

Kathrin Steyer / Annelen Brunner

Das UWV-Analysemodell

Eine korpusgesteuerte Methode zur linguistischen
Systematisierung von Wortverbindungen

OPAL – Online publizierte Arbeiten zur Linguistik
Herausgegeben vom Institut für Deutsche Sprache



Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim
opal@ids-mannheim.de

Technische Redaktion: Norbert Volz

© 2009 IDS Mannheim – Alle Rechte vorbehalten

Das Werk einschließlich seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechts ist ohne Zustimmung der Copyright-Inhaber unzulässig und strafbar. Das zulässige Zitieren kleinerer Teile in einem eigenen selbstständigen Werk (§ 51 UrhG) erfordert stets die Angabe der Quelle (§ 63 UrhG) in einer geeigneten Form (§ 13 UrhG). Eine Verletzung des Urheberrechts kann Rechtsfolgen nach sich ziehen (§ 97 UrhG). Dies gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Die zugänglichen Daten dürfen von den Nutzern also nur zu rein wissenschaftlichen Zwecken genutzt werden. Eine darüber hinausgehende Nutzung, gleich welcher Art, oder die Verarbeitung und Bearbeitung dieser Daten mit dem Zweck, sie anschließend selbst oder durch Dritte kommerziell zu nutzen, bedarf einer besonderen Genehmigung des IDS (Lizenz). Es ist nicht gestattet, Kopien der Textdateien auf externen Webservern zur Verfügung zu stellen oder Dritten auf sonstigem Wege zugänglich zu machen. Bei der Veröffentlichung von Forschungsergebnissen, in denen OPAL-Publikationen zitiert werden, bitten die Autoren und Herausgeber um eine entsprechende kollegiale Information an opal@ids-mannheim.de.

Kathrin Steyer / Annelen Brunner

Das UWV-Analysemodell

Eine korpusgesteuerte Methode zur linguistischen Systematisierung von Wortverbindungen

Inhalt:

1. Vorbemerkung
2. Methodische Prämisse und Beschreibungsgegenstand
3. Ausgangsbasis: Kookkurrenzprofile
4. Das UWV-Analysemodell
 - 4.1 Die Ebene der konkreten Oberflächenmuster
 - 4.1.1 Methodische Begründung
 - 4.1.2 WV-Suchmuster – Aufbau, Syntax und Suchstrategien
 - 4.2 Die Ebene der lexikalischen Realisierungen
 - 4.2.1 Methodische Begründung
 - 4.2.2 Das Konzept der lexikalischen Realisierungen (LR)
 - 4.2.2.1 Spezifische vs. unspezifische Lücken
 - 4.2.2.2 LR-Gruppe und Struktur des LR-Baums
 - 4.2.2.3 Kern-LR
 - 4.2.2.4 Kern-LR-Variante
 - 4.2.2.5 Erweiterungs-LR
 - 4.2.2.6 Kontext-LR
 - 4.2.2.7 Zusammenfassung anhand des Syntagmas *Musik in X Ohren*
 - 4.3 Die Ebene der Wortverbindungen und Wortverbindungsmuster
 - 4.4 Narrative Beschreibungselemente
5. Resümee und Ausblick
6. Literaturverzeichnis
 - 6.1 Sekundärliteratur und Nachschlagewerke
 - 6.2 Internetseiten

1. Vorbemerkung

Die im Folgenden dargestellte korpusgesteuerte Methode ‘UWV-Analysemodell’ wurde auf der Basis der Forschungen zu usuellen Wortverbindungen (UWV) (vgl. Steyer 2000, 2003, 2004, Steyer/Lauer 2007, Brunner/Steyer 2007, Steyer 2008, Steyer demn.) und zahlreicher, exhaustiver Analysen in den letzten Jahren entwickelt. Ziel war ein empirisches Vorgehensmodell, das es ermöglicht, die Differenziertheit und Vernetztheit von Wortverbindungen auf verschiedenen Abstraktionsebenen ausgehend von Kookkurrenzdaten angemessen darzustellen. Daher ging es in dieser Arbeitsphase nicht darum, usuelle Wortverbindungen des Deutschen möglichst umfassend und in großer Menge zu inventarisieren, sondern die ‘innere Natur’ von Wortverbindungen zwischen Varianz und Invarianz mit unterschiedlichen Graden an lexikalischer Spezifiziertheit sowie ihre wechselseitigen Verbindungen im Detail zu erfassen und zu beschreiben.

Die hier vorgestellte Methode soll helfen, rekurrente Strukturen auf der Basis von großen Sprachausschnitten *bottom up* zu identifizieren, qualitativ zu interpretieren und auf der Basis der Befunde Generalisierungen über einzelne Vorkommen zu treffen. Es geht darum, erstens Muster ähnlicher Art zu erkennen und zu bündeln, zweitens diesen Bündeln Merkmale und Eigenschaften zuzuweisen und drittens die Ergebnisse in einer entsprechend visualisierten Form zu präsentieren. Die Entwicklung der Pilotversion erfolgte auf der Basis der systematischen Auswertung von Wortformenprofilen von Bezugswörtern aus den Bereichen ‘Kausativ’ und ‘Körperteilbezeichnungen’, die mit der [statistischen Kookkurrenzanalyse](#) (vgl. [Belica 1995](#)) (im Folgenden KA abgekürzt) berechnet wurden. Eine erste Ergebnisform der Anwendung des UWV-Analysemodells stellen die Wortverbindungsfelder ‘Grund’ und ‘Körper I’ dar, die auf der Projektwebseite ‘[Wortverbindungen online](#)’ (vgl. Steyer/Brunner 2008) verfügbar sind. Diese Version dient dazu, eine Anwendungsmöglichkeit des UWV-Analysemodells exemplarisch zu veranschaulichen. Sie enthält systematisierte und linguistisch kommentierte Wortverbindungen und Wortverbindungsmuster mit einer Vielzahl an hierarchisch angeordneten lexikalischen Realisierungen und den entsprechenden Kontextzeilen aus den Korpora (*Key Word in Context*, abgekürzt KWIC). Die Tragfähigkeit dieses Vorgehens wurde darüber hinaus auch an anderen Wortformenprofilen getestet, u.a. zu adjektivischen Bezugswortformen (z.B. *faul*) oder Funktionswörtern (z.B. *warum*).

Die Idee, die Systematisierung von Wortverbindungen in einem ersten Analyseschritt auf der Basis von KWIC-Konkordanzen vorzunehmen und diese auch in systematisierter Form als Kontextstellen zu hinterlegen, ist für unseren Arbeitszusammenhang von Cyril Belica eingebracht worden. Gemeinsam mit Rainer Perkuhn haben wir anhand des von ihm entwickelten Prototyps eines abstrakten Vorgehensmodells zur Erschließung der Kookkurrenzanalyse (vgl. dazu Perkuhn 2007) erste Pilotstudien zu Möglichkeiten der Systematisierung und Annotierung von Kookkurrenzclustern durchgeführt (vgl. Steyer/Lauer 2007). Die in diesem Beitrag beschriebene empirische Methode zur linguistischen Systematisierung von usuellen Wortverbindungen wurde dann von Kathrin Steyer und Annelen Brunner im Projekt ‘Usuelle Wortverbindungen’ in den vergangenen zwei Jahren entwickelt.¹ Annelen Brunner hat begleitend dazu die entsprechende Arbeitsumgebung programmiert. Das UWV-Tool unterstützt die Strukturierung und Systematisierung der Kookkurrenzdaten, die linguistische Kommentierung auf den verschiedenen Beschreibungsstufen und die Verwaltung von KWIC-Belegen und Abstraktionsknoten.

Dieser Beitrag wird sich schwerpunktmäßig mit der Darstellung des methodischen Vorgehens und der diesem zu Grunde liegenden Kriterien befassen. Genaue Darlegungen des methodischen Rahmens stellen u.E. nach wie vor ein Desiderat in der Linguistik dar. Gerade für eine genuin empirisch geprägte Richtung wie die Korpuslinguistik ist eine Methodentransparenz jedoch von besonderer Wichtigkeit. Mit der stürmischen Entwicklung dieser Disziplin in den letzten Jahren sind maßgebliche Ausdifferenzierungsprozesse vonstatten gegangen, die eine genaue Verortung dessen, was man mit welcher Methode erklären kann und will und was auch nicht, nahezu unumgänglich machen. Wir versuchen bei unserem Ansatz, diesem Anspruch methodischer Transparenz insofern gerecht zu werden, als die einzelnen Analyseschritte auch in der Präsentation der aufbereiteten Daten für den Rezipienten stets rekonstruier- und damit nachprüfbar bleiben.

¹ Die Methode der Systematisierung von Konkordanzzeilen nach linguistischen Kriterien wird auch von anderen korpuslinguistischen Forschungsvorhaben eingesetzt, vgl. z.B. Duffner (2006). Verwiesen sei vor allem auf das Forschungsprojekt ‘Corpus Pattern Analysis’ (Masaryk Universität Brno/Brandeis, Universität Waltham/Massachusetts), das ein syntagmatisches Wörterbuch englischer Verben erarbeitet. Auch hier werden Konkordanzzeilen semantisch gruppiert und linguistisch interpretiert (vgl. [CPA](#)).

Der Erkenntnisfortschritt unserer Methodik beruht auf der Identifikation bisher verdeckter Strukturen und Phänomene. Die automatisch ermittelten Daten liefern dabei erste starke Evidenzen. Jedoch greifen wir auf jeder Analyseebene linguistisch-interpretierend ein. Das bedeutet, dass alle Ergebnisse, die wir darstellen, vor dem Hintergrund unserer subjektiven Interpretation zu bewerten sind. Dennoch ist unser Vorgehen – unter systematischer Ausnutzung automatischer Verfahren – sehr viel strenger daten- und oberflächenorientiert als das bei den meisten vergleichbaren lexikologischen und lexikografischen Projekten der Fall ist. Daher können wir gesichertere Aussagen über den tatsächlichen Sprachgebrauch treffen.

Die forschungstheoretischen Implikationen des UWV-Ansatzes auch im Kontext der aktuellen Mehrwort- und Konstruktionsforschung sowie die exemplarische Beschreibung eines kausalen Wortverbindungsfeldes, vor allem unter funktionalen Gesichtspunkten, werden in einer gesonderten Monografie behandelt. Daher finden diese Aspekte hier nur am Rande Berücksichtigung.

2. Methodische Prämissen und Beschreibungsgegenstand

Sowohl die methodischen Prämissen als auch Grundlegendes zum UWV-Ansatz wurden in den letzten Jahren in zahlreichen Publikationen dargestellt und diskutiert (vgl. u.a. Steyer 2000, 2003, 2004 (Hg.), 2008, demn.). Daher sollen in diesem Beitrag nur einige für das Verständnis des Vorgehensmodells unverzichtbare Aspekte noch einmal zusammengefasst werden.

Den Terminus **‘korpusgesteuert’** verwenden wir im Sinne von Tognini-Bonelli als Bezeichnung für ein streng von den empirischen Korpusdaten geleitetes analytisches induktives Vorgehen, bei dem die linguistische Interpretation nachgelagert – a posteriori – erfolgt.² Das korpusgesteuerte Prinzip lässt sich folgendermaßen zusammenfassen: Beobachtung von aus sehr großen Korpora gewonnenen Sprachdaten – Akzeptanz aller Evidenzen – Bildung von Hypothesen auf Basis der Evidenzen – empirisches Prüfen der Hypothesen (Zusammenspiel von Induktion und Deduktion) – deskriptive Aussagen, die die Evidenzen reflektieren – Verallgemeinerung im Sinne von Gebrauchsregeln (vgl. Steyer/Lauer 2007, S. 494)

Unser Ansatz geht von der Annahme eines Bezugsworts aus, dessen syntagmatische Strukturen untersucht werden sollen. Damit setzen wir folgende Minimalhypothesen a priori an:

- Es gibt ein Ausgangslemma X, dessen Distributionsverhalten eine Relevanz für die zu untersuchende Fragestellung hat.
- X ist Bestandteil des deutschen Wortschatzes.
- X hat eine spezifische morphologische Struktur.
- X wird gespeist aus einer Menge von Wortformen in einer spezifischen orthografischen Form.

Alle weiteren Verallgemeinerungen werden erst bei der Analyse der vorgefundenen Daten, also bei der Auswertung der Kookkurrenzen, getroffen. Auch weitere automatische Analysen, die zur Verifizierung von Hypothesen notwendig sind, leiten sich aus dieser ab. Unser methodisches Vorgehen lässt sich als **induktiv**, **qualitativ** und **iterativ** charakterisieren:

² Vgl. ‘Corpus-Driven-Approach’ bei Tognini-Bonelli (2001); zu verschiedenen korpusgesteuerten Perspektiven vgl. auch Perkuhn (2007), Keibel/Belica (2007), Steyer/Lauer (2007), Brunner/Steyer (2007), Steyer (demn.), Bubenhofer (2008). Einen Überblick über korpusmethodische Zugänge geben Lemnitzer/Zinsmeister (2006) und Sailer (2007). Wenn wir im Folgenden von korpusgesteuerten Analysen sprechen, meinen wir immer eine von einer linguistischen Interpretation geleitete und nicht eine allein auf automatische Berechnungen ausgerichtete Vorgehensweise.

- **induktiv**, weil die Methode von den konkreten Vorkommen ausgeht und dann Generalisierungen a posteriori vornimmt;
- **qualitativ**, weil die automatisch berechneten Daten einer intellektuellen Interpretation unterzogen werden;
- **iterativ**, weil dies ein wechselseitiger Prozess ist, bei dem quantitative Daten qualitativ analysiert werden, diese Ergebnisse wiederum zu neuen automatischen Systematisierungen führen, die dann wiederum qualitativ interpretiert werden.

Dem UWV-Analysemodell liegt die Hypothese der ‘funktionalen Verfestigung’ von Wortfolgen durch den massenhaften Gebrauch zu Grunde, der Konstitution von funktionalen Mustern des Sprechens über viele ähnliche Erfahrungsfälle. Diese Hypothese hat bereits Feilke in seinem Konzept der idiomatischen Prägung ausführlich diskutiert (vgl. u.a. Feilke 1998 und 2004). Tomasello beschreibt dieses Prinzip im Kontext aktueller Spracherwerbsforschung folgendermaßen:

Wenn Menschen wiederholt ‘ähnliche’ Dinge in ‘ähnlichen’ Situationen sagen, entwickelt sich daraus mit der Zeit ein sprachliches Verwendungsmuster, das in den Köpfen der Benutzer als neue Kategorie oder Konstruktion schematisiert wird – mit unterschiedlichen Abstraktionsgraden. [...] Es gibt keine sprachlichen Elemente – ob lexikalische oder syntaktische, konkrete oder abstrakte, reguläre oder idiomatische – die nicht in diesem Sinne symbolisch sind; alle haben einen kommunikativen Bedeutungsgehalt, weil sie alle direkt aus dem Sprachgebrauch abgeleitet sind. (Tomasello 2006, S. 21).

Korpusgesteuert vorzugehen bedeutet nun, diese Verfestigungsprozesse **bottom up** anhand großer Sprachauschnitte zu rekonstruieren. Man geht von konkreten lexikalischen Instanzierungen an der sprachlichen Oberfläche aus und trifft Verallgemeinerungen hin zu abstrakteren Konstruktionsmustern. Entscheidend dabei ist, dass diese lexikalischen Instanzierungen nicht okkasionell oder singulär sind, sondern durch ihre Rekurrenz eine Prototypikalität aufweisen. An der sprachlichen Oberfläche anzusetzen heißt, sich auf der Ebene des konkreten Ausdrucks, der tatsächlichen lexikalischen Realisierungen im Text zu bewegen und also rekurrente Folgen von **Wortformen** eines Lemmas zu betrachten. Für die Aufwertung von Wortformen als eigene lexikalische Einheiten hat schon Sinclair argumentiert:³

There is a good case for arguing that each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity. (Sinclair 1991, S. 8)

Das UWV-Analysemodell folgt konsequent dieser Prämisse vom Primat der Wortform gegenüber dem Lemma, indem zuerst die spezifischen syntagmatischen Umgebungen jeder einzelnen Wortform eines Lemmas untersucht und die entsprechenden wortformenbezogenen Wortverbindungen und Wortbindungsmuster bestimmt werden. Erst nach dieser Detailanalyse können Verallgemeinerungen darüber getroffen werden, welche Wortverbindungen wortformengebunden sind, welche sich über einige Flexionsformen verallgemeinern lassen und welche sich auf das ganze Lemma beziehen.

Unser Beschreibungsgegenstand sind **usuelle Wortverbindungen** (UWV). Als UWV – erstmals eingeführt bei Steyer (2000) – werden auf der Ebene der konkreten Lexikalisierung rekurrente Verbindungen zwischen mindestens zwei lexikalischen Einheiten verstanden, die eine auffällige Affinität zueinander aufweisen und darüber hinaus auch in **rekurrente syntaktische Strukturen** eingebettet sind. Die Komponenten einer usuellen Wortverbindung können unterschiedliche Ausformungen haben, von einer konkreten Wortform eines Lemmas bis zu ganzen Ausdrucksklassen auf der abstrakteren Verallgemeinerungsebene. Ihre Übersummativität ist primär pragmatisch motiviert. Das heißt, sie sind auch ohne weiteren Kontext als Einheit versteh- und einsetzbar. UWV können völlig regulär, aber auch idiosynkratisch sein oder

³ Vgl. dazu auch Stubbs (2001, S. 26ff.); Tognini-Bonelli (2001, S. 92ff.).

im Sinne einer grammatischen Phrase unvollständig erscheinen. Sie sind als ‘Halbfertigprodukte’ der Sprache anzusehen (Hausmann 1985, S. 118). Unabhängig davon, welchen strukturellen Status sie haben, sind sie immer als **kommunikative Einheiten** zu verstehen, als funktionale Chunks, als verfestigte Bausteine des Sprachgebrauchs. Den Status einer usuellen Wortverbindung können (1) syntagmatische Minimaleinheiten,⁴ (2) binäre/trinäre Relationen zwischen autosemantischen Komponenten und (3) mehrgliedrige – auch satzwertige – Phrasen aufweisen.

(1)

syntagmatische Minimaleinheiten*im Grunde/im Grunde genommen*

= ‘eigentlich’

mit Grund

= ‘berechtigterweise’

auf Grund

= ‘wegen’

(2)

autosemantische Binär- und Triplestrukturen*zu Ohren kommen*

= ‘erfahren, (eher zufällig) mitbekommen’

Grund und Boden

= ‘Grundstücks-/Landbesitz’; ‘Ressource’

in Grund und Boden

= ‘vollkommen, völlig, endgültig’

öffentlicher Grund

= ‘kommunaler, nicht privater Grundstücksbesitz’

(3)

komplexere phrasale Einheiten*aus welchem Grund/welchen Gründen auch immer*

= ‘Etwas ist zu konstatieren, obwohl die Ursachen/Motive nicht bekannt oder nicht nachvollziehbar sind.’

