŪSHA-WAEI-DAIJITEN-PW und SANSEIDŌ-DAIJIRIN-IW. Vereinzelt tritt jedoch auch die *kanji*-Form 納まらない in Erscheinung und in einem Werk werden sogar beide *kanji*-Formen verwendet: Das KŌDANSHA-BASIC JAPANESE IDIOMS verwendet die häufigere *kanji*-Version 治まらない bei einem Verweis unter *hara* und 納まらない für den tatsächlichen Eintrag unter *mushi*.

4.5 Semantik-Angabe

Aus den Kapiteln 3.1.8 und 3.1.8.4 wurde deutlich, dass sich die Vorstellungen der idealen Form der Bedeutungsangabe nicht nur unterscheiden, in Teilen werden widersprüchliche Anforderungen formuliert. Das betrifft sowohl strukturelle als auch inhaltliche Aspekte. Im Folgenden geht es darum, eine Auswahl an Anforderungen umzusetzen, die speziell für die Modellierung des japanisch-deutschen Idiom-Wörterbuchs sinnvoll sind und sich mit den verfügbaren Mitteln realisieren lassen.

Die Bedeutungsangabe ist ein Konglomerat verschiedener Teil-Angaben zu wörtlicher Bedeutung, idiomatischer Bedeutung, Pragmatik und Etymologie, welche im Folgenden eingehender erläutert werden. Als Modellierungsbeispiel für die Bedeutungsangabe wurde der Komplex somatischer Phraseme der Körpersprache gewählt. Diese auf sprachlicher Kodierung konventionellen Verhaltens basierenden *kanyōku* erwiesen sich als eine Herausforderung, die die Struktur der Bedeutungsangabe entscheidend geprägt hat. Ihre semantische Vielschichtigkeit und die Tatsache, dass sie als sprachliche Kodierung der Konventionen interkulturelle Gemeinsamkeiten und Unterschiede zum Vorschein bringen, macht sie zu einem interessanten Anwendungsfall.

4.5.1 Wörtliche Bedeutung der Komponenten

Die Angabe der wörtlichen Bedeutung dient dazu, L2-Lernern das Verstehen und Memorieren des Phrasems mit seiner idiomatischen Bedeutung zu erleichtern.¹⁰ Im ersten Ansatz des Wörterbuchs wurde für jedes *kanyōku* eine wörtliche Übersetzung hinterlegt. In der Praxis erwies sich diese jedoch als problematisch, weshalb sie durch Verknüpfung mit den Bedeutungsangaben der Komponenten ersetzt wurde. Zwar ist eine wörtliche Bedeutung der Gesamtphrase für den Benutzer leichter zu erfassen, jedoch bringt die Verlinkung auf die Bedeutung einzelner Komponenten insgesamt mehr Vorteile mit sich, die im Folgenden kurz erläutert werden.

¹⁰ Beispielsweise stellt Zhu (1998, S. 168) in seinem Modell eines deutsch-chinesischen Wörterbuchs der Bedeutungserklärung eine wörtliche Übersetzung der Phraseme voran. Auch Zhu argumentiert, dies soll den Nutzern den Zugang zu Phrasemen erleichtern, da aus der wörtlichen Bedeutung hervorgeht, um welchen Bereich (in diesem Fall Körperteil) es sich handelt.

Existiert keine wörtliche Lesart des Phrasems, ergibt eine wörtliche Übersetzung ein künstliches Konstrukt, dem unter Umständen der Bezug zu der sprachlichen Realität fehlt. In Abhängigkeit der Ausprägung des Merkmals Idiomaticität ist eine wörtliche Übersetzung von *kanyōku* mit polysemen oder kulturspezifischen Komponenten ohne gesicherte etymologische Fundierung arbiträr. Beispielweise gibt es keine angemessene wörtliche Übersetzung des kulturell geprägten Wortes *mushi* in dem Phrasem *mushi-ga ii* (虫がよい、いい、egoistisch, selbstsüchtig). Übersetzt man *mushi* in seinem kulturellen Kontext mit Wurm, führt das zu einer wenig aussagekräftigen wörtlichen Übersetzung »der Wurm ist gut«. Überdies könnte eine künstliche wörtliche Übersetzung von einem L2-Lerner fälschlicherweise als korrektes Deutsch beziehungsweise als Äquivalent interpretiert werden.

Verfügt das *kanyōku* sowohl über eine kompositionelle als auch eine idiomatische Lesart, ist die wörtliche Übersetzung des *kanyōku* quasi identisch mit der Angabe seiner kompositionellen Bedeutung. Das verstärkt die Gefahr der Verwechslung von Angaben, die sich auch durch die üblichen Mittel (Markierung, Erklärung in den Umtexten) nicht vollständig ausschließen lässt. Umgekehrt hat die isolierte Anzeige der Komponentenbedeutungen diverse Vorteile. Bei polysemen Komponenten gibt es die Möglichkeit, auf eine bestimmte Lesart Bezug zu nehmen. Ist keine eindeutige Zuordnung möglich (beispielsweise im Falle phraseologisch gebundener Bedeutungen), muss diese auch nicht erzwungen werden, wie dies bei der Angabe der wörtlichen Übersetzung der Fall ist. Die Angabe einer metaphorischen Bedeutung einzelner Komponenten kann bei teil-idiomatischen Phrasemen als Hinweis auf sprachlich-kulturelle Besonderheiten zusätzliche Hilfestellung leisten.

Ein weiterer wesentlicher Vorteil bei der Angabe der Komponentenbedeutung ist die Möglichkeit, auf existierenden lexikalischen Ressourcen aufzubauen, die als externe Quellen eingebunden werden. WADOKUIDIOMS-IW verlinkt für die Angabe der Komponentenbedeutungen die lexikalischen Einträge des APEL-WADOKUJT-IW (vgl. Abbildung 4.1). Aufgrund des Umfangs der Daten im APEL-WADOKUJT-IW liegt die Abdeckungsrate bei hundert Prozent, wodurch sich auch die redaktionelle Arbeit wesentlich reduzieren lässt.

In einzelnen Fällen ist jedoch eine redaktionelle Anpassung notwendig, um die Inhalte aufeinander abzustimmen. Ein Problem stellen beispielsweise Bedeutungserklärungen in der lexikalischen Ressource dar, die die phraseologische Bedeutung bereits vorweg nehmen. Dies ist insbesondere bei unikalen Elementen der Fall. So wird beispielsweise *yodatsu* (弥立つ) mit »Schauer erregen; einen erschauern lassen« übersetzt, ohne Hinweis darauf, dass es ausschließlich



Abbildung 4.1: Einbindung der Komponentenbedeutungen im Wörterbuch WADOKUIDIOMS-IW

in der Form *mi-no ke-ga yodatsu* (身の毛が弥立つ) verwendet wird. Ein Ergänzung um diesen Hinweis ist für die inhaltliche Konsistenz des Wörterbuchs wichtig.

Ein weiterer wichtiger Punkt ist die direkte Referenzierung von Lesarten, um beispielsweise die Lesart einer polysemen Komponente optisch hervorzuheben, sofern diese für das Phänomen relevant ist. Zwar sind in der derzeitigen Struktur des APEL-WADOKUJT-IW Lesarten markiert, ihre Aufteilung basiert jedoch auf internen Auszeichnungen (Markup), die bei jeder Referenzierung neu geparkt werden muss. Da die Möglichkeit, Lesarten direkt zu referenzieren, nicht vorgesehen ist, fehlt eine eindeutige und nachhaltige Kennzeichnung über *persistent identifiers*.¹¹ Konkret bedeutet dies, dass durch inhaltliche Änderungen in der Datenquelle Referenzen auf einzelne Lesarten verloren gehen können.

Für die Optimierung der Angabe der Komponentenbedeutung würde es sich überdies anbieten, die Anzeige der Lesarten zu reduzieren. Dies gilt insbesondere für die mobile Nutzungsmöglichkeit der Ressource. Eine Gewichtung der Lesarten ist im APEL-WADOKUJT-IW jedoch

¹¹ Als *persistent identifiers* (PIDs) bezeichnet man dauerhafte Referenzen zu digitalen Objekten (Broeder u. a. 2008).

nicht vorgesehen und die sequentielle Anordnung entspricht nicht immer dem synchronen Gebrauch. So listet das APEL-WADOKUJT-IW beispielsweise für *kubi* (首) drei Lesarten:

- 1 Kopf {m}; Haupt {n}.
- 2 Hals {m}; Nacken {m}.
- 3 (Metapher für) Entlassung {f}.

Die erste Lesart, die »alles vom Hals aufwärts« beinhaltet, ist überwiegend in historischen Kontexten und für eine Anzahl somatischer Phraseme relevant. Im IWANAMI-KŌJIEN₆ steht sie an vierter Stelle nach den Bedeutungen:

1. Bei Wirbeltieren der Teil zwischen Kopf und Körper.
2. Teil der Kleidung, der den Hals (*kubi*) berührt.
3. Teilbereich von Dingen, der eine Hals-Form (*kubi-no katachi*) hat.
4. Der Teil des Körpers vom Hals (*kubi*) aufwärts. [...]

Sowohl aus der Anordnung als auch aus der Begrifflichkeit der Definitionen wird deutlich, dass der synchrone Gebrauch von *kubi* überwiegend der ersten Lesart im IWANAMI-KŌJIEN₆ entspricht.

Um die entsprechenden Änderungen an der Zielressource vornehmen zu können, mussten ihre Inhalte gespiegelt werden (siehe auch Kapitel 4.10.4). Dadurch wurde die Datenkonsistenz gesichert und eine redaktionelle Bearbeitung des Datenbestandes ermöglicht.

4.5.2 Idiomatiche Bedeutung und Äquivalenz

Die idiomatiche Bedeutung adäquat wiederzugeben, ist wohl die schwierigste Aufgabe für den Lexikographen. Im Vordergrund steht dabei die funktionale Äquivalenz, die verlangt, dass Lemma und die Entsprechung in der Zielsprache vergleichbare semantische und pragmatische Merkmale aufweisen (Filipenko 2009).

In vergleichbaren bilingualen Arbeiten mit distanten Sprachen wird davon ausgegangen, dass bei Vorliegen eines vollständig äquivalenten Ausdrucks in der Zielsprache keine Notwendigkeit für die paraphrasierende Angabe der Bedeutung besteht (Kim-Werner 1996, S. 87–89;

Zhu 1998, S. 169–171). Die vorliegende Arbeit präferiert einen Ansatz, nach dem eine nicht-idiomatische Bedeutungserklärung unabhängig von der Existenz eines voll-äquivalenten Phrasems notwendig ist.

Es ist aus mehreren Gründen sinnvoll, die idiomatische Bedeutung eines Phrasems zu paraphrasieren. Idiome erfordern ein relativ hohes Sprachverständnis und sind nicht selten in ihrer Verbreitung auf bestimmte Regionen beschränkt. Dies kann für manche Nutzer bedeuten, dass eine Bedeutungsangabe durch eine idiomatische Entsprechung nicht ohne weitere Recherchen verständlich ist. Überdies kann die paraphrasierende Bedeutung zur Disambiguierung beitragen, wenn die Entsprechung in der Zielsprache neben der idiomatischen auch eine kompositionelle Lesart hat.

Nur in einzelnen Fällen ließ sich ein idiomatisches Äquivalent finden, das vergleichbare semantische und pragmatische Merkmale aufweist. Selbst in dem Fall *me-o tsuburu* weisen Korpusbeispiele darauf hin, dass der Aspekt des »Wohlwollens« im Japanischen nicht zwingend ist. Beispiel 25 zeigt, dass das *kanyōku* in manchen Kontexten mit »die Augen vor etwas verschließen« übersetzt werden kann:

(25) この社会的な不公平に目をつぶる政治の無神経さは絶望的だ。

Die Gleichgültigkeit der Regierung, die vor dieser sozialen Ungerechtigkeit die Augen verschließt, ist zum Verzweifeln. (Asahi Shinbun, Morgenausgabe 27.6.1976)

Die Beschreibung im Wörterbuch unterscheidet zwischen paraphrasierenden und idiomatischen Elementen.

me-o tsuburu

1. (慣/idiom.)

etwas absichtlich übersehen; über etwas hinweg sehen

(慣/idiom.) die Augen vor etwas verschließen; (慣/idiom.) ein Auge zudrücken

2. (連/kollok.)

blinzeln, (aufgrund der Lichtverhältnisse) die Augen zukneifen

Das Beispiel zeigt die Realisierung der idiomatischen und kompositionellen Lesart im Wörterbuch. Aus der idiomatischen Bedeutungsangabe geht hervor, dass der Aspekt des »Wohlwollens« nicht zwingend Teil der Bedeutung von *me-o tsuburu* ist. Durch die Trennung in idiomatische und kompositionelle Bedeutung wird auf einen Blick ersichtlich, dass sich im Deutschen

die wörtliche Bedeutung von der idiomatischen unterscheidet, während das Japanische *me-o tsuburu* beide Bedeutungsebenen umfasst.

Idiomatischen Entsprechungen (z. B. die Augen vor etwas verschließen) werden als solche markiert. In vielen Fällen zeigt sich durch die Belege, dass es zwar idiomatische Entsprechungen gibt, diese jedoch auf bestimmte Kontexte beschränkt sind. Sie als mögliche äquivalente Ausdrücke zu nennen birgt die Gefahr, dass der Nutzer die Kontextbedingungen nicht berücksichtigt. Deshalb werden Teil-Äquivalente nicht erwähnt, wenn die Datenlage keine allgemeinen Rückschlüsse erlaubt. Statt dessen wird ihre Verwendung in Anwendungsbeispielen illustriert. Darüber hinaus wird von der Möglichkeit Gebrauch gemacht, sie mittels der Kategorisierung zueinander in Beziehung zu setzen.¹²

Zur Unterscheidung zwischen Teil-Äquivalenten und Äquivalenten dienen insbesondere die Belege. Unter Umständen kann auch die Kategorisierung Aufschluss darüber geben, ob Phraseme in ihrem Bedeutungskern gleich sind. So kann beispielsweise *kubi-o tsukkomu* in der Domäne Medien in manchen Kontexten auch mit »die Nase in etwas hinein stecken« übersetzt werden. Die Analogie durch die Ähnlichkeit des bildlichen Ausdrucks »die Nase/den Kopf in etwas hinein stecken« legt eine Äquivalenzbeziehung nahe. Aus der Kategorisierung werden jedoch grundlegende Unterschiede ersichtlich, denn das SÖTAKUSHA-REIKAIKANYÖKUJITEN-PW weist *kubi-o tsukkomu* dem Kategorienkomplex EINMISCHUNG, HINEINREDEN (干涉、口出 *kanshō, kuchidashi*) zu, »seine Nase in etwas hinein stecken« führt ETTINGERPHRASEOLOGIEDE-IW jedoch im Kapitel UNGEDULD, NEUGIER an. Diese Zuordnungen repräsentieren die Kernbedeutung der Phraseme. Im Kontext von Medien treffen zufällig die Motivation der Neugier, die für das deutsche Idiom Voraussetzung ist, und das Merkmal »Einmischung«, das für das japanische *kanyōku* prägend ist, zusammen. Die Kategorisierung zeigt jedoch, dass der Bedeutungskern der beiden Phraseme differiert und somit die Übereinstimmung der semantischen Merkmale für eine Äquivalenz nicht ausreicht.

Strukturelle Vorgaben in Bezug auf die Bedeutungserklärung gibt es nicht. Sie kann durch Einzelwortlexeme, durch Phrasen oder ganze Sätze erfolgen. Längere Texte und komplexe Strukturen werden im Sinne der Verständlichkeit jedoch gemieden, da für ausgiebigere Erklärungen

¹² Diese Annäherung mittels Kategorisierung auf Begriffsebene kann beispielsweise als Ausgangspunkt für eine Analyse von Phrasem-Gruppen zur Findung potentieller funktionaler Äquivalente sein, wie sie in Dobrovolskij (2000) beschrieben wird.

zur Verwendung andere Angaben (z. B. Etymologie) zur Verfügung stehen. Setzt sich die Bedeutungserklärung aus mehreren Elementen zusammen, wird versucht, sie nach ihrer inhaltlichen Komplexität aufsteigend anzuordnen. Auch wird darauf geachtet, kurze Angaben voranzustellen, so dass die Bedeutungserklärung auch in einer Kurzfassung Sinn ergibt. Dies dient der Kurzanzeige in der Trefferliste, kann jedoch auch für die Anzeige auf mobilen Endgeräten von Bedeutung sein.

Das Beispiel *me-o tsuburu* zeigte bereits die Möglichkeit der Kodierung idiomatischer und kompositioneller Bedeutung eines *kanyōku*. Polysemie ist ein häufiges Phänomen bei der sprachlichen Kodierung non-verbaler Verhaltensweisen, wenn die ursprüngliche, kompositionelle Bedeutung noch in Gebrauch ist. Idiomatische und kompositionelle Lesarten werden grundsätzlich in einem Lexikon-Eintrag zusammengefasst. Die Unterschiede werden in Analogie zur Markierung der lexikalischen Einträge als Idiom beziehungsweise Kollokation gekennzeichnet (siehe Kapitel 4.3). Auch mehrere idiomatische Lesarten sind keine Seltenheit. Das Beispiel *te-o ageru* (手を上げる) vereint mindestens¹³ vier distinkte Bedeutungen:

1. (慣/idiom.) die Hände heben (als Zeichen der Kapitulation)
2. (連/kollok.) die Hand heben (als Geste der Partizipation), sich melden
3. (慣/idiom.) gewalttätig werden, Gewalt androhen
4. (慣/idiom.) die eigenen Fähigkeiten verbessern

Die in den Punkten 1-3 beschriebenen Gesten haben jeweils eine spezifische kommunikative Funktion, die als übertragene Bedeutung interpretiert werden kann. Die Ergänzungen in Klammern sind Teil der Pragmatik-Angabe, welche Thema des nächsten Kapitels (4.5.3) ist. Im vorliegenden Wörterbuch wird davon ausgegangen, dass nur somatische Phraseme, die nicht zwingend auch die faktische Bedeutungsebene (Ausführung der Bewegung) aktivieren, als idiomatisch gelten können. Die Markierung der zweiten Lesart als Kollokation bedeutet, dass ihre übertragene Bedeutung von non-verbaler Komponenten abhängig ist.

¹³ Die Analyse der Korpusbelege ergab zusätzlich zu den vier lexikographisch erfassten Bedeutungen eine Deutung in Zusammenhang mit einer Geste der Beschwichtigung.

Auch wenn die Gebärden in den ersten drei Lesarten unterschiedlich sind, stehen jeweils Bedeutung und Körperbewegung in einem Zusammenhang. Die vierte Bedeutung ist jedoch vollkommen abstrakt und es lässt sich kein Zusammenhang zu den diversen Gesten herstellen. *Te* (Hand) ist hier in seiner metaphorischen Bedeutung durch *ude* (Arm) ersetzbar, einer häufigen Metapher für Fähigkeit oder Kunstfertigkeit.

4.5.3 Pragmatik

Angaben zur Pragmatik umfassen überwiegend stilistisch-zeitliche Angaben, Sprechereinstellungen und Angaben zur Illokution sowie zu Restriktionen auf bestimmte Personengruppen. So wird beispielsweise im Eintrag *me-o hosoku suru* (目を細くする sich freuen) spezifiziert, dass dieser Ausdruck in der Regel mit dem Gesichtsausdruck älterer Menschen in Verbindung gebracht wird.

Im Zusammenhang mit somatischen Phrasemen wird die Pragmatik-Angabe verwendet, um die verschiedenen Bedeutungsebenen und deren Verhältnis zueinander zu beschreiben. Konkret geht es um die Fragen, ob es sich um eine reale, eine semi-reale oder eine imaginäre Körperbewegung handelt. Für reale Gestik und Mimik kann man unterscheiden, ob die Körperbewegung integraler Bestandteil der Bedeutung ist, oder ob diese in ihrer sprachlich kodifizierten Form losgelöst von dem non-verbalen Verhalten die Bedeutung der Körperbewegung denotieren kann. Aus der Perspektive der Bedeutungs-genese handelt es sich um die Unterscheidung zwischen Phrasemen, bei welchen die Bedeutung des non-verbalen Verhaltens als idiomatische Bedeutung Bestandteil des sprachlichen Ausdrucks geworden ist, und solchen, deren übertragene Bedeutungsebene als »Idiomatik der Gestik/Mimik« an das non-verbale Verhalten gekoppelt ist.¹⁴

Im bilingualen Wörterbuch ist eine derartige Unterscheidung insofern relevant, da sich zwischen den Sprachen Abweichungen in der Bedeutung non-verbaler Kommunikation auch in der Regel in fehlenden Möglichkeiten der sprachlichen Kodifizierung widerspiegeln. Der pragmatische Kommentar bietet den Rahmen um die Unterschiede zwischen den Sprachen zu verdeutlichen, da nicht davon ausgegangen werden kann, dass der L2-Lerner das non-verbale Verhalten anderer Kulturen fehlerfrei interpretieren kann (H. Hashimoto 1993; Otsuka 1994).

¹⁴ Ishida (2003) beschreibt in ihrer Analyse von *kanyōku* der Überraschung diesen Unterschied anhand der Merkmale »Innerlichkeit« (*naimensei* 内面性) und »Äußerlichkeit« (*gaimensei* 外面性).

Die Bedeutungserklärungen japanischer mono- und bilingualer Wörterbücher zeigen, dass es durchaus möglich ist, die genannte Information sprachlich festzuhalten, eine systematische Kodierung jedoch die Ausnahme ist (Vogel und Kuchikata im Druck). Bei Kodierung der pragmatischen Aspekte der etwa 30 im Beispielp corpus enthaltenen Phraseme werden die Unterschiede im non-verbalen Verhalten und seiner sprachlichen Kodifizierung in folgender Form verdeutlicht:

reale Geste, Ausführung obligatorisch: ... als Ausdruck von..., Verhalten, wenn..., Geste, die...

reale Geste, Ausführung optional: ... Auch die damit verbundene Geste.

reale Geste, Ausführung unüblich: ... Sagt man beispielhaft für...

semi-reale Geste: Sagt man, wenn ...

imaginäre Geste: sinnbildlich für..., in der Bedeutung von...

In der Praxis ist der pragmatische Kommentar nicht zwingend notwendig, wenn vollständige Äquivalenz herrscht. Dies bedeutet, dass sowohl die übertragene Bedeutungsebene als auch der Realitätswert überein stimmen müssen, wie die Beispiele 26 bis 28 zeigen:

(26) 目を 伏せる
me-o fuseru
Augen-AKK niederschlagen
(beschämt) den Blick senken

(27) 目を 向ける
me-o mukeru
Augen-AKK wenden
den Blick/die Aufmerksamkeit auf etwas richten

(28) 眉を 顰める
mayu-o hisomeru
Augenbrauen-AKK zusammenziehen
(missbilligend) die Stirn runzeln

Aus den Beispielen 26 und 28 wird ersichtlich, dass in der deutschen Entsprechung die Bedeutung des non-verbalen Verhaltens auch durch Attributierung spezifiziert werden kann. Derartige Hinweise sind notwendig, da der Kontext zur Disambiguierung der Mimik fehlt. Beispielsweise beschreibt »die Stirn runzeln« einen Gesichtsausdruck, für welchen als auslösende Emotionen Überraschung, Zweifel und Missbilligung in Frage kommen.

