

Von der sprachlichen Oberfläche zum Muster: **Zur qualitativen Interpretation syntagmatischer Profile**

Kathrin STEYER

Institut für Deutsche Sprache, Mannheim

This paper discusses a corpus-driven approach to the study of multi-word expressions (MWE) (in our terminology: Usuelle Wortverbindungen UWV). Our approach is based on collocation data and syntagmatic profiles. Several interpretative steps lead from the language surface structure to MWE to more abstract multi-word patterns (MWP). MWP contain fixed components as well as slots, which are filled by elements with similar semantic or pragmatic characteristics. Like simple MWE, MWP can be considered units of the lexicon and patterns of language use with a holistic meaning and function. The formation of patterns, the semantic and pragmatic characteristics of the fillers and the restrictions on usage cannot be described by rules or language competence alone, but require bottom-up analysis on the basis of very large corpora.

1. Einleitung

Der Beitrag zeigt am Beispiel 'Wortverbindungen', inwieweit induktive, datengeleitete Analyseverfahren zur Rekonstruktion verfestigter Sprachgebrauchsmuster beitragen können bzw. eine solche Rekonstruktion überhaupt erst ermöglichen. Als empirische Grundlage dient das Deutsche Referenzkorpus *DeReKo*. Es ist "mit über 4,3 Milliarden Wörtern (Stand 27.09.2011) die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit." (vgl. *DeReKo*).

Der Gegenstandsbereich 'Wortverbindungen' ist nicht zufällig gewählt, sondern resultiert aus der Überzeugung, dass usualisierte Mehrwort-einheiten das eigentliche Gerüst des Sprachgebrauchs darstellen. Wortverbindungen können mehrgliedrige Fragmente jeglicher Art sein, so wie sie uns im Sprachgebrauch entgegentreten: ungeordnet, oft unvollständig, auch idiosynkratisch. Es spielt zunächst keine Rolle, ob diese Fragmente binäre oder mehrgliedrige Strukturen aufweisen, ob sie eine idiomatische oder eine nicht-idiomatische Bedeutung haben, ob sie Nominal- oder Verbalphrasen sind oder ob sie als Sprichwort oder Kollokation qualifiziert werden können. Diese kategorialen Einordnungen sind A-Posteriori-Zuschreibungen, die u. a. abhängig von der jeweils zu Grunde liegenden Theorie oder Schule sind. Der Leitgedanke ist vielmehr, dass es sich bei diesen Fragmenten immer um verfestigte Muster des Sprachgebrauchs

handelt. Es gibt keine besseren oder schlechteren Kandidaten. Es gibt keinen Kern und keine Peripherie. Der Unterschied liegt nur im Grad der vordergründigen Auffälligkeit für den Betrachter. Damit ist gemeint, dass das Augenmerk in der Phraseologie und Mehrwortforschung in der Vergangenheit vor allem auf solchen Mehrwortklassen wie 'Idiom', 'Sprichwort' oder ganz allgemein auf phraseologisch gebundenen Einheiten lag, weil sie z. B. durch ihre Nicht-Kompositionalität oder transformationelle Defektivität ins Auge sprangen. Die meisten mehr oder weniger festen Wortverbindungen weisen jedoch keine idiosynkratischen Merkmale auf, sondern sind allein durch rekurrenten Gebrauch zu sprachlichen Halbfertigprodukten¹ geworden. Wortfolgen unterliegen Verfestigungsmechanismen, weil sie als Kondensat komplexe Bedeutungen und kommunikative Funktionen auf besonders sprachökonomische Weise realisieren können.

Im Folgenden wird dargelegt, welche Perspektiven eine primär induktive Vorgehensweise für die Rekonstruktion und Beschreibung musterhaften Sprachgebrauchs, wie er sich in usuellen Wortverbindungen (UWV) manifestiert, eröffnet. Die Beispiele stammen überwiegend aus Feldern von 'Körperteilbezeichnungen', die dem UWV-Analysemodell als experimentelle Beschreibungsobjekte zugrunde gelegt wurden (vgl. Steyer & Brunner 2008-2013). Der hier vorgestellte Ansatz weist Schnittstellen zu Forschungen im Bereich der *Corpus Pattern Analysis*, der *Pattern Grammar* und der *Konstruktionsgrammatik*² v. a. in Hinblick auf die Annahme eines Form-Funktionskontinuums auf. Darauf kann an dieser Stelle jedoch nicht weiter eingegangen werden (vgl. Steyer in Vorb.).

2. Eine Anmerkung zu korpusmethodischen Zugängen

Es soll nun kurz auf die Frage eingegangen werden, wie korpusbasierte Zugänge zu klassifizieren sind und welchen jeweiligen Erkenntniswert sie haben. Korpusanalyse kann zum einen zur Überprüfung von *bekanntem Wissen* dienen, d.h. dass ein "Mehr" an empirischer Beweiskraft durch eine Vielzahl von Korpusdaten hinzukommt und dieses Wissen ausdifferenziert bzw. modifiziert wird, aber auch in Teilen verworfen werden kann. Dieser Zugang dominiert immer noch in weiten Teilen der Linguistik. Das ist im Prinzip nicht zu kritisieren. Denn auch eine empirisch abgesicherte Validierung tradierter linguistischer Theorien und Modelle auf der Basis von

¹ Die Bezeichnung "sprachliche Halbfertigprodukte" bezog Hausmann ursprünglich auf Kollokationen (1985: 118). Im Kontext unseres Modells können quasi alle mehr oder weniger festen Wortverbindungen als solche aufgefasst werden.

² Vgl. z. B. Hunston & Francis (2000); Fischer & Stefanowitsch (2006) bzw. Stefanowitsch & Fischer (2008) sowie Engelberg & Holler & Proost (2011). In den genannten Arbeiten finden sich auch ausführliche Darstellungen und Diskussionen aller Strömungen der *Konstruktionsgrammatik* bzw. *Construction Grammar*.

sprachlichen Massendaten stellte lange Zeit ein Desiderat dar. Korpusanalyse kann zum anderen aber auch *neues Wissen* generieren. Neues Wissen zu generieren bedeutet, auf der Basis riesiger Datenmengen auf lange Sicht Verallgemeinerungen für die Gesamtheit vorzunehmen, um zeigen zu können, dass Sprache in bestimmten Bereichen doch etwas anders funktioniert als über die tradierten, oft aber starren bzw. kleinteiligen Regelapparate der Grammatik. Korpusempirisch können diese Phänomene über die Analyse vieler ähnlicher Verwendungsfälle rekonstruiert werden. Und das bedeutet eben nicht, das zu beschreiben, was man ohnehin erwartet, sondern das, was den menschlichen Wahrnehmungshorizont eigentlich übersteigt, was sich unserem analytischen Blick zunächst verschließt. Dabei geht es nicht nur um ein Mehr an Daten, sondern es sind andere Vernetzungen, ungewöhnliche Querverbindungen, überraschende Verbindungen zwischen eigentlich als nicht als zusammengehörig angesehenen Phänomenen. Eine Grundsatzfrage scheint dabei immer durch: Wie viel A-Priori-Wissen lässt man zu? Wie ausgeprägt ist der Einfluss der Korpusdaten auf den Analyseprozess? Fast alle Vorschläge zur Klassifikation von Korpusparadigmen oder – besser gesagt – zu "korpuslinguistischen Philosophien" orientieren sich an diesen Fragen: So wird die Unterscheidung zwischen 'corpus-based' vs. 'corpus-driven' bereits bei Sinclair (1991) und Tognini-Bonelli (2001) eingeführt und diskutiert. Auch wir haben Vorschläge unterbreitet, z. B. mit der Dichotomie 'Konsultationsparadigma' (das Korpus befragen) und 'Analyseparadigma' (das Korpus analysieren) (vgl. Steyer 2004, Belica & Steyer 2008) oder 'korpusbasiert' vs. 'korpusgesteuert' (vgl. u.a. Steyer & Brunner 2009)³. Die scharfe Trennung unterschiedlicher korpusanalytischer Herangehensweisen war in der Anfangsphase in dieser Radikalität sicherlich sinnvoll. Aber die Erfahrungen der empirischen Praxis haben mittlerweile gezeigt, dass es eher um ein graduelles Phänomen mit fließenden Grenzen geht, um einen skalaren Einfluss korpusanalytischer Befunde auf die linguistische Interpretation: von der reinen Suche nach Belegen als Beweis oder Widerlegung von A-Priori-Annahmen über eine Zugrundelegung von Korpusdaten als empirische Ausgangsbasis bis hin zu einer hart an den Daten orientierten induktiven Vorgehensweise.