X Wort in Gottes Ohr

= ‘Mit der Wortverbindung wird die Hoffnung ausgedrückt, dass bestimmte Aussagen oder erklärte Absichten von anderen wahrgenommen und akzeptiert werden. Häufig ist mit dieser Äußerung aber ein gewisser Zweifel verbunden, das diese auch wirklich auf Akzeptanz stoßen und sich daraus die gewünschten Folgen ergeben.’

hinter die Ohren schreiben

= ‘Jemand soll einen Rat/eine Kritik beherzigen.’

sich die Nacht/Nächte um die Ohren schlagen

= ‘Jemand verbringt eine ganze Nacht ohne Schlaf. Dabei wird ausgedrückt, dass dieser Sachverhalt mit einem besonderen Einsatz und Willen verbunden ist, z.B. beim Arbeiten. Die Wortverbindung wird auch im Kontext von Feiern, Party machen usw. verwendet.’

⁴ Syntagmatische Minimaleinheiten sind statistisch signifikante Verbindungen zwischen einem Inhaltswort und einem/mehreren Funktionswörtern. Das Kriterium für den UWV-Status solcher Minimaleinheiten ist ihre kommunikative Autonomie als sprachlicher Ausdruck. Dieser autonome Status kann sich sowohl auf eine eigenständige Bedeutung im engen Sinne als auch auf eine eigenständige Funktion in der Kommunikation beziehen.

UWV können lexikalisch voll spezifizierte Wortverbindungen sein wie in den Beispielen (1)-(3). Sie können aber auch minimale lexikalische Varianzen aufweisen wie in Beispiel (4) oder Muster höheren Grades (Wortbindungsmuster) im Sinne einer paradigmatischen Varianz von Komponenten konstituieren wie in (5).⁵

(4)

um die Ohren + VERB (**nur**: *hauen/schlagen*)

= ‘Jemand wird verbal in besonders harter Weise mit etwas konfrontiert, z.B. mit Argumenten oder Kritik.’

In einigen Verwendungen beziehen sich diese Wortverbindungen darauf, dass Menschen mit tatsächlichen physischen – meist herumfliegenden – Objekten konfrontiert werden.

im Grunde + DETERMINATIVUM (des/der/seines...) + NOMEN (**nur**: *Herzens/Seele/Wesens*)

= ‘Grundbedeutung: *eigentlich*. Sprecher verleihen dieser Bedeutung aber einen besonderen Nachdruck, z.B. als Ausdruck besonders ausgeprägter, aber oft nur geheimer Vorstellungen, Wünsche, Ideen.’

(5)

ins Ohr + VERB (potenziell alle Verben, die auf eine menschliche Äußerungsart referieren, z.B. *flüstern/raunen/brüllen/schreien* usw.)

= ‘jemandem mit unterschiedlicher Lautstärke etwas mitteilen’

aus Gründen + DETERMINATIVUM (*der/des...*) + NOMEN (potenziell alle Nomina, die auf einen als notwendig angesehenen Sachverhalt oder eine entsprechende Maßnahme referieren wie *Verkehrssicherheit/Umweltschutzes/Selbstschutzes/Immunität*)

= ‘Ursachen und Motive werden explizit benannt.’

In den folgenden Kapiteln wird nun dargestellt, wie diese unterschiedlichen Festigkeiten und variablen Ausfüllungen mit dem UWV-Analysemodell ermittelt und die Wortverbindungen ausgehend von den Korpusdaten in eine systematisierte, erkenntnisleitende Ordnung gebracht werden können.

3. Ausgangsbasis: Kookkurrenzprofile

Die Analysebasis bildet das „Deutsche Referenzkorpus“ (DeReKo) am IDS.

Die Korpora geschriebener Gegenwartssprache des IDS bilden mit über 3.4 Milliarden Wörtern (Stand 28.07.2008) die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit. ([KLa](#))

Da unser Ziel darin besteht, allgemeine Bildungsmuster der Sprachverwendung zu rekonstruieren, ist es sinnvoll, die größte Stichprobe der deutschen Sprache zu Grunde zu legen, die uns derzeit zur Verfügung steht. Eingriffe in die Korpuszusammensetzung wurden für die Entwicklung des Prototyps des UWV-Analysemodells zunächst nicht vorgenommen. Natürlich sind alle Aussagen, die wir treffen, folglich auch nur in Bezug auf dieses Korpus zu interpretieren (vgl. Belica/Steyer 2008). Die Zukunft wird zeigen, für welche Wortbindungsfelder sich die IDS-Korpora in besonderer Weise eignen, ob es nötig erscheint, aus diesen ein

⁵ Zum Konzept lexikalischer Spezifiziertheit und zu konstruktionellen Perspektiven vgl. u.a. Deppermann (2006), Engelberg (2007), Steyer (demn.).

virtuelles Spezialkorpus zusammenzustellen und inwiefern andere Korpora zum Vergleich herangezogen werden sollten.

Den empirischen Ausgangspunkt unseres Analysemodells bilden KWIC-Cluster von Wortformen, die mit Hilfe der statistischen Kookkurrenzanalyse (KA) ([Belica 1995](#)) berechnet wurden.⁶

Die Kookkurrenzanalyse fasst Originaltextausschnitte zu hierarchischen Gruppen, den (...) Clustern, zusammen, z.B. aufgrund eines gleichen bzw. ähnlichen Kontextverhaltens eines zentralen Trefferobjekts. (Perkuhn 2007, S. 482)

Das Funktionsspektrum der KA wird auf der Homepage des Projekts ‘Methoden der Korpusanalyse und -erschließung’ wie folgt zusammengefasst: Sie

- ermöglicht das Aufdecken von signifikanten Regelmäßigkeiten bei der Verwendung von Wortkombinationen in den Korpora
- wertet mit Hilfe mathematisch-statistischer Analyse- und Clusteringverfahren den definierbaren Kontext eines vorgegebenen Suchobjekts in beliebigen virtuellen Korpora aus
- liefert Hinweise auf systematisches gemeinsames Auftreten von Wörtern (Kookkurrenzen) und ein Maß für deren Affinität (Kohäsion)
- fasst Belege, die ähnliches Kookkurrenzverhalten des Wortes dokumentieren, zu Gruppen/Clustern zusammen
- strukturiert diese Belegmengen ggf. hierarchisch
- bietet eine entsprechende synoptische Präsentation der Belege
- erfasst neben binären Wortrelationen auch usuelle phrasale Muster bis hin zu (idiomatischen) Mehrworteinheiten ([KLb](#))

Ein weiteres Charakteristikum der KA ist die Möglichkeit, Analyseparameter variabel einzustellen und so die Kookkurrenzanalyse auf das jeweilige Forschungsinteresse hin zu fokussieren (zu den Analyseparametern im Einzelnen siehe [KA-Tutorial](#)).

Die Kookkurrenzanalyse ist kein einzelnes, starres Verfahren, vielmehr eröffnet sie eine ganze Bandbreite von Möglichkeiten. So wie es verschiedene Phänomene in der Sprache gibt, die untersuchenswert sind, so gibt es unterschiedliche Arten, Kookkurrenzfragen zu formulieren. Nach der Korpusauswahl und der Suchanfrage ist der zu analysierende Kontext zu definieren und es sind die Werte verschiedener Parameter festzulegen. ([KA-Tutorial](#))

Für die Pilotversion haben wir uns für folgende Parameter entschieden:

Spanne: 5 5
Autofokus: ein
Granularität: fein
Zuverlässigkeit: analytisch
Clusterzuordnung: mehrfach
Lemmatisierung: nein
Funktionswörter: nein

Auf die beiden letztgenannten Parameter soll etwas genauer eingegangen werden. Die Entscheidung für den Ausschluss der **Lemmatisierung** der Kookkurrenzpartner hat folgenden Grund: Eine Einbeziehung der Lemmatisierung von Kookkurrenzpartnern (Lemmatisierung: ja), also die automatische Zusammenfassung aller Flexionsformen der Kookkurrenzpartner zu einer Grundform, würde unser oberflächenbasiertes Vorgehen insoweit konterkarieren, als spezifische Vorkommensbesonderheiten einzelner Flexionsformen nicht mehr gut erkennbar wären. Als Resultat erhielte man nämlich KWIC-Cluster, die nicht nach einzelnen Wortformen unterscheiden und deshalb eine genaue Analyse des jeweiligen distributiven Verhaltens

⁶ Zum Konzept der Kookkurrenzanalyse und der Bedeutung der statistischen Angaben verweisen wir auch auf das entsprechende Tutorial <http://www.ids-mannheim.de/kl/misc/tutorial.html>, des Weiteren auf Perkuhn (2007), Keibel/Belica (2007); zur linguistischen Interpretation auf die Arbeiten von Steyer.

einer affinen Wortform zum Bezugsobjekt erschweren. So werden bei dieser Parametereinstellung beispielsweise alle Vorkommen des Adjektivs *gut* in Bezug zu *Grund* in ein- und dasselbe Cluster zusammenfasst: *Grund – gut*. Man findet sowohl KWIC zu *guter Grund* als auch zu *gutem Grund*, zu *guten Grund* usw. Wendet man die KA ohne Lemmatisierung der Kookkurrenzpartner an, erhält man gesonderte Cluster für alle statistisch relevanten Flexionsformen, bei der Bezugswortform *Gründe* z.B. für Groß- und Kleinschreibung (*Gründe – gute*; *Gründe – Gute*). Dass die großgeschriebene Variante *Gute* ein eigenständiger Kookkurrenzpartner ist, indiziert bereits eine Auffälligkeit in der Verwendung dieser sprachlichen Form. Geht man diesen Indizien nach, zeigt sich, dass die Wortverbindung sehr häufig formelhaft als satzeinleitendes Element verwendet wird.

Die Entscheidung, ob man **Funktionswörter** als mögliche Kookkurrenzpartner einbezieht oder ausschließt, hat vor allem Folgen in Hinblick darauf, welche Phänomene bei der KA primär in den Vordergrund treten. Das UWV-Analysemodell sieht prinzipiell sowohl Analysen mit Funktionswörtern als auch ohne vor. Je nach Einstellung ergeben sich andere Resultate oder besser gesagt, andere Fokussierungen. Wenn man die Funktionswörter ausschließt, lenkt man die Analyserichtung hin zu Relationen zwischen autosemantischen Partnern. Man erhält Informationen, die sich zunächst auf deren Affinität zueinander und ihre typischen syntaktischen Realisierungen beziehen und oft konzeptuell-semantic sind.

Nehmen wir als Beispiele die Bezugswortformen *Grund* und *Ohren*. Schließt man die Funktionswörter aus, erhält man als affinste Kookkurrenzpartner von *Gründen* autosemantische Einheiten wie *gutem/Boden/genug/gibt/Feiern*. Entsprechende Wortverbindungen sind u.a. *aus gutem Grund*; *Grund und Boden*; *in Grund und Boden*; *Grund genug*; *Grund geben*; *Grund zum Feiern*. Bei *Ohren* findet man autosemantische Kookkurrenzpartner wie *taube/offene/abstehenden/Augen/klingen* am oberen Ende der Affinitätsskala. Entsprechende Wortverbindungen wären hier *taube/offene Ohren*; *mit abstehenden Ohren*; *Genuss für Augen und Ohren*; *in den Ohren klingen*.

Bezieht man die Funktionswörter ein – und eigentlich entspricht dies der wirklich oberflächenbasierten Analyse sprachlicher Vorkommen – verschiebt sich der Fokus hin zu den syntaktischen Umgebungen. Man erkennt durch den hohen Stellenwert in der statistischen Rangfolge jetzt auch, welche synsemantic Partner eine besondere Affinität zur Bezugswortform aufweisen. So lassen sich typische Strukturmuster identifizieren. Signifikante Kookkurrenzpartner für *Grund* sind bei dieser Parametereinstellung z.B. *für/Der/auf/zur/Als*. Die entsprechenden Strukturmuster sind *Grund für X*; *Der Grund*; *auf Grund*; *Grund zur X*; *als Grund*. Signifikante Kookkurrenzpartner für *Ohren* sind z.B. die Präpositionen *um/auf/hinter/über/für/zu/mit*. Die entsprechenden Strukturmuster sind *um die Ohren*; *auf X Ohren*; *in den Ohren*; *hinter die/den Ohren*; *über die/beide Ohren*; *für X Ohren*; *zu Ohren (kommen)*; *mit X Ohren*.

In der Pilotversion wurde die Systematisierung der Wortverbindungen mit der Parametereinstellung ‚ohne Funktionswörter‘ erarbeitet, weil so die autosemantischen Partner besser herausgearbeitet und zusammengefasst werden können. Analysen mit Funktionswörtern dienen der Validierung der identifizierten signifikanten Strukturmuster. In einem nächsten Schritt ist jedoch geplant, weitere Ergebnisse dieser Analysen, die wir mit der ersten Parametereinstellung nicht abdecken konnten, in die Wortverbindungsfelder zu integrieren.

Abbildungen 1 und 2 stellen Ausschnitte aus den Ergebnissen der KA dar, wie sie über die COSMAS-II-Webschnittstelle angezeigt werden. (vgl. [CII](#))

LLR	kumul.	Häufig	links	rechts	Kookkurrenzen	syntagmatische Muster
104016	41	41	-1	-1	gesundheitlichen vorzeitig Ruhestand	97% aus gesundheitlichen Gründen vorzeitig in den Ruhestand
	42	1	-1	-1	gesundheitlichen vorzeitig zurückgetretenen	100% gesundheitlichen Gründen vorzeitig zurückgetretenen
	222	180	-1	-1	gesundheitlichen vorzeitig	95% aus gesundheitlichen Gründen [...] vorzeitig in/aus den
	369	147	-1	-1	gesundheitlichen Ruhestand	89% aus gesundheitlichen Gründen vorzeitig in den vorzeitigem Ruhestand
	421	52	-1	-1	gesundheitlichen zurückgetretenen	98% den den des aus gesundheitlichen Gründen [...] zurückgetretenen
	7042	6621	-1	-1	gesundheitlichen	88% aus gesundheitlichen [...] Gründen
71022	7101	59	-1	-1	finanziellen personellen	49% aus Aus finanziellen und personellen Gründen nicht ...
	7143	42	-1	-1	finanziellen abgesetzt	90% aus finanziellen Gründen [...] abgesetzt worden.
	7203	60	-1	-1	finanziellen zuletzt	50% nicht zuletzt aus finanziellen Gründen
	13322	6119	-1	-1	finanziellen	87% aus finanziellen [...] Gründen
56100	13323	1	-1	-1	beruflichen Amt Rücktritt	100% beruflichen Gründen ... Rücktritt ... Amt
	13324	1	-1	-1	beruflichen Amt kürzer	100% beruflichen Gründen kürzer ... Amt
	13588	264	-1	-1	beruflichen Amt	59% der aus beruflichen [...] Gründen [sein das] Amt
	13673	85	-1	-1	beruflichen Rücktritt	72% aus beruflichen [...] Gründen ... seinen Rücktritt aus
	13727	54	-1	-1	beruflichen kürzer	100% muss aus beruflichen [...] Gründen [...] kürzer treten treten. und
	18184	4457	-1	-1	beruflichen	89% aus beruflichen [...] Gründen
48666	18215	31	-1	-1	politischen rassischen	61% aus politischen oder und rassischen [...] Gründen
	18237	22	-1	-1	politischen ethnischen	54% und aus politischen oder und ethnischen Gründen Verfolgte eingesetzt
	18254	17	-1	-1	politischen Opportunität	100% aus Gründen der politischen Opportunität und ...
	24376	6122	-1	-1	politischen	79% aus politischen [...] Gründen
34338	24592	216	-1	-1	organisatorischen Anmeldeschluss Anmeldungen	100% den Anmeldeschluss können aus organisatorischen Gründen keine Anmeldungen mehr entgegenen
	25013	421	-1	-1	organisatorischen Anmeldung	87% aus organisatorischen Gründen [bitte ist wir eine um ...] Anmeldung
	25266	253	-1	-1	organisatorischen Anmeldungen	97% können aus organisatorischen Gründen [keine] Anmeldungen mehr entgegenen entgegen
	27661	2415	-1	-1	organisatorischen	92% aus Aus organisatorischen [...] Gründen
33451	27682	1	-1	-1	wirtschaftlichen umweltpolitischen zuletzt	100% wirtschaftlichen Gründen ... zuletzt ... umweltpolitischen
	27689	7	-1	-1	wirtschaftlichen umweltpolitischen	57% aus wirtschaftlichen und umweltpolitischen Gründen
	27698	9	-1	-1	wirtschaftlichen sicherheitspolitischen	66% aus wirtschaftlichen und aus sicherheitspolitischen Gründen
	27726	28	-1	-1	wirtschaftlichen zuletzt	60% Nicht nicht zuletzt aus wirtschaftlichen Gründen
	31197	3471	-1	-1	wirtschaftlichen	82% aus wirtschaftlichen [...] Gründen

Abb. 1: Auszug aus dem Kookkurrenzprofil für *Gründen*, Ansicht in COSMAS II Web

LLR	kumul.	Häufig	links	rechts	Kookkurrenzen	syntagmatische Muster
46459	222	222	-1	-1	taube stieß	77% stieß ... auf taube Ohren.
	341	119	-1	-1	taube stießen	68% stießen ... auf taube Ohren.
	468	127	-1	-1	taube stößt	55% stößt ... auf taube Ohren.
	2770	2302	-1	-1	taube	46% auf taube Ohren.
29575	2772	2	-1	-1	offene stiess Anliegen	50% stiess ... Anliegen ... offene Ohren.
	2834	62	-1	-1	offene stiess	54% stiess ... auf offene Ohren.
	2835	1	-1	-1	offene gestossen Anliegen	100% Anliegen ... offene Ohren gestossen
	2857	22	-1	-1	offene gestossen	100% auf offene Ohren [...] gestossen
	2902	45	-1	-1	offene Anliegen	33% mit ... Anliegen [auf] offene Ohren.
	5210	2308	-1	-1	offene	50% auf offene [...] Ohren
17667	5211	1	-2	-2	Augen offenen wachen	100% wachen Augen ... offenen Ohren
	5284	73	-2	-2	Augen offenen	76% mit den offenen Augen und Ohren durch die
	5391	107	-2	-2	Augen offen	91% Augen und Ohren [...] offen halten.
	5401	10	-2	-2	Augen wachen	90% mit wachen Augen und Ohren
	8016	2615	-2	-2	Augen	67% Augen [und] Ohren
4359	8252	236	-3	-3	faustdick	47% hat es faustdick hinter den Ohren.
4289	8264	12	-1	-1	abstehenden großen	33% großen abstehenden Ohren
	8268	4	-1	-1	abstehenden Lippen	75% abstehenden [...] Ohren, ... und ... Lippen
	8271	3	-1	-1	abstehenden anlegen	100% abstehenden Ohren [...] anlegen
	8506	235	-1	-1	abstehenden	67% mit den abstehenden [...] Ohren und ...
3355	8537	31	-3	-3	Musik klingt	51% klingt wie Musik in den Ohren der ...
	8551	14	-3	-3	Musik klingen	42% wie Musik in den Ohren [der ...] klingen
	8557	6	-3	-3	Musik Kino	83% +Film +Musik+ Kino für die Ohren + Musik für die
	9353	796	-3	-3	Musik	55% Musik [in den] Ohren der ...
3251	9406	53	-3	-3	Fell gezogen	98% das Fell über die Ohren gezogen wird.