Ausschließlich im pragmatischen Kommentar kodiert werden Abweichung auf faktischer Ebene (Gestik oder Mimik). Beispielsweise beschreibt *te-o awaseru* (手を合わせる wörtl.: »die Hände zusammen bringen«) ebenso wie »die Hände falten« Gesten, die jeweils durch ihren kulturellen und religiösen Kontext geprägt sind. Im Japanischen denotiert *te-o awaseru* die typische Geste des Bittens/Dankens durch das Aufeinanderlegen der Handflächen (in der Regel mit einer angedeuteten Verbeugung), wie sie bei Schrein-Besuchen üblich ist. Der Ausdruck kann jedoch auch für das Zusammenführen von Händen verschiedener Personen verwendet werden. Im Deutschen wird »die Hände falten« meist mit der christlich-religiösen Geste der Andacht in Verbindung gebracht, die häufig durch ein Verschränken der Finger gekennzeichnet ist.

Das Beispiel zeigt, dass eine verbale Beschreibung non-verbaler Kommunikation mühsam ist. Gerade in diesem Bereich bietet sich deshalb der ergänzende Einsatz multimedialer Elemente an, welcher Thema des Kapitels 4.5.6 ist.

4.5.4 Etymologie

Die Angabe zur Etymologie dient der Darstellung der Motiviertheit idiomatischer Bedeutungen. Sie beinhaltet überwiegend Erläuterungen zu Komponenten (oder Lesarten von Komponenten), die archaisch oder metaphorisch sind. Um die Gefahr der Remotivierung gering zu halten, wird sie von anderen Elementen der Bedeutungsangabe deutlich getrennt und entsprechend markiert. Die Anzeige erfolgt auf Anfrage, um in oberflächlichen Konsultationssituationen nicht von den wesentlichen Inhalten abzulenken.

kubi-ni naru

1. (慣/idiom)

(paraphrase) seinen Posten/Job verlieren

(etymology) Metapher, die vermutlich aus dem Puppentheater stammt. Die nicht eingesetzten Puppen »werden zu Köpfen« (siehe auch 首).

Etymologische Angaben für *kanyōku* sind in den untersuchten Wörterbüchern relativ selten und es finden sich in unterschiedlichen Quellen teilweise auch widersprüchliche Theorien. Meist wird *kubi-ni naru* auf »enthaupten; köpfen« zurückgeführt, die Metapher aus der Domäne Puppentheater eignet sich jedoch besser um zu erklären, weshalb es meist im Zusammenhang mit dem Verlust einer Position oder eines Arbeitsplatzes verwendet wird. Da das Ziel dieser Angabe ist, die Bildhaftigkeit zu erklären und damit das Memorieren des Phrasems zu erleichtern, wird bei differierenden Informationen eine Auswahl getroffen.

4.5.5 Synopse-Artikel

Im Fall von teil-idiomatischen Phrasemen beziehen sich die etymologischen Angaben zumeist auf mehrere lexikalische Einträge gleichermaßen. Mit Hilfe von Synopse-Artikeln können Herkunft oder Metaphorik in einer Form kodiert werden, die für alle damit assoziierten Artikel gleich ist und einen Überblick über das Gesamtphänomen gibt.

Mit diesen Synopse-Artikeln werden alle Idiome assoziiert, die auf derselben Metapher beruhen oder aus etymologischer Perspektive inhaltlich zusammengehörig sind. Für den Nutzer dienen sie auch als Übersicht und zum besseren Verständnis der Metaphorik in ihrem kulturellen oder historischen Kontext. Im Unterschied zu Synonymgruppen werden hier übergeordnete Zusammenhänge erläutert, die gegebenenfalls auch gegensätzliche Bedeutungen zum Ausdruck bringen können.

Strukturell gliedert sich die Zusatzinformation in eine allgemeine Bedeutungsangabe, einen etymologischen Kommentar und eine Liste von *kanyōku*, die mit dieser konzeptuellen Metapher assoziiert sind. Da im vorliegenden Ansatz die Synopseartikel als übergeordnete Zusatzinformation in Form von Info-Boxen in unterschiedlichen Kontexten angezeigt werden, müssen sie in sich geschlossene Informationseinheiten bilden (zum Thema Verteilungsstrukturen siehe auch H. Bergenholtz und Tarp (2005, 121–122q)). Aufgrund ihres optionalen Charakters werden Synopseartikel nur für weiterführende Information eingesetzt, die nicht in den Einzelartikeln kodiert wird. Der Zugriff auf die Synopse-Artikel erfolgt über die mit ihnen assoziierten Wörterbuchartikel, Kategorien oder die Stichwortsuche.

Ein Beispiel für einen Synopse-Artikel ist die Darstellung der Gruppe von *kanyōku* mit der Komponente *mushi*, zu der auch *hara-no mushi-ga osamaranai* gehört (vgl. Abbildung 4.2). Zunächst wird die Bedeutung der Komponente *mushi* in ihrem kulturellen Zusammenhang

Metaphor 虫

Der Begriff 虫 (*mushi*) steht für eine Reihe von Kriechtieren (Würmer, Maden, Larven) und Insekten.
 (1) Im übertragenen Sinn wird *mushi* (bzw. *hara no mushi*) häufig als eine Metapher für Verärgerung verwendet.
 ⇨ 虫を殺す ⇨ 虫の居所が悪い ⇨ 腹の虫が治まらない ⇨ 虫が取り上す ⇨ 虫に当たる ⇨ 虫を煩う ⇨ 腹の虫が承知しない

(2) Launenhaftigkeit, insbesondere bei kleinen Kindern, die in diesem Fall auch auf Bauchschmerzen/Parasitenbefall zurück geführt wird. ⇨ 虫を起こす ⇨ 虫がおきる ⇨ 虫気 ⇨ 虫が悪い ⇨ 虫が換わる ⇨ 虫が動じる ⇨ 虫を病む

(3) Vergleich von einem Menschen mit einem kleinen Tier. Bildlich auch für jemanden, der wie ein Parasit mit etwas verbunden ist. Meist despektierlich. ⇨ 虫が付く ⇨ 本の虫 ⇨ 虫の息 ⇨ 泣き虫 ⇨ 弱虫 ⇨ 虫の這うよう

(4) Metapher für Selbstzentriertheit oder Schamlosigkeit. ⇨ 虫がいい

Etymologie: Nach daoistischer Vorstellung leben im Körper jedes Menschen drei *mushi* (*sanshi*), die das Wohl und das Schicksal der Menschen beeinflussen. Auch die traditionelle chinesische Medizin kennt das Bild von neun Würmern, die im Körper des Menschen für Krankheiten verantwortlich sind. Daher stammt möglicherweise die Vorstellung, dass diese kleinen Tiere im Inneren für (unerklärliche) schlechte Stimmung verantwortlich sind, wie sie häufig bei kleinen Kindern auftritt. Am Beispiel des *kōshin*-Brauchs ist überliefert, dass die Menschen in Japan früher glaubten, dass drei *mushi*, *sanshi* genannt, im Körper des Menschen leben (eines im Kopf, eines im Bauch (*hara*) und eines in den Füßen). Diese drei waren dafür zuständig, die guten und insbesondere die schlechten Taten der Menschen in regelmäßigen Abständen dem Himmelsgott

Abbildung 4.2: Synopse-Artikel zu *mushi* im Wörterbuch WADOKUIDIOMS-IW

erklärt¹⁵. Darauf folgen vier distinkte metaphorische Deutungen des Begriffs und Verweise auf die jeweils mit ihnen assoziierten *kanyōku*. Da sich alle Lesarten aus der Ethno-Kategorie *mushi* herleiten lassen und die Übergänge zwischen ihnen fließend sind (Vogel 2013), werden sie in einem Synopse-Artikel zusammengefasst.

Auch konzeptuelle Metaphern können durch Synopse-Artikel repräsentiert werden. Matsuki (1995) nennt als Beispiel für eine kulturell geprägte konzeptuelle Metapher die Vorstellung von *hara* als Kontainer für Emotionen, die kontrolliert werden. Geraten diese außer Kontrolle, steigen sie in die Brust und schlimmstenfalls in den Kopf (Kontrollverlust). Bei einem Eintrag, der nicht eine Komponente sondern ein Konzept (z. B. [hara is the container of the emotions])

¹⁵ Der Begriff *mushi* bezeichnete ursprünglich die übrigen Tiere, die nicht zu Bestien, Vögeln, Fischen und Muscheln gehörten, insbesondere Kriechtiere, Larven, Würmer und manche Insekten. Auch heute wird der Begriff unterschiedlich weit gefasst und lässt sich nicht mit den Kategorien der westlichen Zoologie erklären (Laurent 1995).

fokussiert, ist eine Unterteilung in Lesarten hinfällig. Aber auch hier gilt, dass Phraseme in einem größeren Zusammenhang erklärt werden können, was insbesondere in Lernsituationen von Nutzen sein kann (Boers und Demecheleer 2001).

4.5.6 Multimedia

Multimediale Inhalte können das Verständnis der Idiome verbessern, bergen jedoch die Gefahr, von wesentlichen Informationen abzulenken. In der vorliegenden Arbeit beschränkt sich der Einsatz derzeit auf die Illustration non-verbaler Verhaltens, da darin ein deutlicher Mehrwert gegenüber der rein textuellen Beschreibung gesehen wird.

In den Kapiteln 4.5.2 und 4.5.3 wurde bereits darauf hingewiesen, dass die schriftliche Notation non-verbaler Kommunikation umständlich und schwer verständlich ist. Hinzu kommt, dass aus der wörtlichen Bedeutung nicht immer alle Details des non-verbaler Verhaltens zu Tage treten. Beispielsweise wird *te-o awaseru* meist von einer kleinen Verbeugung begleitet, wenn es sich um die Geste des Bittens oder Betens handelt, die in der sprachlichen Fassung der Geste unerwähnt bleibt. Auch können multimediale Inhalte der Disambiguierung dienen, wenn die Form einer Geste entscheidend für die Bedeutung ihrer sprachlichen Kodierung ist.

Ein weiteres Einsatzgebiet ist die Darstellung der Unterschiede ähnlicher Gesten als Ergänzung zu ihrer sprachlichen Kodierung. So ist es einfacher einem Lerner mit divergierendem kulturellem Hintergrund die Unterschiede der Kopfbewegungen *den Kopf neigen* und *den Kopf (an)drehen* (siehe auch die Beispiele 29 und 30) durch ein Video zu verdeutlichen als durch die Angabe der wörtlichen Komponentenbedeutung.

(29) 首を 傾げる
kubi-o *kashigeru*
 Hals/Kopf-Akk neigen
zweifeln

(30) 首を 捻る
kubi-o *hineru*
 Hals/Kopf-Akk drehen
grübeln, etw. anzweifeln

Für alle *kanyōku* des Beispielpokpus, die auf der sprachlichen Kodierung (realen) non-verbaler Verhaltens basieren, illustriert das Wörterbuch die faktische Ebene durch kurze Filmsequenzen

(Vogel 2011). Diese sollen helfen, eine Verbindung zu dem möglicherweise bereits bekannten non-verbalen Verhalten herzustellen und implizieren automatisch, dass es sich bei dem Gezeigten um eine reale Gebärde handelt. In diesen Fällen kann durch eine Videosequenz ein Gesamteindruck der non-verbalen Kommunikationssituation vermittelt werden, der zugänglicher ist als die verbale Umschreibung. Neben der sprachlichen kann dies auch die interkulturelle Kompetenz unterstützen.

Der Einsatz multimedialer Elemente beschränkt sich nicht auf Videos. In manchen Fällen genügen Bilder zur Illustration non-verbaler. Beispielsweise kann *kuchi-o he-no ji-ni suru* (口をへの字にする verärgert sein wörtl.: den Mund zu dem Schriftzeichen »e« formen) durch eine Zeichnung illustriert werden, da es sich anders als bei den Beispielen 29 und 30 um einen statischen Gesichtsausdruck handelt.

Das multimediale Material ist jeweils nur einer Lesart zugeordnet, da den Bedeutungen polysemer *kanyōku* in der Regel auch unterschiedliches non-verbales Verhalten zu Grunde liegt (siehe *te-o ageru* in Kapitel 4.5.2). Um nicht von den wesentlichen Inhalten abzulenken, wird die Multimedia-Komponente als optionales Element realisiert.

4.5.7 Beispiele und Belege

Beispiele und Belege verfolgen jeweils unterschiedliche Ziele im Wörterbuch. Während Beispiele der Präsentation typischer Merkmale eines Phrasems dienen, haben Belege üblicherweise auch eine Nachweisfunktion. Aus diesem Grund werden diese beiden Formen der Kontext-Angabe im Wörterbuch von einander getrennt kodiert.

Bei der Angabe von Beispielen liegt der Fokus auf präzisen Angaben, die sich auf die wesentlichen Merkmale des *kanyōku* konzentrieren. Die Angabe von Übersetzungen dient dem besseren Verständnis der Beispielsätze.

Auf die Notwendigkeit der Integration von Kontextbelegen wurde bereits mehrfach hingewiesen. Durch die Verfügbarkeit des ausgewogenen BCCWJ-Korpus ist zumindest für das geschriebene Japanisch die Basis geschaffen, umfangreiche Belegsammlungen anzulegen. Auch wenn man davon ausgeht, dass der Nutzer in der spezifischen Nutzungssituation Belege konsultiert, sind nicht alle Belege gleichermaßen zur Bedeutungserklärung geeignet. So können einzelne Belege, die als Nachweis einer bestimmten Bedeutung oder Verwendung dienen, sehr umfangreich sein oder schwierige Begriffe enthalten. Auch sind Belege häufig nur im Rahmen des Begleitkontextes verständlich. Das Beispiel von ETTINGERPHRASEOLOGIEDE-IW zeigt, dass

umfangreiche Korpusbelege die Gefahr bergen, die Artikelanzeige mit Information zu überfrachten. Aus den genannten Gründen dienen Belege insbesondere der lexikographischen Arbeit und werden nur in der redaktionellen Umgebung angezeigt.

Derzeit ist die Annotation der Belege auf die Hervorhebung des *kanyōku* beschränkt. Eine Erweiterungsmöglichkeit wäre, für Belege und Beispiele phonetische und semantische Annotationen zur Verfügung zu stellen. Eine Kombination aus MeCAB-Analyse und Annotation durch die Inhalte des APEL-WADOKUJT-IW wäre möglicherweise eine nützliche Hilfestellung bei der redaktionellen Arbeit. Die Integration und Optimierung dieser automatischen Annotation sind jedoch ein separater Themenkomplex, welcher den Rahmen der vorliegenden Arbeit sprengen würde. Auch stehen bereits diverse Erweiterungen für Internet-Browser zur Verfügung, die eben diese Funktionalität vorhalten.¹⁶

4.5.8 Kategorisierung

Die Kategorisierung soll in erster Linie den differenzierten Zugriff auf semantische Information ermöglichen, der sowohl über eine Suche als auch über eine hierarchische Navigationsstruktur erfolgen kann. Beispielsweise sollte sie den Nutzer beim sprachlichen Ausdruck eines bestimmten Kommunikationsbedürfnisses, wie eines bestimmten Gefühls, eines Zustands, eines Charakterzugs etc. unterstützen. Darüber hinaus wird die Kategorisierung eingesetzt, um ähnliche Phraseme über die Sprachgrenzen hinweg zu clustern und zu alinieren.

Wie die Kapitel 3.1 und 3.3 gezeigt haben, ist die Frage des Sinns einer onomasiologischen Zugriffsmethode ein Thema, bei welchem die Meinungen stark divergieren. Dass es auch kommerzielle Wörterbücher gibt, die sich diesem Ansatz verschreiben, kann als Indiz gewertet werden, dass diese Zugriffsmethode nachgefragt wird. Wie der onomasiologische Zugriff sinnvoll in eine lexikalische Ressource mit multiplen Zugriffsstrukturen integriert werden kann, ist Thema in Kapitel 4.9. Im Folgenden werden die Prinzipien erläutert, nach welchen die Kategorisierung durchgeführt wurde.

¹⁶ Beispiele sind PopJisho <http://www.popjisyo.com>, Moji <http://moji.mozdev.org> und ReadingTutor <http://language.tiu.ac.jp>.

Zunächst mussten die Elemente der Kategorien-Menge eingegrenzt werden. Dem Problem, den Objektbereich in eine intersubjektiv akzeptable und für den Benutzer nachvollziehbare Ordnung zu bringen (siehe Kapitel 3.1.6.3), kann sich auch die digitale Lexikographie nicht entziehen. Zunächst wurde eine Unterscheidung in Begriffs- und Sachebene eingeführt (vergleiche Burger (2010, S. 200)). Um diese möglichst allgemein verständlich zu halten, werden hierfür die Begriffe »Sinnkategorie« für die onomasiologische und »Sachkategorie« für die semasiologische Zuordnung verwendet. So ist beispielsweise *hara-ga tatsu* (腹が立つ *sich ärgern*) BAUCH als Sachkategorie und ÄRGER als Sinnkategorie zugeordnet, wobei die Sachkategorie mit dem Lemma assoziiert ist, während die Sinnkategorie mit der entsprechenden Lesart verbunden ist. Wie die gedruckten Wörterbücher mit onomasiologischer Makrostruktur zeigen, bietet sich eine hierarchische Gliederung in Untergruppen an. Eine effektive Benutzerführung ist davon abhängig, dass die Strukturierung der Kategorien-Hierarchie nachvollziehbar ist. Dies gilt insbesondere, wenn der Benutzer sich nicht über die genaue sprachliche Beschreibung der gesuchten Kategorie im Klaren ist und somit die textuelle Suche wenig Aussicht auf Erfolg hat.

Realisiert wurde im Rahmen dieser Arbeit die Kategorisierung anhand der Sachkategorie KÖRPER und der Sinnkategorie EMOTION. Die an sich einfache Strukturierung in Über- und Unterkategorien für die Kategorisierung ist nicht immer trivial. Selbst für den Bereich KÖRPER, der auf gemeinsamen biologischen Grundlagen fußt, gibt es Grauzonen und kulturell geprägte Unterschiede. So stellt sich die Frage nach dem Umgang mit Übergangsbereichen, genauer gesagt Stellen, an welchen verschiedene Körperteile aufeinander treffen (z. B. die Achsel). Organe und Körperflüssigkeiten können entweder eine eigene Gruppe bilden, oder dem mit ihnen assoziierten Körperteil zugeordnet werden. Bei der Alinierung der Sprachen kommt zusätzlich die Übersetzungsproblematik zum Tragen. Dass *kubi* (首) sowohl den Hals als auch alles vom Hals aufwärts bezeichnen kann, wurde bereits erwähnt. Ein ähnliches Problem ist *ashi* (足), welches den Fuß oder das gesamte Bein denotiert. Einen separaten Begriff für Fuß gibt es in der Umgangssprache nicht. Hinzu kommen kulturell geprägte Begriffe, wie beispielsweise *kimo* (肝), welches für Leber, Gallenblase oder allgemein Eingeweide verwendet wird.

Die hierarchische Gliederung von Sinnkategorien ist noch weitaus komplexer als die der Sachkategorien. Sie wurde in Anlehnung an die in Kapitel 3.2.2.4 vorgestellten gedruckten Vorbilder modelliert, mit dem Unterschied, dass in der Kategorisierung der Emotionen eine Zwischenstufe eingeführt wurde, die sich an Forschungsergebnissen zu Basis-Emotionen orientiert (Ekman 1992).

Auch die Zuordnung der Inhalte zu den Sinnkategorien erfolgt in Anlehnung an die erwähnten existierenden onomasiologischen Nachschlagewerke. Dabei fällt auf, dass diese auch in sich nicht immer konsistent sind. So werden die Antonyme *katami-ga semai* (肩身が狭い *sich schämen*) und *katami-ga hiroi* (肩身が広い *stolz sein*) im SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW mit unterschiedlichen Überkategorien assoziiert. Demnach ist *katami-ga hiroi* ein KÖRPER, CHARAKTERZUG, HALTUNG (からだ・性格・態度 *karada, seikaku, taido*) während *katami-ga semai* zu den Emotionen (感情・感覚 *kanjō, kankaku*) zählt. Das dazugehörige Anwendungsbeispiel (31) lässt jedoch darauf schließen, dass es sich zumindest in diesem konkreten Fall um eine Emotion und nicht um einen Charakterzug handelt.

(31) 君が中心となったプロジェクトが成功し、君を推薦した僕も肩身が広いよ。

Da das Projekt unter deiner Leitung Erfolg hat, bin ich, der dich vorgeschlagen hat, auch stolz.

Aufgrund des häufigen Auftretens lexikalischer Variabilität erfolgt eine Zuordnung zu den Sachkategorien nicht auf der Ebene des lexikalischen Eintrags, sondern für einzelne Lemmata. Ähnliches gilt für die Sinnkategorien, die nicht mit dem lexikalischen Eintrag sondern jeweils mit einer Lesart verknüpft werden. In beiden Fällen ist eine Mehrfachzuweisung möglich, da beispielsweise *hara-no mushi-ga osamaranai* sowohl mit *hara* als auch mit *mushi* in Verbindung gebracht werden kann. Im Bereich der Sinnkategorien betrifft dies hauptsächlich *kanyōku*, die Übergänge zwischen Emotionen bezeichnen. Beispielsweise ist *hara-ga ieru* (腹が癒える *sich (wieder) beruhigen*), zwar mit der Sinnkategorie ÄRGER verbunden, im Resultat drückt jedoch das *kanyōku* einen Zustand der GEMÜTSRUHE aus.

4.6 Grammatik-Angabe

Angaben zur Grammatik von Phrasemen können sowohl die interne Verarbeitung stützen als auch dem Nutzer als Orientierungshilfe dienen. Der grammatische Kommentar setzt sich aus den Angaben zur grammatischen Klasse, zur Valenz und zu belegbaren Formen (Restriktionen) zusammen. Für die Berechnung regulärer Varianten wurde als zusätzliche Angabe die interne Phrasenstruktur erfasst. Die Informationen sind unterschiedlich verankert. Während die grammatische Klasse für den gesamten lexikalischen Eintrag gleich ist, ist die interne Valenz von der konkreten Form der zu beschreibenden lexikalischen Einheit (Lemma) abhängig

und Angaben zu Valenz und belegbare Formen können bei polysemen Phrasemen in Abhängigkeit von der Bedeutung variieren (Burger 2010, vgl. Anm. 57, S. 191). Dementsprechend gibt es statt »einer Grammatik-Angabe« im Wörterbuch eine Reihe von Auszeichnungselementen, die ermöglichen, an verschiedenen Stellen grammatische Merkmale zu kodifizieren.

Die vier Elemente der Grammatik-Angabe, die im Folgenden einzeln skizziert werden, zeichnen *kubi-o hineru* (首を捻る grübeln) wie folgt aus:

grammatische Klasse: VP (für »Verbalphrase«)

interne Phrasenstruktur: NoV (für »Nomen« + Akk-Partikel »o« + »Verb«)

externe Valenz: *N-wa N-ni kubi-o hineru*

belegbare Formen: *kubi-o hineru, kubi-o hinete-iru, kubi-o hinete*

Die Anzeige aller Angaben lässt sich durch die Konfiguration der Nutzungssituation manipulieren. Grammatische Klasse und interne Phrasenstruktur werden standardmäßig ausgeblendet, da sie insbesondere für die redaktionelle Arbeit und interne Funktionalität benötigt werden.