Unabhängig vom gewählten Methodendesign und dem Anteil der zugrunde gelegten Hypothesen ist die eigentliche Interpretation nach wie vor vom Linguisten zu leisten. Er muss die Begründungen dafür finden, warum bestimmte Zusammenhänge durch automatische Berechnungen als evident hervorgetreten sein könnten. Die 'Techniken' qualitativer Interpretation unterscheiden sich also nicht von den entsprechenden Phasen

³ Vgl. auch Perkuhn (2007), Sailer (2007) und Lemnitzer/Zinsmeister (2010). Ausführlich und auch kritisch gehen Bubenhofer (2009) und Duffner (2010) auf die unterschiedlichen Paradigmen der Korpusbasiertheit ein.

herkömmlicher empirischer Untersuchungen. Diese sind durch die korpus-linguistische Methodik keineswegs obsolet geworden. Es geht immer um:

- Beobachtung der Sprachdaten
- Betrachtung aller Evidenzen
- Bildung von Hypothesen auf Basis der Evidenzen
- empirisches Prüfen der Hypothesen (Zusammenspiel von Induktion und Deduktion)
- deskriptive Aussagen, die die Evidenzen reflektieren
- Generalisierung im Sinne von Gebrauchsregeln.

3. Usuelle Wortverbindungen – Kurzdefinition und Typen

Voranstellen möchte ich ein Zitat von Tomasello zur allmählichen Verfestigung von sprachlichen Schemata über massenhaften Gebrauch:

"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: 21) (Übersetzung von Wulff & Zeschel in Fischer & Stefanowitsch, 2006)

Usuelle Wortverbindungen (UWV) verkörpern solche sprachlichen Verwendungsmuster par excellence und werden wie folgt definiert: Sie sind mehrgliedrige – zumindest binäre – Wortschatzeinheiten, die in rekurrenten, also häufig wiederkehrenden, syntagmatischen Strukturen vorkommen. UWV müssen minimal lexikalisch spezifiziert sein. Sie fungieren als autonome kommunikative Einheiten, denen eine übersummativ Bedeutung und/oder Funktion zugeschrieben werden kann. Ihre holistische Qualität kann sich auch auf höheren Abstraktionsebenen der zugrunde liegenden Schemata zeigen. UWV unterliegen immer spezifischen Gebrauchsrestriktionen, die sich, so Deppermann (2006: 48), sowohl auf die semantischen Klassen möglicher "lexikalischer Instanzierungen" als auch auf die speziellen, eingeschränkten "syntaktischen oder pragmatischen Kontextbedingungen" beziehen können. In Bezug auf den Grad der lexikalischen Spezifiziertheit wird zwischen 'Wortverbindung' (WV) und 'Wortverbindungsmuster' (WVM) unterschieden. WV weisen einen festen lexikalischen Komponentenbestand mit einer übersummativen Bedeutung und Funktion auf und können als unikale lexikalisierte Wortschatzeinheiten gelten, z. B.:

- (1) *auf Grund laufen*
= 'havariieren'
im Grunde
= 'eigentlich'
Pi mal Daumen
= 'ungefähr|ca.'
hinter verschlossenen Türen
= 'geheim', 'nicht öffentlich'
Der Schein trügt.
= 'etwas ist eigentlich anders, als es auf den ersten Blick wirkt'

Unter WVM werden lexikalisch teilspezifizierte Mehrworteinheiten verstanden, die feste lexikalische Komponenten einerseits und variable Slots andererseits enthalten. Die Besetzung der Slots (Leerstellen) ist durch spezifische Gebrauchsbedingungen beeinflusst, d. h. dass die entsprechenden lexikalischen Füller verwandte semantische und/oder pragmatische Merkmale aufweisen müssen (vgl. Feilke 1996, Steyer in Vorb.). Die Spezifik der jeweiligen Füllergruppen ist nicht allein und nicht immer mit morphosyntaktischen Klassen oder abstrakten semantischen Merkmalen erfassbar, sondern kann aus allen Aspekten usuellen Gebrauchs resultieren. Beispiele für WVM sind:

- (2) $\langle X_{\text{SUB}} \textit{genug} \rangle$
X-Füller: *Grund / Anlass / Motivation ...*
- (3) $\langle \textit{allen} X_{[\text{ADJ}]_{\text{SUB}}} \textit{zum Trotz} \rangle$
X-Füller: *Unkenrufen / Beteuerungen / Warnungen / Dementis / Protesten / anderslautenden Gerüchten / düsteren Prognosen / vollmundigen Erklärungen ...*

Die jeweiligen X-Füller haben nicht nur die morphosyntaktische Eigenschaft, Substantive (SUB) zu sein – in (3) fakultativ durch Adjektive (ADJ) ergänzt, sondern weisen darüber hinaus weitere gemeinsame Merkmale auf: in (2) Ausdruck von Kausalität; in (3) von kommunikativen Handlungen, die häufig konnotativ aufgeladen sind. Auch WVM können übersummativ Bedeutungen und/oder Funktionen zugeschrieben werden: in (2) 'etwas ist hinreichend'; in (3) 'obwohl etwas so schien, ist es doch nicht eingetroffen'.

Aus induktiver Sicht stellen WVM Verallgemeinerungen über WV ähnlicher Art dar, wobei der "Verwandtschaftsgrad" der zu einem WVM gehörenden WV unterschiedlich eng oder weit sein kann: So weisen die WV *Der Ton macht die Musik – Der Unterton macht die Musik – Der Zwischenton macht die Musik* nur minimale Bedeutungsunterschiede auf und konstituieren gemeinsam das monoseme Muster $\langle \textit{Der} \text{SUB}_{\text{Ton-KOMPOSITA}} \textit{macht die Musik} \rangle$ in der Bedeutung 'Es kommt nicht nur darauf an, was man sagt, sondern auf welche Art und Weise man es tut' (vgl. Steyer & Hein 2010). Das WVM $\langle \textit{in} \text{ADJ}_{\text{Zustand_positiv}} \textit{Bahnen} \rangle$ 'etw. befindet bzw. vollzieht sich in einem bestimmten Zustand/abgesteckten Rahmen' stellt dagegen eine Abstraktion vieler WV dar, die sich ihrerseits zumindest in Teilbedeutungen unter-

scheiden: *in geordneten Bahnen – in sicheren Bahnen – in normalen Bahnen – in geregelten Bahnen.*

Die Regularitäten der Slotbesetzungen sind kompetenzbasiert kaum bestimmbar, sondern können nur bottom up über die Analyse rekurrenter Verwendungen in Korpora rekonstruiert werden. Einen heuristischen Zugang stellt das induktive UWV-Analysemodell (vgl. Steyer & Brunner 2009) dar, das im Folgenden vorgestellt wird.

