

Die Sammlung unikalener Wörter des Deutschen

Aufbauprinzipien und erste Auswertungsergebnisse

Manfred Sailer
Universität Göttingen
manfred.sailer@phil.uni-goettingen.de

Beata Trawiński
Universität Tübingen
trawinski@sfs.uni-tuebingen.de

1 Einleitung

Die *Sammlung unikalener Wörter des Deutschen* (SuWD) ist der erste abgeschlossene Teil der Forschungsplattform *CoDII* (*Collection of Distributionally Idiosyncratic Items*), die im Rahmen des Projekts A5 des Sonderforschungsbereichs 441 (*Linguistische Datenstrukturen*) an der Universität Tübingen erstellt wird. Die Grundidee von CoDII ist es, eine Ausgangsbasis zu sein für die linguistische Untersuchung von lexikalischen Elementen mit Distributionsbeschränkungen. Dies beinhaltet, dass CoDII die entsprechenden Elemente auflistet, ihren linguistischen Dokumentationsstand angibt und Möglichkeiten der Datenerhebung zu diesen Elementen aufzeigt. In der ersten Projektphase (2002–2004) wurden die allgemeinen Aufbauprinzipien von CoDII definiert und die hier zu präsentierende Teilplattform SuWD realisiert. In diesem Artikel soll die SuWD in ihren Grundzügen vorgestellt werden.

2 Theoretischer Hintergrund

Das Projekt, in dessen Rahmen die SuWD entstanden ist, beschäftigt sich mit Elementen mit Distributionsidiosynkrasien. Hierunter verstehen wir Wörter, deren

[†]Die vorgestellten Ergebnisse sind im Rahmen des Projekts *Distributionsidiosynkrasien* des Sonderforschungsbereichs 441 an der Universität Tübingen entstanden. Neben den Autoren haben Monique de Jong, Guthrun Love, Carmella Payne und Jan-Philipp Soehn an der Sammlung mitgearbeitet. Wir danken auch Tylman Ule und Andreas Wagner für Unterstützung bei der XML-Umsetzung sowie Cyril Belica, Meike Lauer und Kathrin Steyer vom IDS, Christiane Fellbaum, Undine Kramer, Gerald Neumann und Diana Stancheva von der Berlin Brandenburgischen Akademie der Wissenschaften, für zahlreiche Anregungen. Für Daten und Literaturhinweise danken wir Dmitrij Dobrovolskij, Kurt Feyaerts und besonders Elisabeth Piirainen, die uns großzügig ihre Materialien zur Verfügung gestellt hat.

Verteilung nicht vollständig aus ihren syntaktischen, semantischen und pragmatischen Eigenschaften in Verbindung mit allgemeinen Prinzipien folgt. (Richter and Sailer 2003), (Sailer 2003; Sailer 2004) argumentieren für die Existenz solcher Wörter und zeigen die Schwierigkeiten auf, die Distributionsidiosynkrasien formalen Grammatiksystemen bereiten. Anhand einzelner Beispieltypen konnte im Projekt ein Kontinuum von Distributionsbeschränktheit beobachtet werden. Dieses Kontinuum hat auf der einen Seite weitgehend unbeschränkte Wörter (wie *Buch*). Einen mittleren Grad an Beschränktheit weisen Elemente auf, die systematischen Beschränkungen unterliegen, wie Pronomina (*sie*) und Anaphern (*sich, einander*). Andere Elemente scheinen das Auftreten bestimmter semantischer Operatoren zu verlangen, wie *Polaritätselemente* (Ladusaw 1980; van der Wouden 1997). Deutsche Beispiele sind *beileibe* (nur mit Negation) oder *schon* (nur ohne Negation). Am anderen Ende des Kontinuums liegen Wörter, die nur in einem bestimmten Kontext auftreten können. Diese Wörter werden gemeinhin als *Unikalia* bezeichnet. Typische Vertreter dieser Gruppe sind *Tacheles* (nur in: *Tacheles reden*) oder *abhanden* (nur in: *abhanden kommen*).