Abb. 2: Auszug aus dem Kookkurrenzprofil für *Ohren*, Ansicht in COSMAS II Web

Die Cluster der KA fungieren im UWV-Analysemodell als erste Orientierung und als mögliche Indikatoren für rekurrente syntagmatische Fixiertheit. Sie sind aber nicht der Beschreibungsgegenstand selbst. Das UWV-Analysemodell geht vielmehr von den jeweiligen KWIC-Zeilen der Kookkurrenzcluster aus. Solche Konkordanzzeilen stellen natürlich nur einen Ausschnitt aus größeren textuellen Einheiten dar. Unsere empirischen Analysen haben jedoch gezeigt, dass sich rekurrente Festigkeit innerhalb von Syntagmen mit einem zentralen Bezugswort in den meisten Fällen in dieser Spanne zeigt.

Kookkurrenzcluster lassen sich nach sehr verschiedenen Kriterien ordnen, systematisieren und interpretieren. Die Kriterien richten sich nach Erkenntnisinteresse und linguistischer Schule (vgl. z.B. Stubbs 2001, Blumenthal 2006, *lexiko*). Unser Forschungsinteresse ist auf die syntagmatische Festigkeit von affinen Wortkombinationen, ihre varianten und invarianten Strukturen, ihre wechselseitigen Beziehungen und ihre Funktionen in der Kommunikation, auf ihre Qualität als usuelle Wortverbindungen also, gerichtet. Daher stellt die mikrotextuelle Perspektive auf die rekurrenten syntagmatischen Umgebungen einer Bezugseinheit und ihrer Kookkurrenzpartner den primären Analysezugang dar.

4. Das UWV-Analysemodell

Das UWV-Analysemodell beinhaltet vier aufeinander aufbauende Interpretationsstufen:

- 1) Die Ebene der konkreten Oberflächenmuster in den KWIC-Zeilen (Suchmuster)
- 2) Die Ebene der lexikalischen Realisierungen
- 3) Die Ebene der Wortverbindungen
- 4) Die Ebene der Wortverbindungsmuster

Der Analyseansatz ist darauf ausgelegt, die Hinweise, die die statistische Methode der Kookkurrenzanalyse und die sprachliche Oberflächenform bieten, auf kontrollierte Weise zu verwerten und auszuschöpfen. Auf jeder Ebene nehmen wir jedoch Interpretationen vor, und zwar in Form von Zusammenfassen, Klassifizieren, Zueinander-in-Beziehung-Setzen und Kommentieren von Phänomenen. Die Ergebnisse dieser Analysen dienen zum einen als Heuristik für die Erforschung von Wortverbindungen als lexikalisch teilspezifizierte Konstruktionen. Zum anderen werden bestimmte Ergebnisse in Form von Wortverbindungsfeldern systematisch elektronisch aufbereitet und online publiziert.

4.1 Die Ebene der konkreten Oberflächenmuster

4.1.1 Methodische Begründung

Auf dieser Analyseebene werden die nach statistischer Signifikanz in Cluster zusammengefassten Kontextzeilen, in denen die Bezugswortform und die Kookkurrenzpartner gemeinsam vorkommen, einer weiteren Untersuchung unterzogen. Es sollen gemeinsame Oberflächenmerkmale syntagmatischer Strukturen identifiziert werden, um diejenigen Realisierungen zu bündeln, die für die Beschreibung einer Wortverbindung relevant sind. Gleichzeitig können auf diese Weise Indizien für Formfixiertheiten und paradigmatische Varianzen (Lücken) gewonnen werden. Es werden lexikalische Vorkommen in ihrer unmittelbar einbettenden Satzumgebung betrachtet mit dem Ziel, zusammengehörige syntagmatische Strukturen (also identische Festigkeiten in den Umgebungen des Suchobjektes) zu entdecken und Oberflächenmerkmale für typische Verwendungsweisen zu identifizieren.

Auf der Basis der ersten Untersuchungen der Kookkurrenzcluster formulieren wir Hypothesen über strukturelle Festigkeiten und kontextuelle Indikatoren. Diese versuchen wir dann mit der Formulierung von Suchanfragen (Suchmustern) zu erfassen. Die Suchmuster dienen dazu, die KWIC-Zeilen neu zu sortieren und zu bündeln.

Die Notwendigkeit, die berechneten KWIC-Cluster einer weiteren Analyse zu unterziehen, resultiert aus mehreren Faktoren.

- Die KA, die derzeit via COSMAS II zur Verfügung steht, ist nicht darauf ausgelegt, alle feinen Oberflächenstrukturen in der Umgebung der jeweiligen syntagmatischen Einheiten 1:1 abzubilden. Die zusammengefassten KWIC-Belege weisen viele weitere Merkmale auf, die intellektuell zu identifizieren sind.

Die in folgendem Auszug aus der KWIC-Menge des Clusters *Gründen* – *persönlichen* blau hervorgehobenen Wörter illustrieren Ad-hoc-Annahmen zu kontextuellen Indikatoren in der Umgebung eines rekurrenten Syntagmas,⁷ hier der Verbindung zwischen der Bezugswortform *Gründen* und dem primären adjektivischen Kookkurrenzpartner *per-*

⁷ Der Terminus ‘Syntagma’ wird im Folgenden für eine rekurrente Wortfolge in den KWIC-Zeilen verwendet, auch wenn es sich bei diesen rekurrenten Wortfolgen nicht immer um Phrasen im traditionellen Sinne handelt.

sönlichen. Diese Hypothesen über weitere relevante Umgebungsmuster sind nun durch Analysen der KWIC-Strukturen in Hinblick auf ihre Relevanz zu verifizieren oder auch zu verwerfen.

Gründen – persönlichen:

T05 hingegen schon seit Wochen nicht mehr gestanden. **Aus persönlichen Gründen**, die außer ihm
 T06 macht und die Ehe mit einem eingeht, den sie **aus persönlichen Gründen** nicht hätte
 A98 Fatzer hat nach sechsjähriger Ratstätigkeit aus **persönlichen Gründen** den **Rücktritt**
 A98 seine persönliche Teilnahme **aus «persönlichen und professionellen Gründen»** zurück, teilte
 A98 Diax, Hans Rudolf Wittmer, **tritt** per sofort **«aus persönlichen Gründen»** zurück.
 A98 heisst es, Roland Luchsinger sei **«aus persönlichen und beruflichen Gründen»**
 A99 beteiligt und der BT diese Dienste honoriert. **Aus persönlichen Gründen** und Wegzug aus
 A99 1999. Wie Konrad Schiess begründete, bringe er **aus beruflichen Gründen** sowie
 Veränderungen im **persönlichen** Umfeld die notwendige Zeit
 A99 nach Ablauf der Frist auch Urs Bähler, Herisau, **aus persönlichen Gründen** seine **Demission**
 A00 Niederer schliesslich zog seine **Kandidatur** zwar aus **persönlichen Gründen** zurück, liess
 A00 Nachfolgerin wählen lassen. Gaby Weber hatte sich aus **persönlichen Gründen** zu ihrem
Rücktritt entschlossen, da sie einerseits zeitlich sehr
 A00 Gleich eine halbe Lausanner Mannschaft würde aus **ganz persönlichen Gründen** ihr letztes
 Hemd für einen Uefa-Cup-Erfolg heute Abend über
 V98 Michael Unger tritt **aus persönlichen und gesundheitlichen Gründen** ab,
 V99 am Dienstag seinen Rücktritt erklärt. Er scheidet "aus **persönlichen Gründen**" aus dem **Amt**,
 erklärte er in Washington. Wenn es nach ihm gehe,
 P99 hinaus will, wird schnell klar: Er will, aus **ganz persönlichen Gründen**, dafür sorgen,
 daß der Gangster sobald wie möglich das Zeitliche
 R97 unklar. Alle Kollegen, die aus **dienstlichen oder persönlichen Gründen**, zeitweise
 R97 habe auch die sexuelle Anziehung nachgelassen. **Zu den persönlichen Gründen** für Trennungen
 kommen nach Beobachtung der Familienbildung
 R97 erfüllen soll. Johannes Rau will, wohl auch aus **sehr persönlichen Gründen**, diese
 B03 verdient als der Marketingvorstand. Das **mag** aus **persönlichen Gründen** **nachvollziehbar**
sein. Für

Abb. 3: KWIC-Belege des Kookkurrenzclusters *Gründen – persönlichen* (Ausschnitt)

- Verwandte rekurrente Oberflächenformen sind nicht immer an ein einziges Kookkurrenz-cluster gebunden, sondern können verteilt über mehrere Cluster vorkommen. Deshalb muss auch der Analysefokus entsprechend erweitert werden. Typischerweise sind es Realisierungen einer Wortverbindung, die Variation in ihrem Wortmaterial aufweist (und sei es nur Verbflexion oder orthographische Varianz), die über mehrere Cluster verteilt sind. Beispiele wären:

(6)

Wortverbindung *offenes Ohr*:

Bezugswort: *Ohr* relevante Cluster: *offenes, Offenes, offeneres, offene*

Wortverbindung *übers Ohr hauen*:

Bezugswort: *Ohr* relevante Cluster: *gehauen, hauen, haut, über's*

Wortverbindung *X Ohren trauen*:

Bezugswort: *Ohren* relevante Cluster: *trauen, traute, trauten*

Wortverbindung *X Wort in Gottes Ohr*:

Bezugswort: *Ohr* relevante Cluster: *Gottes, Wort, Worte*

- Umgekehrt kann ein- und dasselbe Kookkurrenzcluster Vorkommen mehrerer Wortverbindungen enthalten, die erst mit Hilfe der Kompetenz unterschieden werden können. Im folgenden Auszug aus den KWIC-Belegen des Clusters *Ohren – klingen* sind Realisierungen der Wortverbindungen *in X Ohren klingen* (schwarz), *für X Ohren Y klingen* (orange) und *die Ohren klingen* (blau) hervorgehoben.

ermöglichen die Zusammenfassung all jener KWIC-Zeilen, die vom Suchmuster erfasst werden. Die ‘Sprache’ der WV-Suchmuster ist angelehnt an reguläre Ausdrücke. Jedes Element repräsentiert mindestens ein ‘Wort’.⁹ Mögliche Elemente sind zum einen konkrete Zeichenketten, zum anderen Lückenelemente, d.h. Platzhalter für nicht fest spezifizierte (Folgen von) Zeichenketten. Folgende Lückenelemente stehen uns momentan zur Verfügung:

- # steht für ein einziges, beliebiges Wort
- #* steht für 0 bis beliebig viele Wörter
- #+ steht für 1 bis beliebig viele Wörter
- N(wort) steht für ein einzelnes Wort, dass nicht mit wort identisch ist („nicht wort“)
- wort1|wort2 steht für „wort1 oder wort2“

Die Lückenelemente erlauben es, Variationen innerhalb der rekurrenten syntagmatischen Einheit zu erfassen und erfüllen damit eine wichtige Funktion bei der Untersuchung von Wortverbindungen. Das UWV-Tool bietet zudem die Möglichkeit, die verschiedenen Lückenfüller in Gruppen mit gleicher Oberfläche zusammenzufassen und ihre Häufigkeit zu zählen.

Obligatorische Elemente eines Suchmusters sind das Bezugswort und ein primärer Kookkurrenzpartner. Fakultativ können weitere Wörter je nach Hypothesen über syntagmatische Festigkeiten und Kontextmarker hinzugefügt werden. Das Suchmuster erfasst alle KWIC-Zeilen eines Clusters, die die festgelegten Wörter (fakultativ mit entsprechenden Lücken) enthalten. Bei der Treffersuche werden nie Satzgrenzen überschritten, da wir davon ausgehen, dass eine Wortverbindung sich in der Regel nicht über mehrere Sätze erstreckt.¹⁰

Bei der Untersuchung der Daten wird versucht, möglichst viele KWIC-Belege eines Clusters durch Suchmuster zu erfassen, sodass eine geringe Menge unspezifizierter Belege übrig bleibt. Es ist jedoch zu betonen, dass eine vollständige Abdeckung eines Clusters mit Suchmustern nicht in jedem Fall sinnvoll ist. Affine Partner innerhalb eines Clusters müssen nicht in jedem Fall eine gemeinsame syntagmatische Bindung aufweisen. Diese Belege sind bei der Beschreibung von Wortverbindungen – wenn überhaupt – nur in Hinblick auf kontextuelle Marker relevant (siehe [Kapitel 4.2.2.6 Kontext-LR](#)).

Das Formulieren von Suchmustern basiert auf Hypothesenbildung und erfolgt iterativ. Zunächst werden heuristische Suchmuster angesetzt und die so strukturierte KWIC-Menge daraufhin betrachtet, ob sie eine sinnvolle Teilmenge im Sinne einer zusammenhängenden Struktur mit gemeinsamen Gebrauchsmarkern ergibt. Auch die Menge der nicht erfassten KWIC-Treffer wird daraufhin überprüft, ob sie noch relevante rekurrente Oberflächenformen enthält. Je nach Befund werden weitere WV-Suchmuster angesetzt, die restriktiver oder allgemeiner formuliert sein können. Ist die Untersuchung abgeschlossen, werden einige der Suchmuster ausgewählt, die die relevanten rekurrenten Oberflächenformen am besten bündeln, und für die weitere Beschreibung gespeichert. Diese können dann in einer nächsten Abstraktionsstufe lexikalischen Realisierungen zugeordnet werden (siehe [Kapitel 4.2](#)).

Das iterative Prinzip bei der Definition von WV-Suchmustern soll nun anhand eines Beispiels illustriert werden. Es geht um WV-Suchmuster, die letztendlich der Beschreibung der Wortverbindung *X Wort in Gottes Ohr* dienen werden.

In [Kapitel 4.1.1](#) wurde bereits erwähnt, dass sich Realisierungen dieser Wortverbindung im Profil *Ohr* in den Clustern *Gottes*, *Wort* und *Worte* finden lassen. Näher dargestellt werden

⁹ Im Kontext der WV-Suchmuster ist mit ‘Wort’ stets eine von Leerzeichen und/oder Satzzeichen umgebene Zeichenkette gemeint.

¹⁰ Aus technischen Gründen wird momentan keine echte Satzendeerkennung durchgeführt, sondern nur auf typischerweise satzbeendende Zeichen hin geprüft. Die Korrektheit der Ergebnisse ist dabei nicht immer zu garantieren, aber ausreichend für den Prototyp.

hier Untersuchungen des Clusters *Ohr – Wort*. Ein kurzer Auszug aus der KWIC-Menge dieses Clusters sieht wie folgt aus:

Unerreichbare Einheit Euer Wort in Gottes	Ohr.	Doch warum, so ist frau versucht zu
jüngsten Ausgabe der 'Zitty'. Sein Wort in Gottes	Ohr.	Die innere Konstruktion von RIAS -TV gibt indes wenig Anlaß
Lande getrieben oder gezerzt hat, endlich zurückkönnen." Sein Wort ins	Ohr	des Neuen Forums, das die Forderung nach Rückkehrerlaubnis für die
Bei Anruf Gottes Wort im	Ohr	
Schnee und am meisten Sonne". Sein Wort in des Wettergottes	Ohr.	Arno Mayer
zur deutschen Einheit zu beteiligen. - Ihr Wort in Kohls	Ohr.	B.D.
dafür bisher keine Krankheit verantwortlich gemacht. Sein Wort in Bacchus	Ohr.	Die Önologen des Bordelaisgebiets sind anderer Auffassung. Sie schlagen schon
raunt ihren früheren Dienstherrn via Telefon ein geheimnisvolles Wort ins	Ohr.	Ihr "Shanthi" treibt die Männer in den Selbstmord. Antoine (Francis
lles Binsenweisheiten: Daß ein Gespräch, ein liebes Wort , ein offenes	Ohr,	ein anteilnehmender Händedruck oft für die Gesundung einer Kranken wichtiger
hört ihr denn ihr Frauen das Wort des Herrn, euer	Ohr	vernehme das Wort seines Mundes! Lehrt eure Töchter die Klage,
eben auf ihn unseren Eid ablegen." Sein Wort in Gottes	Ohr.	Michael Harms
und sich neu zu orientieren." Sein Wort in des Theatergottes	Ohr!	Ein Opernkritiker aus dem benachbarten Halle, der mit den Verhältnissen
Ein Elfjähriger kichert einem Klassenkameraden das Wort Reizwäsche ins geneigte	Ohr.	Der zuständige Pädagoge verhängt dafür keinen Elfmeter, sondern einen Strafaufsatz

Abb. 5: KWIC-Belege des Clusters *Ohr – Wort* (Ausschnitt)

Die automatisch berechneten syntagmatischen Muster der KA liefern erste Hinweise auf die dominante Form der Wortverbindung. Die erste Hypothese wird nun zunächst mit einem relativ allgemeinen Suchmuster *Wort #* Gottes #* Ohr* überprüft.

	Wort	#*	Gottes	#*	Ohr	
Kein	Wort	in	Gottes		Ohr	Da ein katholischer Seelsorger die Aussage verweigert, droht ihm Beugehaft.
das Geld für nützlichere Dinge auszugeben. Sein	Wort	in	Gottes		Ohr	- oder hält er sich für Gott? Denn göttliche Fähigkeiten
ihren sexuellen Spaß haben wollten. Daher ein	Wort	in	Gottes		Ohr	: Für einen solchen Spaß hört sich der Spaß auf! KURT
den Kunstwerken bewußter wahrzunehmen." Sein	Wort	nicht nur in	Gottes		Ohr	, sondern auch in jenem seiner Kolleginnen und Kollegen aus der
einer von ihnen dem Fernsehpublikum versprochen. Sein	Wort	in	Gottes		Ohr	. WERNER ADAM W.A. Es ist kaum zu glauben, doch wenn
(Vandenhoeck & Ruprecht) Hirsch, Eike Christian: "Mein	Wort	in	Gottes		Ohr	". Ein Glaube, der Vernunft annimmt. 380 S., geb., 39,80 (Hoffmann
hilft." Derartiges hat auch Gorbatschow versprochen. Sein	Wort	in	Gottes		Ohr	! Franzosen und Italiener klagen über eine verpaßte Chance, dem Kremelchef
schon bald der Fall sein... SCHWAB: Ihr	Wort	in	Gottes		Ohr	.
zu dem steht, was man sagt." Sein	Wort	in	Gottes		Ohr	bzw. in den Ohren jener, die für den Umbau in
nach der WM als Testprofis ausprobieren!" Sein	Wort	in	Gottes		Ohr	.
schon Gründungsvater Wilhelm Exner 1908 gewarnt. Sein	Wort	verfehlte	Gottes		Ohr	. Beim Ausmisten der Petrefakte kann Werner mit Rük kenwind aus

Abb. 6: Treffermenge für Suchmuster *Wort #* Gottes #* Ohr* (Ausschnitt)

Die entsprechenden Lückenfüller für die Position zwischen *Wort* und *Gottes* bestätigen die Festigkeit der Präposition:

in	73
in des Fußball-	2
ins	1
nicht nur in	1
der Richter in	1
in (Fußball)-	1
verfehlte	1

Abb. 7: Ausgezählte Lückenfüller für die Lücke zwischen *Wort* und *Gottes* im Suchmuster *Wort #* Gottes*

Abbildung 7 zeigt, dass die Präposition *in* die mit Abstand häufigste Füllung (73 Treffer) darstellt und auch die wenigen anderen Füller bis auf eine Ausnahme mit ihr kombiniert vorkommen.