4. Das UWV-Analysemodell

Das Ziel dieses korpugesteuerten Vorgehens ist die Rekonstruktion verfestigter Syntagmen, ihrer invarianten Kerne, der varianten Realisierungen, der zugrunde liegenden Muster und Schemata und der Restringertheit ihrer Slotbesetzungen. In unserem Analysemodell gehen wir – bottom up – von den konkreten sprachlichen Oberflächen rekurrenter Syntagmen aus und nehmen sukzessive Generalisierungen bis zu Mustern höherer Ordnung vor. Den empirischen Ausgangspunkt bilden so genannte statistisch berechnete syntagmatische Profile sprachlicher Objekte, in unserem Falle berechnet durch die Kookkurrenzanalyse (KA) (Belica 1995). Die Vorgehensweise ist iterativ, d. h. wir werten Ergebnisse statistischer Berechnungen aus. Auf dieser Basis formulieren wir Hypothesen, z. B. ob ein Syntagma eine WV indiziert oder nicht. Anschließend setzen wir neue Suchprozeduren und Berechnungen im Korpus ein, um mit verfeinerten Methoden differenziertere Ergebnisse zu erhalten. Ein wichtiges Prinzip unseres Modells ist der Fokus auf das syntagmatische Verhalten von Wortformen. An der sprachlichen Oberfläche anzusetzen heißt, sich auf der Ebene des konkreten Ausdrucks, der tatsächlichen lexikalischen Realisierungen im Text zu bewegen. Für die Aufwertung von Wortformen als eigene lexikalische Einheiten hat schon Sinclair argumentiert:

"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: 8)

Bereits ein Vergleich der oberen statistischen Ränge in den Kookkurrenzprofilen von *Ohr* und *Ohren* zeigt gewisse Unterschiede⁴:

© Institut für Deutsche Sprache, Mannheim
COSMAS II-Server, C2API-Version 4.1.14 - 22. Nov. 2011

Datum : Mittwoch, den 7. Dezember 2011, 13:56:38
Korpus : W-gesamt - alle Korpora des Archivs W (mit Neuakquisitionen)
Archiv-Release: Deutsches Referenzkorpus (DeReKo-2011-I)
Suchanfrage : *Ohr*
Ergebnis : 10.000 Treffer, zufällig reduziert aus 61.609

⁴ Eine Erläuterung zu Kookkurrenzprofilen findet man im entsprechenden Tutorial (Perkuhn & Belica, 2004).

colloc version 3.10 20100718

Kookkurrenzen

Anzahl Kookkurrenzen : 182
 Analyse-Kontext : 5. Wort links bis 5. Wort rechts
 Granularität : grob
 Zuverlässigkeit : hoch
 Clusterzuordnung : eindeutig
 Auf 1 Satz beschränkt : ja
 Lemmatisierung : nein
 Funktionswörter : ignorieren
 Autofokus : aus
 LLR anzeigen : ein
 Nummerierung : ein

Total Anzahl LLR Kookkurrenzen syntagmatische Muster

1 1443 1443 18661 **offenes** 99% immer ein **offenes** [...] **Ohr** für die
 2 1811 368 2387 **Auge** 75% für **Auge** [und] **Ohr**
 3 1992 181 2155 **gehauen** 100% übers **Ohr** **gehauen**
 [...]
 5 2392 155 1483 **Knopf** 98% mit ... **Knopf** im **Ohr**
 [...]
 7 2697 190 1280 **Handy** 96% mit das|dem **Handy** am **Ohr**
 8 2821 124 1241 **hauen** 99% übers **Ohr** [zu] **hauen**
 9 2852 31 855 **stets** 87% **stets** ein ... **Ohr** für ...
 [...]
 11 3010 142 744 **immer** 69% **immer** [...] im] **Ohr**
 12 3136 126 697 **linken** 96% dem|im|am **linken** [...] **Ohr**
 13 3200 64 686 **flüsterte** 65% und **flüsterte** ihm|ihr etwas ins **Ohr**
 14 3258 58 633 **Offenes** 100% **Offenes** **Ohr** für ...
 15 3307 49 628 **geflüstert** 100% ins **Ohr** [...] **geflüstert**
 16 3376 69 596 **Aug** 100% für **Aug** [und] **Ohr**
 17 3427 51 596 **flüstern** 92% ins **Ohr** [zu] **flüstern**

Fig. 1: Ausschnitt aus KA der Wortform *Ohr* (obere statistische Ränge)

© Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.1.14 - 22. Nov. 2011

Datum : Mittwoch, den 7. Dezember 2011, 13:53:44
 Korpus : W-gesamt - alle Korpora des Archivs W (mit Neuakquisitionen)
 Archiv-Release: Deutsches Referenzkorpus (DeReKo-2011-I)
 Suchanfrage : *Ohren*
 Ergebnis : 10.000 Treffer, zufällig reduziert aus 80.851

colloc version 3.10 20100718

Kookkurrenzen

Anzahl Kookkurrenzen : 178
 Analyse-Kontext : 5. Wort links bis 5. Wort rechts
 Granularität : grob
 Zuverlässigkeit : hoch
 Clusterzuordnung : eindeutig
 Auf 1 Satz beschränkt : ja
 Lemmatisierung : nein
 Funktionswörter : ignorieren
 Autofokus : aus
 LLR anzeigen : ein

Nummerierung : ein

Total Anzahl LLR Kookkurrenzen syntagmatische Muster
 1 514 514 7491 **taube** 100% auf **taube** [...] **Ohren**
 2 1364 850 5693 **Augen** 82% **Augen** [und] **Ohren**
 3 1975 611 5440 **offene** 99% auf **offene** [...] **Ohren**
 4 2058 83 1200 **faustdick** 100% hat es **faustdick** hinter den **Ohren**
 5 2161 103 1041 **spitzen** 69% die **Ohren** [...] **spitzen**
 6 2258 97 1032 **gehauen** 98% um die **Ohren** **gehauen**
 7 2260 2 1029 **gestoßen** 50% **Ohren** **gestoßen**
 8 2393 133 997 **klingen** 75% in den **Ohren** [...] **klingen**
 9 2480 87 831 **hauen** 97% um die **Ohren** [zu] **hauen**
 10 2719 239 820 **gekommen** 100% zu **Ohren** **gekommen**
 11 2724 5 804 **stoßen** 60% auf ... **Ohren** [...] **stoßen**

Fig. 2: Ausschnitt aus dem Kookkurrenzprofil von *Ohren* (obere statistische Ränge)

Kookkurrenzlisten ermöglichen eine schnelle und überblicksmäßige Rezeption einer großen Menge unterschiedlich lokalisierter Daten. Sie sind als assoziatives heuristisches Hilfsmittel unentbehrlich. Durch eine solche intelligente Vorstrukturierung sprachlicher Massendaten kann der Linguist auf einer höheren analytischen Ebene ansetzen. Aber letztlich wird kein Linguist z. B. allein auf dieser Basis berechneter Kookkurrenzen Generalisierungen in Bezug auf das Kollokationsverhalten von Substantiven vornehmen. Kein Lexikograf wird nur mit Hilfe dieser Listen oder Netze ein Bedeutungswörterbuch fester Wortverbindungen erarbeiten. Er braucht ab einem bestimmten Interpretationsschritt immer die usuellen, authentischen kotextuellen Umgebungen eines Analyseobjekts für den kontextuellen Hintergrund. An der grundlegenden Kulturtechnik, nämlich Texte zu lesen und zu interpretieren, hat denn auch die Korpuslinguistik u. E. nichts geändert.

In unserem Falle bildet die Analyse der Konkordanzen der jeweiligen Kookkurrenzcluster (KWIC⁵) den primären Zugang zur Ermittlung rekurrenter kotextueller Einbettungen. KWIC dienen der Identifikation von häufig wiederkehrenden identischen oder ähnlichen syntagmatischen Strukturen in der unmittelbaren Umgebung eines sprachlichen Suchobjekts, der Bestimmung von invarianten Kernen und typischen Abwandlungen. Die folgende Abbildung zeigt, auf welche Art und Weise verwandte Vorkommen durch die KA bereits in einem Cluster zusammengefasst werden können:

NUN10/JUN.02578 Auf dass den Deutschen die **Ohren** klingen.

A97/JUN.11880 Kantonesischen in romantischen Pop-Balladen aus, die für westliche **Ohren** meist kitschig klingen.

A08/MAR.02078 Endlich seien Kinder-Lieder auf dem Markt, die auch in den **Ohren** Erwachsener wohl klingen.

⁵ Key Words In Context

B99/OKT.92922 Das sind Sätze in einer Tonlage, die für manche polnischen **Ohren** ungewohnt klingen.