Orthogonal zu diesem Distributionskontinuum gibt es ein *Konstruktionskontinuum*. Hierbei handelt es sich um die Frage, in wieweit ein Syntagma von seiner Struktur her gemäß generellen Prinzipien der Grammatik aufgebaut ist und kompositionell interpretiert wird. Am einen Ende dieses Kontinuums stehen freie Kombinationen (wie: *Bücher lesen*). Sogenannte *Phraseolschablonen* nehmen eine Zwischenstellung ein (*Du und heiraten?!* (Lambrecht 1990)). Am anderen Ende des Kontinuums finden sich Kombinationen, die die normalen Regeln der Syntax verletzen und nur mit festem Wortmaterial zulässig sind (wie: *frei Haus*). Dieses Kontinuum wird beispielsweise im Rahmen der *Construction Grammar* (Fillmore, Kay, and O'Connor 1988) untersucht.

(Sailer 2003) argumentiert, dass die beiden Kontinua unabhängig voneinander sind. So ist der Teilsatz *wo Barthel den Most holt* eine feste Konstruktion. Die obligatorische Verbindung dieses Teilsatzes mit dem Verb *wissen* analysiert Sailer als Distributionsbeschränkung.

Im Gegensatz zu Konstruktionen sind Distributionsidiosynkrasien bislang nicht systematisch in der theoretischen Sprachwissenschaft behandelt. Zwar gibt es für einzelne Gruppen von Wörtern — Pronomina oder Polaritätselemente — jeweils eine umfangreiche Literatur, die im Tübinger Projekt verfolgte integrierende Gesamtperspektive ist jedoch neu.¹

Zur empirischen Fundierung einer Distributionstheorie soll eine möglichst große Zahl von Elementen mit Distributionsidiosynkrasien systematisch erfasst werden. Was unikale Wörter betrifft, entstand im Rahmen der Phraseologiefor-

¹(van der Wouden 1997) stellt einen Zusammenhang her zwischen Unikalia und Polaritätselementen, ohne jedoch auf Unikalia genauer einzugehen.

schung in den vergangenen zwei Jahrzehnten für das Deutsche eine Vielzahl von Arbeiten mit reichen Beispiellisten. Besonders umfangreich, und damit auch der Ausgangspunkt für unsere Arbeit, ist die Sammlung in (Dobrovol'skij 1988). Dort sowie in (Dobrovol'skij 1989), (Dobrovol'skij and Piirainen 1994a) und (Dobrovol'skij and Piirainen 1994b) finden sich verschiedene Klassifikationskriterien für unikale Wörter und die Wendungen, in denen sie auftreten.² Mit der neu erstellten Sammlung wollen wir diese Arbeiten wieder aufgreifen und in einem einheitlichen und gut zugänglichen Format auch für grammatiktheoretische Fragestellungen, wie die nach einer Distributionstheorie, nutzbar machen.

3 Die Sammlung unikalener Wörter des Deutschen

Ausgehend von den existierenden Dokumentationen zu unikalenen Wörtern des Deutschen ((Dobrovol'skij 1988), (Dobrovol'skij 1989), (Dobrovol'skij and Piirainen 1994b), (Dobrovol'skij and Piirainen 1994a), etc.) sowie eigenen Beobachtungen haben wir eine Sammlung von ca. 450 unikalenen Elementen erstellt. Jedes Element in der Sammlung wurde durch vier Informationsblöcke, die unten kurz erläutert werden, charakterisiert.

3.1 Modellierung der Einträge in SuWD

Jeder Eintrag in SuWD wird zunächst durch generelle Informationen charakterisiert, indem das unikale Element selbst und dessen englische Übersetzung, die Wendung, in der es vorkommt, sowie eine Menge von möglichen Paraphrasen der Wendung aufgeführt werden.

Als zweites sind die Klassifikationen gemäß (Dobrovol'skij 1988), (Dobrovol'skij 1989), (Dobrovol'skij and Piirainen 1994b) sowie eine an (Nunberg, Sag, and Wasow 1994) orientierte semantische Klassifikation angegeben. Soweit möglich wird eine Klassifikation auch auf andere in der ursprünglichen Publikation nicht genannte Elemente übertragen. Die Datenstruktur in SuWD ist so gewählt, dass problemlos weitere Klassifikationen hinzugefügt werden können (wie beispielsweise die von (Feyaerts 1994)). Auch weitere Publikationen, in denen das Element diskutiert wird, sind hier aufgelistet.