Für die Position zwischen *Gottes* und *Ohr* ergab die Analyse keine Füllung in allen 80 Treffern des Suchmusters:

	80
--	----

Abb. 8: Ausgezählte Lückenfüller für die Lücke zwischen *Gottes* und *Ohr* im Suchmuster *Wort #* Gottes*

Es konnte also als erstes Zwischenergebnis ein fester Kern an der Oberfläche bestimmt werden: *Wort in Gottes Ohr*. Auf Grund dieser Resultate muss nun ein restriktiveres Suchmuster formuliert werden, das die internen Lücken, da wo sie nicht ausgefüllt sind, nicht mehr zulässt. In einem nächsten Schritt wird mit Hilfe des Suchmusters: *# Wort in Gottes Ohr* überprüft, wie sich die Elemente vor *Wort* verhalten:

	#	Wort	in	Gottes	Ohr	
sie wieder kommen", sagt Maxine Hanson.	Ihr	Wort	in	Gottes	Ohr	, mögen die Hoteliers und Taxifahrer denken, die Restaurantbesitzer, Broadway-Schauspieler, Busunternehmer
ein bißchen etwas genommen werden müssen".	Sein	Wort	in	Gottes	Ohr	.
Tochter! sie wird sich nicht wegwerfen". "	dein	Wort	in	Gottes	Ohr	, meine liebe Leona! ich habe solche Angst, daß unser Kleinod
entgegnete Matthäus. Karl lächelte etwas unbeholfen. „	Dein	Wort	in	Gottes	Ohr	, ich wünsche dir nur von Herzen, dass du recht behältst",
nicht verlieren. "Eine Eingebung", schmunzelt er. "	Sein	Wort	in	Gottes	Ohr	", sagte Trainer Willi Kaipel dazu. Gestern schloß Seidl ein interessantes
Graz Mittwoch im Ländle noch weiterkommt."	Sein	Wort	in	Gottes	Ohr	- Graz-Coach Pöck klagte jedenfalls: "Wir waren immer einen Schritt
Minuten bringt nichts", ist er überzeugt.	Sein	Wort	in	Gottes	Ohr	. Claus Meinert
wird. "Kein Leid mehr", sagt Wagner-Schöppl.	Ihr	Wort	in	Gottes	Ohr	... HARALD BRODNIG
habe Ihre Kolumne sehr genau gelesen.	Ihr	Wort	in	Gottes	Ohr	, daß es nur einmal geschehen sei. Leider muß ich sagen,
Österreich - aber nicht im Ausland."	Sein	Wort	in	Gottes	Ohr	: Unfaire Attacken gegen Klestil hat es schon genug gegeben.
verhindern, so die Philosophie des Norwegers. "	Brogese"	Wort	in	Gottes	Ohr	, darf man da nur wünschen. Brogeland selbst wird seinen Spaß
und werde "als gesellschaftspolitischer Vorreiter" fungieren.	Ihr	Wort	in	Gottes	Ohr	- und Kunstfreunde in die Galerie, z. B. heute um

	#	Wort	in	Gottes	Ohr	
erklärte der Präsident des Hauptverbandes, Sallmutter.	Sein	Wort	in	Gottes	Ohr	. In anderen Staaten hat sich herausgestellt, daß der Zugriff auf
zu werden", sagt Trainer Rudi Hevera.	Sein	Wort	in	Gottes	Ohr	. Aber nach der letzten Seuchensaison muß er der Mannschaft erst

Abb. 9: Treffermenge für Suchmuster # Wort in Gottes Ohr (Ausschnitt)

Abbildung 9 zeigt, dass die Position vor *Wort* typischerweise mit Possessivpronomen oder mit personenbezogenen Genitiven gefüllt wird. Die auffallende Großschreibung der Possessivpronomen weist zudem darauf hin, dass diese Wortverbindung in der überwiegenden Zahl ihrer Realisierungen satzwertig gebraucht wird. Abbildung 10 listet die Häufigkeit der einzelnen lexikalischen Füller auf.

Sein	37
Ihr	15
Mein	3
Dein	3
dein	2
sein	2
Kein	1
Tonis	1

Abb. 10: Ausgezählte Lückenfüller für die Lücke vor *Wort* in Suchmuster # Wort in Gottes Ohr (Ausschnitt)

Im nächsten Schritt wird die Restmenge der KWIC-Belege in diesem Cluster betrachtet, die nicht von dem Suchmuster # Wort in Gottes Ohr erfasst wurden.

Bei Anruf Gottes Wort im	Ohr	
Schnee und am meisten Sonne". Sein Wort in des Wettergottes	Ohr.	Arno Mayer
zur deutschen Einheit zu beteiligen. - Ihr Wort in Kohls	Ohr.	B.D.
dafür bisher keine Krankheit verantwortlich gemacht. Sein Wort in Bacchus	Ohr.	Die Önologen des Bordelaisgebiets sind anderer Auffassung. Sie schlagen schon
raunt ihren früheren Dienstherrn via Telefon ein geheimnisvolles Wort ins	Ohr.	Ihr "Shanthi" treibt die Männer in den Selbstmord. Antoine (Francis
lles Binsenweisheiten: Daß ein Gespräch, ein liebes Wort, ein offenes	Ohr,	ein Anteilnehmender Händedruck oft für die Gesundung einer Kranken wichtiger
hört ihr denn ihr Frauen das Wort des Herrn, euer	Ohr	vernehme das Wort seines Mundes! Lehrt eure Töchter die Klage,
und sich neu zu orientieren." Sein Wort in des Theatergottes	Ohr!	Ein Opernkritiker aus dem benachbarten Halle, der mit den Verhältnissen
Ein Elfjähriger kichert einem Klassenkameraden das Wort Reizwäsche ins geneigte	Ohr.	Der zuständige Pädagoge verhängt dafür keinen Elfmeter, sondern einen Strafaufsatz
der Angreifer Sie dann los." Sein Wort in des Täters	Ohr.	"Na ja, wir sagen allen Frauen, daß wir hier nur
Bayern-Verantwortlichen hätten sich "bestimmt etwas gedacht". Lothars Wort in Ullis	Ohr.	Für Unterhaltung im Münchner Sommerloch ist bestens gesorgt. taz
"im Werbefernsehen so gut gefallen hat". Maskes Wort in Orffens	Ohr.	Geklagt hatte der Mainzer Musikverlag Schott, der die Rechte des
und Clintons Prinzipientreue in Frage gestellt. Ihr Wort in Limbaughs	Ohr.	Der predigt schon seit zwei Jahren, Clinton habe keine Prinzipien,
Berlins Verkehrssenator Herwig Haase (SPD). Doch Kreibichs Wort in Haases	Ohr:	Richtig angewandt und in Kombination mit Behinderungen des Autoverkehrs könnten
schlecht voran, Debuissou führt das Wort, die anderen sind ganz	Ohr.	Indes, Frankreich ist weit, Debuissou im Herzen doch mehr Groß-
der Erforschung ferner Galaxien messen kann." Sein Wort in Rüttgers'	Ohr.	

Abb. 11: Restmenge der nicht durch das Suchmuster # Wort in Gottes Ohr erfasste KWIC-Belege

Die Abdeckung des Clusters durch das Suchmuster # Wort in Gottes Ohr beträgt in diesem Fall ca. 34%. Man kann erkennen, dass in dem Cluster auch Realisierungen anderer Wortverbindungen enthalten sind (z.B. *ganz Ohr sein, ins Ohr raunen, geneigtes Ohr*).

Darüber hinaus zeigt sich jedoch, dass es eine weitere Realisierungsform der Wortverbindung gibt, in der das Element *Gottes* ausgetauscht wird. Um diese Realisierungsform näher zu erfassen, wird die bereits erwähnte Möglichkeit des Ausschlusses bestimmter Wörter im Suchmuster genutzt und folgendes Suchmuster formuliert: # Wort in #* N(Gottes) Ohr.

Die Kombination aus den Lückenelementen #* und N(Gottes) dient dazu, alle Realisierungen zu erfassen, bei denen nicht die Wortform *Gottes* unmittelbar vor *Ohr* steht (N(Gottes)), aber ansonsten beliebig viele Elemente zwischen *in* und *Ohr* auftreten dürfen (#*).

	#	Wort	in	#*	N(Gottes)	Ohr	
Schnee und am meisten Sonne".	Sein	Wort	in	des	Wettergottes	Ohr	. Arno Mayer
zur deutschen Einheit zu beteiligen. -	Ihr	Wort	in		Kohls	Ohr	. B.D.
dafür bisher keine Krankheit verantwortlich gemacht.	Sein	Wort	in		Bacchus	Ohr	. Die Önologen des Bordelaisgebiets sind anderer Auffassung. Sie schlagen schon
und sich neu zu orientieren."	Sein	Wort	in	des	Theatergottes	Ohr	! Ein Opernkritiker aus dem benachbarten Halle, der mit den Verhältnissen
der Angreifer Sie dann los."	Sein	Wort	in	des	Täters	Ohr	. "Na ja, wir sagen allen Frauen, daß wir hier nur
Bayern-Verantwortlichen hätten sich "bestimmt etwas gedacht".	Lothars	Wort	in		Ulis	Ohr	. Für Unterhaltung im Münchner Sommerloch ist bestens gesorgt. taz
"im Werbefernsehen so gut gefallen hat".	Maskes	Wort	in		Orffens	Ohr	. Geklagt hatte der Mainzer Musikverlag Schott, der die Rechte des
und Clintons Prinzipientreue in Frage gestellt.	Ihr	Wort	in		Limbaughs	Ohr	. Der predigt schon seit zwei Jahren, Clinton habe keine Prinzipien,
Berlins Verkehrsminister Herwig Haase (SPD). Doch	Kreibichs	Wort	in		Haases	Ohr	: Richtig angewandt und in Kombination mit Behinderungen des Autoverkehrs könnten
der Erforschung ferner Galaxien messen kann."	Sein	Wort	in		Rüttgers'	Ohr	.
die liegen im Süden und Osten.	Ihr	Wort	in		Tschubais	Ohr	. Jelzins Stabschef hat, wie zuvor Verteidigungsminister Rodionow, mit Konsequenzen gedroht,

Abb. 12: Treffermenge für Suchmuster # Wort in #* N(Gottes) Ohr (Ausschnitt)

Die Füller für die Platzhalter des Suchmusters # Wort in #* N(Gottes) Ohr sehen folgendermaßen aus:

Sein	29
Ihr	15
Gottes	2
Margreiters	1
Maskes	1
Lothars	1
Kreibichs	1

Abb. 13: Ausgezählte Lückenfüller für die Lücke vor *Wort* in Suchmuster # *Wort* in #* *N(Gottes) Ohr* (Ausschnitt)

Der Befund in Abbildung 13 ist vergleichbar mit dem für das Suchmuster # *Wort* in *Gottes Ohr* in Abbildung 9.

	32
des	10
der	5
unser	2
der heimischen	1
Landeshauptmann	1
Kulturreferent	1

Abb. 14: Ausgezählte Lückenfüller für die erste Lücke zwischen *in* und *Ohr* (#*) in Suchmuster # *Wort* in #* *N(Gottes) Ohr* (Ausschnitt)

aller	2
Politikers	2
Schilys	2
Allahs	2
Politiker	1
Volleyballfans	1
Menschen	1

Abb. 15: Ausgezählte Lückenfüller für die zweite Lücke zwischen *in* und *Ohr* (*N(Gottes)*) in Suchmuster # *Wort* in *N(Gottes) Ohr* (Ausschnitt)

Die Tabellen für die beiden anderen Lücken zeigen, dass es vor allem personenbezogene Genitive sind, die anstelle von *Gottes* eingesetzt werden können. Eine Untersuchung der Treffermenge ergibt zudem, dass die Gebrauchsbedeutung dieser Realisierung der Bedeutung der Kernrealisierung sehr ähnlich ist.

Beide Suchmuster zusammen erfassen nun ca. 60% des Clusters. Die Restmenge setzt sich aus Belegen zusammen, die andere Wortverbindungen realisieren. Dies deutet auch darauf hin, dass es sich um eine sehr stabile Wortverbindung handelt, die sich nicht nur in der statistischen Affinität zwischen *Ohr* und *Gottes* zeigt, sondern auch in der zwischen *Ohr* und *Wort*.

Iterativ Suchmuster als Hypothesen für Wortverbindungen zu formulieren, heißt in einem nächsten Schritt, auch andere 'verdächtige' Cluster auf diese Weise zu überprüfen. In unserem Fall betrifft das neben dem Kookkurrenzpartner *Wort* auch die Cluster *Gottes* und *Worte*, die auf die eben dargestellte Weise analysiert werden.

Die vollständige Liste der finalen Suchmuster für diese Wortverbindung, die nach Untersuchung aller relevanter Cluster angesetzt wurden, sieht folgendermaßen aus (die farbliche Kodierung bedeutet: Bezugswort: grün, primärer Kookkurrenzpartner: blau, Lückenelement: grau, restliche Wörter: schwarz):

Wort in Gottes Ohr

Worte in Gottes Ohr

N(Wort|Worte) in Gottes Ohr

Worte in #* N(Gottes) Ohr

Die WV-Suchmuster sind auch in der Präsentation der erarbeiteten Daten, den Wortverbindungsfeldern, sichtbar und es gibt die Möglichkeit, die ihnen zugeordneten KWIC-Treffer als Kontextstellen anzeigen zu lassen. Enthält ein Suchmuster Lückenelemente, sind für diese auch Tabellen mit den automatisch ausgezählten Füllern hinterlegt. Der Rezipient kann also genau nachvollziehen, welche Suchanfrage einem KWIC-Bündel zu Grunde lag. Um die wichtige Information über die statistische Signifikanz zu dokumentieren, werden in der Darstellung der WV-Suchmuster das Bezugswort des zugrunde gelegten Kookkurrenzprofils und die primären Kookkurrenzpartner eingefärbt. Eine Realisierung, bei der statistische Kohäsion zwischen mehreren Elementen besteht, ist durch den hohen Anteil an farbigen Elementen als besonders fest und stabil zu erkennen.

4.2 Die Ebene der lexikalischen Realisierungen

4.2.1 Methodische Begründung

Die lexikalischen Realisierungen (LR) sind eine Zwischenstufe zwischen der Ebene der konkreten Oberflächenmuster – erfasst durch die Suchmuster – und der Ebene der Wortverbindungen. Sie repräsentieren typische Realisierungen einer Wortverbindung in den Korpus-texten und werden auf der Basis der Analyse der KWIC-Strukturen angesetzt. Sie können auf verschiedene Weise – zumeist hierarchisch – miteinander in Beziehung gesetzt werden. LR dienen später dazu, die prototypische Form einer Wortverbindung und typische Varianzen und Kontextualisierungen strukturiert nachvollziehen zu können.

Die Notwendigkeit dieser Analyseebene ist aus den empirischen Erfahrungen eines streng oberflächenorientierten Vorgehens erwachsen: Es hat sich gezeigt, dass Wortverbindungen sehr viel heterogener, verzweigter und auch verschachtelter sind, als das die Mehrwortforschung bisher angenommen hat. Alle vorschnellen Verallgemeinerungen in Bezug auf den invarianten Kern, die kanonische Form oder die Nenn- bzw. Grundform verstellen den Blick dafür, dass Wortverbindungen in hohem Maße variabel sind, sich Komponenten ihrer Realisierungen in vielfältiger Weise überlappen oder wiederum Kombinationen mit Komponenten anderer Wortverbindungen eingehen. Die in der jeweiligen Äußerungssituation verwendeten Kombinationen fokussieren immer nur bestimmte Teilaspekte des ‘Wortverbindungskonzepts’, nie alle.

4.2.2 Das Konzept der lexikalischen Realisierungen (LR)

Das Konzept der lexikalischen Realisierungen dient dazu, Vorkommen einer Wortverbindung mit unterschiedlichen Ausprägungen von Typikalität, Festigkeit und Varianz zu bestimmen und ihre wechselseitigen Abhängigkeiten darzustellen. Wir unterscheiden Kernrealisierungen, Erweiterungen und Kontexte. Kernrealisierungen repräsentieren das Grundkonzept der Wortverbindung. Reduziert man sie, verändert sich die übersummativ Gebrauchsbedeutung. Neben der dominantesten Kernrealisierung kann es alternative Realisierungen geben, die Kern-LR-Varianten. Der Kern und seine Varianten können durch eine Vielzahl von zusätzlichen Elementen ergänzt werden, die sowohl eng an die syntagmatische Struktur des Kerns gebunden sein können (Erweiterungen) als auch nur im weiteren Umfeld ‚lose‘ mit dem Kern verbunden auftreten können (Kontexte). Jede definierte Realisierung fokussiert einen bestimmten

Aspekt der Kombinatorik der Wortverbindung. Abgebildet werden die lexikalischen Realisierungen in einer Baumstruktur mit Wurzelement und Verzweigungen.

Lexikalische Realisierungen fassen WV-Suchmuster (und damit die von diesen erfassten KWIC-Vorkommen) zusammen und erlauben bereits Abstraktionen über konkrete Oberflächenmerkmale wie Schreibung, Flexion oder Stellung. Allerdings muss die linguistische Kompetenz darüber entscheiden, wann Zusammenfassungen zulässig sind. In Ausnahmefällen sind sie nicht sinnvoll, z.B. wenn nur eine Flexionsform in den Vorkommen nachweisbar ist oder zumindest auffällig häufig als autonome Einheit verwendet wird.¹¹

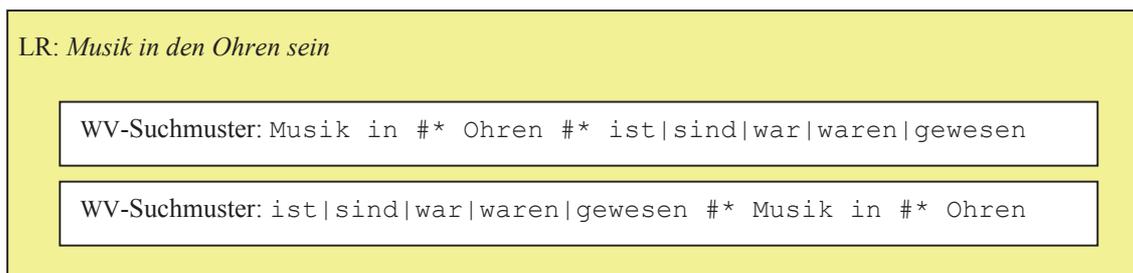


Abb. 16: Schematische Darstellung der LR *Musik in X Ohren sein* und der ihr zugewiesenen Suchmuster

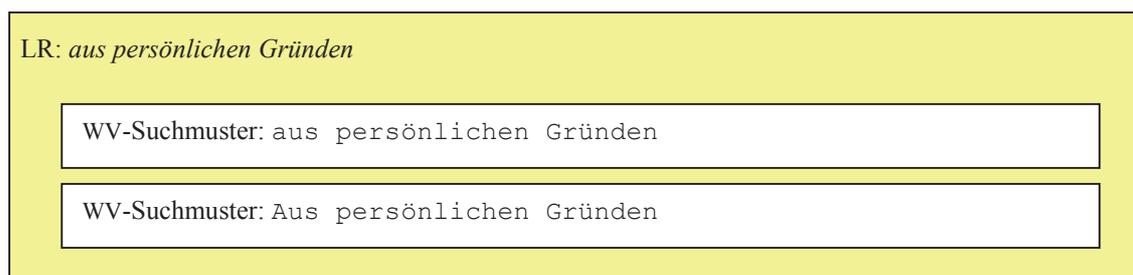


Abb. 17: Schematische Darstellung der LR *aus persönlichen Gründen* und der ihr zugewiesenen Suchmuster

Obligatorische Elemente einer LR sind die **Bezugswortform** (hier *Gründen* und *Ohr*) und mindestens ein primärer **Kookkurrenzpartner** (hier *persönlichen* und *Musik*). Weitere Komponenten, z.B. die Präpositionen *aus/Aus* bei *persönlichen Gründen* und *in* bei *Musik in den Ohren sein*, können hinzugefügt werden, wenn sie sich als fester Bestandteil der Realisierungsform erwiesen haben. LR können auch Lücken enthalten, die mit Hilfe der Suchmuster ermittelt werden, d.h. dass zwischen den obligatorischen Elementen des Syntagmas variable lexikalische Einheiten auftreten können. Die Analyse des Vorkommens dieser Füller ist von zentraler Bedeutung für die beschriebene Realisierungsform der Wortverbindung.