4.6.1 Grammatische Klasse

Die Angabe der grammatischen Klasse (Verbalphrase (VP), Nominalphrase (NP) etc.) gibt allgemein Aufschluss über die Einbindungsmöglichkeiten des Phrasems in den Kontext. Als einziges grammatisches Merkmal ist sie auf der Ebene des lexikalischen Eintrags angelegt und verhindert so indirekt eine Vermischung von Derivation und Variation, da unterschiedliche grammatische Klassen nicht in einem Wörterbucheintrag zusammengefasst werden können.

In ihrer generischen Form eignet sich die grammatische Angabe, um auf formaler Ebene Äquivalente zu vergleichen. Beispielsweise könnte eine Beschränkung der Verlinkung auf Äquivalente desselben Typs als eine Methode der automatischen Qualitätssicherung die redaktionelle Arbeit unterstützen.

4.6.2 Interne Phrasenstruktur

Die Angabe der internen Struktur dient der Erfassung und Berechnung von regulärer Variabilität. Das Variabilitätsmuster berechnet sich aus der POS-Ausgabe des Parsers. Eine redaktionelle

Prüfung ist jedoch notwendig, insbesondere aufgrund der bereits erwähnten Segmentierungsproblematik. Zur Erstellung der Angabe werden die Komponenten durch ihre Wortart ersetzt; Partikeln und Postpositionen werden romanisiert. So entsteht eine menschenlesbare Angabe, die jedoch ausreichend generisch ist, um als Muster für die in Kapitel 4.8.3 erläuterten Transformationen eingesetzt werden.

4.6.3 Externe Valenz

Die Angabe der externen Valenz ergibt sich aus der morphosyntaktischen Analyse der Beispielsätze und Kontext-Belege. Je nach Umfang der Belegsammlung können dabei mehrere Möglichkeiten für die externe Valenz generiert und gegeneinander gewichtet werden. So lassen beispielsweise die Belege für *kubi-o hineru* erkennen, dass das Objekt des Zweifels mit der Partikel *ni* gekennzeichnet werden kann, ebenso häufig jedoch der Ausdruck des Zweifels in Form einer direkten Rede mit der Partikel *to* dem *kanyōku* vorangestellt wird. Da die morphosyntaktische Analyse teilweise sehr umfangreiche Strukturen generiert, welche nur in Teilen als Valenzmuster für den Nutzer relevant sind, wird die Angabe der externen Valenz ebenfalls redaktionell überprüft. Bei einer Ausweitung der Belegsammlung könnte diese redaktionelle Arbeit durch statistische Methoden, die die Relevanz einzelner Elemente berechnen, unterstützt werden.

4.6.4 Belegbare Formen

Da derzeit keine verlässlichen Verfahren zur automatischen Berechnung von Restriktionen für das Japanische verfügbar sind, wird diese durch die Angabe der belegbaren Formen ergänzt. Im Fokus steht hier die Flexibilität der *kanyōku*. Der Stabilitäts-Aspekt wird ausschließlich im Lemma kodiert und nur wenn die Bildung der Grundform davon betroffen ist. Dies hat den klaren Vorteil, dass unabhängig von der Anzahl und Ausgewogenheit der Korpusbelege die Möglichkeit besteht, den Nutzer verlässlich über übliche Formen des Phrasems zu informieren. Auch ist zu vermuten, dass in einer Situation der aktiven Nutzung des Wörterbuchs (Textproduktion) dem Nutzer unter Umständen die Angabe der häufigsten Formen des Phrasems zugänglicher ist als abstrakte Restriktionsregeln.

4.7 Polylexikalität und orthographische Variation

Der Umgang mit orthographischer Variation ist in Bezug auf *kanji-kana*-Variation insofern trivial, als die daraus resultierenden Formen vorhersehbar und regelhaft sind. Für alle *kanyōku* gilt dabei gleichermaßen, dass Lexeme, die durch *kanji*-Zeichenketten repräsentiert werden, durch ihre *kana*-Entsprechung ersetzt werden können. Die orthographischen Varianten werden durch Ersetzungsregeln automatisch generiert.

Die Frage, wie die *kanji*- und *okurigana*-Varianten sinnvoll in das Wörterbuch integriert werden sollten, kann in Abhängigkeit von den Anforderungen unterschiedlich beantwortet werden. Eine Möglichkeit ist, dass sich die Anzeige nach der Suchanfrage des Nutzers richtet. In dem vorliegenden Wörterbuch wird jedoch diese Form der Benutzeradaptivität dem normativen Charakter des Werkes untergeordnet. Das bedeutet, dass bei der Suche gefundene orthographische Varianten in der Lemma-Angabe eines Eintrags durch ihre als Standardform markierte Version ersetzt werden, um dem Nutzer unabhängig von seiner Suchanfrage mit einer repräsentativen Schriftform zu konfrontieren.

Bei der Entscheidung, welche orthographische Form als Standard gilt, wird hinsichtlich des *kanji-kana*-Gebrauchs den *kanji*-Repräsentation der Vorzug eingeräumt, sofern es sich dabei um *jōyō-kanji* handelt. In Bezug auf *okurigana*-Varianten wird die Langform präsentiert, um eine maximale Hilfestellung bei der Lesung der *kanji* zu bieten. Alle Entscheidungen, die darüber hinausgehen, werden mangels korpusbasierter Frequenzstatistiken analog zu der japanischen phraseographischen Praxis kodiert.

4.8 Modellierung struktureller Stabilität und Variabilität

Die Modellierung struktureller Stabilität und Variabilität verfolgt das Ziel, maximale Flexibilität in der Suche zu ermöglichen, ohne die Qualität der Suchergebnisse zu schmälern. Wie bereits in Kapitel 4.6 erwähnt, tritt strukturelle Festigkeit in der Lemma-Angabe in Erscheinung, wenn sie die Bildung der Grundform tangiert, während Variation auf drei unterschiedliche Weisen im Lexikon verankert ist. Den einfachsten Fall stellen indizierte Varianten dar, deren Kodierung ausschließlich den Zweck verfolgt, die Form auffindbar zu machen (Index-Varianten). In der Anzeige wird die indizierte Form durch eine Standard-Form ersetzt. Typischerweise handelt es sich dabei um orthographische oder anhand der grammatischen Transformationen berechnete Varianten.

Die zweite Art der Realisation ist die Zusammenfassung mehrerer Lemma-Formen in einem Wörterbuchartikel. Diese Nennform-Varianten werden bei in der Lemma-Anzeige gleichwertig behandelt. Indem Varianten mit quasi identischen semantischen Eigenschaften zusammengefasst werden, wird Redundanz vermieden, ohne von Verweisen Gebrauch zu machen. Im Folgenden wird hierfür auch von Form-Varianten die Rede sein. Überwiegend sind lexikalische Varianten, bei welchen Komponenten durch synonyme Begriffe ersetzt werden, auf diese Weise kodiert.

Zusätzlich gibt es die Möglichkeit, auf eine variante Form in einem anderen Eintrag im Wörterbuch zu verweisen, wenn sich aufgrund deutlicher inhaltlicher Differenzen die Notwendigkeit ergibt. Diese Form bietet sich beispielsweise bei Varianten an, die sich durch das Merkmal der Transitivität unterscheiden. Die folgenden Kapitel geben einen kurzen Überblick über den Umgang mit morphosyntaktischen Restriktionen und typischen Formen der Variabilität im Wörterbuch.

4.8.1 Morphosyntaktische Restriktionen und Präferenzen

In den Kapiteln 4.4.1 und 4.6.4 wurde gezeigt, dass Restriktionen in der Regel in der Nennform kodifiziert sind, sofern sie die Bildung der Grundform tangieren während Präferenzen als belegbare Formen Teil der Angabe zur Grammatik sind.

Ersetzt die flektierte Form die Grundform in der Nennform, ist in den herkömmlichen Nachschlagewerken der Zugriff nur über die flektierte Form möglich. Für den Benutzer ist jedoch beim Nachschlagen eines unbekanntes Phrasems nicht einsichtig, ob die Bildung der Grundform möglich ist. Ein ähnliches Problem entsteht bei frei flektierbaren Phrasemen, welche eine deutliche Präferenz zu einer Form aufweisen, die fälschlicherweise als Restriktion interpretiert wird. Die Herausforderung bei der Modellierung von Stabilität und Festigkeit im Wörterbuch ist folglich, die Auffindbarkeit der *kanyōku* über das gesamte Kontinuum hinweg zu optimieren. Als Anwendungsszenarien werden dabei berücksichtigt:

1. Die flektierte Form entspricht dem Lemma (Restriktion):
 - a) Der Benutzer gibt eine flektierte Form ein, in der *korrekten* Annahme, dass die flektierte Form auf morphosyntaktischer Restriktion basiert.

- b) Der Benutzer gibt eine unflektierte Form (=Grundform) ein, in der *falschen* Annahme, dass es sich um keine morphosyntaktische Restriktion handelt, sondern das Phrasem frei flektierbar ist.

2. Die Grundform entspricht dem Lemma:

- a) Der Benutzer gibt eine flektierte Form ein, in der *falschen* Annahme, dass die flektierte Form auf morphosyntaktischer Restriktion basiert.
- b) Der Benutzer gibt eine unflektierte Form (=Grundform) ein, in der *korrekten* Annahme, dass das Phrasem frei flektierbar ist.

In der Hälfte der Szenarien (1a und 2b) ist die Annahme des Benutzers korrekt und demnach der Zugriff unproblematisch. Jedoch soll auch in den Fällen 1b und 2a, in welchen der Nutzer unter einer falschen Annahme handelt, die Auffindbarkeit des Phrasems sicher gestellt werden. Das Problem entsteht, da im Falle von morphosyntaktischen Restriktionen in der Regel ausschließlich die flektierte Form als Lemma-Angabe übernommen wird und sich die indizierten (orthographischen) Varianten daraus berechnen. Um die Auffindbarkeit über die Grundform (Szenario 1b) dennoch zu gewährleisten, wird zusätzlich die Grundform mit ihren jeweiligen Varianten in einem separaten Index registriert. Eine Trennung von diesem Index und beispielsweise dem für orthographische Varianten ist notwendig, da der Benutzer bei einem Zugriff über eine ungebräuchliche Form auf die Diskrepanz zwischen Suchanfrage und Suchergebnis hingewiesen wird.

Das Szenario 2a erfordert umgekehrt die Auffindbarkeit von Grundformen bei der Eingabe einer flektierten Form. Diese wird durch die unscharfe Suchfunktionalität gewährleistet, die auf einem Grundformenabgleich der Komponenten basiert (siehe Kapitel 4.9). Eine weitere Möglichkeit ist die Doppel-Kodierung der flektierten und unflektierten Form, wenn die Präferenz sehr deutlich ist, ohne dass eine Restriktion im engen Sinne vorliegt. So ist beispielsweise im Falle von *hara-no mushi-ga osamaranai* die affirmative Form nicht ausgeschlossen, jedoch sehr ungebräuchlich. Hier wird die unmarkierte Form *hara-no mushi-ga osamaru* als separater Eintrag kodiert, der einen Verweis auf die negierte Form enthält. Die Präferenz zu der negierten Form wird in einem entsprechenden pragmatischen Kommentar verankert.

4.8.2 Morphosyntaktische, phonetische und lexikalische Variabilität

Um zu eruieren, wie variante Formen sinnvoll im Lexikon verankert werden können, muss zunächst geprüft werden, welche Bereiche der Mikrostruktur von der Variabilität betroffen sind. Tabelle 4.1 gibt anhand konkreter Beispiele einen Überblick über den Einfluss der in Kapitel 2.2.2.2 angeführten Variationsarten auf die unterschiedlichen Bereiche des lexikalischen Eintrags. Aus der Aufstellung geht hervor, dass alle belegbaren Arten morphosyntaktischer und lexikalischer Variation in erster Linie Auswirkungen auf den Formkommentar haben. Betroffen sind insbesondere Orthographie, Phonetik und Morphologie, mit Ausnahme der relativ seltenen phonetischen Varianten bei gleichbleibender Orthographie.

Variationsart	Orthographie	Phonetik	Komponenten	Grammatik	Semantik	Beispiele/Belege	Variations-Beispiele
Kasus	+	+	=	+	=	+/=	<i>heso-de/ga cha-o wakasu</i>
Reihenfolge	+	+	=	=	=	+/=	<i>he-no kappa / kappa-no he</i>
Transitivität	+	+	+	+	+(?)	+	<i>kata-no ni-ga oriru/kata-no ni-o orosu</i>
Sprachstadium	+	+	=	=	=	=	<i>hara-no mushi-ga osamaranai/nu</i>
Voice	+	+	+	+	+/=	+	<i>kokoro-o utsu/utareru</i>
Polarität	+	+	=	+	+	+/=	<i>shō-ga au/awanai</i>
Aspekt	+	+	=	+	=	+/=	<i>me-ga koeru/koete iru</i>
Phonetik	-	+	=	=	=	=	<i>iyake/iyaki-ga sasu</i>
Substitution	+	+	+	=	=	+/=	<i>hara-o minuku/yomu</i>
Extension	+	+	+	+	=(?)	=	<i>mushi-mo korosanu [kao-o shite iru]</i>
Reduktion/Derivation	+	+	+	+	=	=	<i>hara-ga kuroi/haraguroi</i>

Tabelle 4.1: Auswirkungen der Variations-Arten auf die Mikrostruktur, wobei »+« Abweichungen und »=« Übereinstimmung zwischen den Varianten symbolisiert.

Die Architektur des Wörterbuchs setzt für eine Verankerung von Varianten in einem Lexikon-eintrag die Übereinstimmung des semantischen Kommentars und der grammatischen Klasse voraus. Ist bei zwei ähnlichen *kanyōku* diese Bedingung nicht erfüllt, werden diese in zwei unterschiedlichen Wörterbucheinträgen kodiert, die durch entsprechende Relationen (siehe

Kapitel 4.9.3) miteinander verbunden sind. Dies betrifft durchweg die Phänomene Polarität, Transitivität und Derivation.

Um möglichst viele Varianten in einem Eintrag zusammenfassen zu können, wurden die Merkmale, die sich zwischen den Varianten häufig unterscheiden, der Lemma-Angabe zugeordnet. Dadurch wird ermöglicht, dass für einen einzelnen lexikalischen Eintrag mehrere Lemma-Formen mit den für sie spezifischen Merkmalen (Angaben zu Orthographie, Phonetik, Komponenten und Grammatik) kodiert werden und diese für die Repräsentation vollwertiger und in sich konsistenter Lexikonartikel verwendet werden können. Überwiegend werden *kanyōku*, bei welchen Kasus, Reihenfolge, Sprachstadium oder lexikalische Besetzung von der Variation betroffen sind, auf diese Weise kodifiziert.

Voraussetzung für die Realisierung als sogenannte Lemma-Varianten ist nicht nur, dass die Formen hinsichtlich ihrer semantischen Merkmale invariant sind, die fragliche Form muss sich auch als repräsentative Lemma-Form qualifizieren. Ausgeschlossen von der Realisierung als Lemma-Variante sind deshalb die Beispiele zu Aspekt und Voice in Tabelle 4.1, da die Anzeige der flektierten Form in diesen Fällen fälschlicherweise als Vorliegen einer morphosyntaktischen Restriktion gewertet werden könnte. Statt dessen werden diese Varianten als Präferenzen kodiert.

4.8.3 Modifikation

Die indizierten Varianten lassen sich um reguläre Modifikation aus den Lemma-Angaben erweitern. Typische Variationsmuster bieten redaktionelle Hilfestellungen, die auf der Angabe der internen Phrasenstruktur basieren (siehe Tabelle 4.2).

Muster	Beispiele
NgaVnu ⇔ NgaVnai	<i>hara-no mushi-ga osamarai/ osamaranu</i>
NgaADJi ⇔ NnoADJi	<i>hara-ga kuroi, hara-no kuroi</i>
NgaADJi ⇔ NADJi	<i>hara-ga kuroi, haraguroi</i>
NoNnisuru ⇔ NoNnoyōnisuru	<i>me-o sara-ni suru, me-o sara-no yō-ni suru</i>
NniV ⇔ Nnsuru/ni naru	<i>ki-ni kakaru/kigakari (-da, -ni naru)</i>
NgaADJna ⇔ NADJna	<i>mi-ga karui/ migaru-na</i>

Tabelle 4.2: Typische Phrasenstrukturen und ihre Transformationsmuster

Typische Transformationen sind die Bildung von Komposita durch Weglassung der Partikeln *o*, *ga* oder *ni* und gegebenenfalls Nominalisierung (z. B. *me-ni kado-o tateru* vs. *mekado-o tateru*) oder Attributierung. Auch ist häufig das Ersetzen der Partikeln *ga* und *o* durch die Themenpartikel *wa* möglich.¹⁷

Durch die Transformationsregeln können die Einträge von *kanyōku*, die einem Muster entsprechen, um automatisch berechnete Varianten erweitert werden. Aus dem geringen Datenvolumen geht jedoch bereits hervor, dass nicht in allen Fällen die Ergebnisse der Transformation berechenbar sind. So tritt *haji-o shiru* (恥を知る sich zu benehmen wissen) in seiner negativen Form meist attributiv ohne Partikel in der Form *haji shirazu* (恥知らず schamlos sein) auf. Aus diesem Grund ist es auch hier notwendig, die Ergebnisse der automatischen Berechnung von Varianten redaktionell auf Übereinstimmung mit den Korpusbelegen zu prüfen.

4.8.4 Transitiv-intransitiv Varianten im Wörterbuch

Variation in der Transitivität von Idiomen kann als eine Sonderform lexikalischer Variation betrachtet werden. Die Besonderheit der transitiv-intransitiv Paare besteht in der semantischen und syntaktischen Regelmäßigkeit ihrer Unterschiede zueinander, da die Varianz der *kanyōku* die Eigenschaften ihrer Komponenten widerspiegelt (siehe auch Ishida (1996)). Das heißt, semantisch zeichnet sich ein *kanyōku* mit einer transitiven Verb-Komponente durch eine aktive Bedeutung aus, während sein intransitives Pendant eine gewisse Passivität suggeriert. Der Ersetzungstest zeigt, dass im Fall und von *hara-ga tatsu* die intransitive Form nicht durch die transitive Form *hara-o tateru* austauschbar ist (und umgekehrt):

- (32) 彼は無神経さにいちいち腹を立てていては身がもたない。
kare-wa mushinkeisa-ni ichiichi hara-o tatete-ite-wa mi-ga motanai.
Wenn er sich immer über die Gleichgültigkeit aufregt, wird das seine Gesundheit schädigen.
- (33) *彼は無神経さにいちいち腹が立っていては身がもたない。
kare-wa mushinkeisa-ni ichiichi hara-ga tatte-ite-wa mi-ga motanai.
*Wenn er sich immer die Gleichgültigkeit ärgert, wird das seine Gesundheit schädigen.

¹⁷ Zum Thema korpusbasierte Erstellung von Variationsregeln für Idiome siehe Adachi u. a. (2013), Miyata u. a. (2014) sowie Takeuchi u. a. (2014) .

- (34) 彼のいいかげんな仕事ぶりを見ていると、腹が立って仕方がなかった。
kare-no ii kagen-na shigotoburi-o mite iru to, hara-ga tatte shikata-ga nakatta.
 Wenn ich seine unpräzise Arbeitsweise gesehen habe, habe ich mich immer
 unwillkürlich darüber geärgert.
- (35) *彼のいいかげんな仕事ぶりを見ていると、腹を立てて仕方がなかった。
 **kare-no ii kagen-na shigotoburi-o mite iru to, hara-o tatete shikata-ga nakatta.*
 *Wenn ich seine unpräzise Arbeitsweise gesehen habe, habe ich mich immer
 unwillkürlich darüber aufgeregt.

Aus den Beispielen wird deutlich, dass es sich um eine subtile Unterscheidung zwischen zwei verschiedenen Arten von Ärger handelt. Die transitive Form (Beispiel 32) legt nahe, dass *hara-o tateru* eine kontrollierbare Form des Ärgers denotiert, bei welcher es in der Macht der betreffenden Person liegt, ob sie dem Gefühl nachgibt. Im Gegensatz dazu wird mit der intransitiven Form (Beispiel 34) von *hara-ga tatsu* eine gewisse Machtlosigkeit ausgedrückt. Der Ärger überkommt die Person unwillkürlich in einer bestimmten Situation (*shikata-ga nakatta* bedeutet in etwa »konnte ich nicht anders als«). Die Beispiele 33 und 35 zeigen, dass die Verben nicht gegen ihr transitives beziehungsweise intransitives Pendant ausgetauscht werden können.

In der japanischen Phraseographie werden die transitiven und intransitiven Versionen der *kan-yōku* meist in separaten Einträgen kodiert (SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW, IWANAMI-KŌJIEN⁶, TANNO-KANYŌKUJITEN-PW und ŌBUNSHA-SEIGORIN-PW). Die Bedeutungserklärungen überschneiden sich teilweise und sie verweisen gelegentlich auf gemeinsame Synonyme, in der Regel jedoch nicht aufeinander, wie die Beispiele 36 und 37 aus dem IWANAMI-KŌJIEN⁶ zeigen.

- (36) 腹が立つ : しゃくにさわる。立腹する。
hara-o tateru: shaku-ni sawaru. rippuku suru.
 genervt sein (wörtl.: »von Magenkrämpfen gequält werden«), sich ärgern
- (37) 腹を立てる : 怒る。立腹する。
hara-o tateru: okoru. rippuku suru.
 sich aufregen, sich ärgern

Im Deutschen lassen sich die feinen Unterschiede zwischen transitiver und intransitiver Form nur andeutungsweise wiedergeben. Deutlich wird in jedem Fall, dass sich die Bedeutungserklärungen überschneiden. Das SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW differenziert zwischen *shaku-ni sawaru* und *okoru*, indem Ersteres passivisch umschrieben wird, während die

Lesarten von *okoru* unter anderem ein aktives schimpfen (しかる *shikaru*) enthalten. In den Belegen lässt sich dies entsprechend nachvollziehen, da *hara-o tateru* im Zusammenhang mit der Partikel *to* für die direkte Rede auftritt, um einen verbalen Wutausbruch zu beschreiben.

- (38) 「あいつは失礼だ」と腹を立てる。
 »*Aitsu-wa shitsurei da*« *to hara-o tateru*
 »Was fällt ihm ein« rege ich mich auf. (Katō Taizō; Ki-ga hareru shinrigaku
 [Aufmunternde Psychologie] PHP 1997)

Da die sprachlichen Nuancen im semantischen Kommentar gegebenenfalls übersehen werden könnten, sind Anwendungsbeispiele in diesem Fall besonders wichtig.

Ähnlich komplex ist die Darstellung der grammatischen Unterschiede. Hier schlägt sich die Transitivität meist nicht nur in der internen Phrasenstruktur nieder, auch die externe Valenz ist potentiell davon betroffen. Vergleicht man beispielsweise die Idiome *hara-ga tatsu* und *hara-o tateru* fällt auf, dass die interne Valenz für die Form mit *tatsu* nach dem Subjektpartikel verlangt, während *hara* und *tateru* in einem direkten Objektverhältnis zueinander stehen. In Bezug auf die externe Valenz lässt sich beobachten, dass bei der Verwendung von *hara-ga tatsu* der Aktant mit dem Topikmarker (*wa*) kenntlich gemacht wird, während der sich Ärgernde bei *hara-o tateru* als Subjekt (*ga*) realisiert wird. In beiden Fällen wird jedoch die Person oder Sache, gegen welche sich der Ärger richtet, als indirektes Objekt mit der Partikel *ni* markiert.