B00/FEB.13599 lauter Dinge, die gut in den **Ohren** der Aktionäre klingen sollten:

B03/MAR.17614 Den Spielern klingen noch heute die **Ohren**. "Dat wird ne janz enge Kiste"

B05/FEB.09147 empfangen, aber auch mit deutlichen Worten, die in deutschen **Ohren** leicht harsch klingen.

BRZ09/APR.11918 Worte, die wie Musik in unseren **Ohren** klingen.

DPA06/MAI.14501 So ein Slogan muss bei aller Griffigkeit - denen wie Hohn in den **Ohren** klingen, die angesichts radikaler Übergriffe große Angst haben,

F01/105.25991 Träger dürfte noch lange in den **Ohren** klingen, was sein Knecht nun glaubt, dem Bundeskanzler vorwerfen zu müssen

Fig. 3: KWIC-Cluster: *Ohren – klingen*

Auch in solch automatisch vorstrukturierten Clustern sind noch sehr verschiedene und heterogene Informationen enthalten. So lassen sich in diesem Ausschnitt bereits drei unterschiedliche WV bzw. WVM identifizieren: a. <in X *Ohren klingen*>; b. <für X *Ohren klingen*> und c. jmdm. *klingen die Ohren* (a. 'besonders genau auffallen; nachhaltig behalten; aufmerken'; b. 'in jmds. Wahrnehmung' c. 'etw. wurde jmdm. – evtl. auf unangenehme Weise – sehr deutlich gemacht'). Des Weiteren sind bereits etliche kotextuelle Indikatoren erkennbar, deren Typikalität zu überprüfen ist, z. B.:

- Genitiv-, Dativ- und Akkusativergänzungen, die die Person spezifizieren (*den Deutschen / Erwachsener / der Aktionäre / Klinsmann*)
- nationale Zuschreibungen (z. B. *polnischen / deutschen*)
- Objektergänzungen (z. B. *Kinderlieder / Worte / lauter Dinge*)
- Bewertungen (z. B. *gut / leicht harsch / wie Hohn / wie Musik*)
- temporale Kotextmarker als Ausdruck der nach wie vor bestehenden Geltung (z. B. *noch heute / noch lange*)

Um die spezifischen Verwendungstypen und ihre jeweiligen Gebrauchsbesonderheiten zu erkennen, zu ordnen und zu systematisieren, ist also eine weitere inhaltliche Strukturierung der einzelnen Cluster notwendig. Für diese Zwecke haben wir ein oberflächenbasiertes Vorgehensmodell entwickelt, das eine sukzessive Generalisierung von der sprachlichen Oberfläche bis hin zu Mustern höherer Ordnung ermöglicht. Ein entsprechendes Tool (UWV-Tool) unterstützt diese analytischen Schritte (Programmierung und Implementierung: Annelen Brunner)⁶:

⁶ Die über COSMAS II berechneten Kookkurrenzprofile werden dazu zunächst in ein XML-Format exportiert und dann zur weiteren Bearbeitung in unsere Arbeitsumgebung importiert.

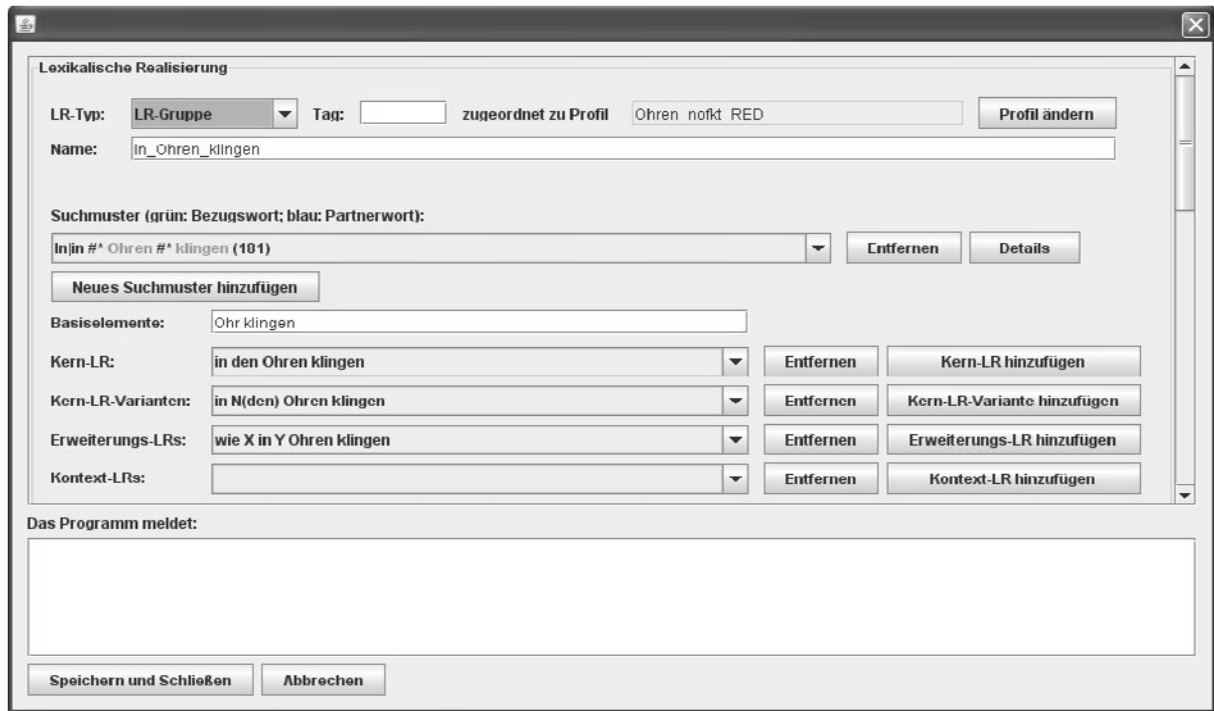


Fig. 4: Suchmaske im UWV-Tool

5. Qualitative Interpretation syntagmatischer Profile

Die qualitative Interpretation syntagmatischer Profile erfolgt in verschiedenen Stufen:

- i. profilinterne Analysen
 - a. Analyse eines KWIC-Clusters
 - b. Analyse des gesamten Wortformenprofils
- ii. profilvergleichende Analysen
 - a. Analyse verschiedener Wortformenprofile eines Lemmas
 - b. Analyse externer Vernetzungen von syntagmatischen Profilen unterschiedlicher Objekte (Wörter oder Syntagmen)

5.1. Profilinterne Analyse

Zunächst beschäftigen wir uns mit auffälligen Gebrauchspänomenen innerhalb eines syntagmatischen Profils ausgehend von der inhaltlichen Auswertung rekurrenter Phänomene innerhalb eines Clusters. Folgende Fragen interessieren uns primär:

- Welcher Grad an syntaktischer und lexikalischer Verfestigung lässt sich bei rekurrenten Syntagmen rekonstruieren?
- Wie baut sich die Hierarchie lexikalischer Realisierungen auf (invariante Kerne, typische Variationen und lexikalische Erweiterungen)?
- Welcher Natur sind die ermittelten Füllergruppen eines Slots?

- Sind rekurrente Syntagmen autonome Wortverbindungen mit eigenständigen Bedeutungen?
- Ist eine Wortverbindung ein einmaliges, also unikales Exemplar oder verkörpert sie eine typische Realisierung eines abstrakteren Schemas oder Musters?
- Welche kommunikativen Funktionen werden auf welcher Abstraktionsebene erfüllt?

Der erste analytische Schritt ist die Neusortierung der KWIC eines Clusters auf der Basis von Suchmustern (SUM), die auf Hypothesen zur Festigkeit und Varianz basieren. Dadurch lassen sich bereits einige wichtige Hinweise gewinnen, welche syntagmatische Form welchen Anteil am Gesamt-vorkommen hat und ob es variable Slots gibt oder nicht.