4.2.2.1 Spezifische vs. unspezifische Lücken

Im Gegensatz zu den Suchmustern, die automatische Gesamtfüllerlisten generieren, wie in [4.1.2](#) dargelegt, werden die Lücken auf der LR-Ebene kompetenzbasiert – also qualitativ – interpretiert, und es wird zwischen **spezifischen und unspezifischen Lücken** unterschieden. Spezifische Lücken sind solche, deren lexikalische Füller ein erkennbares Paradigma bilden. Nur diese werden in den Namen der LR in Form von Platzhaltern aufgenommen.

¹¹ Auch die groß- und kleingeschriebenen Varianten eines Syntagmas können bei der Definition von LR unterschieden werden. Dies ist insbesondere dann sinnvoll, wenn die großgeschriebene Variante sehr häufig ist, da oft interessante Gebrauchsnuancen damit verbunden sind.

Das Vorhandensein unspezifischer Lücken wird im Namen nur in seltenen Fällen durch drei Punkte ... markiert, etwa um zu betonen, dass die Elemente im Namen nicht direkt aufeinander folgen. Systematisch spielen diese Art von Lücken nur bei der Definition von Kontext-LR eine Rolle (vgl. [Kapitel 4.2.2.6](#))

Weist eine LR spezifische Lücken auf, sieht das Modell **Füllerangaben** vor. Im LR-Namen wird der Platzhalter mit einem Buchstaben versehen und eine Beispielliste typischer Elemente, mit denen die Lücken gefüllt werden, angegeben. Häufig werden die frequentesten Füllerwörter aufgelistet. Es können aber auch abstrakte Kategorien (z.B. POSSESSIVPRONOMEN, TEMPORALADVERBIEN) genannt oder Kommentare hinzugefügt werden.

Anhand der Suchmuster für das Cluster *Gründen – persönlichen* soll illustriert werden, wie spezifische Lücken auf iterativem Wege identifiziert und in LR dokumentiert werden.

Untersuchungsfokus sind mögliche Erweiterungen des rekurrenten Musters *aus persönlichen Gründen* (zum Konzept der Erweiterungs-LRs vgl. [Kapitel 4.2.2.5](#)). Um alle möglichen Lücken und die Systematik der Füller zu untersuchen, werden mehrere Suchmuster mit dem Platzhalter für eine obligatorische Lücke (#+) angesetzt, d.h. eine Lücke, die immer von mindestens einem Element gefüllt sein muss.

rein	5
dringenden humanitären oder	3
ganz	3
gesundheitlichen und	2
schwerwiegenden	2
Echzell, die aus	1
Dreesen werde die Bank «aus	1
beruflichen oder	1
wirtschaftlichen, aus	1
beruflichen und auch	1
Berlin kommt, aus	1
irgendwelchen	1
zeitlichen und	1
ideologischen wie	1
nicht näher genannten,	1
diesem, aber auch aus	1
Sorge um die Universität, nicht aus	1
dringenden humanitären und	1
ausschließlich	1
politischen, sondern aus	1
wirklich rein	1
sehr einsichtigen und akzeptablen	1
beruflichen wie	1
sonstigen wichtigen	1
fachlichen und	1
parteilpolitischen und	1
angeblich	1
medizinischen und	1
der Firma 1980 aus	1
persönlichen Gründen Das Zeugnisverweigerungsrecht aus	1
sehr	1
verschiedenen	1

Abb. 18: Ausgezählte Lückenfüller für die Lücke zwischen *Aus|aus* und *persönlichen* in Suchmuster *Aus|aus #+ persönlichen Gründen*

Die Füllerliste in Abbildung 18 zeigt, dass die Lücke vor *persönlichen* durch sehr unterschiedliche lexikalische Vertreter gefüllt wird. Sie wird daher als unspezifisch eingestuft.

Kompetenzbasiert werden nun auf Grund von Annahmen über die spezifischen Strukturmuster restriktivere Suchmuster formuliert, um die jeweiligen Füller und damit die Kontextzeilen besser zu gruppieren:

1. Wenn man nur genau ein Element als Lückenfüller zulässt, erhält man eine Liste von Modifikatoren, die sich direkt auf *persönlichen Gründen* beziehen:

Suchmuster: Aus|aus # persönlichen Gründen

rein	5
ganz	3
schwerwiegenden	2
irgendwelchen	1
sehr	1
verschiedenen	1
ausschließlich	1
angeblich	1

Abb. 19: Ausgezählte Lückenfüller für die Lücke zwischen *Aus|aus* und *persönlichen* in Suchmuster *Aus|aus # persönlichen Gründen* (Ausschnitt)

Auf dieser Basis wird eine LR mit folgenden Merkmalen definiert:

LR-Name: aus X persönlichen Gründen

X: MODIFIKATOR

Typische Füller: *rein/ganz/schwerwiegenden*

2. Es gibt ein Muster der Aufzählung mit anderen Adjektiven, das in der Regel mit Konjunktionen wie *und*, *oder*, *als* und *wie* konstruiert wird.

Suchmuster: Aus|aus #+ und|oder|als|wie persönlichen Gründen

dringenden humanitären	4
gesundheitlichen	2
beruflichen	2
zeitlichen	1
ideologischen	1
medizinischen	1
parteipolitischen	1

Abb. 20: Ausgezählte Lückenfüller für die Lücke zwischen *Aus|aus* und *und|oder|als|wie* in Suchmuster *Aus|aus #+ und|oder|als|wie persönlichen Gründen* (Ausschnitt)

Das Lückenelement *X* umfasst potenziell alle Adjektive, die die Ursache/das Motiv eines Sachverhalts explizit thematisieren. Definiert wird darum folgende LR:

LR-Name: *aus X ... persönlichen Gründen;*

X: ADJEKTIV_{DOMAENE}

Typische Füller: *humanitären/gesundheitlichen*

Analog dazu wird die zweite mögliche Lücke zwischen *persönlichen* und *Gründen* untersucht und auf dieser Basis ergibt sich eine weitere LR.

Suchmuster: Aus|aus persönlichen und|oder|als|wie #+ Gründen

politischen	4
beruflichen	3
familiären	2
wirtschaftlichen	2
auch aus finanziellen	1
nicht aus medizinischen	1
intimen	1
familiären"	1

Abb. 21: Ausgezählte Lückenfüller für die Lücke zwischen *und|oder|als|wie* und *Gründen* in Suchmuster *aus|Aus* persönlichen und|oder|als|wie #+ Gründen (Ausschnitt)

LR-Name: *aus persönlichen ... X Gründen*

X: ADJEKTIV_{DOMAENE}

Typische Füller: *politischen/beruflichen/wirtschaftlichen/familiären*

Durch die kompetenzbasierte Untersuchung konnten auf diese Weise statt der unspezifischen Lücke zwei verschiedene Typen von spezifischen Lücken identifiziert werden, die durch Füllertabellen expliziert werden und aufgrund ihrer unterschiedlichen Funktionen bei der Modifikation des Musters *aus persönlichen Gründen* jeweils eigene LR rechtfertigen.

4.2.2.2 LR-Gruppe und Struktur des LR-Baums

Das Konzept der lexikalischen Realisierungen geht von der Gesamtmenge aller zu einer Wortverbindung gehörenden konkreten Realisierungen aus und wird dann immer spezifischer.

Das Wurzelement eines LR-Baumes bildet die so genannte **LR-Gruppe**. Zweck einer LR-Gruppe ist es, die Gesamtmenge aller möglichen Realisierungsformen einer Wortverbindung so genau wie möglich zu erfassen. Richtlinie ist dabei, dass eher zu viele als zu wenige Oberflächen erfasst werden sollten. Darum sind die der LR-Gruppe zugeordneten Suchmuster in der Regel relativ allgemein. Sie enthalten obligatorisch nur Bezugswort und primären Kookkurrenzpartner, fakultativ weitere Kookkurrenzpartner, falls diese nötig sind, um zwischen verschiedenen Wortverbindungen zu differenzieren, und beliebige Lücken zwischen den Komponenten. Die Suchmuster für eine LR-Gruppe müssen zum einen kein ganzes Cluster abdecken, zum anderen können unter eine LR-Gruppe Suchmuster aus verschiedenen Clustern subsumiert werden, wenn sie zum Realisierungsinventar derselben Wortverbindung gehören.

Im Beispielcluster *persönlichen – Gründen* erhält man in den syntagmatischen Mustern der KA einen Hinweis auf einen weiteren festen Partner, die Präposition *aus*. Diese Festigkeit wird mit Suchmustern verifiziert. In über 97% aller Vorkommen ist die Präposition *aus* Teil der rekurrenten Wortfolge *persönlichen Gründen*. Daher bildet die Folge *aus/Aus_persönlichen_Gründen* die übergeordnete LR-Gruppe mit den entsprechenden Suchmustern:

aus #* persönlichen #* Gründen

Aus #* persönlichen #* Gründen

Dass man die LR-Gruppe nicht automatisch mit allen Clustervorkommen gleichsetzen kann, wird auch hier an der Restmenge von KWIC im Cluster *persönlichen – Gründen* klar. Diese Restmenge (ca. 3%) enthält andere Wortfolgen, die für die weitere Bearbeitung nicht mehr berücksichtigt werden.

August 1999. Wie Konrad Schiess begründete, bringe er aus beruflichen	Gründen	sowie Veränderungen im persönlichen Umfeld die notwendige Zeit für das
Hauptstadt Rangun wurde die Friedensnobelpreisträgerin am frühen Sonnabendmorgen (Ortszeit) aus	"Gründen	der persönlichen Sicherheit nach Hause eskortiert". Sie hatte zuvor gemeinsam
u. teilt gleichzeitig mit, daß er »aus persönlichen u. sachlichen	Gründen«	aus der Leitung der Firma ausscheidet u. mir weitere gute
der Transformation mutiert - selten aus innerer Überzeugung, meist aus	Gründen	des Machterhalts und des persönlichen Profits . Zurück zum alten Repressionssystem
dem Medien- Multi. Laut Presseerklärung gibt er die Leitung "aus	Gründen	der persönlichen Lebensplanung , die auch eigenunternehmerische Tätigkeit vorsieht", ab. Reinhard
Privatwohnung des Diplomaten sind extraterritorial. Im Krieg werden Diplomaten aus	Gründen	der persönlichen Sicherheit in der Regel abgezogen. In Zeiten diplomatischer
auf die Hilfe von Leibgurten weder aus technischen noch aus	Gründen	meiner persönlichen Eitelkeit angewiesen fühle. Denn um einen Gürtel geht
erläutert, daß es (für den Fall einer künftigen Zusammenarbeit) aus	Gründen	der persönlichen Sicherheit sinnvoll wäre, über die Tatsache des Gesprächs
Die Region sei militärische Sperrzone, begründeten die Beamten, die „aus	Gründen	der persönlichen Sicherheit “ keine Israelis durchlassen durften. „Dass sie uns

Abb. 22: Restmenge der nicht durch das Suchmuster Aus|aus #* persönlichen #* Gründen erfassten KWIC-Belege

Neben der Zuordnung von Suchmustern werden in einer LR-Gruppe auch die jeweiligen Basiselemente – die autosemantischen Komponenten – gesondert ausgezeichnet. Die Basiselement-Liste soll der Suche nach Wortverbindungen über ihre Komponenten dienen, z.B. alle Wortverbindungen im *Grund*-Feld, die das Adjektiv *gut* enthalten. Diese Suche kann dazu dienen, verschiedene syntaktische Realisierungen zusammenzufassen, die mit einem spezifischen autosemantischen Kookkurrenzpartner verbunden sind, z.B. alle Realisierungen der Verbindungen mit dem Adjektiv *gut* in den Profilen aller *Grund*-Wortformen: *aus|mit gutem Grund/aus einem guten Grunde/aus|mit guten Gründen/gute Gründe*. Auf diesem Wege lässt sich ermitteln, welche konkreten syntaktischen Realisierungen einer konzeptuellen Relation zwischen zwei Lexemen (hier *Grund*_{lemma} – *gut*_{lemma}) tatsächlich usuell sind und wo die Restriktionen im Gebrauch zu finden sind. Möglich wären auch Verlinkungen über die Basiselemente zu anderen lexikalischen Ressourcen, z.B. über eine ID-Vergabe mit den Einworteinträgen in [elexiko](#).

Die zunächst relativ grobkörnig in einer LR-Gruppe zusammengefassten Vorkommen weisen einen sehr unterschiedlichen Grad an Festigkeit und unterschiedliche syntagmatische Strukturen auf und müssen daher weiter spezifiziert werden.

Das Modell sieht vier Typen von lexikalischen Realisierungen vor, die jeweils eine spezifische Teilmenge von Vorkommen aus der Gesamtmenge der LR-Gruppe nach bestimmten Kriterien extrahieren und bündeln: Kern-LR, Kern-LR-Variante, Erweiterungs-LR und Kontext-LR. Diesen ist gemeinsam, dass sie gezielt bestimmte Aspekte der Oberflächenformen fokussieren und zusammenfassen und diese in Hinblick auf ihren Status bei der Beschreibung einer Wortverbindung klassifizieren:

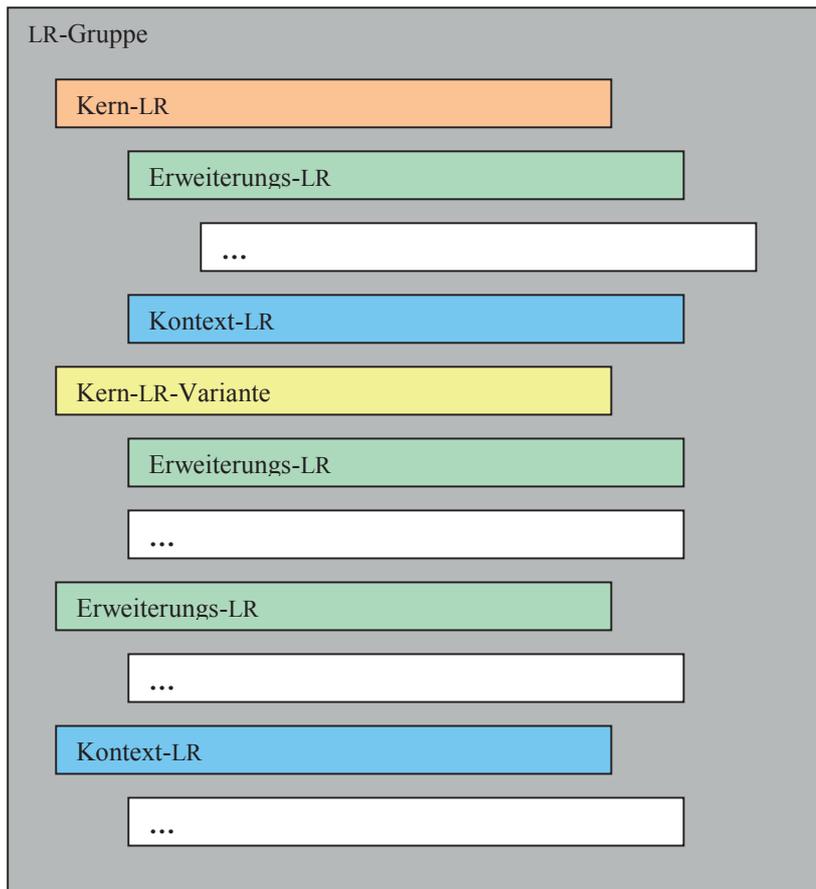


Abb. 23: Schematisierte Darstellung einer LR-Hierarchie

Jedem dieser vier Typen lexikalischer Realisierungen können sowohl eine Gebrauchsbeschreibung als auch ein analytischer Kommentar zugeordnet werden. Auf diese narrativen Beschreibungselemente wird in [Kapitel 4.4](#) ausführlich eingegangen.

Für die LR-Hierarchie gelten folgende Regeln:

- Einer LR-Gruppe ist immer genau eine Kern-LR zugeordnet.
- Einer LR-Gruppe können zusätzlich beliebig viele Kern-LR-Varianten, Erweiterungs-LR oder Kontext-LR zugeordnet werden.
- Den vier LR-Typen (Kern, Kern-Variante, Erweiterung und Kontext) können auf einer weiteren Hierarchieebene wiederum beliebig viele Erweiterungs-LR oder Kontext-LR zugeordnet werden.

Im Folgenden wird die Struktur und Rolle der unterschiedlichen LR-Typen näher erläutert.

4.2.2.3 Kern-LR

Dieser LR-Typ basiert auf der Annahme, dass es in der Oberflächenform einer Wortverbindung einen Kern gibt, der besonders entscheidend für das übersummativem WV-Konzept und die daran geknüpfte Gebrauchsbedeutung ist. Man kann diesen Realisierungstyp als das größtmögliche Kondensat bezeichnen, das noch eine autonome Bedeutung aufweist und auch ohne weitere lexikalische Partner als sprachliche Einheit verwendet werden kann. Die Kern-LR bildet den invarianten syntagmatischen Teil der WV. Sie ist normalerweise die am wenigsten spezifizierte lexikalische Realisierung im LR-Baum und subsumiert die meisten KWIC-Belege.

Im Falle der LR-Gruppe *aus_persönlichen_Gründen* zeigen die Untersuchungen der Belegmenge, die mit den Suchmustern dieser LR-Gruppe erfasst wurde, dass in 90% aller Fälle keine weiteren Elemente zwischen die drei Teile *aus/Aus persönlichen Gründen* des Syntagmas treten. Folglich wird eine Kern-LR *aus persönlichen Gründen* angesetzt, der die Suchmuster *aus persönlichen Gründen* und *Aus persönlichen Gründen* zugewiesen werden.

4.2.2.4 Kern-LR-Variante

Die Kern-LR kann Oberflächenvariationen unterworfen sein, ohne dass die Identität der Wortverbindung dabei verloren ginge. In einem solchen Fall wird für die frequenteste Variante eine Kern-LR angesetzt. Weitere Realisierungsformen werden als **Kern-LR-Varianten** erfasst, die relativ zur Kern-LR definiert werden. Eine Kern-LR-Variante bildet eine Realisierung ab, die an der Oberfläche in mindestens einem Element von der Form der Kern-LR abweicht, so dass nicht dieselben Suchmuster unter diese LR subsumiert werden können wie unter die Kern-LR. Routinemäßige Transformationen (z.B. Verbflexion/-stellung, orthographische Varianten) rechtfertigen nur in Ausnahmefällen eine Kern-LR-Variante, etwa dann, wenn sie eine auffallende Frequenz, Signifikanz oder semantisch-gebrauchsspezifische Besonderheiten aufweisen.

Ein Beispiel für eine Kern-LR-Variante findet sich im Zusammenfassungsbeispiel *Musik in X Ohren* (vgl. [4.2.2.7](#)).