Diese Unterschiede der externen Valenz werden nur im YONEKAWA-NIHONGO KANYŌKUJITEN-PW systematisch thematisiert:

- (39) 腹が立つ : 用法 文型「ダレダレはダレナニに腹が立つ」
hara-ga tatsu: yōrei bunkei »*daredare-wa darenani-ni hara-ga tatsu*«
 sich ärgern: Beispiel Satzmuster »jemandTOP jemand/etwasDAT BauchSUB aufstehen«
- (40) 腹を立てる : 用例 文型「ダレダレがダレナニに腹を立てる」
hara-o tateru: yōrei bunkei »*daredare-ga darenani-ni hara-o tateru*«
 sich aufregen: Beispiel Satzmuster »jemandSUBJ jemand/etwasDAT BauchAKK aufstellen«

Ganz so eindeutig ist es bei der Betrachtung der Korpusbeispiele jedoch nicht. So wird je nach Kontext auch bei der Verwendung von *hara-o tateru* der Aktant mit Topik gekennzeichnet (siehe Beispiel (41)), während in untersuchten Beispielen für *hara-ga tatsu* nur selten ein Aktant im direkten Kontext explizit genannt wurde. Beispiel 42 illustriert, dass dieser auch durch die Fokuspartikel *demo* markiert werden kann, statt mit der Subjektpartikel *ga*.

- (41) 克人は父の冷たさに腹を立てながら校庭を歩きまわった。
Yoshihito-wa chichi-no tsumetasa-ni hara-o tate-nagara kōtei-o arukimawatta
Während Yoshihito sich über die Kälte seines Vaters aufregt, läuft er auf dem Schulhof auf und ab. (Sōda Osamu; »Sora-no gai« [Wolkenküste], Kadokawa 1991)
- (42) 初めは聞き流していた私でも、さすがに腹が立ち言い返してしまう。
Hajime-wa kikinagashite ita watashi-demo, sasuga-ni hara-ga tachi, iikaeshite shimau.
Selbst ich, die ich anfangs nur mit einem Ohr zugehört hatte, wurde wütend und habe dem widersprochen. (Ayumi Oyamada, »Himawari« [Sonnenblume], Nihon Bungakkan 2005)

Der Grund, weshalb externe Valenz in Wörterbüchern überwiegend durch Beispiele oder Belege impliziert wird, liegt vermutlich in der Schwierigkeit ihrer exakten Bestimmung. Jedoch gereicht dies in zweierlei Hinsicht zum Nachteil: zum Einen setzt es einen versierten Benutzer voraus, der aktiv die essentielle Information aus den Beispielsätzen extrahiert, zum Anderen werden fakultative Komponenten nicht zwingend in Beispielsätzen realisiert.

Um die vorhandenen semantischen und strukturellen Unterschiede zu abbilden zu können, werden Varianten, die sich durch das Merkmal Transitivität unterscheiden, in separaten Einträgen im Wörterbuch realisiert. Typisierte Verweise zwischen den Formen können dazu dienen, in einer aktiven Nutzungssituation (Textproduktion) das richtige Phrasem zu wählen (siehe Kapitel 4.9.3).

4.9 Zugriffstrukturen

Ein wesentlicher Vorteil des digitalen Mediums für die Lexikographie ist die Möglichkeit, Inhalte dynamisch in Abhängigkeit von einer Anfrage zu generieren. Mehrere alternative Zugriffsstrukturen können sich ergänzen, da nicht wie im gedruckten Werk die Wahl der Ordnungsstruktur der Haupteinträge automatisch über die Hauptzugriffsstruktur entscheidet. Die verschiedenen Zugriffswege im Wörterbuch dienen dazu, unterschiedliche Konsultationssituationen der Nutzer abzubilden.

Die Zugriffsstruktur gliedert sich in drei Funktionsbereiche. Die einfache Suche erstreckt sich auf die Lemmata und ihre Varianten. Sie entspricht der Standardeinstellung, da davon ausgegangen wird, dass der typische Nutzer sich nicht mit komplexen Suchwerkzeugen auseinandersetzen möchte.

Die erweiterte Suche ermöglicht die Spezifikation der Such-Bereiche. Zur Auswahl stehen: Lemma-Angaben (erweiterbar um orthographische Varianten), Bedeutungsangaben (erweiterbar um Pragmatik- und Etymologie-Angabe), Kategorien und Belege sowie eine unscharfe Suche. Zusätzlich lässt sich die Abfrage in einer Weise justieren, dass nur der Anfang, das Ende, ein Teil der Zeichenkette oder die gesamte Angabe auf Übereinstimmung mit der Suchanfrage geprüft wird.

Die Kategorien-Suche erlaubt den direkten Zugriff auf Inhalte über deren Komponenten, Kategorien und Schlagworte. Von ihr werden Sinn- und Sachkategorien gleichermaßen erfasst. Da es keine nennenswerten inhaltlichen Überschneidungen in den Sinn- und Sachkategorien gibt, erübrigt sich eine Differenzierung zwischen den Kategorientypen. Zusätzlich bietet eine hierarchisch gegliederte Indexliste der Kategorienstichworte die Möglichkeit des onomasiologischen Zugriffs. Mit Hilfe der Navigation durch die Sinnkategorien können *kanyōku* anhand ihrer semantischen Eigenschaften selektiert werden. Die drei Kategorisierungsebenen entsprechen der in Kapitel 4.5.8 beschriebenen Systematik.

4.9.1 Variabilität und Suchfunktionalität

Die Herausforderung in der Modellierung der Suchfunktionalität ist, die unterschiedlichen Ausprägungen von Variabilität und Modifikation zu erfassen und die Ergebnisse gegeneinander zu gewichten. Die vorangegangenen Kapitel haben gezeigt, dass für usuelle Varianten die Möglichkeiten der datenstrukturellen Modellierung im Wörterbuch intensiv genutzt werden. Für kontextabhängige beziehungsweise okkasionelle Modifikation bieten sich hingegen applikative Lösungen an.

Als applikative Möglichkeit sind zwei Arten der unscharfen Suche implementiert. Wie bereits in Kapitel 4.8.1 erwähnt, besteht ein Ansatz darin, flektierte Formen in der Suchanfrage auf ihre Grundform zurückzuführen. Eine Ausweitung der unscharfen Suche ist überdies durch einen reinen Abgleich der Komponenten möglich. Erfolgt dieser mit Hilfe der Einträge im APEL-WADOKUJT-IW, lassen sich dadurch auch die diversen Formen orthographischer Variation erfassen. Da sich unterschiedliche *kanyōku* teilweise nur durch ihre morpho-syntaktische Form voneinander unterscheiden, wie die Beispiele 43 und 44 zeigen, ist die unscharfe Suche im Hinblick auf die Qualität der Suchergebnisse der Varianten-Indizierung unterlegen.

(43) 虫も 殺さぬ
mushi-mo *korosanu*
Wurm-AKK/ADDITIV töten-NEG/ATTRIBUTIV
keiner Fliege etwas zu Leide tun können

(44) 虫を 殺す
mushi-o *korosu*
Wurm-AKK töten
sich beruhigen

Die Beispiele 43 und 44 gehören zu unterschiedlichen Bedeutungs-Paradigmen, unterscheiden sich jedoch nur durch die Verwendung der Partikeln (*mo* beziehungsweise *o*) und die Polarität des Verbs. Bei einer unscharfen Suche über den einfachen Komponentenabgleich wird der Unterschied zwischen den beiden Formen nicht erkannt.

Andererseits kann durch die unscharfe Suche ein Großteil der Variationstypen erfasst werden. Dies gilt insbesondere für Varianten, die bislang noch nicht belegt und deshalb auch nicht im Index hinterlegt wurden. Für den sprachlich versierten Benutzer, der das Suchergebnis kritisch zu interpretieren weiß, ist dies durchaus eine Option.

4.9.2 Ergebnisanzeige

Für den direkten und schnellen Zugriff auf Inhalte ist die Strukturierung der Ergebnisanzeige nahezu ebenso wichtig wie die Suche. Die Ergebnisse werden nach Präzision gruppiert in absteigender Reihung geordnet. Die Treffer werden dabei nach der Art der Übereinstimmung mit der Suchanfrage gewichtet:

1. exakte Übereinstimmung mit einer Lemma-Variante
2. exakte Übereinstimmung mit einer indizierten natürlichsprachlichen Variante
3. teilweise Übereinstimmung mit einer indizierten natürlichsprachlichen Variante
 - a) dem Beginn der Zeichenkette
 - b) dem Ende der Zeichenkette
 - c) einem Teilbereich in der Mitte der Zeichenkette

4. exakte Übereinstimmung mit einer künstlich generierten Zeichenkette (z. B. Grundform bei Restriktionen)
5. Ergebnis der unscharfen Suche

Darüberhinaus wird in der Ergebnisanzeige grundsätzlich zwischen direkten Treffern (Punkte 1-3) und indirekten Treffern (Punkte 4 und 5) unterschieden. Für jeden direkten Treffer wird ein Kurzeintrag generiert, der aus der Lemma-Angabe und einer auf zwei Zeilen reduzierten Bedeutungsangabe besteht. Eine Ausnahme bildet die Anzeige eines einzelnen exakten Treffers. In diesem Fall wird die Listenansicht übersprungen und direkt die Vollanzeige des Artikels aufgerufen. Indirekte Treffer werden durch einen Link auf den referenzierten Eintrag präsentiert. Diese Art der Anzeige gleicht einem Verweis-Eintrag im gedruckten Wörterbuch.

Die Anordnung der Trefferlisten erfolgt nach der im Japanischen üblichen exhaustiven Alphabetisierungsmethode. Aufgrund der agglutinierenden Natur der japanischen Sprache und ihrer relativ starren Satzstruktur ist diese unproblematisch. Die Häufung von Synsemantika an erster Stelle in der Nennform deutscher Mehrwortlexeme stellt kein Problem dar, wie die morphologische Analyse¹⁸ der im IWANAMI-KŌJIEN₆ enthaltenen *kanyōku* in Tabelle 4.3 zeigt.

Die Analyse der ca. 7.000 im IWANAMI-KŌJIEN₆ als *kanyōku* gekennzeichneten Einträge ergibt, dass circa 90% mit einem Nomen beginnen. Pronomina und andere Synsemantika sind mit jeweils weniger als einem Prozent vertreten. Externe Valenzen sind regulär nicht in der Nennform enthalten und spielen deshalb für die Anordnung der Mehrwortlexeme keine Rolle.

¹⁸ Der Auswertung lagen die Nennformen der im Kōjien enthaltenen Mehrwortlexeme zu Grunde, die mit Hilfe des MeCab-Parsers einer morphologischen Analyse unterzogen wurden. Die Zuteilung zu Wortarten entsprechen im Kern der Systematik des Parsers. Bei Nomen (*meishi*) wurde jedoch das sekundäre morphologische Merkmal mit einbezogen um Pronomina (*daimeshi*) und na-Adjektive (*keiyōshi*) separat aufführen zu können.

Wortart (1. Komponente)	Häufigkeit	Anteil (%)
Nomen	6277	89,63
Verb	318	4,54
i-Adjektiv	112	1,60
Adverb	88	1,26
na-Adjektiv	78	1,11
Präfix	62	0,89
Pronomen	55	0,79
Interjektion	5	0,07
Konjunktion	5	0,07
Füllwort	3	0,04
Insgesamt	7003	100

Tabelle 4.3: Statistische Auswertung der im IWANAMI-KŌJIEN₆ enthaltenen *kanyōku* bezüglich Wortart des ersten Lexems

4.9.3 Mediostruktur

Die Verweise¹⁹ im Wörterbuch umfassen interne Relationen zwischen den Einträgen und Referenzen auf externe Ressourcen. Für alle Verweise gilt, dass sie den Inhalt des Wörterbucheintrags um optionale Information ergänzen, die für das Verständnis des Artikelinhalts jedoch nicht zwingend ist.

Die internen Verweise dienen der Sichtbarmachung der Relationen zwischen den Einheiten im Lexikon. Verweise erfolgen auf semantisch assoziierte *kanyōku*, die sich nicht als Formvarianten qualifizieren. Hierzu zählen bedeutungsähnliche Phraseme, wie transitive beziehungsweise intransitive Formen, Derivate und Antonyme. Die Verweise sind typisiert und werden dementsprechend auch im Wörterbuch markiert.

¹⁹ Als Verweise werden die Relationen bezeichnet, die der Benutzer auf der Oberfläche als solche empfindet. Relationen, die sich aus dem modularen Aufbau ergeben und nicht für den Benutzer ersichtlich sind, werden als Merkmale der Datenstruktur in Kapitel 4.10.5 thematisiert.

Für semantisch motivierte Verweise gilt, dass sie bei polysemen Phrasemen in der Regel nur für eine der Bedeutungen gültig sind. Folglich ist vorgesehen, als Ankerpunkt der Verweise einzelne Lesarten zu referenzieren. Verweisziel ist meist ein spezifisches Lemma eines anderen Eintrags, es sind jedoch auch Verweise auf einzelne Lesarten möglich. Durch die modulare Kodierung von Nennform-Varianten wird auch ermöglicht eine spezifische Formvariante für die Anzeige im Artikelkontext zu referenzieren.

Die Phänomene Polarität, Transitivität und Derivation sind zwar in gewissem Ausmaß regelhafte, eine automatische Bestimmung ist aufgrund des geringen Datenvolumens nicht möglich. Auch decken die in Kapitel 4.8.4 erläuterten Mustertransformationen zur Transitivität nicht alle Phänomene ab, wie die folgenden Beispiele zeigen.

(45) 口に出す
kuchi-ni dasu
Mund-DAT
äußern, aussprechen

(46) 口に出る
kuchi-ni deru
Mund-DAT
(unwillkürlich) sagen

(47) 口を出す
kuchi-o dasu
Mund-AKK
ins Wort fallen

Das Beispiel *kuchi-ni dasu* (45) verfügt über die intransitive Variante *kuchi-ni deru* (46), das transitive *kuchi-o dasu* (47) scheint jedoch völlig ohne Bezug zu den verwandt wirkenden *kanyōku* zu sein. Im Wörterbuch werden dementsprechend die transitiv-intransitiv Varianten über Verweise zwischen Lesart und Lemma miteinander assoziiert. Der Eintrag für Beispiel 47 ist weder Anker noch Ziel eines transitiv-intransitiv Verweises.

Verweise zwischen Antonymen werden analog hierzu realisiert. Sie sind auch bei separater Kodierung von affirmativer und negierter Form sinnvoll, wenn unklar ist, welche Form die häufigere ist. So sind bei *shiroi ha-o miseru* zwei konträre Lemmaformen in den Lexika zu finden, wie die Beispiele 48 und 49 zeigen.

- (48) 白い 歯を 見せる
shiroi ha-o miseru
 Weiße Zähne-AKK zeigen
 lächeln
- (49) 白い 歯を 見せない
shiroi ha-o misenai
 Weiße Zähne-AKK zeigen-Neg
 ernst bleiben

Für das Beispiel *shiroi ha-o miseru* (48) ist sowohl affirmative als auch negierte Form in Gebrauch. IWANAMI-KŌJIEN⁶ wählt die negierte Form als Lemma-Angabe, da es eine deutliche Präferenz zu dieser Form gibt. Diese Fälle bilden den Grenzbereich zwischen Präferenzen und Restriktionen und werden bevorzugt durch separate lexikalische Einträge, die jeweils aufeinander verweisen, im Wörterbuch repräsentiert.

Derivation ist ähnlich wie die Modifikation gewissen Regeln unterworfen, aber auch sie lässt sich nicht vollautomatisch generieren, da mit diversen Unregelmäßigkeiten gerechnet werden muss. Typisch ist die Bildung durch das Weglassen der Partikeln, insbesondere bei Attributiv- und Adnominalformen in der Umgangssprache (Miyaji 1985, S. 69). Auch die Hervorhebung einzelner Komponenten durch *wa*, der Wechsel in die formale Sprachebene oder die Verwendung von Formen eines anderen Sprachstadiums kommen hier ins Spiel. Miyaji weist darauf hin, dass bei der Nominalisierung manchmal der idiomatische Charakter verloren geht, da z. B. *nami-ni noru* (波に乗る wörtl.: »auf der Welle reiten«) »eine Mode mitmachen« denotiert, während *naminori* (波乗り wörtl.: »Wellenreiten«) der kompositionellen Bedeutung »Surfen« entspricht.

Eine häufige Form der Derivation ist die Nominalisierung:

- (50) 胸が 騒ぐ
mune-ga sawagu
 Brust-SUBJ aufgeregt
besorgt sein
- (51) 胸騒ぎが する。
munasawagi-ga suru.
 Brust-Aufregung-SUBJ machen.
besorgt sein

In dem Beispiel 51 wird der Gesamtausdruck *mune-ga sawagu* (Beispiel 50) zu einem Nomen *munasawagi* (胸騒ぎ dunkle Vorahnung, dumpfes Gefühl) zusammengezogen, das sich mit dem Verb *suru* zu einem Prädikat verbindet, welches mit dem *kanyōku* (Beispiel 50) bedeutungs- und funktionsgleich ist. Das Nomen *munasawagi* kann als Einzelwortlexem oder Einzelwortphrasem betrachtet werden, dessen Gebrauch nicht auf die Form mit *suru* beschränkt ist. In diesem Sinne passt es nicht in das Schema der verbalen *kanyōku*, die den Kern der vorliegenden Arbeit ausmachen. Folglich wird hierfür kein separater Eintrag im Wörterbuch angelegt und der Verweis von Beispiel 50 führt zu dem Eintrag von *munasawagi* im APEL-WADOKUJT-IW.

Nominalisierung resultiert jedoch nicht zwangsläufig in externen Verweisen. In dem folgenden Beispiel (52) wird nur das Verb *shiraseru* (知らせる etw. wissen lassen; mitteilen) nominalisiert und das Merkmal der Polylexikalität bleibt somit erhalten.

(52) 虫が 知らせる。
mushi-ga shiraseru.
mushi-SUBJ mitteilen.
eine (böse) Vorahnung haben

(53) 虫の 知らせ
mushi-no shirase
mushi-Poss Mitteilung.
die böse Vorahnung

Hinsichtlich der Prädikation ist die nominalisierte Form (Beispiel 53) relativ unflexibel, da dem Ausdruck eine passivische Bedeutung anhaftet. Man kann eine Vorahnung haben, indem man sie »fühlt« wie in *mushi-no shirase-o kanjiru* (虫の知らせを感じる wörtlich: »die Mitteilung der Würmer fühlen«) oder »hat« *mushi-no shirase-ga aru* (虫の知らせがある wörtlich: »die Mitteilung der Würmer haben«), jedoch ist es in diesem Fall nicht möglich eine Vorahnung aktiv hervorzurufen oder zu manipulieren. Um diese Besonderheiten abzubilden, wird ein vollwertiger lexikalischer Eintrag für *mushi-no shirase* im Wörterbuch angelegt, der als Derivat mit dem verbalen *kanyōku* assoziiert ist.

4.10 Technische Realisierung

Um diese unterschiedlichen Vorstellungen, Modelle und Vorlagen zu einem in sich konsistenten Werk zusammenzufügen, war eine Gewichtung der Anforderungen aus den unterschiedlichen Bereichen notwendig. Am deutlichsten wird dies in der Datenstruktur, die als konzeptionelle Basis für die Wörterbuchfunktionalität eine zentrale Rolle einnimmt.

Da weder das LMF noch die TEI Richtlinien passende Modellierungsrahmen bieten, wurde ein eigenes Datenformat entwickelt. Um dennoch Nachhaltigkeit und Interoperabilität mit den genannten Standards zu gewährleisten, stellt das System entsprechende Schnittstellen zur Verfügung (vergleiche Kapitel 4.10.5.9).

4.10.1 Komponenten

Den Kern des Systems bildet die interne Datenbasis und systemimmanente Funktionalität, die durch externe Ressourcen und Werkzeuge ergänzt wird. Diese Modularisierung ermöglicht eine gewisse Flexibilität im Systemumfang, denn für keine der externen Ressourcen und Werkzeuge kann deren nachhaltige Verfügbarkeit und Pflege garantiert werden. Folglich müssen die systemimmanenten Komponenten einen in sich geschlossenen und unabhängigen Grundbestand an Daten und Funktionalität zur Verfügung stellen, während externe Ressourcen ausschließlich der Erweiterung dienen.

4.10.2 Datenbasis

Der zentrale Bestandteil des Systems ist die japanisch-deutsche lexikalische Datenbasis, die Kerndaten zu den idiomatischen Phrasemen in einer relationalen Datenbank verwaltet. Dazu gehören die Informationen rund um einen klassischen Wörterbucheintrag, d. h. Lemmata, Angaben zu Grammatik, Semantik, interne und externe Relationen sowie Belege. Ein onomasiologisches Modul stellt in begrenztem Rahmen Daten zur Verknüpfung von Bedeutungsebenen zur Verfügung.

Gerade für die onomasiologische Komponente wäre die Einbindung einer externen Ressource wünschenswert. Naheliegend wäre die Verwendung des japanischen WortNets, welches sich jedoch in der derzeitig verfügbaren Version nicht für die onomasiologische Kategorisierung

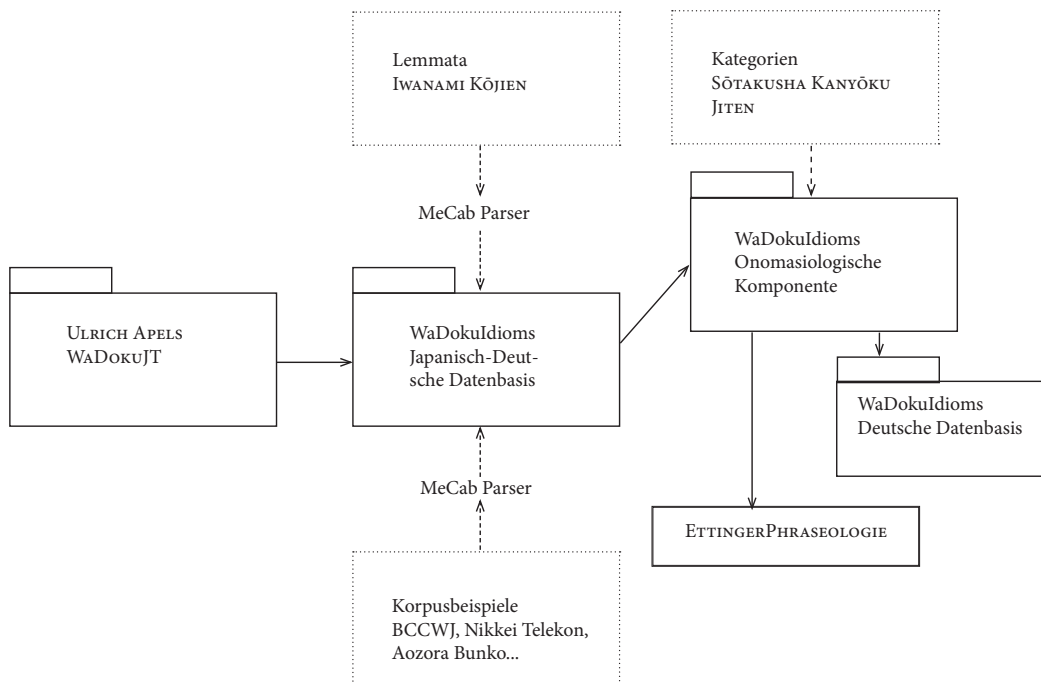


Abbildung 4.3: Überblick über die Systemkomponenten des WADOKUIDIOMS-IW

eignet, da die Relationen zwischen den Begriffen aus dem für das Englische konzipierten Vorbild übernommen wurden.²⁰

²⁰ Das Japanese WordNet ist eine Initiative des National Institute of Information and Computation Technology in Kyōto. Die Entwicklung erfolgte in Anlehnung an das Princeton WordNet, aus welchem in einem ersten Schritt eine Auswahl von 10.000 Synsets mit Hilfe anderer WordNets (Spanisch, Französisch) und elektronischer Wörterbücher automatisch übersetzt wurden. Etwa die Hälfte des Datenbestandes konnte auf diese Weise ins Japanische übertragen werden. Die Relationen im Japanese WordNet über den derzeitigen Datenbestand von 51.000 Synsets wurden vollständig aus dem Princeton WordNet übernommen. Zukünftige Projektvorhaben sehen eine Verlinkung mit der Suggested Upper Merged Ontology sowie dem semantischen Lexikon für Japanisch *Goi Taikei* (The National Institute for Japanese Language 2004) vor, wodurch zusätzliche semantische Netz-Strukturen über die bestehenden Hierarchien gelegt werden sollen (Bond u. a. 2009; Isahara u. a. 2008).