120 Treffer für "in den Ohren #* klingen" (von 287; 41,81%)
 48 Treffer für "In|in N(den) Ohren #* klingen" (von 287; 16,72%)
 35 Treffer für "Für|für #* # Ohren #* klingen" (von 287; 12,20%)
 29 Treffer für "die Ohren #* klingen" (von 287; 10,10%)
 20 Treffer für "klingen #* in den Ohren" (von 287; 6,97%)
 9 Treffer für "klingen #* die Ohren" (von 287; 3,14%)
 8 Treffer für "klingen #* in N(den) Ohren" (von 287; 2,79%)

 21 RestKWIC (von 287; 7,32%)

Fig. 5: Prozentuale Abdeckung einzelner Realisierungen im Cluster *Ohren - klingen*⁷

Diese Übersicht verdeutlicht beispielsweise, dass die lexikalischen Realisierungen *in den Ohren klingen* bzw. *klingen in den Ohren* in ca. 50% aller Vorkommen des Clusters nachzuweisen ist. Des Weiteren kristallisieren sich noch andere Gruppen heraus, z. B. durch Ausschluss des Artikels *den* (*in ... Ohren klingen*: 48 Treffer; ca. 17%) oder durch die Ersetzung der Präposition *in* durch *für* (*für ... Ohren klingen*: 35 Treffer; ca. 12%) usw. Auf dieser Basis lässt sich dann eine Hierarchie lexikalischer Realisierungen für das WVM $\langle in X Ohren klingen \rangle$ ansetzen:

(4) Typische Realisierungsform: *in den Ohren klingen*

Lexikalische Varianten: $\langle in X Ohren klingen \rangle$

X1 = Possesivpronomina (*seinen / ihren*)

X2 = Adjektive, die geografische bzw. nationale Zugehörigkeiten ausdrücken (*europäischen / chinesischen / palästinensischen*)

X3 = Personennamen (*Magaths / Barrichellos / ...*)

Typische lexikalische Erweiterungen:

wie Musik in den Ohren klingen

wie Hohn in den Ohren klingen

Auch die Slotbesetzungen ermitteln wir nicht kompetenzbasiert, sondern mit Hilfe von Suchmustern, indem wir feste lexikalische Komponenten und

⁷ Der Platzhalter # bedeutet genau eine Leerstelle (# # genau zwei usw.), der Platzhalter #* beliebig viele.

Leerstellen definieren. Das UWV-Tool liefert uns beispielsweise folgendes Resultat für das Suchmuster *in* bzw. *In # Ohren klingen* (aber ohne den Artikel *den*):

| Feld 1 | Feld 2 | Feld 3 | Feld 4 | Feld 5 | Feld 6 | Feld 7 | Clusters |
|----------------------------|--------|------------|--------|----------------------|---------|--------------------------------|-----------------|
| Ubuntu! Eine solche ... | in | westlichen | Ohren | komisch | klingen | . Doch keine Angst, sie ... | [klingen] |
| bestraft worden. Die ... | in | westlichen | Ohren | hohl und feindse... | klingen | . Auf der arabischen Ha... | [klingen] |
| daß ihre Behauptung... | in | westlichen | Ohren | glaubwürdiger | klingen | . Die Ukraine wird wahr... | [klingen] |
| ein, dass dies als Be... | in | westlichen | Ohren | seltsam | klingen | mag. "Aber zum ersten ... | [klingen] |
| bewährt sich auch im... | in | vielen | Ohren | | klingen | mag: Holz ist als Baust... | [klingen] |
| sozial verantwortliche... | in | vielen | Ohren | | klingen | mag, so wenig können... | [klingen] |
| wenig spektakulär zu ... | in | vielen | Ohren | banal | klingen | . Und doch ist es gerad... | [klingen] |
| "Menschen für ander... | in | vielen | Ohren | nach Elite | klingen | , nach Bedürfnissen ein... | [klingen] |
| auch noch Hafentsch... | in | vielen | Ohren | | klingen | . Doch die 2000 Mark s... | [klingen H...] |
| ihre Bedeutung bewa... | in | unsern | Ohren | | klingen | . Der ehemalige Amris... | [klingen] |
| trinken, ihr seidiges F... | in | unseren | Ohren | | klingen | lassen, wenn wir jetzt d... | [klingen, l...] |
| Märchen entfalten ihr... | in | unseren | Ohren | anders | klingen | als das Gewohnte. Bild... | [klingen] |
| fuhr im Stalom mit -A... | in | Stolz' | Ohren | der Beifall der T... | klingen | , als er mit leeren Händ... | [klingen H...] |
| - aus moralischen Gr... | in | serbischen | Ohren | wie eine Provok... | klingen | muß. Zudem fragt sich... | [klingen] |
| Pen auf dem Kongre... | in | seinen | Ohren | wie Hohn | klingen | muß; durchsichtiges Zi... | [klingen H...] |
| Geringeres als die Ze... | in | seinen | Ohren | zeitlos | klingen | mag, ist bestenfalls ge... | [klingen] |
| Nachrichten?> Dass ... | in | Schweizer | Ohren | seltsam | klingen | , lässt uns im Land des... | [klingen] |
| Namen wie Bogujki, ... | in | russischen | Ohren | lustig | klingen | , durch die liebeliche klei... | [klingen] |
| kann sich wirklich nic... | in | polnischen | Ohren | | klingen | . Mit Steinbach ist ein V... | [klingen] |

Fig. 6: KWIC-Sortierung nach Leerstellen in Feld 3 (alphabetisch) im UWV-Tool

Die in den Slots vorkommenden lexikalischen Füller können auch automatisch gezählt werden. Diesen Füllertabellen kann man wichtige Hinweise auf die Natur der lexikalischen Füller und die Skala ihrer Typikalität bzw. Produktivität – und damit letztlich auf den Grad der Festigkeit einer WV entnehmen.

Indikatoren, die wir in einem Cluster finden, werden in einem zweiten Analyseschritt im gesamten Wortformenprofil überprüft. Die folgenden drei Füllertabellen sind auf der Basis der Auswertung aller KWIC-Vorkommen des *Ohren*-Profils erstellt worden und verdeutlichen drei unterschiedliche Fälle skalarer Typikalität⁸:

| Füller | Häufigkeit |
|----------------|------------|
| Musik | 62 |
| Hohn | 28 |
| es | 2 |
| Erpressung | 1 |
| Befehlston | 1 |
| Zukunftsmusik | 1 |
| Drachengebrüll | 1 |

⁸ Diese Lückenfüllertabellen zeigen nur Ausschnitte. Die Auslassungen werden mit [...] gekennzeichnet.

| | |
|--------------|---|
| Drohungen | 1 |
| Spott | 1 |
| Nadelstiche | 1 |
| Kauderwelsch | 1 |
| Melodie | 1 |
| Öl | 1 |
| Glocken | 1 |
| Teufelswerk | 1 |
| symphonia | 1 |
| angenehm " | 1 |
| Unterdruck | 1 |
| Geläut | 1 |

Fig. 7: Automatische Füllertabelle SUM: wie|Wie # in den Ohren (gesamtes *Ohren*-Profil)

| Füller | Häufigkeit |
|-------------------|------------|
| die | 330 |
| unsere | 52 |
| deutsche | 46 |
| westliche | 33 |
| heutige | 32 |
| europäische | 30 |
| Augen, | 20 |
| ihre | 12 |
| offene | 12 |
| mitteleuropäische | 10 |
| manche | 10 |
| meine | 9 |
| amerikanische | 9 |
| viele | 7 |
| ungeübte | 7 |
| verwöhnte | 6 |
| Schweizer | 6 |
| fremde | 5 |
| empfindliche | 5 |
| hiesige | 4 |
| feine | 4 |
| [...] | |
| übersättigte | 1 |
| moderne" | 1 |
| Bensheimer | 1 |
| lärmempfindliche | 1 |
| deine | 1 |
| [...] | |