In einem dritten Informationsblock wird die Angabe zur syntaktischen Kategorie des unikalenen Elements gemacht sowie die Beschreibung der syntaktischen Struktur aufgeführt, innerhalb derer das unikale Element auftritt. Zur syntaktischen Charakterisierung unikalener Elemente und der Wendungen, in denen sie

²Für Verweise auf weitere Publikationen zu unikalenen Wörtern siehe die *Bound Words Bibliography*, <http://www.sfb441.uni-tuebingen.de/a5/bwb/>.

vorkommen, wird das in vielen Korpora des Deutschen eingesetzte Stuttgart-Tübingen Tagset (STTS)³ verwendet. Bei der Beschreibung syntaktischer Eigenschaften der Wendungen wird auf mögliche syntaktische Variationen wie z.B. Passivierung, Pronominalisierung, Modifizierung, Auftreten in Raisingkonstruktionen, usw. geachtet. Für alle Kontexte wird auf Belegsätze aus verschiedenen Korpora, dem Internet oder der Fachliteratur verwiesen.

Der linguistischen Erfassung des unikalen Elements folgen Hinweise zur weiteren Datenrecherche durch den Benutzer. Hierzu stellen wir für verschiedene öffentlich zugängliche Korpora des Deutschen optimierte Suchanfragen bereit. Diese Korpora umfassen zurzeit: das Internet mittels Google, die öffentlich zugänglichen Korpora des Instituts für Deutsche Sprache mittels Cosmas II,⁴ das Korpus des Digitalen Wörterbuchs der Deutschen Sprache – DWDS⁵, die Tübinger Sammlung nutzbarer empirischer linguistischer Datenstrukturen – TUSNELDA⁶ sowie der TIGERSearch⁷. Auch hier ist die Aufnahme weiterer Korpora und Suchwerkzeuge problemlos möglich.

Es ist sowohl konzeptuell als auch technisch durchaus möglich, die Struktur der SuWD-Einträge inhaltlich um weitere linguistische Informationen wie Abhängigkeitsstruktur und die logische Form auszubauen.

3.2 Technische Realisierung

3.2.1 Interne Datenstruktur

Die Sammlung wird zunächst intern in XML kodiert, einer Markup-Sprache, die sich in den letzten Jahren als der Hauptstandard zur Datenrepräsentation auch in der Linguistik etabliert hat (Witt 2002). Die DTD wurde so konzipiert, dass das Element `codii`, dessen Instanz durch die Attribute `type` (zur Spezifizierung des Sammlungstyps wie die Sammlung unikalener Elemente oder die Sammlung der Polaritätselemente) und `xml:lang` (zu Spezifizierung der Sprache, aus der die Daten kommen) identifiziert wird, die Dokumentwurzel darstellt. Das Inhaltsmodell des `codii`-Elements besteht aus zwei Elementen: `dii-list`, dessen Inhalt die Liste mit Einträgen für distributionell beschränkte Elemente ist, und `dii-examples`, dessen Inhalt die Liste der Belege ist. Das Inhaltsmodell des `dii-list`-Elements besteht aus einer Liste der `dii-entry`-Elemente, die durch eine Reihe von Attributen charakterisiert werden, die die

³Zu allgemeinen Informationen zu STTS siehe die STTS-Homepage unter <http://www.sfs.uni-tuebingen.de/Elwis/stts/stts.html>.

⁴Siehe <http://www.ids-mannheim.de/cosmas2/>.

⁵Siehe <http://www.dwds.de/>.

⁶Siehe <http://www.sfb441.uni-tuebingen.de/tusnelda.html>.

⁷Siehe <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/>.

distributionell beschränkten Ausdrücke indentifizieren (*dii*), die Wendungen, in denen sie vorkommen beschreiben (*dii-expression*), den Dokumentationsstand zu den einzelnen Elementen schildern (*dii-classification*), die syntaktischen Eigenschaften der Elemente und der Wendungen aufführen (*dii-syntax*) sowie die Abfragemöglichkeiten angeben (*dii-queries*). Abbildung 1 präsentiert ein Fragment des *dii-entry*-Elements, das das unikale Wort *Zampano* beschreibt.

```
<dii-entry id="zampano">
  <dii>
    <ol>Zampano</ol>
    <en>golden boy</en>
  </dii>
  <dii-expression>
    <ol>der gro&#223;e Zampano</ol>
    <en>the big doer</en>
  </dii-expression>
  <dii-classification>
    <dii-class class="dekompo" type="A5">
      <bibliography bib-item="A5"/>
    </dii-class>
  </dii-classification>
  <dii-syntax hits="zampano-Bsp zampano-apposition" cat="NE">
    <dii-expression-syntax cat="NP">
      der/ART gro&#223;e/ADJA Zampano/NE
    </dii-expression-syntax>
    <variation kind="OPEN" hits="zampano-ecclestone">
      <comment status="external">
        Spitzname von Formel-1-Manager Bernie Ecclestone
      </comment>
    </variation>
  </dii-syntax>
  <dii-queries>
    <query type="cosmasII">
      <query-text><![CDATA[Zampano]]>
      </query-text>
    </query>
  </dii-queries>
</dii-entry>
```