4.10.3 MeCab-Parser

MeCab²¹ ist ein Parser zur morphosyntaktischen Analyse japanischer Texte, der an der Universität Kyōto entwickelt wurde. Der Parser unterstützt einerseits die Erstellung von Neueinträgen im Redaktionssystem durch automatische Bestimmung der morphosyntaktischen und phonetischen Merkmale. Darüber hinaus wird er zur Erweiterung der Suchfunktionalität eingesetzt, z. B. um die Tokenisierung und Lemmatisierung der Eingabe für den Grundformen- und Komponentenabgleich durchzuführen (siehe Kapitel 4.9).

4.10.4 Externe lexikalische Daten

Das APEL-WADOKUJT-IW als wichtigste Quelle externer lexikalischer Daten wurde in den Kapiteln 4.5.1 und 4.9 bereits ausführlich beschrieben. Da das WADOKUIDIOMS-IW als Erweiterung des APEL-WADOKUJT-IW konzipiert ist, sollten die Daten beider Datenbestände in eine Oberfläche zu integrierbar sein. Hierbei stellt sich der uneinheitliche Umgang mit *kanyōku* in den Wörterbüchern als Problem dar. Um eine kohärente Ressource aus den beiden Datenquellen zu modellieren, müsste eine Angleichung erfolgen, um eine einheitliche Markierung und Konsistenz im Zugriff zu gewährleisten. Eine automatisierte Extraktion und Typisierung der im APEL-WADOKUJT-IW erfassten Idiome ist eine interessante Herausforderung, auf welche jedoch aufgrund der Schwerpunktsetzung der vorliegenden Arbeit an dieser Stelle nicht weiter eingegangen wird.

Die Verknüpfung mit dem Wörterbuchprojekt ETTINGERPHRASEOLOGIEDE-IW umfasst intersprachliche onomasiologische Relationen auf Begriffsebene. Auch diese birgt noch Erweiterungspotenzial. So könnte die Möglichkeit der Verknüpfung der Bedeutung einzelner idiomatischer Entsprechungen im Deutschen genutzt werden, um dem Nutzer Zugriff auf Definitionen und Korpusbelege der Äquivalente zu gewähren. Um die Nachhaltigkeit der Relationen zu sichern, müsste jedoch zunächst ein dauerhaftes Referenzierungssystem (z. B. mit PIDs) zwischen den Ressourcen etabliert werden.

²¹ Die Projekthomepage ist unter [http : //mecab.googlecode.com/svn/trunk/mecab/doc/index.html](http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html) verfügbar.

4.10.5 Datenstruktur

Die Datenstruktur wurde ausgehend von den Paradigmen existierender lexikographischer Strukturen und computerlinguistischer Standards entworfen. Im Laufe der Entwicklungsarbeit hat sich diese jedoch von den ursprünglichen Vorlagen entfernt, da die existenten Modelle nicht die Ergebnisse der empirischen Datenanalyse in einer Weise abzubilden vermochten, die den theoretischen Anforderungen Genüge leistet.

Der Faktor, dem in der Entwicklung der Datenstruktur am meisten Gewicht beigemessen wurde, sind die Ergebnisse der Datenanalyse selbst. So mussten diverse Anpassungen vorgenommen werden, um die Muster, die sich aus den lexikalischen Daten herauskristallisierten, in einer Weise erfassen zu können, die eine Entwicklung der für das phraseologische digitale Wörterbuch anvisierten Funktionalität ermöglichte. Das Ergebnis ist eine Struktur, die spezifisch ist für ihre Domäne, jedoch in wichtigen Punkten mit etablierten und sich heraus bildenden Standards kompatibel bleibt.

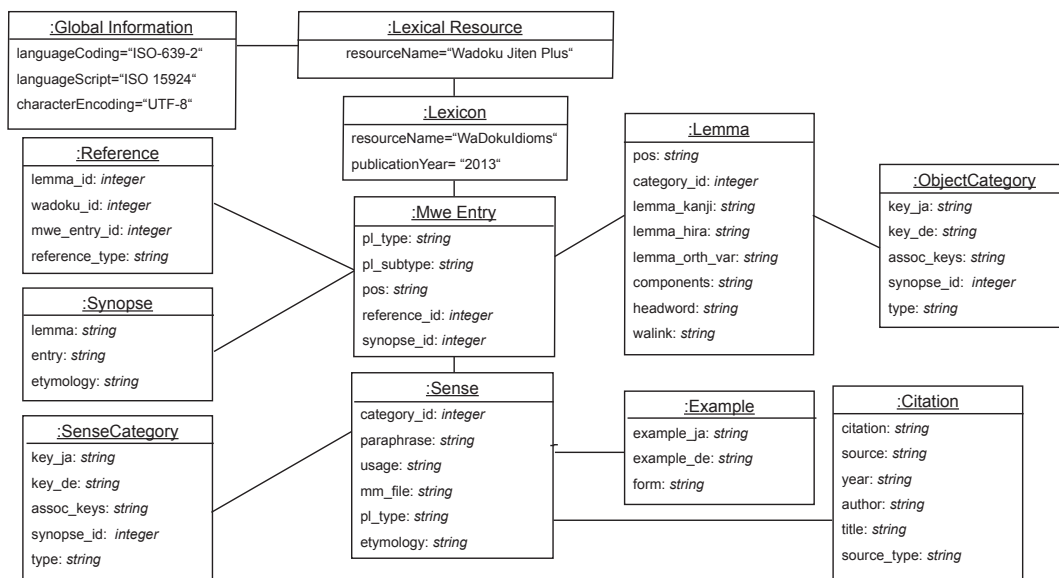


Abbildung 4.4: Überblick über die Datenstruktur des WADOKUIDIOMS-IW

Abbildung 4.4 gibt einen Überblick über die neun Elemente des Kerndatenbestands: Lexikalischer Eintrag (*MweEntry*), Lemma (*Lemma*), Bedeutung (*Sense*), Synopse-Artikel (*Synopse*), Objektkategorie (*ObjectCategory*), Sinnkategorie (*SenseCategory*) Verweis (*reference*), Beispiel (*Reference*) und Beleg (*Citation*). Im Folgenden werden die Eigenschaften dieser Elemente sowie ihre Relationen zueinander skizziert.

4.10.5.1 Der phraseologische Wörterbucheintrag (*Mwe Entry*)

Der Wörterbucheintrag stellt in der lexikographischen Tradition das zentrale Element dar. Die Summe aller *Mwe Entry*-Elemente ergeben den Artikelteil des Nachschlagewerks. Auffällig an der Datenstruktur des WADOKUIDIOMS-IW ist, dass das Element *Mwe Entry* eine sehr geringe Anzahl von Attributen enthält, da alle von morphosyntaktischer Variabilität betroffene Information in das *Lemma*-Element ausgelagert wurde. Wie aus Tabelle 4.4 hervorgeht, beschränken sich die Attribute der Klasse *Mwe Entry* auf die Spezifikationen des Phrasentyps und interne Vermerke für die redaktionelle Arbeit am Wörterbuch.²²

Feld	Erläuterung	Beispiel
<i>synopse_id</i>	Synopsenartikel	⇒ 虫
<i>reference_id</i>	Verweise	(Antonym) ⇒ 腹の虫が治まる
<i>pl_type</i>	Phrasem-Typ	<i>kanyōku</i>
<i>pl_subtype</i>	Phrasem-Subtyp	z. B. <i>emotion</i>
<i>pos</i>	Phrasem-Art	VP

Tabelle 4.4: Attribute der Klasse *MweEntry*

Der Wörterbucheintrag *MweEntry* hat die zentrale Funktion eines Bindeglieds zwischen den Elementen, da alle essentiellen Relationen über dieses Element definiert sind. Mit jedem Wörterbucheintrag sind 1:n Lemma-Formen (*Lemma*) und 1:n Bedeutungsangaben (*Sense*) assoziiert.

²² Hierzu gehört beispielsweise eine Spezifikation der in Kapitel 4.5.3 vorstellten Phrasem-Typen, die als Schablonen für eine konsistente Kodierung bestimmter Merkmale fungieren.

Die Unterscheidung zwischen Phrasentypen dient als Grundlage für die Markierung (siehe Kapitel 4.3) auf Ebene des Wörterbucheintrags. Obwohl sich das vorliegende Werk auf die Kodierung von *kanyōku* konzentriert, ist die Möglichkeit Kollokationen, formelhafte Sprache und Sprichwörter in Analogie zu dem vorgestellten Modell zu kodieren für die zukünftige Weiterentwicklung vorgesehen.

4.10.5.2 Lemma (*Lemma*)

Eine diffizile Entscheidung beim Aufbau des Systems war die Frage nach dem Umgang mit Lemmata und deren Varianten. Ziel der Modellierung ist deren vollständige Erfassung sowie Präsentation, die sich in erster Linie an der Benutzeranfrage und nicht an der Wörterbucharchitektur orientiert. Das Modell, das die Lemma-Angabe aus dem Kernelement Wörterbucheintrag in eine separate *Lemma*-Klasse auslagert, bricht mit der lexikographischen Tradition und den gängigen Standards. Am ehesten ist sie vergleichbar mit der komprimierten Darstellung optionaler und austauschbarer Elemente in der Lemma-Angabe. Im Unterschied zu den existierenden Ansätzen (z. B. dem SHŌGAKUKAN-NIHONKOKUGODAIJITEN-IW) wird statt einer komprimierten eine selektive Darstellungsform implementiert.

Durch die Realisierung in einer eigenen Klasse wurde eine weitgehende Abdeckung des Phänomens Variation bei gleichzeitiger Reduktion von Redundanzen erreicht. Die Lemma-Klasse repräsentiert Form-Varianten, d. h. Varianten, die sich durch einen hohen Grad an Synonymie auszeichnen. Durch die Zuordnung zu einem lexikalischen Eintrag verfügen sie über eine gemeinsame Selektion von assoziierten Bedeutungsangaben (*Sense*). Meist handelt es sich dabei um lexikalische und phonetische Varianten, jedoch wird auch die Veränderung der Reihenfolge der Komponenten sowie gewisse morphosyntaktische Veränderungen in dieser Art und Weise kodifiziert.

Jede Instanz eines Lemmas wird mit spezifischen Form-Angaben (Phonetik, orthographische Varianten, Morphologie, Variationstyp und Komponenten) angereichert. Einen Überblick über die wichtigsten Felder der Klasse *Lemma* mit Angaben zum Beispiel *hara-no mushi-ga osamaranai* führt Tabelle 4.5 an.

Die modulare Kodierung der Form erleichtert überdies die redaktionelle Verwaltung. So ist es trivial, eine Lemma-Form einem anderen (oder neuen) Eintrag zuzuordnen, falls sich im Laufe der lexikographischen Arbeit zeigen sollte, dass sich eine Variante wider Erwarten inhaltlich

deutlich von dem Rest des Eintrags unterscheidet. Auch ermöglicht es die direkte Referenzierung einzelner Lemma-Formen.

Feld	Erläuterung	Beispiel
<i>lemma_kanji</i>	Lemma-Angabe	腹の虫が治まらない
<i>lemma_hira</i>	Phonetik (Hiragana)	はらのむしがおさまらない
<i>lemma_orth_var</i>	orthographische Varianten	はらのむしが治まらない 腹の虫がおまらない
<i>headword</i>	Lemma-Angabe mit Furigana	<small>はら むし おさ</small> 腹の虫が治まらない
<i>components</i>	lexikal. Elemente (lemmatisiert)	腹, の, 虫, が, 治まる
<i>walink</i>	⇒ Komponenten APEL-WADOKUJT-IW	⇒ 腹の虫, ⇒ 治まる
<i>category_id</i>	Referenz zur Sachkategorie	⇒ 虫
<i>pos</i>	Phrasem-Struktur	NnoNgaVnai

Tabelle 4.5: Attribute der Klasse *Lemma*

4.10.5.3 Bedeutungsangabe (*Sense*)

Die Bedeutungsangabe ist zentral für die Erfassung der Phraseme. Sie beinhaltet neben der klassischen Paraphrase beziehungsweise Übersetzungsangabe auch Angaben zur Etymologie und gegebenenfalls pragmatische Hinweise. Auf Bedeutungsebene sind auch die Verweise zu den Sinnkategorien (onomasiologische Komponente) und zu multimedialen Inhalten (Grafiken und Videos) angelegt, da diese bei polysemen Bedeutungen in der Regel nur eine Bedeutung betreffen. Tabelle 4.6 gibt einen Überblick über die Kern-Elemente der Bedeutungsangabe von *kubi-o hineru* (首を捻る).

Die zunächst eingeführte Trennung zwischen definatorischer und paraphrasierender Bedeutungsangabe wurde aufgehoben, da sich eine Definition komplexer semantischer Zusammenhänge als sehr schwierig und unüblich erwies.

Feld	Erläuterung	Beispiel
<i>paraphrase</i>	Bedeutungsumschreibung	grübeln, etw. anzweifeln
<i>usage</i>	Pragmatik	begleitet von einer [..]
<i>equivalents</i>	Übersetzung (idiom.)	
<i>mm_file</i>	Multimedia-Ressource	(Video der Geste)
<i>category_id</i>	(Sinn-)Kategorie	⇒ Zweifel
<i>pl_type</i>	Phrasem-Typ	<i>kan</i>
<i>etymologie</i>	diachronische Angaben zur Verwendung	

Tabelle 4.6: Attribute der Klasse *Sense*

Jedem Wörterbucheintrag ist mindestens eine Bedeutungsangabe der Klasse *Sense* zugeordnet, mit einigen Wörterbuch-Einträgen sind jedoch mehrere disjunkte Lesarten assoziiert. Zusammengefasst werden dabei nicht nur Bedeutungen derselben Art, auch Kombinationen kompositioneller und idiomatischer Lesarten werden in einem Wörterbucheintrag gebündelt. Um die Auszeichnung des Phrasem-Typs auf Ebene einzelner Bedeutungsangaben spezifizieren zu können, gibt es die Möglichkeit, durch das Attribut *pl_type* den vom lexikalischen Eintrag vererbten Wert zu überschreiben.

Die (konstruierten) Beispielangaben und Belegstellen sind in Abhängigkeit zu den Bedeutungen hinterlegt. Die ursprüngliche Modellierung als 1:n-Beziehung erwies sich für die Beispielangaben als unproblematisch, da diese in der Regel auf eine spezifische Lesart gemünzt sind. Im Falle der Belegstellen wurde jedoch im Lauf der Arbeit festgestellt, dass gelegentlich innerhalb des Minimal-Kontextes, welcher zum Verständnis des Belegs notwendig ist, weitere idiomatische Ausdrücke enthalten waren. Zur Vermeidung von Redundanz wurde deshalb für Belege eine m:n-Relation zwischen Bedeutungsangaben und Belegen implementiert.

4.10.5.4 Kategorie (*Category*)

Die Kategorisierung der Sach- und Sinnebene wird durch die zwei Klassen *Object Category* und *Sense Category* repräsentiert. Der entscheidende Unterschied zwischen den beiden Typen

ist rein inhaltlicher Natur. Strukturell gibt es keine wesentlichen Unterschiede zwischen Kategorien, die die wörtliche Bedeutung der Komponenten beziehungsweise die übertragene Gesamtbedeutung repräsentieren, weshalb *ObjectCategory* und *SenseCategory* alle wesentlichen Attribute von der abstrakten Klasse *Categories* erben. Aufgrund von Single Table Inheritance (STI) werden die Elemente der Kindklassen in einer gemeinsamen Tabelle gespeichert werden, die sich nur durch den Feldwert *type* unterscheidet (siehe Tabelle 4.7).

Feld	Erläuterung	Beispiel
<i>key_ja</i>	Begriff (japanisch)	怒り
<i>key_de</i>	Begriff (dt. Übersetzung)	Wut, Ärger
<i>type</i>	Kategorientyp (STI)	Sense Category
<i>add_keys_ja</i>	Schlagworte japanisch	怒る
<i>synopse_id</i>	Link zu Synopse-Artikel	⇒HARA IS THE CONTAINER OF THE EMOTIONS

Tabelle 4.7: Attribute der Klasse *Category* und ihrer Kindklassen

Unterschieden werden die Klassen durch ihre Relationen. *Object Category* bildet aufgrund der ausgeprägten lexikalischen Variation Relationen zur *Lemma*-Klasse aus. *Sense Category* ist mit Bedeutungen assoziiert. Darüber hinaus kann ausgehend von *Category* auch auf Synopse-Artikel referenziert werden.

Zusätzliche Relationen sorgen für eine Hierarchisierung der Kategorien in drei Ebenen. Daraus ergibt sich eine Baumstruktur, die für den onomasiologischen Zugriff eingesetzt wird.

4.10.5.5 Synopseartikel (*Synopse*)

Die Klasse für Synopseartikel ermöglicht einen enzyklopädischen Wörterbucheintrag anzulegen, der sich inhaltlich und relational von *Mwe Entry* unterscheidet. Er dient dazu, mehreren Wörterbucheinträgen eine singuläre Einheit semantischer Information zuzuordnen. Charakteristisch für die Klasse ist neben einer Stichwortangabe (*lemma*) die Angabe einer allgemeinen und einer etymologischen Bedeutung.

Feld	Erläuterung	Beispiel
<i>lemma</i>	Lemma	虫
<i>entry</i>	Bedeutungserläuterungen	Für sich alleine bedeutet [...]
<i>etymology</i>	Etymologie	Nach daoistischer Vorstellung leben [...]

Tabelle 4.8: Attribute der Klasse *Synopse*

4.10.5.6 Verweis (*Reference*)

Verweise repräsentieren die Beziehungen zwischen lexikalischen Einträgen (*Mwe Entry*) und einzelnen Lesarten (*Sense*) oder Einträgen im APEL-WADOKUJT-IW. Sie sind typisiert, um eine Unterscheidung zwischen Verweisen zu Synonymen, Antonymen etc. zu ermöglichen. Tabelle 4.9 zeigt einen Referenzeintrag anhand des Beispiels für einen Verweis zwischen der zweiten Lesart von *te-o ageru* (手を上げる) und dem Eintrag für *ude-o migaku* (腕を磨く seine Fähigkeiten verbessern).

Feld	Erläuterung	Beispiel
<i>sense_id</i>	Anker (Lesart)	⇒ Fähigkeiten verbessern [<i>te-o ageru</i>]
<i>mwe_entry_id</i>	Ziel (Artikel)	⇒ <i>ude-o migaku</i>
<i>wadoku_id</i>	Link APEL-WADOKUJT-IW	-
<i>reference_type</i>	Verweis-Typ	Synonym

Tabelle 4.9: Attribute der Klasse *Reference*

4.10.5.7 Beispiel (*Example*)

Bei Beispielangaben (*Example*) handelt es sich um konstruierte Kurztexte. Tabelle 4.10 gibt einen Überblick über die Attribute anhand des Beispiels *aita kuchi-ga fusagaranai* (開いた口が塞がらない staunen; jmd. die Sprache verschlagen). Im Gegensatz zu Belegen ist für Beispiele eine deutsche Übersetzung vorgesehen. Überdies dient die Valenzstruktur des Beispiels als Hilfestellung für die strukturelle Erfassung der *kanyōku*.

Feld	Erläuterung	Beispiel
<i>example_ja</i>	Beispiel (japanisch)	[...]今度は自民党とは <mwe id='243'>開いた口が ふさがらない。</mwe>
<i>example_de</i>	Beispiel (dt. Übers.)	[...] das Verhalten der LDP in letzter Zeit verschlägt einem die Sprache.
<i>form</i>	interne Form und externe Valenz	Ntowa <i>aita kuchi-ga fusaganarai</i>

Tabelle 4.10: Attribute der Klasse *Example*

4.10.5.8 Beleg (*Citation*)

Belege werden aufgrund ihres Umfangs in der Regel nicht übersetzt. Die Mehrzahl der Felder dienen dem bibliographischen Nachweis. Die Belegstelle selbst ist annotiert, um eine farbige Hervorhebung der Textstelle zu ermöglichen. Da in Belegen teilweise auch unterschiedliche *kanyōku* enthalten sind, geht aus der Annotation auch hervor, auf welches Lemma sich die Auszeichnung bezieht. Tabelle 4.10 zeigt eine Auswahl der Felder der *Citation*-Klasse.

Feld	Erläuterung	Beispiel
<i>source</i>	Referenz Quelle	OB4X_00273 3260
<i>source_type</i>	Genre	mag
<i>year</i>	Publikationsjahr	1993
<i>author</i>	Autor	小沢一郎
<i>title</i>	Titel	
<i>citation</i>	Beleg	多くの政治家は一分でも一秒でも長く気持ちのいい[...]

Tabelle 4.11: Attribute der Klasse *Citation*

4.10.5.9 Kompatibilität zu Standards und Interoperabilität

Interoperabilität soll einerseits den Datenaustausch mit Werkzeugen und Ressourcen ermöglichen, auf der anderen Seite knüpft sich daran die Hoffnung, durch die Konformität mit Standards die Nachhaltigkeit des Nachschlagewerks zu sichern. Dabei bauen die unterschiedlichen Ebenen der Datenmodellierung aufeinander auf.