Fig. 8: Ausschnitt aus der Füllertabelle SUM: für # *Ohren* (gesamtes *Ohren*-Profil)

| Füller | Häufigkeit |
|-----------------|------------|
| | 769 |
| den | 108 |
| viel um die | 9 |
| grün hinter den | 7 |

| | |
|------------------------------------|---|
| beiden | 4 |
| und | 4 |
| sehen und | 3 |
| ihren | 3 |
| mit tausend | 3 |
| laut in den | 3 |
| großen | 3 |
| merken, weil die auch die größeren | 2 |
| den abstehenden | 2 |

Fig. 9: Ausschnitt aus der Füllertabelle SUM: *zu #* Ohren* (gesamtes *Ohren*-Profil)

Die Slotbesetzungen in Fig.7 (*wie/Wie # in den Ohren*) zeugen von der Dominanz zweier lexikalischer Füller, nämlich *Musik* und *Hohn*, auf der einen Seite und der Existenz zahlreicher so genannter Einmalfüller auf der anderen Seite, z. B. *Befehlston*, *Drachengebrüll* oder *Zukunftsmusik*. Diese Einmalfüller sind okkasionelle Bildungen, tragen aber ebenso zur Konstitution des WVM *<wie X in den Ohren>* bei. Die Tabelle in Fig.8 (*für # Ohren*) zeigt, dass es neben dem häufigsten lexikalischen Füller, dem Artikel *die*, eine ganze Gruppe von zahlreichen vergleichbaren Füllern des WVM *<für X Ohren>* gibt. Je nach Textbasis oder auch aktuellen thematisch-diskursiven Trends können die einzelnen lexikalischen Füller austauschbar sein. Die dritte Füllertabelle (*zu #* Ohren*) in Fig.9 stellt schließlich ein Beispiel für hochgradige Geprägtheit dar, da die Leerstelle zwischen den Komponenten *zu* und *Ohren* am häufigsten überhaupt nicht besetzt ist (WV: *zu Ohren kommen*). Die anderen Füller sind entweder ganz reguläre Realisierungen oder gehören zu anderen WV.

Die Analyse solcher Profile kann auf viele Aspekte der Phraseologie ein neues Licht werfen. Eine Erkenntnis ist beispielsweise, dass WV bzw. WVM selten in ihrer Reinform vorkommen, sondern häufig miteinander kombiniert sind und sich überlappen. So stellen die lexikalischen Realisierungen des Kovorkommens *Ohren* und *klingen* häufig fragmentarische Überlappungen der WVM *<in X Ohren VERB>*; *<in X Ohren klingen>*, *<wie X in den Ohren>*, *<für X Ohren>*:

- (5) WV: *wie blanker Hohn in deutschen Ohren klingen*
 WV: *wie Hohn VERB* (VERB = *klingen / wirken / vorkommen / erscheinen / anmuten*)
 WV: *wie ADJ Hohn* (ADJ = *blanker / reiner / purer*)
 WV: *wie SUB in den Ohren klingen* (SUB = *Hohn / Musik*)
 WV: *in ADJ_{nation} Ohren* (ADJ = *deutschen / europäischen / chinesischen*)

Des Weiteren ermöglicht die Systematisierung rekurrenter Kotextpartner einen empirischen Zugang zu Bedeutungen und Funktionen einer Analyseinheit⁹. Dies soll am Beispiel der präpositionalen Syntagmen *um die Ohren* vs. *bis über beide Ohren* deutlich gemacht werden. Das rekurrente

⁹ Im Grunde werden die Ideen des britischen Kontextualismus auf eine neue qualitative Stufe gehoben, indem nicht nur Bedeutungen einzelner Wörter primär über ihre Distribution erklärt werden, sondern auch die von komplexeren Syntagmen.

Syntagma *um die Ohren* hat isoliert keine autonome Bedeutung. Erst durch weitere lexikalische Ergänzungen, hier die rekurrent kovorkommenden Verben *hauen* und *haben*, wird eine eigenständige Sinneinheit konstituiert. Je nach verbaler Ergänzung handelt es sich dabei um unterschiedliche WV, z. B.: *um die Ohren hauen* 'jemanden mit harter Kritik konfrontieren' oder *um die Ohren haben* 'sehr viel zu tun haben; äußerst beschäftigt sein'. In diesen Fällen leisten die Verben *hauen* und *haben* den entscheidenden Beitrag zur Sinnkonstitution. In anderen Fällen reichen nicht einmal die verbalen Ergänzungen aus, um die Bedeutungen der Wortverbindungen eindeutig disambiguieren zu können. Für das Syntagma *um die Ohren fliegen* sind beispielweise noch weitere substantivische Ergänzungen nötig, um von UWV sprechen zu können: Jemandem *fliegen die Bälle um die Ohren* bedeutet 'im Fußballspiel unterlegen sein' und Die *Wrackteile fliegen jemandem um die Ohren* heißt 'die Folgen einer Explosion erleiden'. Im Gegensatz dazu weist das Syntagma *bis über beide Ohren* selbst bereits eine autonome Bedeutung auf, nämlich 'absolut, vollkommen, total'. Die WV *bis über beide Ohren* gehört zu einer Gruppe von WV, die die Funktion mehrgliedriger Intensitätspartikel erfüllen. Andere Beispiele sind *in Grund und Boden* oder *von Grund auf*. Obwohl alle dieselbe Kernbedeutung haben, unterscheiden sie sich in ihrem je spezifischen Gebrauchskontext. *Bis über beide Ohren* kann nur die Ausprägung eines Zustands sein, der einem Menschen selbst zugeschrieben wird, *in Grund und Boden* kann auch die Ausprägung der Wirkung eines Verhaltens in Bezug auf andere ausdrücken. *Bis über beide Ohren* wird zumeist mit den beiden Verben *verlieben* bzw. *verschulden*, aber nicht mit *kritisieren* / *verdammten* / *fahren* / *bomben* / ... Diese Verben stellen rekurrente Kotextpartner der WV *in Grund und Boden* dar. Diese lexikalische Restriktion ist wechselseitig. Die verbalen Ergänzungen der WV *von Grund auf* weisen wiederum eine andere Musterhaftigkeit auf und sind nicht mit den anderen mehrgliedrigen Intensivierern kompatibel, z. B. *von Grund auf modernisieren* / *reformieren* / *restaurieren* / *erneuern* / ...¹⁰

5.2. Profilvergleichende Interpretationen

Um eine wirkliche Sicherheit bezüglich des Lexikalisierungsgrades oder der Varianz einer Einheit zu gewinnen, sind Profilvergleiche nötig. Ein zentraler methodischer Zugang ist die Reziprokanalyse, bei der der Kookkurrenzpartner bzw. das zugrunde liegende Syntagma unter Ausschluss des Ausgangsobjekts selbst einer Kookkurrenzanalyse unterzogen wird¹¹. Es

¹⁰ Die Beispiele mit *Grund* habe ich aus meiner umfangreichen Feldstudie zum Distributionsverhalten dieses Lemmas in meiner Monografie entnommen (vgl. Steyer, in Vorb.)

¹¹ Duffner setzt diese Methode auch für die korpusbasierte Analyse von Satzadverbien ein (2010).

geht darum, den Grad wechselseitiger Determiniertheit der Komponenten zu verifizieren.

Bei der WV *faustdick hinter den Ohren haben* – das Adjektiv ist eine obligatorische Komponente (**es hinter den Ohren haben*) – ergibt die reziproke KA des Adjektivs *faustdick* beispielsweise, dass es einigen wenige andere rekurrente Realisierungen gibt: *es kommt faustdick; eine faustdicke Überraschung / Lüge*. Das gemeinsame Konzept für alle WV ist der Ausdruck eines extrem ausgeprägten Sachverhalts, Zustands oder Verhaltens. Allerdings ist die Kombinatorik stark eingeschränkt auf einige wenige Substantive und Verben.