Abbildung 1: Ein Fragment der CoDII-XML-Kodierung von *Zampano*

Das Inhaltsmodell des Elements *dii-examples* besteht aus einer Liste von *example*-Elementen, die mittels der Attribute *dii* und *id* mit entsprechenden Einträgen verlinkt werden. Abbildung 2 zeigt ein *example*-Element, das einen Korpusbeleg für *Zampano* beschreibt.

```

<example dii="zampano" id="zampano-Bsp">
  <source corpus="cosmasII">
    R97/APR.32703 Frankfurter Rundschau, 29.04.1997, S. 15,
    Ressort: WIRTSCHAFT; F&#252;r eine lohnende &#220;bernahme
    sind einige H&#252;rden zu nehmen
  </source>
  <ol>
    "Ich glaube nicht, da&#224; da Manna vom Himmel f&#228;llt
    und der gro&#224;e Zampano f&#252;r diverse neue Stellen
    sorgt", meint der Betriebsratschef der Vegesacker Werft,
    Wolfgang Dettmer.
  </ol>
</example>

```

Abbildung 2: Die CoDII-XML-Beschreibung eines Korpusbeleges für *Zampano*

Die Datenstruktur in SuWD und generell in CoDII wurde also so gestaltet, dass weitere Sammlungen distributionell beschränkter Ausdrücke in beliebigen Sprachen unproblematisch modelliert werden können.

3.2.2 Visualisierung und Abfragemöglichkeiten

Die SuWD ist in Form von einer mit XSL-Transformationen generierten Menge von XHTML-Dateien unter der URL

<http://www.sfb441.uni-tuebingen.de/a5/codii>

über das Internet öffentlich zugänglich. Abbildung 3 zeigt ein Beispiel der Browser-Darstellung eines SuWD-Eintrages.

Bislang können nur Informationen zu einzelnen Lemmata abgerufen werden, die Suchmöglichkeiten können jedoch jederzeit verfeinert werden und den Zugriff auf die Daten durch eine Überführung der Sammlung in eine Datenbank dynamisiert werden.⁸

4 Andere Sammlungen

Die SuWD versteht sich als ein Hilfsmittel zur linguistischen Forschung. Das bedeutet, dass vorhandene linguistische Analysen zusammengetragen und ergänzt

⁸Dabei muss noch entschieden werden, ob eine relationale Datenbank notwendig ist, oder ob mit dem bisherigen Datenformat gearbeitet werden kann und ob XQuery als die Sprache zur Abfrage der CoDII-XML-Daten leistungsfähig genug ist.