Auf unterster Ebene wird dies durch die Zeichenkodierung in Unicode umgesetzt. Eine REST-Schnittstelle ermöglicht den Datenaustausch in den gängigen Formaten (XML und JSON). Das interne Datenformat wird dabei in XML-Fragmente umgeformt, die einzelne Datensätze in den in Kapitel 3.4 erläuterten Formaten TEI und LMF ausliefern.

```
<lexical_entry id="198" pos="phraseological unit">
  <lemma id="196">
    <FormRepresentation>
      <feat attr="script" val="kanji"/>
      <feat attr="written_form" val="首を捻ねる"/>
    </FormRepresentation>
    <FormRepresentation>
      <feat attr="script" val="hiragana"/>
      <feat attr="written_form" val="くびをひねる"/>
    </FormRepresentation>
    <list_of_components>
      <component>
        <feat attr="written_form" val="首"/>
      </component>
      <component>
        <feat attr="written_form" val="を"/>
      </component>
      <component>
        <feat attr="written_form" val="捻ねる"/>
      </component>
    </list_of_components>
    <mwe_pattern/>
  </lemma>
  <sense id="197">
    <SemanticDefinition>
      <feat att="text" val="grübeln; etw. anzweifeln"/>
    </SemanticDefinition>
  </sense>
</lexical_entry>
```

Listing 4: Beispiel für die LMF-Schnittstelle im WADOKUIDIOMS-IW

Listing 4 zeigt anhand eines kurzen Beispiels die XML-Version des LMF-Formats. An der singulären Lemma-Angabe wird deutlich, dass das Austauschformat gegebenenfalls im Dialog mit

anderen Projekten weiter ausdifferenziert werden muss, um den Umfang der enthaltenen Informationen sinnvoll nutzen zu können. Am Deutlichsten wird dies in Bezug auf die Frage der Variabilität beziehungsweise Synonymität von Phrasemen sowie bezüglich der Äquivalenzrelationen.

Die TEI-Richtlinien bieten überwiegend semantisches Markup. Phrasem-spezifische und strukturelle Vorgaben sind nur rudimentär spezifiziert. Aus diesem Grund sind die Datensätze, die durch die TEI-Schnittstelle zur Verfügung gestellt werden, auf die in Kapitel 3.4.2.1 erläuterten Kerninformationen reduziert.

4.11 Fazit

In der vorliegenden Arbeit wurde exemplarisch gezeigt, welche Anforderungen an die Phraseographie sich aus dem derzeitigen Forschungsstand ableiten und in der Modellierung eines elektronischen Wörterbuchs für die Idiomatik Japanisch-Deutsch in die Praxis umsetzen lassen. Die Anforderungsanalyse ergab, dass der Repräsentation von Variabilität und der Beschreibung semantischer Eigenschaften vorrangig Aufmerksamkeit gewidmet werden muss. Insbesondere das Themengebiet Variabilität stellt eine der größten Herausforderungen bei der Kodierung im elektronischen Medium dar. Dies ist in Teilen der Komplexität des japanischen Schriftsystems geschuldet. Die Analyse einschlägiger japanischer Nachschlagewerke gab hier wichtige Impulse, zeigte jedoch auch, dass die derzeitigen digitalen lexikalischen Ressourcen den Bereich noch nicht systematisch erfassen. Um den Anforderungen nach flexiblem Zugriff auf Phraseme im digitalen Wörterbuch gerecht zu werden und gleichzeitig die Datenkonsistenz zu gewährleisten, ist eine Typisierung der Variationsformen zwingend erforderlich. Das Herauslösen der Lemma-Angabe aus dem Wörterbuchartikel selbst ermöglicht dabei die Kombinatorik verschiedener Variationsformen. Dadurch kann sicher gestellt werden, dass die Variabilität in vollem Umfang kodiert und gleichzeitig Redundanz in den Daten vermieden werden kann. Insgesamt lässt sich sagen, dass sich das Thema Variabilität als weitaus vielschichtiger entpuppte als ursprünglich vermutet.

Umgekehrt erschöpfte sich die Gestaltung der onomasiologischen Komponente in der Frage nach der Gliederung des Objektbereichs und der Art des Zugriffs. Selbst die Modellierung des relativ gut erforschten Teilbereichs der somatischen Idiome erwies sich als nicht trivial. Die

Ergebnisse der Analyse onomasiologischer Wörterbücher zeigen, dass eine Erfassung aller Objektbereiche in einer intuitiv nachvollziehbaren Ordnung eine hochgradig komplexe Aufgabe ist.

Die immensen Unterschiede, die sich in den untersuchten Bereichen widerspiegeln, deuten darauf hin, dass eine Einfachkategorisierung nicht ausreichend sein kann. Für den intuitiven Zugriff müssen Kategorien über mehrere alternative Strukturen vernetzt werden. Aus technischer Perspektive stellt dies zwar keine besondere Herausforderung dar, inhaltlich bedeutet dies jedoch einen wesentlichen Mehraufwand. Offen ist auch die Frage, wie dem Nutzer die alternativen Zugriffsstrukturen präsentiert werden sollen. L'Homme, Robichaud und Leroyer (2013, S. 226–231) zeigen in einer Pilotstudie, dass die Mehrheit der Probanden von der onomasiologischen Zugriffsmöglichkeit profitiert. Eine Klärung, welchen Anforderungen dieses Modul genügen muss, um lexikalische Ressourcen um einen greifbaren Mehrwert zu bereichern, steht noch aus. Eine Weiterentwicklung ist daher nur vor dem Hintergrund begleitender Nutzerstudien sinnvoll. Beispielsweise muss sich erst noch zeigen, ob Sinn- und Sachkategorien gleichermaßen nachgefragt werden und welche Stellung eine Stichwortsuche in aktiven Nutzungskontexten einnehmen kann.

Mit weit weniger Aufwand ließe sich die Verschlagwortung durch Beteiligung der Wörterbuchnutzer ausweiten. Hierfür könnten auf Thesauren basierende Vorschlagslisten als Hilfestellung und Mechanismus der Qualitätskontrolle eingesetzt werden. Eine Vernetzung der Kategorien untereinander könnte WordNet-ähnliche Strukturen wachsen lassen, die die Spezifika des Anwendungsgebiets reflektieren.

Der Themenkomplex adäquate semantische Beschreibung gehört wohl zu den umstrittensten in der Phraseographie. Die Analyse hat gezeigt, dass kollokative und idiomatische Bedeutungskomponenten nicht immer deutlich von einander abgrenzbar sind. Insbesondere motivierte Idiome verfügen meist auch über eine kollokative Lesart. Deren systematische Kodierung ist für Sprachlerner besonders wichtig. Das digitale Medium eröffnet eine Reihe interessanter Anwendungsgebiete, darunter beispielsweise die Frage nach der Modellierung der Schnittstelle zwischen verbaler und non-verbaler Kommunikation. Hier wurde durch die multimediale Erfassung von Körpersprache eine Dimension der Bedeutungskodierung in phraseologische Wörterbücher eingefügt, die es meines Erkenntnisstandes in dieser Form bislang nicht gab.

Zu den vernachlässigten Bereichen in der Phraseographie zählt allen voran die Beschreibung der Grammatik. Lediglich eines der analysierten japanischen Nachschlagewerke war um die

systematische Beschreibung der Valenzen bemüht. Für den Einsatz in der natürlichen Sprachverarbeitung ist eine Lesarten-bezogene Kodierung grammatischer Besonderheiten von Phrasemen jedoch unabdingbar. In Zusammenhang mit korpuslinguistischen Methoden eröffnen sich hierdurch neue Perspektiven für die Lesartendisambiguierung und die Eruiung belegbarer Formen.

Deutlich wurden in der vorliegenden Arbeit jedoch auch die Grenzen, an welche die Neukonzeption eines phraseologischen Wörterbuchs für das Sprachpaar Japanisch-Deutsch derzeit stößt. Zwar hat sich die Verfügbarkeit von Korpusdaten für das Japanische im Laufe des Entstehens dieser Arbeit durch die Veröffentlichung des BCCWJ deutlich verbessert, dennoch fehlen verlässliche Daten für das gesprochene Japanisch. Die Argumentation, dass das Internet als Ressource für die gesprochene Umgangssprache dienen kann, auf welcher das Online-Wörterbuch ETTINGERPHRASEOLOGIEDE-IW aufbaut, haben sich nicht bestätigt. Die Belege des BCCWJ zeigten keine deutliche Tendenz in diese Richtung, obwohl auch Internettex-te zu einem wesentlichen Teil das Korpus konstituieren. Auch konnte die Palette linguistischer Werkzeuge des BCCWJ systematisch nur Inhalte abfragen, die sich innerhalb der prognostizierten Variabilitätsformen bewegen. Es fehlen Methoden der verlässlichen Extraktion und Lesartendisambiguierung unbekannter *kanyōku*. Aufgrund dieser Einschränkungen können Aussagen über Erscheinungsformen in der Regel nur affirmativ getroffen werden. Belastbaren Ergebnisse zu Restriktionen und Frequenz sind folglich nicht aus der Korpusanalyse zu erzielen.

Ernüchternd fällt auch das Fazit zur Einbindung externer Ressourcen aus. Die Notwendigkeit von Standards in der Lexikonkodierung wird hier ebenso deutlich wie die einer einheitlichen Typisierung im Bereich Phraseologie. Auch offene Fragen nach der Integrität und Verfügbarkeit der externen Ressourcen sind ein Unsicherheitsfaktor, der sich negativ auf die Interoperabilität von lexikalischen Ressourcen auswirkt.

Das digitale Zeitalter lässt die Menge der Anforderungen und Wunschvorstellungen, die an die Phraseographie herangetragen werden, exponentiell anwachsen. Um innovativ und systematisch Sprache zu beschreiben, muss man sich von bestehenden Strukturen lösen (Atkins, Kegl und Levin 1988). Die rasche technologische Weiterentwicklung ist in gewisser Weise gleichzeitig treibende Kraft und bremsender Faktor, da dem raschen Paradigmen-Wechsels auf technischer Seite eine lange Inkubationszeit bei der inhaltlichen Erarbeitung lexikalischer Ressourcen

gegenübersteht. So bedarf es sorgfältiger Abwägung, welche Entwicklungen tatsächlich nachhaltigen Nutzen erbringen, denn vermutlich wird sich dieser Trend fortsetzen. Die digitale Lexikographie kann damit nur Schritt halten, wenn auch die Hilfswerkzeuge zur Erschließung von Primär- und Sekundärdaten von dem technologischen Fortschritt profitieren werden und der Gewährleistung von Integrität, Interoperabilität und Nachhaltigkeit der Daten zukünftig mehr Gewicht beigemessen wird.

Literatur

Sekundärliteratur

- Adachi, Ryōko, Kōichi Takeuchi, Ryō Murayama, Wolfgang Fanderl, Rei Miyata, Iris Vogel, Ulrich Apel und Kyō Kageura (2013). »Development and Use of a Platform for Defining Idiom Variation Rules«. In: *Proceedings of the 5th International Language Learning Conference (ILLC 2013)*. Penang, Malaysia, S. 1–19.
- Almind, Richard (2005a). »Designing internet dictionaries«. In: *Journal of Linguistics* 34, S. 37–54.
- Almind, Richard (2005b). »First impressions last: layout in dictionaries«. In: *Finnische Beiträge zur Germanistik*, Bd. 14. New York: Peter Lang, S. 93–103.
- Almind, Richard, Henning Bergenholtz und Vibeke Vrang (2006). »Theoretical and Computational Solutions for Phraseological Lexicography«. In: *Linguistik online* 27.23, S. 159–181.
- Alsleben, Brigitte, Hrsg. (2002). *Der Duden in zwölf Bänden*. 2., neu bearb. und aktualisierte Aufl. Bd. 11: Duden, Redewendungen: Wörterbuch der deutschen Idiomatik. Mannheim: Dudenverlag.
- Apel, Ulrich (2011). »Enhancing a Japanese-German Dictionary through the Integration of Multimedia Data«. In: *LEXICOGRAPHY: Theoretical and Practical Perspectives. Proceedings of Asialex 2011*. Kyoto: The Asian Association of Lexicography, S. 39–45.
- Arnold-Kanamori, Horst (2001). *Großer Zampano: japanische Sprichwörter*. Bd. 6. Ulmer Sprachstudien. Hamburg: Kovač.
- Asahi Shinbun (16. Jan. 2008). 追隨許さぬ広辞苑人気 (*Beliebtheit des Kōjien ist nicht zu übertreffen*). URL: <http://book.asahi.com/clip/TKY200801160127.html>, aufgerufen am 12.05.2012. Archiviert unter <http://www.webcitation.org/67bY4gO9Q>.
- Atkins, Beryl T., Judy Kegl und Beth Levin (1988). »Anatomy of a Verb Entry: from Linguistic Theory to Lexicographic Practice«. In: *International Journal of Lexicography* 1.2, S. 84–126. eprint: <http://ijl.oxfordjournals.org/cgi/reprint/1/2/84.pdf>.

- Barz, Irmhild (1992). »Phraseologische Varianten: Begriff und Probleme«. In: *Deutsche Phraseologie in Sprachsystem und Sprachverwendung*, S. 25–47.
- Barz, Irmhild, Henning Bergenholtz und Jarmo Korhonen (2005). *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. Finnische Beiträge zur Germanistik, Bd. 14. New York: Peter Lang.
- Bergenholtz, Henning (2011). »Access to an Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum.
- Bergenholtz, Henning und Inger Bergenholtz (2011). »A Dictionary Is a Tool, a Good Dictionary Is a Monofunctional Tool«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum, S. 187–207.
- Bergenholtz, Henning, Sandro Nielsen und Sven Tarp, Hrsg. (2009). *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Bd. 90. Linguistic Insights: Studies in Language and Communication. Bern: Peter Lang.
- Bergenholtz, Henning und Sven Tarp (2002). »Die moderne lexikographische Funktionslehre Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen«. In: *Lexicographica. Series maior* 18, S. 253–263.
- Bergenholtz, Henning und Sven Tarp (2005). »Verteilungsstrukturen in Wörterbüchern«. In: Finnische Beiträge zur Germanistik, Bd. 14. New York: Peter Lang.
- Berndt, Jürgen, Hrsg. (1984). *Sprichwörter aus Japan*. 1. Aufl. Berlin: Verlag Volk u. Welt.
- Berndt, Jürgen und Suiko Simon, Hrsg. (1992). *Gutes ist am besten gleich getan: 100 Sprichwörter aus Japan*. Berlin: Ed. q.
- Boers, Frank und Murielle Demecheleer (2001). »Measuring the impact of cross-cultural differences on learners' comprehension of imageable idioms«. In: *ELT Journal* 55.3, S. 255–262.
- Boers, Frank, Murielle Demecheleer und June Eyckmans (2004). »Etymological elaboration as a strategy for learning idioms«. In: *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Amsterdam: John Benjamins, S. 53–78.

- Boers, Frank, June Eyckmans und H  l  ne Stengers (2007). »Presenting figurative idioms with a touch of etymology: more than mere mnemonics?« In: *Language Teaching Research* 11.1, S. 43–62.
- Boers, Frank, Ana Mar  a Piquer P  riz, H  l  ne Stengers und June Eyckmans (2009). »Does pictorial elucidation foster recollection of idioms?« In: *Language Teaching Research* 13.4, S. 367–382.
- Bond, Francis, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi und Kyoko Kanzaki (2009). »Enhancing the Japanese WordNet«. In: *ALR7: Proceedings of the 7th Workshop on Asian Language Resources*. Stroudsburg, PA, USA: Association for Computational Linguistics, S. 1–8.
- Bothma, Theo JD (2011). »Filtering and Adapting Data and Information Needs in an Online Environment in Response to User Needs«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum.
- Bowker, Lynne (2013). »Meeting the needs of translators in the age of e-lexicography: Exploring the possibilities«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 379–397.
- Broeder, Daan, Malte Dreyer, Marc Kemps-Snijders, Andreas Witt, Marc Kupietz und Peter Wittenburg (2008). *Persistent and Unique Identifiers. CLARIN-Deliverable D2.2 EC FP7 project no. 212230*. Techn. Ber. MPI for Psycholinguistics /CLARIN - Common Language Resources und Technology Infrastructure.
- Burger, Harald (1983). »Phraseologie in den W  rterb  chern des heutigen Deutsch«. In: *Germanistische Linguistik*, S. 13–66.
- Burger, Harald (1989). »Phraseologismen im allgemeinen einsprachigen W  rterbuch«. In: *W  rterb  cher / Dictionaries / Dictionnaires*. Hrsg. von Franz-Josef Hausmann, Oskar Reichmann, Herbert E. Wiegand und Ladislav Zgusta. Handb  cher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science [HSK] 5.1. Berlin, New York: Walter de Gruyter, S. 593–599.
- Burger, Harald (2002). »Die Charakteristika phraseologischer Einheiten: Ein   berblick«. In: *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von W  rtern und Wortsch  tzen*. Hrsg. von D. Alan Cruse, Franz Hundsnurscher, Michael Job und Peter Rolf Lutz-eier. Bd. 1. de Gruyter, S. 392–401.

- Burger, Harald (2003). *Phraseologie - Eine Einführung am Beispiel des Deutschen*. Grundlagen der Germanistik 36. Berlin: Erich Schmidt.
- Burger, Harald (2007). *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung*. Bd. 28. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter.
- Burger, Harald (2009). »Semantische Aspekte der deutschen Phraseografie: die aktuelle Praxis – allgemeine und phraseologische Wörterbücher im Vergleich«. In: *Theorie und Praxis der idiomatischen Wörterbücher*. Bd. 135. de Gruyter, S. 23.
- Burger, Harald (2010). *Phraseologie: eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt.
- Burger, Harald, Dmitrij Dobrovolskij, Peter Kühn und Neal R. Norrick (2007). »Einführung/ Subject area, terminology and research topics«. In: *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung = Phraseology: an international handbook of contemporary research*. Bd. 28. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter.
- Calzolari, Nicoletta (1989). »Computer-aided lexicography: dictionaries and word data bases«. In: *Hausmann (1989), Wörterbücher*, S. 510–837.
- Bunkachou (1978). 送り仮名の付け方 (*Die Verwendung von okurigana*). URL: <http://www.bunka.go.jp/kokugo/frame.asp?tm=20091108161240>, aufgerufen am 06. 11. 2009. Archiviert unter: <http://www.webcitation.org/64boR2qM4> am 11.01.2012.
- Colson, Jean-Pierre (2010). »The Contribution of Web-based Corpus Linguistics to a Global Theory of Phraseology«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren.
- Copestake, Ann, Fabre Lambeau, Aline Villavicencio, Francis Bond, Timothy Baldwin, Ivan Sag und Dan Flickinger (2002). »Multiword expressions: Linguistic precision and reusability«. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, S. 1941–1947.
- Cowie, Anthony Paul (1999). »Phraseology and Corpora: Some Implications for Dictionary-Making«. In: *International Journal of Lexicography* 12.4, S. 307–323.

- Dobrovol'skij, Dmitrij (2000). »Contrastive Idiom Analysis: Russian and German Idioms in Theory and in the Bilingual Dictionary«. In: *International Journal of Lexicography* 13.3, S. 169–186.
- Dobrovol'skij, Dmitrij (2002). »Phraseologismen im de Gruyter Wörterbuch Deutsch als Fremdsprache«. In: *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des "de Gruyter Wörterbuch Deutsch als Fremdsprache"*. Hrsg. von Herbert E. Wiegand, S. 363–374.
- Dobrovol'skij, Dmitrij (2009). »Zur lexikografischen Repräsentation der Phraseme (mit Schwerpunkt auf zweisprachigen Wörterbüchern)«. In: *Theorie und Praxis der idiomatischen Wörterbücher*. Bd. 135. Berlin: de Gruyter, S. 149–168.
- Dobrovol'skij, Dmitrij und Tat'jana Filipenko (2007). »Russian Phraseology«. In: *Phraseologie: ein internationales Handbuch zeigenössischer Forschung*. Berlin: de Gruyter.
- Dobrovol'skij, Dmitrij und Elisabeth Piirainen (2010). »Idioms: Motivation and etymology«. In: *Yearbook of Phraseology (2010)*. Berlin: de Gruyter, S. 73–96.
- Dräger, Marcel und Britte Juska-Bacher (2010). »Online-Datenerhebungen im Dienste der Phraseographie«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren, S. 165–179.
- Eerten, Laura van (Feb. 2010). *Data Category: Lexicon*. URL: <http://www.isocat.org/rest/dc/2811>, aufgerufen am 28. 10. 2012. Archiviert mit <http://www.webcitation.org/6Bk8LqHd6> am 28.10.2012.
- Eerten, Laura van (2012). *Data Category: Dictionary*. URL: <http://www.isocat.org/datcat/DC-2812>, aufgerufen am 28. 10. 2012. Archiviert unter <http://www.webcitation.org/6Bk8Wr0kq> am 28.10.2012.
- Ehegötz, Erika (Apr. 1990). »Versuch einer Typologie von Entsprechungen im zweisprachigen phraseologischen Wörterbuch«. In: *Zeitschrift für Slawistik* 4, S. 499–504.
- Ehmann, Paul (1893-1897). »Japanische Sprichwörter und sprichwörtliche Redensarten. Gesammelt und übersetzt.« In: *Mittheilungen der deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens*, S. 70–102.

- Ehmann, Paul (1897). *Japanische Sprichwörter und sprichwörtliche Redensarten. Gesammelt und übersetzt*. Tokyo: Deutsche Gesellschaft für Natur- und Völkerkunde Ostasiens (Supplement III der Mittheilungen der Gesellschaft).
- Ekman, Paul (1992). »Are there basic emotions?« In: *Psychological Review* 99, S. 550–553.
- Engel, Michael (2012). *Furigana*. Beitrag in Mailingliste: J-Studien. URL: <https://listserv.shuttle.de/mailman/private/j-studien/2012/006052.html>.
- Engelberg, Stefan und Lothar Lemnitzer (2009). *Lexikographie und Wörterbuchbenutzung*. Bd. 14. Stauffenburg Einführungen. Tübingen: Stauffenburg, XI, 327 S.
- Eschbach-Szabo, Viktoria (2001). »Lust und Last: Japanisch, die Komplizierteste Schrift der Welt im Elektronischen Zeitalter«. In: *Szines eszmék nem alszanak. (Farbige Ideen schlafen nicht)*. Bd. 1. Szépe György 70. születésnapjára. (Szépe-Festschrift). József Andor, Tibor Szücs und Terts István, S. 307–325.
- Espinal, Theresa (2005). »A conceptual dictionary of Catalan idioms«. In: *International Journal of Lexicography* 18.4, S. 509–540.
- Ettinger, Stefan (2009a). »Haben die Männer am Grill die Hosen an?« Phraseografie und Sprachwirklichkeit«. In: *Theorie und Praxis der idiomatischen Wörterbücher*. Hrsg. von Carmen Mellado Blanco. Bd. 135. Lexicographica. Series maior. Tübingen: Niemeyer, S. 45–64.
- Ettinger, Stefan (2009b). *Meine persönliche Phraseologiesammlung*. URL: <http://www.ettinger-phraseologie.de/pages/vorwort/vorwort5.php>, aufgerufen am 05.01.2012. Archiviert unter: <http://www.webcitation.org/64bsopi2e> am 11.1.2012.
- Ettinger, Stefan (2009c). *Phraseologie und Phraseodidaktik*. URL: <http://www.ettinger-phraseologie.de/>, aufgerufen am 05.01.2012. Archiviert unter: <http://www.webcitation.org/64buhobM0> am 11.1.2012.
- Ettinger, Stefan (2009d). *Warum Belege aus dem Internet?* URL: <http://www.ettinger-phraseologie.de/pages/vorwort/vorwort3.php>, aufgerufen am 05.01.2012. Archiviert unter: <http://www.webcitation.org/64blnA8lz> am 11.1.2012.
- Ettinger, Stefan (2009e). *Was hat sich bei der Neubearbeitung für das Internet (2007 bis 2009) gegenüber der Buchform (1997) geändert und welche Teile sind gleich geblieben?* URL: <http://www.ettinger-phraseologie.de/pages/vorwort/vorwort2.php>, aufgerufen am 05.01.2012. Archiviert unter: <http://www.webcitation.org/64bpi9UMG> am 11.1.2012.