Anders stellt sich der Fall bei den WV *auf offene Ohren stoßen* bzw. *auf taube Ohren stoßen* dar. Hier ergibt die KA des Syntagmas *auf... stoßen* ein ganzes Bündel von Realisierungen. Die Teilbedeutungen ändern sich je nach lexikalischer Füllgruppe, wobei in bestimmten Fällen erst die einbettenden Adjektive entsprechende Konnotationen ausdrücken:

(6) WVM: <*aufX stoßen*>

- Gegenreaktion: Verständnis
 - *auf offene Ohren stoßen / auf[großes] Interesse stoßen / auf Wohlwollen stoßen*
- Gegenreaktion: Zustimmung
 - *Auf[ungeteilte] Zustimmung stoßen / auf[große] Resonanz stoßen*
- Gegenreaktion: Zweifel
 - *auf Skepsis stoßen / auf geteiltes Echo stoßen*
- Gegenreaktion: Kritik
 - *auf[heftige / scharfe] Kritik stoßen*
- Gegenreaktion: Unverständnis
 - *auf taube Ohren stoßen / nicht auf Gegenliebe stoßen / auf[einhellige] Empörung stoßen*
- Gegenreaktion: Abwehr
 - *auf[erbitterten] Widerstand stoßen / auf Gegenwehr stoßen*

Auffällig ist, dass fast alle adjektivischen Ergänzungen als fakultative Intensivierer fungieren. Nur bei *auf ADJ Ohren stoßen* und *auf geteiltes Echo stoßen* sind sie obligatorisch, was den herausgehobenen Status dieser Einheit als feste WV untermauert. Die WV *auf Gegenliebe stoßen* ist die mit Abstand häufigste Realisierung und wird prototypisch in einem Negationskontext verwendet. Dies kann man empirisch validieren, indem man eine Analyse unter Ausschluss von Negationsmarkern wie *kein, nicht, wenig, selten, nirgendwo, ohne* usw. vornimmt. Es bleibt nur eine minimale Anzahl affirmativer Verwendungen der WV *auf Gegenliebe stoßen* übrig. Alle diese musterhaften Verwendungen eint jedoch ihre übergeordnete Musterbedeutung 'eine Gegenreaktion auf etwas erhalten'. Damit sind sie klar zu unterscheiden von anderen Bedeutungen von *auf etw. stoßen*, z. B.:

'etwas zufällig entdecken'

T90/JAN.01809 gegraben haben, um unter März -Readern und Thesenpapieren **auf** ein paar verstaubte Bände Expressionismus zu **stoßen**.

'jemanden zufällig treffen'

NUN91/FEB.01843 gering, werden schwere Kämpfe erwartet, sobald die Alliierten **auf** Saddams Elitetruppe, die Republikanische Garde, **stoßen**, die im

Die syntagmatischen Profile liefern auch Evidenzen für Sequenzmuster, die nicht im engen Sinne Wortverbindungen sind, bei denen aber die Vermutung besteht, dass sie gleichfalls verfestigte Einheiten mit nicht beliebig austauschbaren Elementen darstellen und spezifischen Kontextrestriktionen unterliegen.

Im Wortformenprofil von *Daumen* tritt im Cluster *Daumen drücken* das syntagmatische Muster *Jetzt heißt es (Daumen drücken)* besonders hervor. Analysiert man das Syntagma *Jetzt heißt es ...*, findet man u. a. folgende Nachfeldergänzungen:

- (7) <Jetzt heißt es X>
Daumen drücken
Ärmel hochkrempeln
Nerven bewahren
einen kühlen Kopf bewahren
Mund abputzen
Kopf hoch
Augen zu und durch

Diese Gruppe der X-Ergänzungen ist nicht zufällig, sondern sie ist semantisch restringiert, da die Füller alle zum Körperfeld gehören. Sie schreiben dem Syntagma *Jetzt heißt es* eine eigenständige Bedeutung und Funktion zu und zwar als Ermutigung oder Aufforderung, etwas durchzustehen bzw. auszuhalten oder etwas anzupacken. Das 'somatische Muster' stellt wiederum ein Teilkomponente eines übergeordneten funktionalen Musters dar, das auch durch andere sprachliche Füllergruppen konstituiert wird, z. B. verbale Ergänzungen wie *Jetzt heißt es anpacken / üben / punkten* usw. Die funktionale Gebrauchsbeschränkung wird auch deutlich, wenn man andere Verwendungen betrachtet, die eine solche nicht aufweisen:

RHZ05/JUL.10527 **Jetzt heißt es**, die Entscheidung in Mainz abzuwarten

R98/SEP.70364 **Jetzt heißt es** wieder Schlange stehen auf der verwinkelten Steintreppe

Das folgende Beispiel verdeutlicht, wie weit gebrauchsbasierte Restringiertheit gefasst werden kann. Die WV *seinen Ohren nicht trauen* in der Bedeutung 'über etwas äußerst verwundert oder verärgert sein' ist häufig in negierte Infinitiv-mit-zu-Konstruktionen mit dem Verb *glauben* eingebettet, z. B. *er glaubte, seinen Ohren nicht zu trauen*. Nun stellt sich die Frage, ob es sich beim wiederholten Kovorkommen der lexikalischen Komponenten *glauben* und *trauen* in der syntaktischen Struktur [glauben nicht, zu trauen] um eine restringierte Konstruktion handelt oder nicht. Die Kookkurrenz-

analyse von [glauben nicht, zu trauen]¹² zeigt, dass nur *Ohren* und *Augen* ein rekurrentes Korvorkommen aufweisen (516 Treffer), alle anderen Realisierungen im statistisch unspezifischen Bereich liegen (184). Die Annahme, dass letztere keine Musterhaftigkeit aufweisen, lässt sich durch die Suche nach dieser Konstruktion unter Ausschluss der Lemmata *Ohr* und *Auge* verifizieren. Diese ergab durchweg nur Realisierungen, die in völlig heterogenen syntaktischen Strukturen vorkommen:

B00/JAN.04113 Zum Schluss gibt es doch noch politische Anerkennung: "Ich **glaube nicht**, dass sich jemand im Westen **getraut** hätte, einfach die CIA-Zentrale zu stürmen", zollt der Grünenpolitiker Wolfgang Wieland den Akteuren des 15. Januar seinen Respekt.

B01/NOV.85647 60er-Jahre konnte es sehr lange dauern, bis ein junger Mann sich **traute**, seinen Eltern **zu** sagen, dass er **nicht** an Gott **glaubt**.

Daher scheint die Konstruktion [glauben nicht, zu trauen] ebenso semantisch-funktional eingeschränkt zu sein wie das bei *Jetzt heißt es* zumindest für einen Teilbereich der lexikalischen Füller zutrifft. Gerade diese zwar nicht im strengen Sinne festen WV oder Syntagmen, aber doch auch restringierten Sequenzmuster sind nur durch Analysen, wie sie hier vorgestellt wurden, rekonstruierbar.

6. Perspektiven

Abschließend werden einige grundsätzliche Gedanken noch einmal zusammengefasst:

- Die wenigen Beispiele sollten deutlich gemacht haben, dass unser Ansatz einer lexikalistischen Perspektive verpflichtet ist. Das bedeutet, dass wir Erklärungen dafür, wie sich Sprache im Gebrauch konstituiert, primär über die Erforschung lexikalischer Verfestigungs- und Musterbildungsprozesse finden wollen. Dabei geht es nicht um eine Ausblendung der Syntax. Denn natürlich sind alle mit unseren Methoden identifizierbaren rekurrenten sprachlichen Einheiten in ihrem konkreten Textzusammenhang in reguläre grammatische Strukturen wie Sätze oder Phrasen eingebettet oder bilden selbst eigenständige syntaktische Einheiten (z. B. Sprichwörter als Typ satzwertiger Ausdrücke). Als feste Bausteine des Sprachgebrauchs werden aber oft nur lexikalische Fragmente weitertransportiert und dann an den neuen Kontext angepasst. Die Kontexte solcher Chunks sind, wie wir gesehen haben, häufig restringiert und das primär funktional. Es ist also anzunehmen, dass Sprecher über ein stark ausgeprägtes lexikalisches Wissen bezüglich solcher sprachlichen Bausteine und der zugrunde liegenden Muster verfügen, die nicht unbedingt mit syntaktischen Regularitäten einhergehen müssen.