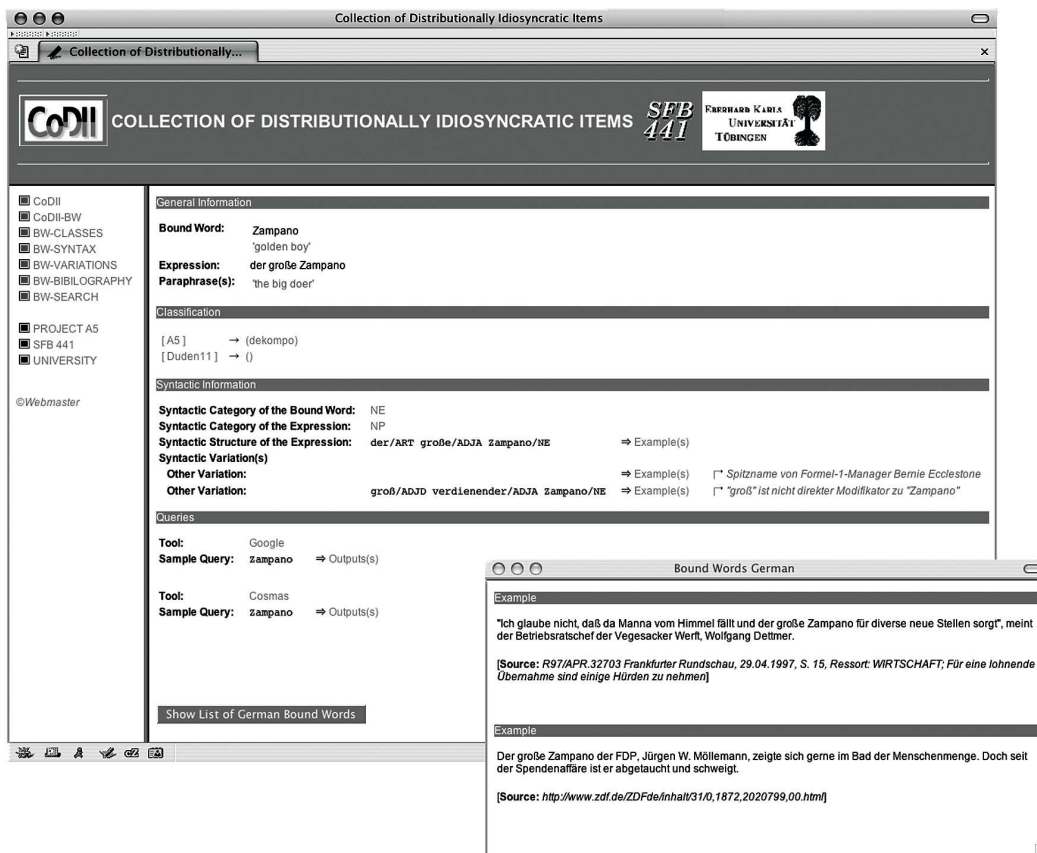


Abbildung 3: Die Browser-Darstellung des SuWD-Eintrages für *Zampano*

werden, eigene Analyseansätze werden als solche gekennzeichnet. Die SuWD unterscheidet sich prinzipiell von anderen Sammlungen zu Wortkombinationen, die im Moment im Aufbau sind. Zu erwähnen sind dabei besonders: das Projekt *Usuelle Wortverbindungen* (UWV) im Rahmen von *lexiko* am IDS-Mannheim und das Projekt *Kollokationen im Wörterbuch* (KiW) an der Berlin-Brandenburgischen Akademie der Wissenschaften.

Im Rahmen von UWV (Steyer 2004) wird von hochfrequenten Wörtern ausgegangen. Diese Wörter werden einer Kookkurrenzanalyse unterworfen, deren Ergebnisse dann die Basis für eine linguistische und lexikographische Beschreibung des typischen Gebrauchs eines Wortes bilden. Diese Auswertung wird in ein elektronisches Informationssystem integriert und kann dort zum Beispiel von Deutschlernern genutzt werden. Als Datenbasis dienen die am IDS gesammelten Korpora. Das methodologisch Besondere an UWV ist, dass die gesamte Datenauswahl rein statistischen Kriterien unterliegt und die linguistische In-

terpretation im Vergleich zu anderen Projekten relativ spät ansetzt. Im Gegensatz dazu beruht die Zusammenstellung der SuWD-Einträge auf einer primär kompetenzbasierten Klassifizierung von Elementen als potenziell unikal. Interessante Überschneidungen ergeben sich beispielsweise im Bereich der gebundenen/unikalen Lesarten von hochfrequenten Wörtern.

Auch im Rahmen des Projekts *Kollokationen im Wörterbuch* ist eine Nutzung für ein breiteres Publikum anvisiert. Die Datenbasis bietet hier der Sprachgebrauch des Deutschen im 20. Jahrhundert, wie er in den Korpora des Projekts *Digitales Wörterbuch der deutschen Sprache des 20. Jahrhunderts* erfasst ist. Bei der Untersuchung von Mehrwortkombinationen werden wie bei der SuWD syntaktische Flexibilität der Wendungen und andere linguistische Kriterien mit berücksichtigt. Es geht im Projekt KiW jedoch um eine lexikographische Erfassung eines etablierten Wortkombinationenbestands des Deutschen unter Beschränkung auf verbale Phraseologismen. Im Gegensatz dazu soll die SuWD eine Basis liefern, um den Möglichkeitenraum für Distributionsbeschränkungen lexikalischer Elemente zu erforschen. Dabei ist das Prinzip der "Offenheit" des Korpus relevant, d.h., dass Linguisten die SuWD jederzeit verwenden können, um eigene Datenrecherche auf ständig aktualisierten Korpora (wie dem Internet) zu betreiben, und so zu aktuelleren Erkenntnissen kommen können, als die Klassifikationen und Variationsbeobachtungen, die in die SuWD Eingang gefunden haben.