- Ettinger, Stefan (2009f). *Weshalb diese Internetausgabe?* URL: <http://www.ettinger-phraseologie.de/pages/vorwort/vorwort1.php>, aufgerufen am 05.01.2012. Archiviert unter: <http://www.webcitation.org/64buTE3N9> am 11.1.2012.
- Ettinger, Stefan (2013). *B.2.5 Missmut - Ärger - Wut*. URL: <http://www.ettinger-phraseologie.de/pages/buchinhalt/b/b.2/b.2.5.php>, aufgerufen am 18.01.2013. Archiviert mit unter: <http://www.webcitation.org/6DILhLR8am> 11.1.2012.
- Fellbaum, Christiane (2004). »Idiome in einem digitalen lexikalischen System«. In: *Philologie auf neuen Wegen*. Stuttgart u.a.: Metzler.
- Filipenko, Tat'jana (2009). »Darstellung von deutschen und russischen Idiomen in zweisprachigen Wörterbüchern«. In: *Theorie und Praxis der idiomatischen Wörterbücher*. Bd. 135. Berlin: Walter de Gruyter, S. 168–186.
- Fjeld, Ruth Vatvedt, Lars Nygaard und Eckhard Bick (2010). »Semi-Automatic Retrieval of Phraseological Units in a Corpus of Modern Norwegian«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren, S. 165–179.
- Fleischer, Wolfgang (1997). *Phraseologie der deutschen Gegenwartssprache*. 2., durchgesehene und ergänzte Auflage. Niemeyer Studienbuch. Tübingen: Niemeyer.
- Földes, Csaba (1994). »Mehrsprachige phraseologische Wörterbücher – als Herausforderung für die Lexikographie«. In: *SANDIG, Barbara (Hg.): Tendenzen der Phraseologieforschung (Europhras 92)*, S. 175–201.
- Francopoulo, Gil, Nuria Bel, Monte George, Nicoletta Calzolari, Monica Monachini, Mandy Pet und Claudia Soria (2006). »Lexical markup framework (LMF) for NLP multilingual resources«. In: *Proceedings of the workshop on multilingual language resources and interoperability*. Association for Computational Linguistics, S. 1–8.
- Francopoulo, Gil, Nuria Bel, Monte George, Nicoletta Calzolari, Monica Monachini, Mandy Pet und Claudia Soria (2009). »Multilingual resources for NLP in the Lexical Markup Framework (LMF)«. In: *Language Resources and Evaluation* 43.1, S. 57–70.
- Fraser, Bruce (2011). *The TEI Dictionary Schema: its worth and weakness for the lexicographer*. URL: <http://people.ds.cam.ac.uk/blf10/links/TEI.html>, aufgerufen am 01.01.2013. Archiviert unter <http://www.webcitation.org/6DLsnr2bS>.

- Fuertes-Olivera, Pedro A. und Henning Bergenholtz, Hrsg. (2011). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London, New York: Continuum.
- Fuertes-Olivera, Pedro A. und Marta Niño-Amo (2011). »Internet Dictionaries for Communicative and Cognitive Functions: *El Diccionario Inglés-Español de Contabilidad*«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum, S. 168–186.
- Gehweiler, Elke (2006). »Going to the Dogs? A Contrastive Analysis of S.th. is Going to the Dogs and jmd./etw. geht vor die Hunde«. In: *International Journal of Lexicography* 19.4, S. 419–438.
- Gouws, Rufus H. (2009). »Dictionaries as Innovative Tools in a New Perspective on Standardisation«. In: *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Hrsg. von Henning Bergenholtz, Sandro Nielsen und Sven Tarp. Bd. 90. Linguistic Insights: Studies in Language and Communication. Bern: Peter Lang, S. 265–283.
- Granger, Sylviane und Magali Paquot, Hrsg. (2013). *Electronic lexicography*. Oxford: Oxford University Press.
- Halpern, Jack (2000). »Lexicon-based orthographic disambiguation in CJK intelligent information retrieval«. In: *Proc. of Workshop on Asian Language Resources and International Standardization*.
- Hashimoto, Chikara, Satoshi Satō und Takehito Utsuro (2006-05-19). »Classification and Lexical Knowledge of Japanese Idioms for Automatic Idiom Recognition«. In: *IPSJ SIG Notes* 2006.53, S. 59–66.
- Hashimoto, Hiroko (1993). »Teaching Japanese Body Language Phrases to JFL Students«. In: *Journal of College of International Studies* 11, S. 141–159.
- Hausmann, Franz-Josef (1983). »Was taugen die Wörterbücher des heutigen Deutsch«. In: *Wortschatz und Verständigungsprobleme: Was sind »schwere Wörter« im Deutschen?* Hrsg. von Helmut Henne und Wolfgang Mentrup. Düsseldorf: Pädagogischer Verlag Schwann-Bagel, S. 195–217.
- Heid, Ulrich (2007). »Computational linguistic aspects of phraseology II«. In: *Phraseologie-Phraseology. Ein internationales Handbuch der zeitgenössischen Forschung. An International Handbook of Contemporary Research*. Hrsg. von Harald Burger, Dmitrij Dobrovolskij, Peter Kühn und Neal R. Norrick. Berlin: Mouton de Gruyter, S. 1036–1044.

- Heid, Ulrich (2011). »Electronic Dictionaries as Tools: Towards an Assessment of Usability«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum, S. 287–304.
- Heine, Antje (2009). »Möglichkeiten und Grenzen der Korpusanalyse für die Lexikografie am Beispiel eines Wörterbuchs deutscher Funktionsvergefüge mit finnischen Äquivalenten«. In: *Theorie und Praxis der idiomatischen Wörterbücher*. Hrsg. von Carmen Mellado Blanco. Bd. 135. *Lexicographica. Series maior*. Tübingen: Niemeyer, S. 233–250.
- Heinecke, Ilse (1999). »Rezension zu: Hessky, Regina; Ettinger, Stefan: Deutsche Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene.« In: *Informationen Deutsch als Fremdsprache: Info DaF* 29.29, S. 210.
- Herbst, Thomas (2004). *Lexikografie, ihre Basis- und Nachbarwissenschaften: (englische) Wörterbücher zwischen »common sense« und angewandter Theorie*. *Lexicographica. Series maior* 118. Tübingen: Niemeyer.
- Herold, Axel (2007). »Corpus Queries«. In: *Idioms and Collocations*. Hrsg. von Christiane Fellbaum. *Research in corpus and discourse*. London: Continuum.
- Hessky, Regina (1993). *Durch die Blume — Arbeitsbuch zur deutschen Phraseologie für Fortgeschrittene — Virágnyelven*. Nemzeti Tankönyvk.
- Hessky, Regina und Stefan Ettinger (1997). *Deutsche Redewendungen: ein Wörter- und Übungsbuch für Fortgeschrittene*. *Narr-Studienbücher*. Tübingen: Narr.
- Hrisztova-Gotthardt, Hrisztalina (2010a). »Methoden und Ergebnisse einer korpusbasierten Untersuchung zur Vorkommenshäufigkeit bulgarischer Sprichwörter in zeitgenössischen Zeitungstexten«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. *Phraseologie und Parömiologie*. Baltmannsweiler: Schneider Hohengehren, S. 55–65.
- Hrisztova-Gotthardt, Hrisztalina (2010b). *Vom gedruckten Sprichwörterbuch zur interaktiven Sprichwortdatenbank*. Bern [u. a.]: Lang.
- Hümmer, Christiane und Katerina Stathi (2006). »Polysemy and Vagueness in Idioms: A Corpus-based Analysis of Meaning«. In: *International Journal of Lexicography* 19, S. 361–377.
- Ide, Nancy und Laurent Romary (2004). »A registry of standard data categories for linguistic annotation«. In: *Proceedings of the Fourth Language Resources and Evaluation Conference*, S. 135–39.

- Ilsou, Robert (1987). »Illustrations in dictionaries«. In: *The Dictionary and the Language Learner — Papers from the EURALEX Seminar at the University of Leeds, 1—3 April 1985*. Hrsg. von Anthony Cowie. *Lexicographica. Series maior* 17. Tübingen: Niemeyer, S. 120–125.
- Inoue, Ai (2011). »Updating English-Japanese dictionaries from the phraseological viewpoint«. In: *LEXICOGRAPHY: Theoretical and Practical Perspectives. Proceedings of Asialex 2011*. Kyoto: The Asian Association of Lexicography, S. 608–617.
- Isahara, Hitoshi, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama und Kyoko Kanzaki (2008). »Development of the Japanese WordNet«. In: *The sixth international conference on Language Resources and Evaluation (LREC2008)*.
- Ishida, Priscilla (1996). »日英語の対照研究 : 「目」の慣用句を中心として (Kontrastive Studien Japanisch-Englisch: Idiome mit der Komponente ›Auge‹)«. In: *Tsukuba öyōgengogaku kenkyū*, S. 49–63.
- Ishida, Priscilla (2003). »慣用句の意味を分析する方法 (Methode zur Analyse von Idiomen)«. In: *Nihongo to Nihongaku* 37, S. 13–26.
- Itoh, Makoto (2002). »慣用句の分析の特異性 (Besonderheiten der Phraseologismen-Analyse)«. In: *Tsukuba daigaku »Tōsei gengo bunka no ruikeiron«* 5, S. 757–776.
- Itoh, Makoto (2004). »Phraseologie aus kontrastiver Sicht - Entsprechungen und pragmatische Einschränkungen der deutschen und der japanischen Phraseologismen«. In: *Über die Grenzen hinweg*. Hrsg. von Shūichi Inoue und Koji Ueda. München: Iudicium.
- Itoh, Makoto (2005). *Deutsche und japanische Phraseologismen im Vergleich*. Tübingen: Julius Groos.
- Ješensek, Vida (2009). »Phraseologische Wörterbücher auf dem Weg zu Phraseologiedatenbanken«. In: *Lexicographica. Series maior* 135, S. 65–81.
- Jónsson, Jón Hilmar (2009). »Lemmatisation of Multi-word Lexical Units: Motivation and Benefits«. In: *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Hrsg. von Henning Bergenholtz, Sandro Nielsen und Sven Tarp. Bd. 90. *Linguistic Insights: Studies in Language and Communication*. Bern: Peter Lang, S. 165–194.
- Kanetake, Nobuya (2000). 「広辞苑」は信頼できるか (*Kann man dem Kōjien glauben?*). Tōkyō: Kōdansha.
- Keil, Martina (1998). *Wort für Wort: Repräsentation und Verarbeitung verbaler Phraseologismen (Phraseo-Lex)*. *Sprache und Information* 35. Tübingen: Niemeyer.

- Kemps-Snijders, Marc, Menzo Windhouwer, Peter Wittenburg und Sue Ellen Wright (2008). »ISOCat: Corraling Data Categories in the Wild«. Englisch. In: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Hrsg. von Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis und Daniel Tapias. Marrakech, Morocco: European Language Resources Association (ELRA), S. 28–30.
- Kerde, Ortrud und Silke-Susann Otto (1996). *Einführung in die Hilfsmittel der Japanologie*. Marburger Japan-Reihe: Sonderband. Marburg: Förderverein Marburger Japan-Reihe.
- Keysar, Boaz und Bridget Martin Bly (1995). »Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans?« In: *Journal of Memory and Language* 34.1, S. 89–109.
- Keysar, Boaz, Yeshayahu Shen und Sam Glucksberg William S. Horton (2000). »Conventional language: How metaphorical is it?« In: *Journal of Memory and Language* 43.4, S. 576–593.
- Kilgarriff, Adam und Iztok Kosem (2013). »Corpus tools for lexicographers«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 32–55.
- Kim-Werner, Samhwa (1996). *Phraseologisches Wörterbuch: Deutsch-Koreanisch. Am Beispiel der somatischen Phraseologismen*. Seoul: Yulin-Madang Publishing.
- Kjaer, Anne Lise (1987). »Zur Darbietung von Phraseologismen in einsprachigen Wörterbüchern des Deutschen aus der Sicht ausländischer Textproduzenten«. In: *Beiträge zur allgemeinen und germanistischen Phraseologieforschung. Internationales Symposium in Oulu*. Hrsg. von Jarmo Korhonen. Bd. 13. Oulu: Oulun Yliopisto, S. 165–181.
- Klappenbach, Ruth (1980). »Probleme der Phraseologie«. In: *Studien zur modernen deutschen Lexikographie: Ruth Klappenbach (1911-1977), Auswahl aus den lexikographischen Arbeiten*. Amsterdam: Benjamins.
- Knobloch, A. (1874). »Japanische Sprichwörter und Redensarten.« In: *Mittheilungen der deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens* 1, S. 23–26.
- Kokuritsu Kokugo Kenkyūsho (2009). 辞書を知る (Wörter, die nicht im Wörterbuch stehen). Shin-kotoba shiriizu 22. Tōkyō: Gyōsei.
- Koller, Werner (1987). »Überlegungen zu einem Phraseologie-Wörterbuch für Fremdsprachenunterricht und Übersetzungspraxis«. In: *Aktuelle Probleme der Phraseologie*. Hrsg. von Harald Burger und Robert Zett. Frankfurt am Main [u. a.]: Peter Lang, S. 109–120.

- Konecny, Christine (2010). »Lexikalische Kollokationen und der Beitrag der Internet-Suchmaschine Google zu ihrer Erschließung und Beschreibung«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren.
- Konieczna, Joanna (2005). »Einige Anmerkungen zur Phraseographie in zwei Wörterbüchern für Deutsch als Fremdsprache: KDaF und LDaF«. In: *Lingua ac Communitas* 15, S. 71–84.
- Korhonen, Jarmo (1990). »Zur (Un-) Verständlichkeit der lexikographischen Darstellung von Phraseologismen«. In: *BudaLEX '88 proceedings: papers from the 3rd International EURALEX Congress, Budapest, 4-9 September 1988*. Hrsg. von Tamás Magay und J. Zigány. Bd. 3. Heidelberg: Groos, S. 197–206.
- Korhonen, Jarmo (1992). »Morphosyntaktische Variabilität von Verbidiomen«. In: *Deutsche Phraseologie in Sprachsystem und Sprachverwendung*. Wien: Ed. Praesens, S. 49–88.
- Kövecses, Zoltán (1986). *Metaphors of anger, pride, and love: A lexical approach to the structure of concepts*. John Benjamins.
- Kühn, Peter (1983). »Pragmatische und lexikographische Beschreibung phraseologischer Einheiten: Phraseologismen und Routineformeln«. In: *Studien zur neuhochdeutschen Lexikographie* 4, S. 175–235.
- Kühn, Peter (2003). »Phraseme im Lexikographie-Check: Erfassung und Beschreibung von Phrasemen im einsprachigen Lernerwörterbuch«. In: *Lexicographica* 19, S. 97–118.
- Kunihiro, Tetsuya (1985). »慣用句論 (Phraseologie)«. In: *Nihongogaku* 4.1, S. 4–14.
- Kurashima, Tokihisa (2002). *辞書と日本語・国語辞典を解剖する (Wörterbücher und Japanisch: Analyse einsprachig japanischer Wörterbücher)*. Tōkyō: Kōbunsha shinsho.
- Lakoff, George und Mark Johnson (1980). *Metaphors we live by*. Chicago [u.a.]: University of Chicago Press.
- Lange, Rudolf (1875). »Noch einige Sprichwörter und sprichwörtliche Redensarten der Japaner«. In: *Mittheilungen der deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens* 1,8, S. 50–52.
- Lange, Rudolf (1876). »Noch einige Sprichwörter und sprichwörtliche Redensarten«. In: *Mittheilungen der deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens* 1,9, S. 59–60.
- Lange, Rudolf (1880). »Japanische Sprichwörter«. In: *Mittheilungen der deutschen Gesellschaft für Natur- und Völkerkunde Ostasiens* 2,20, S. 415–421.

- Laurent, Érick (1995). »Definition and Cultural Representation of the Category *Mushi* in Japanese Culture«. In: *Society and Animals* 3(1), S. 61–77.
- Lew, Robert (2011). »Online Dictionaries of English«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum, S. 230–250.
- Lew, Robert und J Doroszewska (2009). »Electronic dictionary entries with animated pictures: Lookup preferences and word retention«. In: *International Journal of Lexicography* 22.3, S. 239–257.
- Lew, Robert und Renata Szczoaniak (2011). »The role of imagery in dictionaries of idioms«. In: *Applied Linguistics* 32.3, S. 323–347.
- Lewin, Bruno (1991). »Japanische Lexikographie«. In: *Wörterbücher / Dictionaries / Dictionnaires*. Hrsg. von Franz-Josef Hausmann, Oskar Reichmann, Herbert E. Wiegand und Ladislav Zgusta. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science [HSK] 5.3. Berlin, New York: Walter de Gruyter, S. 2617–2623.
- L’Homme, Marie-claude, Benoît Robichaud und Patrick Leroyer (2013). »Encoding collocations in DiCoInfo: From formal to user-friendly representations«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 211–236.
- LMFWorking Group (2008). »Language resource management – lexical markup framework (LMF)«. In: *Technical Report ISO/TC 37/SC 4 N453 (N330 Rev.16)*, ISO.
- Martín, Luis (2001). *Phraseologie im zweisprachigen Wörterbuch Deutsch-Spanisch*. Bd. 28. Hispano-Americana. Frankfurt am Main [u.a.]: Peter Lang.
- Matsuki, Keiko (1995). »Metaphors of anger in Japanese«. In: *Language and the Cognitive Construction of the World*. Hrsg. von John R. Taylor; Robert E. MacLaury. Berlin: de Gruyter, S. 137–151.
- Mellado Blanco, Carmen (2009). »Einführung. Idiomatiche Wörterbücher und Metafraseografie: zwei Realitäten, eine Herausforderung«. In: *Theorie und Praxis der idiomatiche Wörterbücher*. Hrsg. von Carmen Mellado Blanco. Lexicographica. Series maior 135. Tübingen: Niemeyer, S. 1–22.
- Miura, Ichiro (Nov. 1996). »Comprehension of idiomatic expressions in Japanese–auditory and visual presentations.« In: *Journal of Psycholinguistic Research* 25.6, S. 659–76.

- Miyaji, Yutaka (1982). 慣用句の意味と用法 (*Bedeutung und Gebrauch von Idiomen*). Meiji shoin.
- Miyaji, Yutaka (1985). »慣用句の周辺 — 連語・ことわざ・複合語 (Die Peripherie der Idiome: Kollokationen, Sprichwörter und Komposita)«. In: *Nihongogaku* 4 (1), S. 62–75.
- Miyata, Rei, Ryōko Adachi, Ulrich Apel, Iris Vogel, Wolfgang Fanderl, Ryō Murayama, Kōichi Takeuchi und Kyō Kageura (2014). »The Use of Corpus Evidence and Human Introspection to Create Idiom Variations.« In: *Proceedings of the the Second Asia Pacific Corpus Linguistics Conference (APCLC 2014)*. Hong Kong, S. 201–202.
- Momiyama, Yosuke (1997). »慣用句の体系的分類 — 隠喩・換喩・提喩に基づく慣用的意味の成立の中心に (Systematische Klassifikation der Idiome: Metaphern, Metonymie und Allegorien in der Genese der idiomatischen Bedeutung)«. In: *Nagoya Daigaku Kokugo Kokubungaku* 80, S. 29–43.
- Moon, Rosamund (2004). »On Specifying Metaphor: An Idea and its Implementation«. In: *International Journal of Lexicography* 17.2, S. 195–222.
- Müller-Spitzer, Carolin (2003). »Ordnennde Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen«. In: *Lexicographica* 19, S. 140–168.
- Müller-Spitzer, Carolin, Alexander Koplenig und Antje Töpel (2013). »Online dictionary use: Key findings from an empirical research project«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 426–457.
- Müller-Yokota, Wolfgang (1989). »Schrift und Schriftgeschichte«. In: *Allgemeines, Sprache und Schrift: Sprache und Schrift Japans-in Zusammenarbeit mit Kay Genenz, Wolfgang Müller-Yokota*, S. 185.
- Naciscione, Anita (2010). »Visual representation of phraseological image«. In: *Yearbook of Phraseology (2010)*. de Gruyter Mouton, S. 19–44.
- Nakajima, Heizō und Shigeo Tonoike (1994). 言語学への招待 (*Einführung in die Linguistik*). Tōkyō: Taishūkan shoten.
- Nakao, Keisuke (1998). »The State of Bilingual Lexicography in Japan: Learner's English-Japanese/ Japanese-English dictionaries«. In: *International Journal of Lexicography* 11.1, S. 35–50.
- Nesi, Hilary (2013). »Alternative e-dictionaries: Uncovering dark practices«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 363–378.

- Nikkei Digital Media, Inc (2012). [日経テレコン] (*Nikkei Telecom*). URL: <http://t21.nikkei.co.jp/>, aufgerufen am 26. 10. 2012. Archiviert durch WebCite <http://www.webcitation.org/6BhH9vHGk>.
- Odijk, Jan (2004). »A proposed standard for the lexical representation of idioms«. In: *EURALEX 2004 Proceedings*. Université de Bretagne Sud, S. 153–164.
- Oppentocht, Lineke Oppentocht und Rik Schutz (2003). »Developments in electronic dictionary design«. In: *A Practical Guide to Lexicography*, S. 215–227.
- Otani, Akihiro Yonekawa; Itsuko, Hrsg. (2005). *日本語慣用句辞典 (Japanisches Idiom-Wörterbuch)*. Tōkyō: Foresuto.
- Ōtsubo, Yoshiko (1985). »名詞慣用句 (*Nominale Phraseme*)«. In: *Nihongogaku* 4 (1), S. 54–61.
- Ōtsuka, Yōko 大塚容子 (1994). »身振りと身体ことば—日本語教育の立場から (*Gestures and Body-part Phrases*)«. In: *Bulletin of Gifu College of Education* 27, S. 51–65.
- Oxford Times (30. Aug. 2010). *Oxford Dictionary* »not going online-only«. URL: http://www.oxfordtimes.co.uk/news/8360446.Oxford_Dictionary___not_going_online_only___/, aufgerufen am 08. 03. 2013.
- Parizoska, Jelena (2010). »The Canonical Form in Murky Waters: Idiom Variation and the Croation National Corpus«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren, S. 95–106.
- Petermann, Jürgen (1983). »Zur Erstellung ein- und zweisprachiger phraseologischer Wörterbücher: Prinzipien der formalen Gestaltung und der Einordnung von Phrasemen«. In: *Phraseologie und ihre Aufgaben* 1, S. 172–193.
- Petrova, Oksana (2010). »Computer-Mediated Discourse vs. Traditional Text Corpora as a Data Source for Idiom Variation Research in Finnish«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren, S. 133–150.
- Piirainen, Elisabeth (2012). *Widespread Idioms in Europe and Beyond. Toward a Lexicon of Common Figurative Units*. International Folkloristics 5. New York: Peter Lang.
- Pilz, Klaus Dieter (1978). *Phraseologie / Versuch einer interdisziplinären Abgrenzung*. Göppingen: Kümmerle.