4.2.2.5 Erweiterungs-LR

Erweiterungs-LR dienen dazu, weitere Aspekte der Kombinatorik, die mit einer rekurrenten Wortfolge verbunden sind, in den Fokus zu rücken. Diese zusätzlichen Elemente sind auffällig, jedoch nicht obligatorisch für die Identität des Wortverbindungskerns. Es handelt sich dabei also um Erweiterungen dieses Kerns in seinen verschiedenen Realisierungen. Zum einen können dies **interne Erweiterungen** sein, die zwischen die Elemente des Kerns treten, zum anderen **externe Erweiterungen**, die vor oder hinter dem Kern hinzutreten. Es handelt sich dabei z.B. um typische Verbkonstruktionen, Subjekt- oder Objektergänzungen, Spezifikatoren wie Temporaladverbien oder Intensivierer wie Modalpartikel. Erweiterungen müssen stets syntaktisch mit dem Wortverbindungskern verknüpft und in ein stabiles Muster eingebunden sein.

Bei den so gebündelten KWIC-Belegen kann es sich um echte Teilmengen der Belege der Kern-LR handeln, wenn ihre Suchmuster zusätzliche Elemente vor oder hinter dem Muster des Kerns spezifizieren (Kern-LR: *offenes Ohr*; Erweiterungs-LR: *offenes Ohr für Belange*). Die Menge der KWIC-Belege kann sich aber auch komplementär zu der der Kern-LR verhalten, wenn die Suchmuster der Kern-LR z.B. ohne Lücken definiert sind, die der Erweiterungs-LR aber gerade die Fälle fokussiert, in denen doch Elemente zwischen die Kern-Komponenten treten können (z.B. Kern-LR *aus persönlichen Gründen*, Erweiterungs-LR *aus X persönlichen Gründen*).

Bei der Entscheidung, ob eine Erweiterungs-LR definiert wird, sind wiederum statistische Signifikanz und Frequenz zentrale, aber nicht zwingende Kriterien. Daneben können auch pragmatische und gebrauchsspezifische Besonderheiten als Begründung dafür herangezogen werden, dass eine bestimmte Form als eigenständige LR angesetzt wird.

Die Erweiterungs-LR werden auf verschiedene Weise bestimmt: In der Regel wird für interne Erweiterungen zunächst eine allgemeine Erweiterungs-LR mit dem Platzhalter X angesetzt. Diese deckt alle lexikalischen Elemente dieser Lücke ab. In einigen Fällen ist die Typik der lexikalischen Besetzung der Lücke dann bereits durch die automatischen Füllerlisten ausrei-

chend beschrieben oder aber die Füller stellen eine in sich heterogene Menge dar, aus der keine einzelnen Vertreter als prototypisch isolierbar sind. In bestimmten Fällen jedoch werden darüber hinaus Erweiterungs-LR mit ausgewählten lexikalischen Vertretern aus dieser Liste noch einmal gesondert angesetzt. Darüber hinaus können auch Erweiterungs-LR angesetzt werden, deren Elemente keiner Füllergruppe angehören, aber auf Grund ihrer Häufigkeit ein rekurrentes Erweiterungsmuster mit dem Syntagma bilden.

Entsprechend der bereits beschriebenen Suchmuster ergeben sich beispielsweise für die Kern-LR *aus persönlichen Gründen* folgende Erweiterungs-LR:

aus X persönlichen Gründen

aus X ... persönlichen Gründen

aus persönlichen ... X Gründen

Eine relativ komplexe Struktur von Erweiterungs-LR bildet die LR-Hierarchie für die Wortverbindung *ein offenes Ohr* ab. Diese soll hier auszugsweise dargestellt werden:

ein offenes Ohr (Kern-LR)

ein X offenes Ohr (Erweiterungs-LR)

ein offenes X Ohr (Erweiterungs-LR)

immer ein offenes Ohr (Erweiterungs-LR)

stets ein offenes Ohr (Erweiterungs-LR)

offenes Ohr für X (Erweiterungs-LR)

offenes Ohr für Belange (Erweiterungs-LR)

offenes Ohr für Sorgen (Erweiterungs-LR)

offenes Ohr für Anliegen (Erweiterungs-LR)

offenes Ohr für Probleme (Erweiterungs-LR)

offenes Ohr haben (Erweiterungs-LR)

offenes Ohr finden (Erweiterungs-LR)

offenes Ohr finden bei (Erweiterungs-LR)

4.2.2.6 Kontext-LR

Den vierten Typ von lexikalischen Realisierungen stellen die kontextuellen lexikalischen Realisierungen dar, die so genannten Kontext-LR. Kontext-LR geben Hinweise auf Verwendungszusammenhänge der rekurrenten Wortfolgen. Sie bilden den assoziativen Rahmen der zu beschreibenden Wortverbindung. Sie sind wichtig, um den Gebrauchskontext und spezielle pragmatische Nuancen festzuhalten und können auch Indikatoren für einen typischen diskursiven Hintergrund liefern. So ermöglichen sie auf assoziative Weise ein besseres Verständnis der Bedeutung und Funktion der Wortverbindung in spezifischen Kommunikationssituationen. Kontext-LR umfassen typischerweise die Muster der Kern-LR (und ggf. der Kern-LR-Varianten) sowie noch ein zusätzliches Wort, das auffällig häufig in der Umgebung vorkommt. Im Gegensatz zu den Erweiterungs-LR muss dieses Einzelwort jedoch nicht in einer stabilen syntaktischen Beziehung zur übergeordneten LR stehen. Kontext-LR werden zum

einen dann angesetzt, wenn das Kontext-Wort ein signifikanter Partner des Bezugsworts ist, man also von einer auffallenden Kohäsion zwischen den Elementen sprechen kann.

Im Beispielcluster *persönlichen – Gründen* ist z.B. das Nomen *Amt* ein Kookkurrenzpartner. Zum anderen ermöglichen die WV-Suchmuster weitergehende Verifizierungen in den KWIC, indem Lücken im Vor- und Nachfeld der rekurrenten Wortfolge definiert und ausgezählt werden können:

```
# # # aus #* persönlichen #* Gründen
Aus|aus #* persönlichen #* Gründen # # #
```

Man erhält eine Reihe weiterer Kontextpartner, die das Szenario ‘etwas aus persönlichen Gründen tun / etwas tritt aus persönlichen Gründen ein / etwas wird mit persönlichen Gründen begründet’ beleuchten: *Amt/Funktion/Posten/Angebot/Wahl/Mandat/Rücktritt/Rückzieher*.

Für diese allgemeinen Suchanfragen wird entsprechend eine sehr allgemeine Kontext-LR angesetzt: *x ... aus persönlichen Gründen ... x*. Die automatischen Lückenfüller-Tabellen deuten dann diesen Rahmen an. Das Wort *Amt* hat sowohl im Vor- als auch im Nachfeld die höchste Frequenz. Deshalb wird diese Kontext-LR als eine Spezifikation gesondert aufgeführt: *aus persönlichen Gründen ... Amt*

Ein zweites Beispiel: Die Kern-LR *das menschliche Ohr* weist in ihrer Umgebung folgende, auch statistisch signifikante Kontextmarker auf: *Töne, Frequenzen, wahrnehmbare*. Die entsprechenden Kontext-LR sind:

Töne ... das menschliche Ohr

(Suchmuster: *Töne #* das menschliche Ohr, das menschliche Ohr #* Töne*)

Frequenzen ... das menschliche Ohr

(Suchmuster: *Frequenzen #* das menschliche Ohr, das menschliche Ohr #* Frequenzen*)

wahrnehmbare ... das menschliche Ohr

(Suchmuster: *wahrnehmbare #* das menschliche Ohr, das menschliche Ohr #* wahrnehmbare*)

Die drei Punkte stehen für die unspezifische Lücke zwischen den Elementen und kennzeichnen, dass die Kontextelemente in keiner festen Beziehung zu den restlichen Elementen stehen. Alle drei Kontext-LR weisen darauf hin, dass die Wortfolge *das menschliche Ohr* typischerweise gebraucht wird, wenn von der menschlichen Fähigkeit der akustischen Wahrnehmung die Rede ist. Es kann aber auch Fälle geben, bei denen Umgebungspartner zwar nicht durch die statistische Berechnung als besonders auffällig erkannt wurden, die aber trotzdem eine wichtige Kontextualisierungsfunktion in Bezug auf das Syntagma leisten. Auch diese werden als Kontext-LR bestimmt.

4.2.2.7 Zusammenfassung anhand des Syntagmas *Musik in X Ohren*

Anhand des Beispiels *Musik in X Ohren* soll die Genese einer Wortverbindung von den Ausgangsdaten im KWIC-Cluster bis zum LR-Baum noch einmal im Zusammenhang rekapituliert werden.

Die Ausgangsbasis ist das Kookkurrenzcluster *Ohren – Musik*.

Ohren – Musik

M06	"klingt wie Musik in meinen Ohren ", sagte die Sozialbürgermeisterin und	Ohren ", sagte die Sozialbürgermeisterin und
MLD	"Das ist Musik für unsere Ohren ", sagt Ursula Schmitz. Musik, von	Ohren ", sagt Ursula Schmitz. Musik, von
S93	etwas ist Musik in den Ohren ungezählter Chaosjünger in den Chefetagen	Ohren ungezählter Chaosjünger in den Chefetagen
S94	Die Musik kann für westliche Ohren eine Qual sein; sie ist	Ohren eine Qual sein; sie ist
O94	Chancen... " Musik in unseren Ohren . Doch auch wenn uns wohltut,	Ohren . Doch auch wenn uns wohltut,
O94	klingen wie Musik in den Ohren von Flugzeug-Fans: "Tiger Moth", "Boeing	Ohren von Flugzeug-Fans: "Tiger Moth", "Boeing
O94	Uhren, Ohren , Musik In der Grazer ESC	Ohren , Musik In der Grazer ESC
O94	wahre Musik in Horst Fendrichs Ohren . "Es leben die PS..."	Ohren . "Es leben die PS..."
O94	das war Musik in den Ohren des Akustik-Sachverständigen. Auch der	Ohren des Akustik-Sachverständigen. Auch der
O95	Motoren wieder Musik in seinen Ohren . Die ersten vier	Ohren . Die ersten vier
O95	Sonderling, Künstler, der "die Ohren voll Musik hat und den	Ohren voll Musik hat und den
O95	sind auch Musik in den Ohren des Liebespartners. Was gehört noch	Ohren des Liebespartners. Was gehört noch
O95	Lust sind Musik in den Ohren des Liebespartners	Ohren des Liebespartners
O96	ist wahre Musik in seinen Ohren : Seine Liebe zur Eisenbahn ließ	Ohren : Seine Liebe zur Eisenbahn ließ
O96	Gästen. Musik in Carsten Kelms Ohren , denn wenn`s bei denen im	Ohren , denn wenn`s bei denen im
O96	narrative Musik halten Augen und Ohren immer wieder lustvoll auf Trab.	Ohren immer wieder lustvoll auf Trab.
O96	Schlummersong wirklich Musik in den Ohren der Babys, "dann hörten sie	Ohren der Babys, "dann hörten sie
O97	ist süße Musik in den Ohren der Tourismusstrategen: Murau/Kreischberg	Ohren der Tourismusstrategen: Murau/Kreischberg
O97	- zeitgenössische Musik für junge Ohren (Ossiach, 13. bis 16. 7.).	Ohren (Ossiach, 13. bis 16. 7.).
O97	Während es etwa in den Ohren vieler wie Musik klingt, wenn	Ohren vieler wie Musik klingt, wenn
O97	Musik in des versöhnten Ombudsmans Ohren ! Die Starparade ist der Höhepunkt	Ohren ! Die Starparade ist der Höhepunkt
O97	Dramatik bestimmen. Anregende Musik für Ohren und Augen! V.K.	Ohren und Augen! V.K.
O97	euch" wie Musik in den Ohren . Aber die Papst-Visite wurde ebenso	Ohren . Aber die Papst-Visite wurde ebenso
O97	angekündigt. Musik in meinen Ohren .	Ohren .
O98	ganz gut. Musik in den Ohren gibt wiederum Kraft für das	Ohren gibt wiederum Kraft für das
O98	Musik des "Studio Percussion" die Ohren über. Bernd Schmidt	Ohren über. Bernd Schmidt
O98	Erl vor den Augen und Ohren des Besuchers Musik erarbeitet:	Ohren des Besuchers Musik erarbeitet:
O98	Musik in den Ohren der Politiker! Der Jazz-Sommer der	Ohren der Politiker! Der Jazz-Sommer der
O98	das ist Musik in den Ohren der Kinder! Viele Jugendliche absolvieren	Ohren der Kinder! Viele Jugendliche absolvieren
O99	Das ist Musik in der Ohren der Motorsport-Freaks, die dem	Ohren der Motorsport-Freaks, die dem
O99	Ton nach. "Kino für die Ohren + Musik für die Augen",	Ohren + Musik für die Augen",
O99	der lauten Musik kurz die Ohren zuhielt... Gefeierte wird in St.	Ohren zuhielt... Gefeierte wird in St.
O99	wohl wie Musik in den Ohren der Wähler klingen, nicht aber	Ohren der Wähler klingen, nicht aber
O99	sogar symphonische Musik in meinen Ohren hören. Auch dann, wenn sich	Ohren hören. Auch dann, wenn sich
O99	dem Gefängnis Musik für heile Ohren produzieren.	Ohren produzieren.
O00	Musik in den Ohren : Sting tritt doch auf! Die	Ohren : Sting tritt doch auf! Die
O00	Klang ist Musik in meinen Ohren ; manchmal traurig, manchmal sehr fröhlich.	Ohren ; manchmal traurig, manchmal sehr fröhlich.
O00	- Musik auch in den Ohren von drei Salzburger Seglern, die	Ohren von drei Salzburger Seglern, die

Abb. 24: KWIC-Belege des Kookkurrenzclusters *Ohren – Musik* (Ausschnitt)

Untersucht und strukturiert man die Belegmenge dieses KWIC-Clusters wie in [Kapitel 4.1](#) beschrieben mit Hilfe der WV-Suchmuster, lassen sich folgende Beobachtungen anstellen:

- Das Suchmuster *Musik in #* Ohren* erfasst über die Hälfte der Belege. Man kann hier von einem stabilen, rekurrenten Muster sprechen. In der Restmenge finden sich Realisierungen diverser anderer Wortverbindungen, die jedoch strukturell und semantisch abweichend sind (z.B. *X für die Ohren*, *europäische Ohren*, *Ohren zuhalten* usw.)
- In ca. drei Viertel dieser Teilmenge ist die Lücke *#** mit dem bestimmten Artikel *den* gefüllt (Suchmuster: *Musik in den Ohren*)
- In der Restmenge wird die Lücke häufig durch verschiedene Possessivpronomen (z.B. *meinen/unseren/seinen*) oder gelegentlich auch personenbezogene Genitivphrasen (z.B. *des Bürgermeister/Kohls*) realisiert (Suchmuster: *Musik in #* N(den) Ohren*).
- In ca. 50% der Vorkommen tritt das Verb *sein* in verschiedenen Flexionsformen vor oder nach der Wortfolge *Musik in #* Ohren* auf (Suchmuster: *Musik in #* Ohren #* ist|sind|war|waren|gewesen, ist|sind|war|waren|gewesen #* Musik in #* Ohren*).
- In einem knappen Drittel dieser Teilmenge tritt zusätzlich noch das Demonstrativpronomen *das* vor die Flexionsformen *ist* oder *war*. (Suchmuster: *Das|das ist|war #* Musik in #* Ohren*)
- In ca. einem Viertel der Treffer für *Musik in #* Ohren* tritt der Konjunktiv *wie* vor die Einheit (Suchmuster: *Wie|wie Musik in #* Ohren*)

- Erweitert man dieses Suchmuster durch das Verb *klingen*, zeigt sich, dass dieses Muster über zwei Drittel der Treffermenge von *Wie|wie Musik in #* Ohren* erfasst (Suchmuster: *klingen|klingt|klinge|klang|klängen|geklungen wie Musik in #* Ohren, Wie|wie Musik in #* Ohren klingen|klingt|klinge|klang|klängen|geklungen*).
- Erweitert man das Suchmuster *Wie|wie Musik in #* Ohren* durch Formen des Verbs *sein*, ergibt sich ebenfalls eine, recht geringe, Treffermenge (Suchmuster: *Wie|wie Musik #* in #* Ohren #* ist|war|sind|waren|gewesen, ist|war|sind|waren #* wie Musik #* in #* Ohren*).

Dieser komplexe Befund zeigt bereits die vielfältigen Variationen, denen die rekurrente Wortfolge *Musik in #* Ohren* unterliegt. Durch eine einfache Grundform wie etwa *Musik in jmds. Ohren sein*, wie im Duden 11 verzeichnet (2008, S. 534), wird dies nur unzureichend erfasst. Das LR-Konzept hingegen erlaubt es, mehrere Realisierungsformen zu definieren, welche dann in Beziehung zueinander gesetzt werden, um das Verhalten und die Besonderheiten der Wortverbindungen abzubilden.

Ein weiterer interessanter Aspekt in diesem Beispiel ist der Status der Realisierungen, die von dem Muster *wie Musik in X Ohren klingen* erfasst werden. Diese realisieren gleichzeitig sowohl die rekurrente Wortfolge *Musik in X Ohren* als auch die rekurrente Wortfolge *in den Ohren klingen* und sind damit quasi zwei Wortverbindungen gleichzeitig zuzuordnen. In unserem Ansatz würde dieser Befund auch so abgebildet, d.h. die entsprechende LR (und die ihr untergeordneten Belege) finden sich bei der Beschreibung beider Wortverbindungen.

Die LR-Hierarchie für dieses Beispiel *Musik in X Ohren* wird folgendermaßen strukturiert:

- **LR-Gruppe**

Der LR-Gruppe wird das allgemeine Suchmuster *Musik #* in #* Ohren* zugeordnet, welches alle relevanten Realisierungsformen gut erfasst.

- **Kern-LR**

Als Kern-LR wird die LR *Musik in den Ohren* (Suchmuster: *Musik in den Ohren*) definiert, da die Realisierung der Lücke zwischen *in* und *Ohren* in mehr als der Hälfte der Fälle mit *den* gefüllt wird.

- **Kern-LR-Variante**

Relativ zur Kern-LR *Musik in den Ohren* wird die Kern-LR-Variante *Musik in N(den) Ohren* angesetzt (Suchmuster: *Musik in #* N(den) Ohren*). Diese LR erfasst also die Fälle, in denen statt des Artikels *den* ein Possessivpronomen oder Genitiv steht. Da die Frequenz von *den* so auffällig ist, ist in diesem Fall das Ansetzen von zwei verschiedenen LR sinnvoll, jedoch ändert der Unterschied in der Realisierung dieser Lücke nichts an dem grundsätzlichen Status der Wortfolge als definierendem Element der Wortverbindung. Aus diesem Grund werden die beiden LR als Kern-LR und Kern-LR-Variante definiert.

Die Lücke zwischen *in* und *Ohren* ist spezifischer Natur, die deshalb markiert und durch eine Füllerliste näher beschrieben wird. Diese enthielte in diesem Fall die Elemente: POSSESSIVPRONOMEN, GENITIV (PERSON).