¹² Die zugrunde liegende Suchanfrage lautet: &glauben /s0 &trauen /s0 "nicht" /s0 zu.

- Der UWV-Ansatz wirft Fragen bezüglich des Sinns und des Erklärungspotenzials tradierter semantischer Taxonomien auf. Je abstrakter diese sind, desto weniger sind sie überhaupt auf usuellen Sprachgebrauch abbildbar. Die aus Korpora ermittelten rekurrenten lexikalischen Füller eines Musters können eben nicht nur strukturell oder semantisch sehr unterschiedlichen Klassen angehören, sondern sie können Eigenschaften aufweisen, die tief in die Pragmatik hineinreichen und für die bisher keine Kategorien bereitgestellt werden. Bei der Entwicklung neuartiger **Gebrauchsentologien** wird es eine spannende Frage sein, auf welcher Abstraktionsebene die Merkmale solcher Füllergruppen noch der Distinktivität eines Musters gerecht werden können oder ob diese Abstraktionen nicht vielmehr die tatsächlichen kontextuellen Beschränkungen des Gebrauchs nivellieren, z. B. zugunsten einer besseren Formalisierung. Die Analyse der 'Füllersystematik' kann auch zu einer **gebrauchsbasierten Paradigmatik** führen, da unterschiedlich versprachlichte mono- und polylexematische Füller unter dem holistischen Aspekt ihres Musters zusammengefasst werden und damit eine wie auch immer geartete Gebrauchsverwandtschaft zu Tage tritt.
- Dieser Beitrag ist darüber hinaus als Plädoyer für eine qualitativ ausgerichtete Korpuslinguistik zu verstehen. Die WVM zeigen beispielsweise, dass quantitative Kriterien wie 'Frequenz' oder auch 'statistischen Affinität' in Bezug auf konkrete lexikalische Exemplare sehr relativ zu sehen sind, da diese in höchstem Maße durch andere ersetzt werden können. Auch seltene Füller tragen zur Musterbildung bei und können von Sprechern dementsprechend eingesetzt werden. Die Usualität zeigt sich eben erst im übergeordneten Muster. Sinnvoll scheint es daher zu sein, eher von 'Häufigkeitstrends' und 'prototypischen Vertretern' zu sprechen. Generell sollte eine Argumentation also nicht auf Zahlen und Statistiken beruhen, sondern auf der Extrapolierbarkeit der Ergebnisse außerhalb der Welt der statistischen Berechnungen.
- Schließlich muss sich auch die Phraseologie der Diskussion über ihren genuinen Gegenstand stellen: Die traditionelle Fokussierung auf stark lexikalisierte, zumeist idiosynkratische, Wortverbindungen hat zu einer Überbewertung ihres unikalen Status im Wortschatz geführt. Die wenigsten usuellen Wortverbindungen erfüllen jedoch dieses Kriterium. Die meisten sind Konstituenten abstrakterer Muster, teilen Eigenschaften mit anderen – auch monolexematischen – Wortschatzeinheiten und werden in der Regel nicht als formal abgeschlossene Einheiten verwendet, sondern fragmentarisch, sich überlappend, fuzzy. Diese Erkenntnisse müssen letzten Endes auch weitreichende

Folgen für Mehrwortlexikografie haben, z. B. in Form neuartiger syntagmatischer Musterwörterbücher.

Wir stehen erst am Anfang.

Literatur

- Belica, C. (1995): Statistische Kollokationsanalyse und -clustering. Korpuslinguistische Analyseverfahren. Institut für Deutsche Sprache, Mannheim. <http://corpora.ids-mannheim.de/>
- Belica, C. & Steyer, K. (2008): Korpusanalytische Zugänge zu sprachlichem Usus. In: Vachková, M. (éd.): Beiträge zur bilingualen Lexikographie. Prag, 7-24.
- Bubenhof, N. (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin/New York.
- Deppermann, A. (2006): Construction Grammar – Eine Grammatik für die Interaktion? – In: Deppermann, A. & Fiebler, R. & Spranz-Fogasy, T. (éds.), Grammatik und Interaktion. Untersuchungen zum Zusammenhang zwischen grammatischen Strukturen und Gesprächsprozessen. Radolfzell, 43–65.
- DeReKo: Das Deutsche Referenzkorpus - DeReKo. Institut für Deutsche Sprache. Mannheim. <http://www.ids-mannheim.de/kl/projekte/korpora>
- Duffner, R. (2010): Die Satzadverbien im Deutschen: eine korpusbasierte Untersuchung. Dissertation. Online abrufbar unter <http://doc.rero.ch/record/25013>. Université de Neuchâtel.
- Engelberg, S., Holler, A. & Proost, K. (éds.) (2011): Sprachliches Wissen zwischen Lexikon und Grammatik. (Jahrbuch des Instituts für Deutsche Sprache 2010). Berlin/New York.
- Feilke, H. (1996): Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik. Frankfurt am Main.
- Fischer, K. & Stefanowitsch, A. (éds.) (2006): Konstruktionsgrammatik. Von der Anwendung zur Theorie. (Stauffenburg Linguistik, 40). Tübingen
- Hausmann, F. J. (1985): Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: Bergenholtz, H. & Mugdan, J. (éds.): Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984. (Lexicographica 3). Tübingen, 118-129
- Hunston, S. & Francis, G. (2000): Pattern Grammar. A corpus-driven approach to the lexical grammar of English. (Studies in Corpus Linguistics, 4). Amsterdam/Philadelphia.
- Lemnitzer, L. & Zinsmeister, H. (2010): Korpuslinguistik. Eine Einführung. 2., durchges. und aktualisierte Aufl. (Narr Studienbücher). Tübingen
- Perkuhn, R. (2007): "Corpus-driven": Systematische Auswertung automatisch ermittelter sprachlicher Muster. In: Kämper, H. & Eichinger, L. M. (éds.): Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache. (Studien zur deutschen Sprache, 40). Tübingen, 465-491.
- Perkuhn, R. & Belica, C. (2004): Eine kurze Einführung in die Kookkurrenzanalyse und syntagmatische Muster. Institut für Deutsche Sprache, Mannheim. <http://www.ids-mannheim.de/kl/misc/tutorial.html>
- Sailer, M. (2007): Corpus linguistic approaches with German corpora. In: Burger, Harald/Dobrovolskij, Dmitrij/Kühn, Peter/Norricks, Neal R. (éds): Phraseologie/Phraseology.

- Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research. (HSK 28.2). Berlin/New York, 1060-1071.
- Sinclair, J. (1991): *Corpus, Concordance, Collocation*. Oxford.
- Stefanowitsch, A. & Fischer, K. (éds.) (2008): *Konstruktionsgrammatik II: Von der Konstruktion zur Grammatik*. (Stauffenburg Linguistik, 47). Tübingen.
- Steyer, K. (2004): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: Steyer, K. (Hg.) (2004): *Wortverbindungen □ mehr oder weniger fest*. (Jahrbuch des Instituts für deutsche Sprache 2003). Berlin/New York, 87–116.
- (in Vorb.): *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. (Studien zur deutschen Sprache). Tübingen.
- Steyer, K. & Brunner, A. (2009): *Das UWV-Analysemodell. Eine korpusgesteuerte Methode zur linguistischen Systematisierung von Wortverbindungen*. (OPAL 1/2009). Mannheim. Online verfügbar unter: <http://pub.ids-mannheim.de/laufend/opal/pdf/opal2009-1.pdf>
- Steyer, K. & Brunner, A. (2008 - 2013): *Wortbindungsfelder*. Abrufbar unter "Wortverbindungen online" <http://www.ids-mannheim.de/ll/uwv/wv-online/wv-online.html>
- Steyer, K. & Hein, K. (2010): *Deutsche Sprichwortartikel*. In: Datenbank des EU-Projekts *SprichWort*. <http://www.sprichwort-plattform.org/sp/Sprichwort>
- Tognini-Bonelli, E. (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, K. & Stefanowitsch, A. (éds.), 19–37.