- Pilz, Klaus Dieter (1987). »Allgemeine und phraseologische Wörterbücher. Brauchen wir überhaupt phraseologische Wörterbücher?« In: *Beiträge zur allgemeinen und germanistischen Phraseologieforschung. Internationales Symposium in Oulu, 13. - 15. Juni 1986*. Hrsg. von Jarmo Korhonen. Oulu: Oulun Yliopisto, S. 129–153.
- Pilz, Klaus Dieter (2002). »Vorschläge für ein Phraseolexikon der deutschen Sprache oder: Vorschläge für ein Lexikon der deutschen Phraseme/Phraseologismen«. In: *Wer A sägt, muss auch B sägen: Beiträge zur Phraseologie und Sprichwortforschung aus dem Westfälischen Arbeitskreis*. Hrsg. von Dietrich Hartmann und Jan Wirrer. Bd. 9. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren, S. 289–310.
- Ptashnyk, Stefaniya, Erla Hallsteinsdóttir und Noah Bubenhofer, Hrsg. (2010). *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren.
- Quasthoff, Uwe, Fabian Schmidt und Erla Hallsteinsdóttir (2010). »Häufigkeit und Struktur von Phraseologismen«. In: *Korpora, Web und Datenbanken: computergestützte Methoden in der modernen Phraseologie und Lexikographie*. Hrsg. von Stefaniya Ptashnyk, Erla Hallsteinsdóttir und Noah Bubenhofer. Bd. 25. Phraseologie und Parömiologie. Baltmannsweiler: Schneider Hohengehren.
- Reinirkens, Hubert (1955). *Sprichwörter und Redensarten Deutsch-Japanisch. Kotowaza to Seigo* [諺と成語]. Tokyo: Deutsche Gesellschaft für Natur- und Völkerkunde Ostasiens (Supplement 21 der Mittheilungen der Gesellschaft).
- Rothkegel, Annely (2007). »Computerlinguistische Aspekte der Phraseme I«. In: *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung*. Hrsg. von Harald Burger. Bd. 28. Handbücher zur Sprach- und Kommunikationswissenschaft. Berlin: de Gruyter, S. 1027–1035.
- Rundell, Micheal (5. Nov. 2012). *Stop the presses – the end of the printed dictionary*. URL: <http://www.macmillandictionaryblog.com/bye-print-dictionary>, aufgerufen am 08. 03. 2013.
- Rundell, Micheal (2013). »The road to automated lexicography: An editor's viewpoint«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 15–30.
- Satake, Hideo (2000). »使いやすさを巡る戦い・国語辞書編 (Kampf um die Nutzerfreundlichkeit - einsprachig japanische Wörterbücher)«. In: *Gengo* 29.5, S. 42–47.

- Schemann, Hans (1991). »Die Phraseologie im zweisprachigen Wörterbuch«. In: *Dictionaries*. Hrsg. von Franz-Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta. Berlin [u.a.]: Walter de Gruyter, S. 2789–2794.
- Schemann, Hans (2009). »Zur Anlage idiomatischer Wörterbücher — einige Maximen und Reflexionen«. In: *Theorie und Praxis der idiomatischen Wörterbücher*, S. 101.
- SCHEMANN-DEUTSCHEIDIOMATIK-PW: Schemann, Hans (1993). *Deutsche Idiomatik: die deutschen Redewendungen im Kontext*. PONS.
- Schmidt, Ingrid und Carolin Müller (2001). »Entwicklung eines lexikographischen Modells: Ein neuer Ansatz«. In: *Lexicographica* 107, S. 29–52.
- Scholze-Stubenrecht, Werner (1988). »Phraseologismen im Wörterbuch«. In: *Das Wörterbuch*. Hrsg. von Gisela Harras, S. 284–302.
- Schryver, Gilles-Maurice de (2003). »Lexicographers' Dreams in the Electronic-Dictionary Age«. In: *International Journal of Lexicography* 16.2, S. 143–199.
- Sharpe, Peter (1995). »Electronic Dictionaries with Particular Reference to the Design of an Electronic Bilingual Dictionary for English-speaking Learners of Japanese«. In: *International Journal of Lexicography* 8.1, S. 39–54.
- Stalph, Jürgen und Harald Suppanschitsch (1999). *Wörterbücher und Glossare: eine teilanno-tierte Bibliographie japanisch-deutscher und deutsch-japanischer Nachschlagewerke*. Bibliographische Arbeiten aus dem Deutschen Institut für Japanstudien der Philipp-Franz-von-Siebold-Stiftung 5. München: Iudicium.
- Stantcheva, Diana (2002). *Phraseologismen in deutschen Wörterbüchern*. Hamburg: Dr. Kovač.
- Sterkenburg, Piet van, Hrsg. (2003). *Practical guide to lexicography*. Bd. 6. Terminology and lexicography research and practice. Amsterdam: John Benjamins.
- Storrer, Angelika (2001). »Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertext-konzepts in der Lexikographie«. In: *Chancen und Perspektiven computergestützter Lexikographie. Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Hrsg. von Ingrid Lemberg, Bernhard Schröder und Angelika Storrer. Tübingen: Niemeyer, S. 53–69.
- Takashima, Toshio (2003). *広辞苑の神話 (Mythos Kōjien)*. Tōkyō: Bungei shunju.
- Takeuchi, Kōichi, Takahiro Shiraishi, Ulrich Apel, Rei Miyata, Ryōko Adachi, Wolfgang Fanderl, Ryō Murayama, Iris Vogel und Kyō Kageura (2014). »簡単なイデオム異形規則の

- 作成: プラットフォームと日本語の異形規則 (Die einfache Erstellung von Variationsregeln für Idiome: Die Plattform und Variationsregeln für das Japanische)«. In: *Gengoshori Gakkai dai-nijūnen jidaikai happyō ronshū*. Hokkaido, S. 488–491.
- Tarp, Sven (2009). »Beyond Lexicography: New Visions and Challenges in the Information Age«. In: *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Hrsg. von Henning Bergenholtz, Sandro Nielsen und Sven Tarp. Bd. 90. Linguistic Insights: Studies in Language and Communication. Bern: Lang.
- Tarp, Sven (2011). »Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum.
- Tarp, Sven (2013). »Theoretical challenges in the transition from lexicographical p-works to e-tools«. In: *Electronic lexicography*. Hrsg. von Sylviane Granger und Magali Paquot. Oxford: Oxford University Press, S. 107–118.
- TEI Consortium, Hrsg. (2008). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. URL: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>, aufgerufen am 27. 04. 2010. Archiviert am 11.01.2012 unter <http://www.webcitation.org/64bpAdHij>.
- The National Institute for Japanese Language, Hrsg. (2004). 分類語彙表 一 増補改訂版 (*Word List by Semantic Principle, Revised and Enlarged Edition*). Bd. 14. Kokuritsu kogoku kenkyūsho shiryoshū. Dai-Nippon Tosho.
- Tomita, Micho et al. (1997). 青空文庫 (*Aozora Bunko*). URL: <http://www.aozora.gr.jp>, aufgerufen am 06. 10. 2012. Archiviert unter <http://www.webcitation.org/6BhFM8JIm>.
- Tono, Yukio (2001). *Research on dictionary use in the context of foreign language learning: focus on reading comprehension*. Lexicographica. Series maior 106. Tübingen: Niemeyer.
- Tono, Yukio (2009). »Pocket Electronic Dictionaries in Japan: User Perspectives«. In: *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Hrsg. von Henning Bergenholtz, Sandro Nielsen und Sven Tarp. Bd. 90. Linguistic Insights: Studies in Language and Communication. Bern: Peter Lang, S. 33–67.
- Trippel, Thorsten (2006). *The Lexicon Graph Model A generic model for multimodal lexicon development*. Saarbrücken: AQ-Verlag.
- Ueda, Yasunari (2002). »日独イディオム比較・対象 一 ”Mund”と「口を構成要素とするイディオム表現 (*Kontrastive Idiomatik Japanisch-Deutsch: idiomatische Ausdrücke mit den*

- Komponenten ›Mund‹ und ›kuchi‹*«. In: *The Hiroshima University studies, Graduate School of Letters* 62, S. 79–97.
- Ueda, Yasunari (2003). »日独イディオム比較・対象 — 「耳」と”Ohr”を構成要素とするイディオム表現 (*Kontrastive Idiomatik Japanisch-Deutsch: idiomatische Ausdrücke mit den Komponenten ›mimi‹ und ›Ohr‹*)«. In: *Hiroshima Daigaku Daigakuin Bungaku Kenkyūka ronshū* 63, S. 51–69.
- Ueda, Yasunari (2004). »日独イディオム比較・対象 — 「鼻」と”Nase”を構成要素とするイディオム表現 (*Kontrastive Idiomatik Japanisch-Deutsch: idiomatische Ausdrücke mit den Komponenten ›hana‹ und ›Nase‹*)«. In: *Doitsugo Doitsubungaku Kenkyū Nenpō = Nenpo. Jahresbericht des Germanistischen Seminars der Hokkaido Universität* 31, S. 298–313.
- Unicode (2010). *The Unicode Standard: A Technical Introduction*. URL: <http://www.unicode.org/standard/principles.html>, aufgerufen am 06. 01. 2010.
- Verlinde, Serge (2011). »Modelling Interactive Reading, Translation and Writing Assistants«. In: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. Hrsg. von Pedro A. Fuertes-Olivera und Henning Bergenholtz. London, New York: Continuum, S. 275–286.
- Vogel, Iris (2011). »Towards the encoding of multilayered information in digital lexicography«. In: *LEXICOGRAPHY: Theoretical and Practical Perspectives. Proceedings of Asialex 2011*. Kyoto: The Asian Association of Lexicography, S. 549–557.
- Vogel, Iris (2013). »The role of mushi in Japanese idioms: encoding conceptual information in an electronic dictionary«. In: *Research on Phraseology Across Continents*. Hrsg. von Joanna Szerszunowicz, Bogusław Nowowiejski, Katsumasa Yagi und Takaaki Kanzaki. Intercontinental Dialogue on Phraseology 2. University of Bialystok Publishing House, S. 163–175.
- Vogel, Iris und Shuichi Kuchikata (im Druck). »Phraseologismen der Körpersprache im elektronischen Wörterbuch – aktuelle Praxis und Perspektiven der japanisch-deutschen Lexikographie«. In: *Linguistik, Informations- und Ressourcenwissenschaften: Referate des 14. Deutschsprachigen Japanologentages*. Hrsg. von Martina Ebi und Robert Horres. Berlin: Lit-Verlag, S. 17–40.
- Wible, David (2008). »Multiword expressions and the digital turn«. In: *Phraseology in foreign language learning and teaching*, S. 163–181.
- Wiegand, Herbert Ernst (1998). *Wörterbuchforschung / Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Bd. 1. Berlin: Walter de Gruyter.

- Wiegand, Herbert Ernst (2010). »Makro- und mikrostrukturelle Präsentationsprobleme bei Phrasemen«. In: *Feste Wortverbindungen und Lexikographie: Kolloquium zur Lexikographie und Wörterbuchforschung*. Hrsg. von Peter Ďurčo, Monika Banášová und Astrid Astrid Hanzlíčková. 2. Aufl. Berlin [u.a.]: Walter de Gruyter, S. 161–180.
- Wright, Sue-Ellen (2009a). *Data Category: idiom*. URL: <http://www.isocat.org/rest/dc/351>. Archiviert am 28.10.2012 unter <http://www.webcitation.org/6Bk98cDYP>.
- Wright, Sue-Ellen (2009b). *Data Category: phraseological unit*. URL: <http://www.isocat.org/datcat/DC-339>. Archiviert am 28.10.2012 unter <http://www.webcitation.org/6Bk9JXizs>.
- Yagi, Katsumasa (2011). »Updating English-Japanese dictionaries from the phraseological viewpoint«. In: *LEXICOGRAPHY: Theoretical and Practical Perspectives. Proceedings of Asialex 2011*. Kyoto: The Asian Association of Lexicography, S. 597–607.
- Yihua, Zhang (2009). »A Bilingual Dictionary Generation System«. In: *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*. Hrsg. von Henning Bergenholtz, Sandro Nielsen und Sven Tarp. Bd. 90. Linguistic Insights: Studies in Language and Communication. Bern: Peter Lang.
- Zhu, Kaifu (1998). *Lexikographische Untersuchung somatischer Phraseologismen im Deutschen und Chinesischen: eine kontrastive Analyse unter interkulturell-kommunikativen Aspekten*. Europäische Hochschulschriften: Reihe 21, Linguistik ; Bd. 206. Frankfurt am Main: Lang.

Wörterbücher

- APEL-WADOKUJT-IW: Ulrich Apel (2009). *Ulrich Apel's WaDokuJiten*. URL: <http://www.wadoku.de>, aufgerufen am 05. 01. 2012.
- Duden: *Redensarten ; Herkunft und Bedeutung* (1999). Bd. 29. Duden-Taschenbücher. Mannheim [u.a.]: Dudenverlag, S. 256.
- DUDEN-REDEWENDUNGEN-PW (1998). *Duden 11, Redewendungen und sprichwörtliche Redensarten: Wörterbuch der deutschen Idiomatik*. Nach den Regeln der neuen deutschen Rechtschreibung überarb. Nachdr. der 1. Aufl. Der Duden in zwölf Bänden. Mannheim: Dudenverlag.
- DUDEN-ZITATE REDEWENDUNGEN-PW: Alsleben, Brigitte (2007). *Duden, das große Buch der Zitate und Redewendungen*. 2. überarb. und aktualisierte Aufl. Mannheim [u.a.]: Dudenverlag.

- ETTINGERPHRASEOLOGIEDE-IW: Stefan Ettinger (2009). *Phraseologie und Phraseodidaktik*. URL: <http://www.ettinger-phraseologie.de/>, aufgerufen am 05. 01. 2012.
- FRIEDERICHIDIOMATIK-PW: Friederich, Wolf (1976). *Moderne deutsche Idiomatik: Alphabet. Wörterbuch mit Definitionen u. Beispielen*. Ismaning: Hueber.
- GOERNERREDENSARTEN-PW: Görner, Herbert (1982). *Redensarten: Kleine Idiomatik der deutschen Sprache*. Leipzig: VEB Bibliographisches Institut Leipzig.
- GROSSESJAPANISCHDEUTSCHESWÖRTERBUCH-IW: Stalph, Jürgen et. al., Hrsg. (2009). *Großes japanisch-deutsches Wörterbuch*. URL: <http://www.japan.de/>, aufgerufen am 17. 03. 2013.
- GROSSESJAPANISCHDEUTSCHESWÖRTERBUCH-PW: Stalph, Jürgen et. al., Hrsg. (2009). *Großes japanisch-deutsches Wörterbuch*. Bd. 1 / A-I. München: Iudicium.
- IWANAMI-KÖJIEN₅: Shinmura, Izuru, Hrsg. (1999). *広辞苑 (Kōjien)*. 5. Auflage. Tōkyō: Iwanami shoten.
- IWANAMI-KÖJIEN₆: Shinmura, Izuru, Hrsg. (2008). *広辞苑 (Kōjien)*. 6. Auflage. Tōkyō: Iwanami shoten.
- IWANAMI-KÖJIEN₆-CW: Shinmura, Izuru, Hrsg. (2008). *広辞苑 (Kōjien)*. DVD-ROM. 6. Auflage. Tōkyō: Iwanami shoten.
- Kamei, SANSEIDŌ-GENGOGAKU-DAIJITEN-PW: Takashi, Rokurō Kōno und Eichi Chino, Hrsg. (1989). *言語学大辞典 (Großes Wörterbuch der Linguistik)*. Bd. 6. Tōkyō: Sanseidō.
- KARADAKOTOBAJITEN: Tōgō, Yoshio (東郷吉男), Hrsg. (2003). *からだことば辞典 (Körperwörter-Wörterbuch)*. Tōkyō: Tōkyōdō Shuppan.
- KEMPCKE-WÖRTERBUCH-DAF-PW: Günter Kempcke and Barbara Seelig (2000). *Wörterbuch Deutsch als Fremdsprache*. Berlin, New York: de Gruyter.
- KENKYŪSHA-WAEIDAIJITEN-PW: Koine, Yoshio, Hrsg. (1991). *研究社新英和大辞典 (Kenkyūsha's New English-Japanese Dictionary)*. 5. Aufl. Tōkyō: Kenkyūsha.
- Kindaiichi, Hideho, Hrsg. (2007). *知っておきたい日本語コロケーション辞典 (Japanese Collocation Dictionary)*. Tōkyō: Gakken.
- Kitahara, Yasuo, Hrsg. (2007). *ことわざ成句使い方辞典 (Wörterbuch zur Verwendungsweise von Sprichwörtern und Redensarten)*. Tōkyō: Taishūkan shoten.
- KŌDANSHA-BASICJAPANESEIDIOMS: Garrison, Jeff et. al, Hrsg. (2003). *Kodansha's Dictionary of Basic Japanese Idioms*. Tōkyō, New York, London: Kodansha.
- Lange, William de und Hiromi Miyagi-Lusthaus, Hrsg. (2005). *A Dictionary of Japanese Idioms*. Warren: Floating World.

- LOGOVISTA-KANYŌKUJITEN-CW (2012). 用例でわかる慣用句辞典 (*Idiomwörterbuch mit Beispielen*). Tōkyō: Logovista.
- ŌBUNSHA-SEIGORIN-PW, Hrsg. (1992). 成語林—故事ことわざ慣用句 (*Phrasem-Hain – Weisheiten, Sprichwörter und Idiome*). Tōkyō: Ōbunsha.
- RÖHRICH-REDENSARTEN-CW: Röhrich, Lutz (2004). *RÖHRICH-REDENSARTEN-CW: Lexikon der sprichwörtlichen Redensarten*. Parallel als Buchausg. erschienen u.d.T.: Röhrich, Lutz: Lexikon der sprichwörtlichen Redensarten.
- SCHEMANN-DEUTSCHEIDIOMATIK-PW: Schemann, Hans (2011). *Deutsche Idiomatik: die deutschen Redewendungen im Kontext*. Berlin: de Gruyter.
- SCHEMANN-IDIOMS-PW: Schemann, Hans (1997). *PONS Idiomatik Deutsch - Englisch = German - English dictionary of idioms*. Stuttgart: PONS.
- SHŌGAKKAN-NIHONKOKUGODAIJITEN-IW (2012). ジャパンナレッジ プラス (JapanKnowledge+). NetAdvance Inc. URL: <http://www.jkn21.com/top/corpdisplay>, aufgerufen am 18. 11. 2012.
- SHŌGAKUKAN-NIHONKOKUGO-DAIJITEN₂ (2001 (1972)). 日本国語大辞典 (*Nihon kokugo daijiten*). Tōkyō: Shōgakukan.
- SŌTAKUSHA-REIKAIKANYŌKUJITEN-PW: Inoue, Muneo, Hrsg. (2003). 例解慣用句辞典—言いたい内容から逆引きできる (*Idiomwörterbuch mit Erklärungen anhand von Beispielen: Mit der Möglichkeit nach dem Inhalt nachzuschlagen, den man ausdrücken möchte.*) Tōkyō: Sōtakusha.
- SŌTAKUSHA-SHIBATA'S JAPANESE ENGLISH ROMAJI-PW: Shibata, Masaaki (1998). *English-Japanese Dictionary in Rōma-Ji: for one million foreigners*. Tōkyō: Sōtakusha, 445 S.
- TANNO-KANYŌKUJITEN-PW: Tannō, Akira, Hrsg. (1998). 意味から引ける慣用句辞典 (*Idiomwörterbuch zum Nachschlagen anhand der Bedeutung*). Tōkyō: Nihon Jitsugyō Shuppansha.
- WADOKUDE-IW: Thomas Latka (2013). *Ulrich Apel's WaDokuJiten*. URL: <http://www.wadoku.de>, aufgerufen am 05. 01. 2012.
- WADOKUEU-IW: Ulrich Apel (2009). *Ulrich Apel's WaDokuJiten*. URL: <http://www.wadoku.eu>, aufgerufen am 05. 01. 2012.
- WADOKUIDIOMS-IW: Vogel, Iris (2013). 慣用句辞典 *Japanisch-Deutsches Idiom-Wörterbuch*. URL: <http://www.wadoku-idioms.de>, aufgerufen am 20. 03. 2013.
- YONEKAWA-NIHONGO KANYŌKUJITEN-PW: Yonekawa, Akihiro und Itsuko Otani, Hrsg. (2005). 日本語慣用句辞典 (*Japanisches Idiom-Wörterbuch*). Tōkyō: Foresuto.

Ressourcen

- AOZORA BUNKO: (17. März 2013). 青空文庫 (*Aozora Bibliothek*). URL: <http://www.aozora.gr.jp>.
- BCCJW: National Institute for Japanese Language and Linguistics (17. März 2013). *Balanced Corpus of Contemporary Written Japanese*. URL: <http://www.ninjal.ac.jp/english/products/bccwj/>.
- BIBIKO: Bibiko, Hans-Jörg (7. Okt. 2011). 和独字典 *Japanisch-Deutsches Kanji-Lexikon*. URL: <http://lingweb.eva.mpg.de/kanji/>.
- CROSSASIA: Ostasienabteilung der Staatsbibliothek zu Berlin (17. März 2013). *Virtuelle Fachbibliothek CrossAsia*. URL: <http://crossasia.org/>.
- DCR: Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands (17. März 2013). *Data Category Registry*. URL: <http://www.isocat.org>.
- DWDS: Berlin-Brandenburgischen Akademie der Wissenschaften (2013). *Digitales Wörterbuch der Deutschen Sprache*. URL: www.dwds.de.
- FREEDICT: Eyermann, Horst und Michael Bunk (2013). *FreeDict - freie zweisprachige Wörterbücher*. URL: <http://www.freedict.org/de/>.
- KOLLOKATIONEN: Fellbaum, Christiane und Alexander Geyken und Undine Kramer (2. Mai 2007). *Kollokationen im Wörterbuch*. URL: <http://kollokationen.bbaw.de/>.
- MECAB: Kudō, Taku (17. März 2013). 和布蕪 *MeCab: Yet Another Part-of-Speech and Morphological Analyzer*. URL: <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>.
- MINNA NO HON'YAKU: Minna no Hon'yaku Project (17. März 2013). みんなの翻訳 (*Übersetzung für Alle*). URL: <http://trans-aid.jp/>.
- MOJI: Loisant, Erwant und Gerald Vogt (2007). *Moji*. URL: <http://moji.mozdev.org/>.
- POP JISYO: Foster, Michael (17. März 2013). *Pop Jisyo*. URL: http://www.popjisyo.com/WebHint/Portal_e.aspx.
- READING TUTOR: Kawamura, Yoshiko und Tatsuya Kitamura und Rei Hobara (2007). *Reading Tutor*. URL: http://language.tiu.ac.jp/index_e.html.
- TATOEBA: Ho, Trang und Allan Simon (2006). *Tatoeba: Collecting example sentences*. URL: tatoeba.org.