- **Erweiterungs-LR**

Der LR-Baum der Wortverbindung *Musik in den Ohren* weist gleich mehrere externe Erweiterungs-LR auf. Zunächst gibt es die LR *wie Musik in X Ohren* (Suchmuster: *Wie|wie Musik in #* Ohren*) und die LR *Musik in X Ohren sein* (Suchmuster: *Musik*

in #* Ohren #* ist|sind|war|waren|gewesen und ist|sind|war|waren|gewesen #* Musik in #* Ohren). Beide zeichnen sich dadurch aus, dass jeweils ein zusätzliches Element – der Konjunktoren *wie* bzw. eine Form des Verbs *sein* – zum Kern hinzutritt, und erfassen eine deutliche Menge der KWIC-Belege der LR-Gruppe. Hervorzuheben ist, dass diese beiden LR sowohl Realisierungen der Kern-LR als auch der Kern-LR-Variante erfassen – in den Suchmustern, die ihnen zugeordnet sind, wird nicht festgelegt, welches Element vor *Ohr* steht. Diese Variabilität ist auch am Namen der LR zu erkennen, der ein X als Markierung für ein variables Element enthält. Da es sich um eine spezifische Lücke handelt, gibt es auch zu dieser eine Füllerliste, welche ‘den POSSESSIVPRONOMEN, GENITIV (PERSON)’ enthält. Beide Erweiterungs-LR sind darum auch direkt der LR-Gruppe untergeordnet und nicht einer der beiden Kern-Realisierungen.

Ein solches Vorgehen ist typisch für die Strukturierung des LR-Baums: Die Unterscheidung zwischen den beiden Lückenfüllern *den* vs. POSSESSIVPRONOMEN/GENITIV wurde bereits mit Hilfe der Kern-LR und der Kern-LR-Variante beleuchtet und muss somit auf der Erweiterungsebene nicht mehr beachtet werden.

Beiden Erweiterungs-LR können zudem weitere Erweiterungen zugeordnet werden, die die Muster noch weiter spezifizieren. Die LR *wie Musik in X Ohren* ist erweiterbar durch die LR *wie Musik in X Ohren klingen* (Suchmuster: klingen|klingt|klinge|klang|klängen|geklungen wie Musik in #* Ohren, wie Musik in #* Ohren klingen|klingt|klinge|klang|klängen|geklungen), die LR *Musik in X Ohren sein* durch die beiden LR *Das ist/war Musik in X Ohren* (Suchmuster: Das|das ist|war #* Musik in #* Ohren) und *wie Musik in X Ohren sein* (Suchmuster: Wie|wie Musik in #* Ohren #* ist|sind|war|waren|gewesen und ist|sind|war|waren|gewesen #* wie Musik in #* Ohren). Die beiden neuen Erweiterungs-LR werden direkt den jeweiligen ‘Eltern-LR’ zugeordnet und erfassen in diesem Fall eine echte Untermenge der KWIC-Belege von letzteren. In der Tat hätte die Erweiterungs-LR *wie Musik in X Ohren sein* ebenso der LR *wie Musik in X Ohren* untergeordnet werden können, da sich in ihr zwei Erweiterungs-LR überschneiden. Aufgrund der engen semantisch-pragmatischen Nähe zwischen *Musik in X Ohren sein* und *wie Musik in X Ohren sein* erschien es jedoch sinnvoller, diese beiden LR zusammen zu gruppieren. Die Grafik in Abbildung 23 illustriert noch einmal, wie sich die durch die jeweiligen Suchmuster erfassten KWIC-Mengen zueinander verhalten und zeigt die Struktur des daraus entwickelten LR-Baumes.

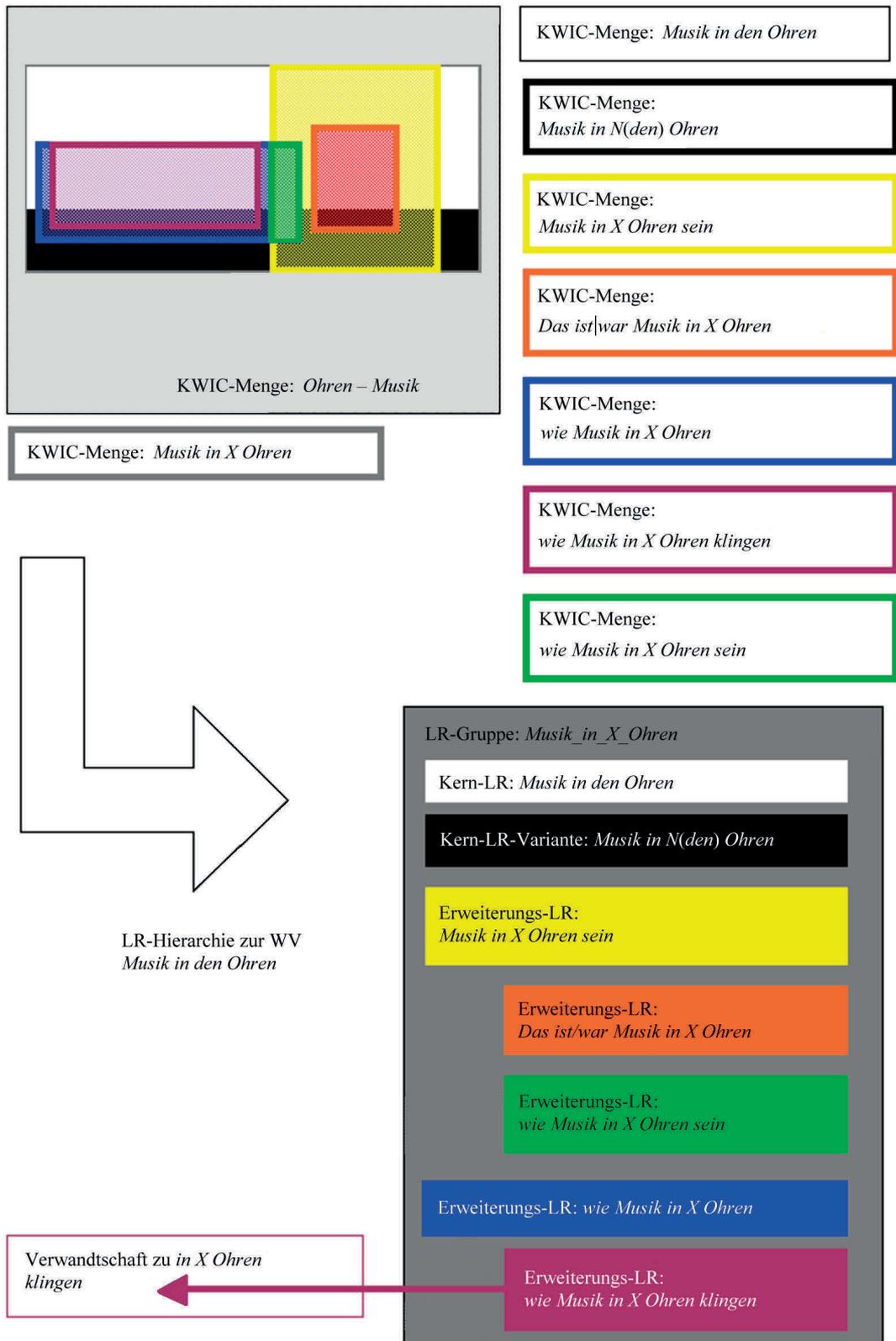


Abb. 25: Von den KWIC-Belegen zum LR-Baum – Beispiel *Musik in den Ohren*

4.3 Die Ebene der Wortverbindungen und Wortverbindungsmuster

Auf dieser Ebene werden alle lexikalischen Realisierungen zunächst zu Wortverbindungen zusammengefasst. Das entscheidende Kriterium für die Bestimmung einer Wortverbindung und die Zuordnung der entsprechenden lexikalischen Realisierungen ist die übersummativ Gebrauchsbedeutung. Alle lexikalischen Vorkommen müssen jeweils einem gemeinsamen semantisch-pragmatischen Konzept zuzuordnen sein, ob sie nun den Kern der Wortverbindung abbilden oder Teilaspekte des Gebrauchs in irgendeiner Weise modifizieren.

Wenn LR definiert werden, haben wir natürlich Hypothesen darüber, welche Wortverbindungen sie konstituieren, und die LR-Hierarchie wird in Hinblick darauf aufgebaut. Einer Wortverbindung wird jeweils eine LR-Gruppe mit dem gesamten Baum lexikalischer Realisierungen und den entsprechenden KWICs zugeordnet. In der Regel sind die Wortverbindungsnamen identisch mit den Namen der Kern-LR, da diese den prototypischen Kern illustrieren.

WV: *aus persönlichen Gründen*

Kern-LR: *aus persönlichen Gründen*

Erweiterungs-LR: *aus X persönlichen Gründen*

Erweiterungs-LR: *aus X ... persönlichen Gründen*

Erweiterungs-LR: *aus persönlichen ... X Gründen*

Kontext-LR: *Amt ... aus persönlichen Gründen*

Als Ergebnis entsteht ein Inventar von typischen lexikalischen Wortverbindungen, das jedoch einer weiteren analytischen Untersuchung bedarf, um deren distinktiven Gebrauch wirklich zu erhellen. Die grundsätzliche Frage lautet: Inwieweit weisen mehrere Wortverbindungen identische Strukturmuster mit variabler lexikalischer Besetzung (paradigmatischer Varianz) auf?

Es gibt zum einen Wortverbindungen, die einen relativ eigenständigen Status haben. Das heißt, sie sind hinsichtlich ihrer lexikalischen Komponenten und syntaktischen Struktur weitgehend fest, zumindest bezogen auf das untersuchte Wortformenprofil. Weitgehend fest bedeutet, dass es zwar Variationen an der sprachlichen Oberfläche (z.B. in Bezug auf Schreibung, Flexion und Stellung) und auch in Bezug auf synonyme lexikalische Vertreter geben kann, diese aber die übersummativ Gebrauchsbedeutung nicht antasten. In der Präsentation der Wortverbindungsfelder werden sie in einer Liste angeboten:

Bezugswortform *Grund*

auf Grund laufen

Grund genug

kein Grund zur Panik

Grund und Boden

in Grund und Boden

Bezugswortform *Ohr*

das menschliche Ohr

mit halbem Ohr

offenes Ohr

übers Ohr hauen

Andere Wortverbindungen dagegen lassen sich Strukturmustern mit variabler lexikalischer Füllung zuordnen. Diese Strukturmuster bilden die den Wortverbindungen jeweils zu Grunde liegenden morpho-syntaktischen Strukturen ab und bestehen aus der Bezugswortform als obligatorischem lexikalischem Element, fakultativ weiteren festen lexikalischen Partnern (z.B. nicht austauschbare feste Kookkurrenzpartner, Funktionswörter) und den jeweiligen Ausdrucksklassen der paradigmatisch austauschbaren Kookkurrenzpartner, eingebettet in eine feste syntagmatische Struktur. Ausdrucksklassen sind in vielen Fällen Wortarten, aber nicht immer. Beispiele für Strukturmuster sind die folgenden:

aus + ADJEKTIV + *Gründen*

aus + *Gründen* + DETERMINATIVUM + NOMEN

im Grunde + DETERMINATIVUM + NOMEN

Gründe + VERB + ADJEKTIVPRÄDIKATIV

auf + ADJEKTIV + *Grund*

NOMEN + *im Ohr*

NOMEN + *für* + NOMEN + *und Ohr*

ins Ohr + VERB

Die in der Mehrwortforschung und -lexikografie nach wie vor übliche Strukturierung nach morpho-syntaktischen Klassen (Wortarten) **allein** ist jedoch u.E. für eine lexikologische Feldstudie relativ unbefriedigend.¹² Deshalb werden im UWV-Analysemodell darüber hinaus den Ausdrucksklassen Eigenschaften zugeordnet, die diese weiter ausdifferenzieren. Diese Merkmalszuschreibungen sind interpretative Setzungen und abhängig vom jeweiligen linguistischen Beschreibungsrahmen. In unserem Ansatz werden zunächst gemeinsame semantische und/oder pragmatische Merkmale der autosemantischen Kookkurrenzpartner herangezogen, um die Strukturmuster weiter zu spezifizieren, z.B.:

aus + ADJEKTIV_{DOMÄNE} + *Gründen*

(Adjektive, die eine Ursache/ein Motiv explizit benennen, z.B. *politischen/finanziellen/technischen*)

im Grunde + DETERMINATIVUM (z.B. *seines/meines/des/der*) + NOMEN_{WESENSKERN} (*Herzens/Seele/Wesens*)

Gründe + *sind* + ADJEKTIVPRÄDIKATIV_{NACHVOLLZIEHBARKEIT_POSITIV}

(Adjektive prädikativer Verwendung, die die Nachvollziehbarkeit von Ursachen/Motiven thematisieren, z.B. *ausschlaggebend/handfest*)

auf + ADJEKTIV_{FARBE} (Farbadjektive, z.B. *weißem/rotem*) + *Grund*

NOMEN_{PHYSISCHES OBJEKT} *im Ohr*

(Nomen, die physische Objekte bezeichnen, z.B. *Ring, Knopf, Stöpsel, Kopfhörer*)

Genuss/Schmaus + *für* NOMEN_{KÖRPERTEILE} (z.B. *Auge...*) + *und Ohr*

ins Ohr + VERB_{ÄUSSERUNGSAKT}

(potenziell alle Verben, die auf eine menschliche Art des Äußerns referieren, z.B. *flüstern, raunen, hauchen, brüllen*)

¹² Einräumen muss man jedoch, dass diese morpho-syntaktischen Strukturtypen im lexikografischen Alltag nach wie vor eine leicht zu praktizierende Vorgehensweise für eine schnelle und möglichst umfassende Inventarisierung von Wortverbindungen darstellen. Beispielhaft wird diese Praxis im WICOL-Projekt umgesetzt (vgl. Āurĉo 2008).

Nach unseren empirischen Erfahrungen lassen sich induktiv ermittelte Ausdrucksgruppen nicht immer in bestehende semantische Taxonomien und Ontologien einordnen. Die Markierungen sind zunächst eher als heuristische Kommentare zu einem typischen Gebrauchsmerkmal von Komponenten der Wortbindungsmuster zu verstehen. Die Frage, inwieweit diese heuristischen Label künftig bestehende Kategorieninventare ergänzen bzw. modifizieren können bzw. neue bilden, wird die Zukunft zeigen.

Die Ebene ‘Wortbindungsmuster’, die die zuvor identifizierten Wortverbindungen zusammenfasst, ist nicht allein als analytische Heuristik zum besseren Verständnis der Zusammenhänge für den späteren Rezipienten zu verstehen. Vielmehr bildet sie kognitive Mechanismen des Sprechens und Verstehens ab, und zwar das Phänomen der dynamischen Aktualisierung von konkreten Gebrauchsbedeutungen einer Wortverbindung und dem zu Grunde liegenden Muster in einer Äußerungssituation. Die Fähigkeit von Sprechern, konkrete Äußerungen vor der Folie eines Musters höherer Ordnung zu verstehen, wird in der Phraseologie vor allem als Erklärung für das Verstehen von sprachspielerischen Modifikationen herangezogen.¹³ Sie scheint aber sehr viel weitreichender für das Funktionieren von Kommunikation zu sein und eben nicht nur im engen Sinne für den Gebrauch von Phraseologismen zu gelten. Ein Beispiel ist die Wortverbindung *aus finanziellen Gründen*. Wird diese Wortverbindung in einer konkreten Äußerungssituation eingesetzt, können die Hörer natürlich ihre Bedeutung erschließen, nämlich dass die Ursachen/Motive für einen Sachverhalt finanzieller und nicht etwa technischer oder politischer Natur sind oder im privaten Bereich liegen. Gleichzeitig verstehen sie jedoch die pragmatische Funktion mit: Es wird im Vagen gelassen, welcher Art die *finanziellen Gründe* und was die genauen Umstände sind. Die Sprecher können diese Vagheitsfunktion sowohl einsetzen als auch decodieren, weil sie viele Verwendungsfälle ähnlicher Art bereits als verfestigte Muster im Kopf haben, und zwar als restringiertes Muster *aus X Gründen*. Die Restriktion liegt im rekurrenten Gebrauch begründet und bezieht sich zum einen auf den Referenzbereich der Adjektive (domänenbezogen), wobei die einzelnen lexikalischen Vertreter frei austauschbar sind, zum anderen auf die rekurrente Verwendung der Pluralform. Singularverwendungen wie *aus einem politischen/finanziellen/usw. Grund* sind zwar bildbar und auch völlig regulär, aber eben – zumindest nach den Befunden auf Grundlage des verwendeten umfangreichen Korpus – nicht usualisiert.

Die mit dem UWV-Analysemodell erarbeitete Hierarchie der Wortbindungsmuster soll solche Prozesse rekonstruierbar machen. Deshalb ist im UWV-Tool die Möglichkeit vorgesehen, neben den konkreten Wortverbindungen, denen die Hierarchien des LR-Baums zugeordnet sind, auch abstrakte Knoten zu definieren. Abstrakte Knoten werden quasi als Etikett (Überschrift) angelegt und subsumieren die zusammengehörigen Wortverbindungen als ‘Kinder’. Mithilfe von abstrakten Knoten können mehrere Hierarchieebenen gebildet werden, die helfen, Abstraktionsstufen strukturiert zu erfassen.

In der Pilotversion bilden die abstrakten Knoten ‘Wortverbindungen’ und ‘Wortbindungsmuster’ die oberste ‘Einstiegs’-Ebene, die Unterscheidung in relativ feste Wortverbindungen und Wortbindungsmuster dient also als primärer Zugang für den Rezipienten.

¹³ Zu Mechanismen phraseologischer Modifikationen vgl. Burger (2007, S. 160ff.).

[Wortverbindungen]

WV: *aus welchen/was für Gründen auch immer*

[Wortbindungsmuster]

aus + ADJEKTIV + *Gründen*

aus + ADJEKTIV_{DOMÄNE} + *Gründen*

WV: *aus politischen Gründen*

WV: *aus finanziellen Gründen*

WV: *aus technischen Gründen*

aus + ADJEKTIV_{NACHVOLLZIEHBARKEIT_POSITIV} + *Gründen*

WV: *aus guten Gründen*

WV: *aus nachvollziehbaren Gründen*

WV: *aus triftigen Gründen*

aus + ADJEKTIV_{NACHVOLLZIEHBARKEIT_OFFEN_NEGATIV} + *Gründen*

WV: *aus unbekannten Gründen*

WV: *aus unerklärlichen Gründen*

WV: *aus unerfindlichen Gründen*

Zusammenfassend lässt sich festhalten: Wortbindungsmuster verdeutlichen das Phänomen, dass Wortverbindungen in einem äußerst fein gesponnenen Netz mit vielen Knoten und Verzweigungen vorkommen, bei denen situationsspezifisch einige Möglichkeiten der Kombination von lexikalischen Einheiten in den Vordergrund, andere in den Hintergrund treten. Um 'universale' Wortverbindungen zu finden, müsste man die Gesamtheit aller Verwendungen quer zu allen beteiligten Wortformen und -lexemen betrachten. Da eine solche allumfassende Analyse nicht zu verwirklichen ist, bieten automatische Vorstrukturierungen wie Kookkurrenzcluster und eine Beschränkung auf bestimmte Bezugswortformen einen praktikablen Bezugsrahmen. Alle Aussagen zu Wortverbindungen sind dementsprechend auf dieser Basis zu verstehen und zu interpretieren.

4.4 Narrative Beschreibungselemente

Das erklärte Hauptziel unseres Modells ist eine erkenntnisleitende Systematisierung in Wortverbindungshierarchien mit den entsprechend strukturierten Belegzeilen. Der Rezipient einer solchen Analyse soll in erster Linie durch die Logik der wechselseitigen Vernetzungen und Zusammenhänge und aus den jeweils zusammengefassten KWIC-Mengen relevante Aspekte des Gebrauchs einer Wortverbindung ableiten können. Allerdings ist eine begleitende Kommentierung trotzdem unverzichtbar, und zwar aus zwei Gründen: Zum einen erfüllt diese Kommentierung wichtige Dokumentationsaufgaben. Methodentransparenz bedeutet auch die Offenlegung der jeweils angesetzten Kriterien für eine LR, eine Wortverbindung oder ein Wortbindungsmuster. Zum anderen muss man konzedieren, dass die Aufbereitung in Wortverbindungsfelder einen neuen, ungewöhnlichen Rezeptionsmodus mit sich bringt. Die Erschließung der typischen Verwendungsmerkmale allein auf Grundlage der strukturierten Daten ist nur für einen sehr kompetenten Sprecher möglich. Erläuterungen zum typischen Gebrauch im eher klassischen Sinne einer Gebrauchsbeschreibung und analytische Kommentare sind daher nach wie vor nützlich und sinnvoll. Welche Aufbereitungs- und Beschreibungsform für welche Rezeptionssituation sinnvoll ist, wird noch näher zu untersuchen sein.

Linguistische Kommentierungen erfolgen auf den Systemisierungsebenen LR, Wortverbindung und Wortbindungsmuster. Im UWV-Analysemodell gibt es zwei Typen von Kommentierungen: zum einen Gebrauchskommentare im engen Sinne, die vergleichbar sind mit lexikografischen Bedeutungsbeschreibungen; zum anderen Kommentare zu Auffälligkeiten im Korpusgebrauch, zu den angesetzten Definitionskriterien usw.

Einige Beispiele für Kommentare / Gebrauchsbeschreibungen auf verschiedenen Ebenen

Beispiel: WV: *aus persönlichen Gründen*

Kommentar zur Kern-LR *aus persönlichen Gründen* mit den Suchmustern *aus persönlichen Gründen/Aus persönlichen Gründen*

„Die syntagmatische Einheit ‘aus persönlichen Gründen’ deckt über 82 % der Vorkommen in dieser LR-Gruppe ab, die großgeschriebene Variante noch einmal über 8%. Auf Grund dieser klaren Häufigkeit von 90% aller Vorkommen wird das Syntagma ohne Lücke als Kern-LR angesetzt.“

Kommentar zur Erweiterungs-LR *aus X persönlichen Gründen*

„Durch die ergänzenden Modifikatoren wird die Aussage, dass es sich um **persönliche** Gründe handelt, zusätzlich verstärkt, z.B. durch solche Intensivierer wie *rein* oder *ganz*. Hier wird stillschweigend ein gewisser Zweifel antizipiert und vorsorglich entkräftet. Mit weiteren Modifikatoren wie *angeblich* wird dieser Zweifel sogar explizit thematisiert.“

Gebrauchskommentar zur Wortverbindung *aus persönlichen Gründen*

„Mit dieser Wortverbindung wird ausgedrückt, dass ein eingetretener Sachverhalt nicht äußeren Gründen geschuldet ist, sondern die Ursache im privaten Bereich liegt. Damit scheint der Sachverhalt weniger beeinfluss- oder kalkulierbar gewesen zu sein. Anders als bei der direkten Benennung von Gründen wie *finanziellen*, *technischen* usw. wird hier bewusst die subjektive Komponente ins Spiel gebracht, die eine entsprechende Handlung als eher nachvollziehbar und gegebenenfalls entschuldigbar erscheinen lässt.“

Beispiel: Kommentar zum Wortbindungsmuster *aus* + ADJEKTIV_{DOMAENE} + *Gründen*

„Dieses Wortbindungsmuster kann potenziell alle Adjektive enthalten, die die Domäne eines Grundes explizit benennen und nicht deren Nachvollziehbarkeit in irgendeiner Weise thematisieren oder bewerten. Diese domänenspezifischen Adjektive treten nur für die Pluralwortformen *Gründen* und *Gründe* als signifikante Kookkurrenzpartner auf. Adjektive, die die Nachvollziehbarkeit thematisieren oder bewerten, werden auch mit den Singularformen verwendet. Die Pluralrestriktion für die Kombination mit ADJEKTIV_{DOMAENE} ist funktional motiviert: Sprecher benennen zwar den allgemeinen Bereich, dem eine Ursache zuzuordnen ist, müssen sich aber durch die mit dem Plural verbundene Vagheit nicht genau auf eine Begründung festlegen. Innerhalb dieser Gruppe gibt es auch eine Skala der Vagheit: So ist die Wortverbindung *aus statischen Gründen* sehr viel konkreter auf die wirklichen Ursachen bezogen als das bei der WV *aus persönlichen Gründen* der Fall ist.“

5. Resümee und Ausblick

Abschließend ist noch einmal zu fragen, welchen Nutzen ein solch detailliertes Vorgehen haben kann. Zu betonen ist zunächst, dass sich das UWV-Analysemodell in erster Linie für die Auswertung von hochfrequenten Vorkommen eignet, bei denen sich die Usualität und Musterhaftigkeit von Syntagmen auf herkömmliche Weise nicht erschließen lässt. Man erhält auf diesem Wege exaktere Informationen

- zur inneren syntaktischen Struktur von Wortverbindungen,
- zu frequenten lexikalischen Realisierungen,
- zum Grad stereotyper lexikalischer Füllungen bis hin zur abstrakten Ebene,
- zu typischen Gebrauchsmustern und ihrer pragmatischen Relevanz.

Das hier dargestellte Vorgehen zur Systematisierung von Wortverbindungen eines Wortformenprofils stellt ein empirisches Fundament dar. Der profil- und wortformenbezogene Ansatz der ersten Entwicklungsphase des UWV-Analysemodells bringt es mit sich, dass Wortverbindungen im UWV-Tool bisher nur in Hinblick auf eine feste Bezugswortform definiert werden können. Momentan kann noch kein Wortbindungsmuster angesetzt werden, das über diese abstrahiert. Entsprechende Verweise werden deshalb zunächst manuell eingetragen. Das UWV-Analysemodell ist aber prinzipiell darauf ausgerichtet, auch wortformenprofilübergreifende und – perspektivisch – lexemübergreifende Verallgemeinerungen zur Usualität von Wortbindungskonzepten und Konstruktionsmustern treffen zu können. Die inhaltliche Konzipierung dieser Verallgemeinerungsebene und die Implementierung der entsprechenden Funktionalität in das UWV-Tool werden zu den nächsten Entwicklungsschritten unseres Modells gehören. In diesem Zusammenhang werden auch Überlegungen eine Rolle spielen, wie Verknüpfungen und Verallgemeinerungen zu Profilen ganz anderer Bezugswörter herzustellen sind (also Musterverweise des Typs: *aus welchen Gründen auch immer* → *X auch immer* → *X = aus welcher* Motiven/Ecke/Richtung/in welcher Form/Art und Weise/für welchen Zweck/ warum/wann/wieso/wie*). In methodisch sauberer Form würde eine Einbeziehung solcher Verallgemeinerungen jedoch zu einem exponentiellen Wachstum der Menge der zu untersuchenden Profile führen. Aus diesem Grund werden hypothetische Extensionen auf profilübergreifende Muster momentan nur durch COSMAS-Anfragen überprüft und bei der Wortverbindung vermerkt. Perspektivisch ist unser Ansatz natürlich so weiter zu entwickeln, dass ein systematischeres Modell zur Behandlung dieser Fälle geschaffen wird. Weitere Überlegungen beziehen sich auf die Art und Weise, wie Anschlussstellen an die jeweiligen Volltextstellen in den IDS-Korpora, an externe Korpora oder andere Analyseverfahren sinnvoll und realisierbar sind.

Konkrete Überlegungen, wie sich eine solche Methodik auch für die exhaustive Bearbeitung vieler Wörter adaptieren lässt, stehen noch aus (vgl. dazu ausführlich Steyer demn.). Wenn man in absehbarer Zeit zu einem großen Inventar usueller Wortverbindungen gelangen will, muss man sich im Grunde auf die wirklich auffälligen Phänomene konzentrieren und kann nicht jede lexikalische Realisierung in der Weise interpretieren, wie wir das in unserem Vorgehensmodell vorsehen. Eine Fokussierung auf frequente Vorkommen bringt aber die Gefahr mit sich, dass man nur Resultate erhält, die auch auf der Basis der reinen Introspektion oder einfacher korpusbasierter Suchen im Sinne eines reinen Befragens des Korpus bzw. einer Belegsuche zu erzielen wären. Diese Vorkommen springen auf Grund ihrer Häufigkeit ins Auge und werden natürlich auch entsprechend erkannt und beschrieben (z.B. in Wörterbüchern, beispielhaft im Duden 11). Mit unserer Methode ist es jedoch nicht nur möglich, feine Ausdifferenzierungen von Varianzen vorzunehmen, sondern auch vordergründig seltene Vorkommen auf einer höheren Ebene zu usuellen Mustern zusammenzufassen. So kommen die Ad-

jektiv-Nomen-Verbindungen *ersichtlicher Grund* oder *triftiger Grund* für sich genommen relativ selten vor. Sie würden bei einer Reduktion auf die häufigsten Vorkommen möglicherweise durchs Raster fallen. Für die Rekonstruktion des Wortverbindungsmusters ADJEKTIVNACHVOLLZIEBARKEIT_POSITIV *Grund* sind sie jedoch als typische lexikalische Vertreter von absoluter Relevanz.

Mit dem hier dargestellten UWV-Analysemodell und der entsprechenden Präsentation der Ergebnisse in Wortverbindungsfeldern versuchen wir in erster Linie, Wortverbindungen in ihrer sprachlichen Form, ihrem Gebrauch und ihrer Bedeutung verstehbar zu machen. Die Art der Darstellung erlaubt dem späteren Benutzer aber auch, die Interpretationsschritte vom Einzelvorkommen zu einer (teilweise abstrakten) Wortverbindung mit explizierter Gebrauchsbedeutung nachzuvollziehen. Gleichzeitig kann man dieses Modell als einen methodischen Ansatz verstehen, mit dem sprachliche Massendaten intelligent strukturiert zugänglich gemacht werden, um als empirische Basis für weitere linguistische Untersuchungen im Bereich der Erfassung und Beschreibung allgemeiner sprachlicher Konstruktionen zu dienen. Unser Prinzip der möglichst schulenneutralen, auf allgemeinen Prinzipien lexikalischer Festigkeit beruhenden Systematisierung sollte eine Adaption für verschiedene linguistische Erkenntnisinteressen erleichtern.

6. Literatur

6.1 Sekundärliteratur und Nachschlagewerke

- Belica, Cyril (1995): Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analyseverfahren. Mannheim. Internet: <http://www.ids-mannheim.de/kl/projekte/methoden/ur.html> (Stand: November 2008).
- Belica, Cyril/Steyer, Kathrin (2008): Korpusanalytische Zugänge zu sprachlichem Usus. In: Vachková, Marie (Hg.): Beiträge zur bilingualen Lexikographie. Prag, S. 7-24.
- Blumenthal, Peter (2006): Wortprofil im Französischen. (= Beihefte zur Zeitschrift für Romanische Philologie 332). Tübingen.
- Brunner, Annelen/Steyer, Kathrin (2007): Corpus-driven study of multi-word expressions based on collocations from a very large corpus. In: Proceedings of the 4th Corpus Linguistics conference, Birmingham. Internet: http://corpus.bham.ac.uk/corplingproceedings07/paper/182_Paper.pdf (Stand: Oktober 2008).
- Bubenhöfer, Noah (2008): Muster an der sprachlichen Oberfläche. Methoden einer korpuslinguistischen Diskurs- und Kulturanalyse am Beispiel des Sprachgebrauchs in der 'Neuen Zürcher Zeitung' von 1995-2005. Univ. Diss. Zürich.
- Burger, Harald (2007): Phraseologie. Eine Einführung am Beispiel des Deutschen. 3., neu bearb. Aufl. (= Grundlagen der Germanistik 36). Berlin.
- Deppermann, Arnulf (2006): Construction Grammar – Eine Grammatik für die Interaktion? In: Deppermann, Arnulf/Fiehler, Reinhard/Spranz-Fogasy, Thomas (Hg.): Grammatik und Interaktion. Untersuchungen zum Zusammenhang zwischen grammatischen Strukturen und Gesprächsprozessen. Radolfzell, S. 43-65. Internet: <http://www.verlag-gespraechsforschung.de/2006/pdf/grammatik.pdf> (Stand: Februar 2009).
- Duden 11 (2008): Duden. Band 11: Redewendungen. Wörterbuch der deutschen Idiomatik. Mehr als 10 000 feste Redensarten und Sprichwörter. 3., überarb. u. aktual. Aufl. Herausgegeben von der Dudenredaktion. Mannheim/Leipzig/Wien/Zürich.
- Duffner, Rolf (2006): Satzadverbien – korpusbasiert. In: Linguistik Online 28, 3/08, S. 133-155. Internet: http://www.linguistik-online.de/28_06/duffner.html (Stand: November 2008).
- Ďurčo, Peter (2008): Zum Konzept eines zweisprachigen Kollokationswörterbuchs. Prinzipien der Erstellung am Beispiel Deutsch ↔ Slowakisch. In: Lexicographica 24 (2008), Thementeil: Kollokationen in der europäischen Lexikographie und Wörterbuchforschung. Hrsg. von Franz Josef Hausmann. Tübingen, S. 69-89.
- Engelberg, Stefan (2007): Konstruktionelle Varianten zwischen Wörterbuch und Grammatik. In: Germanistische Mitteilungen 66, S. 11-27.

- Feilke, Helmuth (1998): Idiomaticische Prägung. In: Barz, Irmhild/Öhlschläger, Günther (Hg.): Zwischen Grammatik und Lexikon. (=Linguistische Arbeiten 390). Tübingen, S. 69-80.
- Feilke, Helmuth (2004): Kontext – Zeichen – Kompetenz. Wortverbindungen unter sprachtheoretischem Aspekt. In: Steyer (Hg.), S. 41-64.
- Fischer, Kerstin/Stefanowitsch, Anatol (Hg.) (2006): Konstruktionsgrammatik. Von der Anwendung zur Theorie. (= Stauffenburg Linguistik). Tübingen.
- Hausmann, Franz Josef (1985): Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: Bergenholtz, Henning/Mugdan, Joachim (Hg.): Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984. (= Lexicographica 3). Tübingen, S. 118-129
- Kämper, Heidrun/Eichinger, Ludwig M. (Hg.) (2007): Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache. (= Studien zur deutschen Sprache 40). Tübingen.
- Keibel, Holger/Belica, Cyril (2007): CCDB: A corpus-linguistic research and development workbench. In: Proceedings of the 4th Corpus Linguistics conference, Birmingham. Internet: http://corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf (Stand: Oktober 2008).
- Lemnitzer, Lothar/Zinsmeister, Heike (2006): Korpuslinguistik. Eine Einführung. (= Narr Studienbücher). Tübingen.
- Perkuhn, Rainer (2007): „Corpus-driven“: Systematische Auswertung automatisch ermittelter sprachlicher Muster. In: Kämper/Eichinger (Hg.), S. 465-491.
- Sailer, Manfred (2007): Corpus linguistic approaches with German corpora. In: Burger, Harald/Dobrovolskij, Dimitrij/Kühn, Peter/Norricks, Neal R. (Hg.) (2007): Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research. 2 Halbbde. (= HSK 28.1/2). Berlin/New York, S. 1060-1071.
- Sinclair, John (1991): Corpus, Concordance, Collocation. Oxford.
- Steyer, Kathrin (2000): Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten. In: Deutsche Sprache 28, 2, S. 101-125.
- Steyer, Kathrin (2003): Korpus, Statistik, Kookkurrenz. Lässt sich Idiomatics „berechnen“? In: Burger, Harald/Häcki Buhofer, Annelies/Gréciano, Gertrud (Hg.) (2003): Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifität der Phraseologie. (= Phraseologie und Parömiologie 14). Baltmannsweiler, S. 33-46.
- Steyer, Kathrin (2004): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: Steyer (Hg.), S. 87-116.
- Steyer, Kathrin (2008): Kollokationen in deutschen Wörterbüchern und in der deutschen Wörterbuchforschung. In: Lexicographica 24 (2008), Thementeil: Kollokationen in der europäischen Lexikographie und Wörterbuchforschung. Hrsg. von Franz Josef Hausmann. Tübingen, S. 185-207.
- Steyer, Kathrin (demn.): Zwischen theoretischer Modellierung und praxisnaher Anwendung. Zur korpusgesteuerten Beschreibung usueller Wortverbindungen. In: Mellado Blanco, Carmen (Hg.): Theorie und Praxis idiomaticer Wörterbücher. (=Lexicographica. Series Maior). Tübingen.
- Steyer, Kathrin (Hg.) (2004): Wortverbindungen – mehr oder weniger fest. Jahrbuch des Instituts für Deutsche Sprache 2003. Berlin/New York.
- Steyer, Kathrin/Brunner, Annelen (2008): Wortverbindungsfelder. Internet. <http://www.ids-mannheim.de/ll/uwv/wv-online/wv-online.html> (Stand: Januar 2009).
- Steyer, Kathrin/Lauer, Meike (2007): „Corpus-Driven“: Linguistische Interpretation von Kookkurrenzbeziehungen. In: Kämper/Eichinger (Hg.), S. 493-509.
- Stubbs, Michael (2001): Words and Phrases. Corpus Studies of Lexical Semantics. Oxford.
- Tognini-Bonelli, Elena (2001): Corpus Linguistics at Work. (= Studies in Corpus Linguistics 6). Amsterdam/Philadelphia.
- Tomasello, Michael (2006): Konstruktionsgrammatik und früher Erstspracherwerb (übersetzt von Stefanie Wulff und Arne Zeschel). In: Fischer/Stefanowitsch (Hg.), S. 19-37.

6.2 Internetseiten (Stand: Januar 2009)

CII: COSMAS II (<http://www.ids-mannheim.de/cosmas2/>)

KA-Tutorial: Eine kurze Einführung in die Kookkurrenzanalyse und syntagmatische Muster. Programmbereich Korpuslinguistik. (<http://www.ids-mannheim.de/kl/misc/tutorial.html>)

KL a: Projektseite 'Ausbau und Pflege der Korpora geschriebener Gegenwartssprache. Das Deutsche Referenzkorpus - DeReKo'. (<http://www.ids-mannheim.de/kl/projekte/korpora>)

KL b: Projektseite 'Methoden der Korpusanalyse- und erschließung. Korpuslinguistische Methodik und die Natur sprachlicher Daten'. (<http://www.ids-mannheim.de/kl/projekte/methoden/>)

UWV: Projektseite 'Usuelle Wortverbindungen' (<http://www.ids-mannheim.de/ll/uwv/>)

elexiko: Projektseite '*elexiko*: Ein Online-Wörterbuch zur deutschen Gegenwartssprache' (<http://www.ids-mannheim.de/ll/elexiko/>)

CPA: Projektseite 'Corpus Pattern Analysis' (<http://nlp.fi.muni.cz/projekty/cpa/>)