5 Erste Auswertungsergebnisse

Die in der SuWD zusammengetragenen Daten erlauben die Untersuchung vielfältiger Fragestellungen. Aufgrund der theoretischen Orientierung des Projekts haben wir uns bislang auf die Frage konzentriert, wie der jeweils von einem unikal Element geforderte Kontext zu beschreiben ist. Als Arbeitsdefinition betrachten wir dabei als unikale Elemente, Elemente, die nur zusammen mit einem bestimmten weiteren *Lexem* auftreten. Damit unterscheiden wir beispielsweise zwischen Unikalia und Polaritätselementen. Bei letzteren wird das Ko-vorkommen eines bestimmten semantischen Operators (wie Negation) gefordert.

Zur Bestimmung des möglichen Abstandes zwischen einem unikal Element und seinem obligatorischen Kontextlexem (LEX) hat es sich, analog zu (O'Grady 1998) und (Everaert and Kuiper 1996), als sinnvoll erwiesen, die Dependenzbeziehungen zu betrachten, die zwischen den beiden Elementen bestehen.

In (Soehn and Sailer 2003) wurde die folgende Generalisierung formuliert: Ein unikales Element selegiert das LEX, oder das unikale Element ist ein Komplement von LEX. Diese Generalisierung hat jedoch Fälle ausgeschlossen, in denen LEX

ein Adjunkt zum unikalen Element ist, wie in *armer Schlucker* und *der große Zampano*. Verbindungen mit Unikalia von diesem Typ sind selten, aber, wie die Beispiele zeigen, möglich.

In einer Verfeinerung der Generalisierung aus (Soehn and Sailer 2003) formulieren wir nun die folgende Hypothese: Ein unikales Element selegiert sein obligatorisches Kontextlexem (i), oder es wird von ihm selegiert (ii), oder es gibt maximal ein weiteres Lexem, das in der Dependenzstruktur zwischen dem unikalen Element und LEX steht (iii). Hier einige Beispiele:

- (i) Angst einjagen (LEX ist Komplement)
- (ii) Tacheles reden (Unikales Element ist Komplement)
der große Zampano (LEX ist Adjunkt)
- (iii) sich freuen wie ein Schneekönig
(LEX: *freuen*; intervenierendes Lexem: *wie* selegiert das unikale Element)

Diese Hypothese erfasst die Unikalia, die bisher in der SuWD erfasst sind.⁹ Wir gehen davon aus, dass ein unikales Element genau ein Lexem fordert, und dass dieses Lexem in einem lokalen Dependenzbereich auftreten muss. Diese Beschränkung ist stärker als der in (O’Grady 1998) formulierte *Continuity Constraint* für idiomatische Wendungen. O’Grady nimmt lediglich an, dass Idiombestandteile über eine (beliebig lange) Dependenzkette mit dem Kopf des Idioms verbunden sein müssen.

Sollte sich die Hypothese für Unikalia als tragfähig erweisen, so ergeben sich zwei logisch folgende Untersuchungsfragen: zum einen ist für Idiombestandteile und Kollokationsbestandteile zu prüfen, ob sie auch unter diese Generalisierung fallen. In diesem Falle wäre es eine wichtige Einschränkung des *Continuity Constraints*. Zum anderen ist der Frage nachzugehen, wie diese Generalisierung in eine formale Grammatikarchitektur integriert werden kann.

6 Zusammenfassung und Ausblick

Die in SuWD zusammengetragenen Informationen erlauben eine systematische Studie von unikalen Elementen. Das Ziel der Sammlung ist es, den vorhandenen linguistischen Forschungsstand zu diesen Elementen zu dokumentieren und damit zur weiteren Forschung einzuladen. Wie anhand der diskutierten Hypothesen gezeigt, ist eine solche Evaluation bereits jetzt möglich. Neben der deutschen Sammlung ist, im Sinne der in Abschnitt 2 angedeuteten Forschungsperspektive, die Ausweitung auf Teilsammlungen zu anderen Sprachen und zu anderen Typen

⁹Von dieser Hypothese unberührt bleibt die Klasse von Unikalia, die in Paarformeln auftreten, wie klipp und klar, aber vergleiche (Soehn 2003), Abschnitt 5.1 hierzu.

von distributionell beschränkten Elementen vorgesehen. Besonders aussichtsreich erscheint hier die Aufnahme von Polaritätselementen des Deutschen (Kürschner 1983; Zwarts 1997) sowie das Erstellen einer vergleichbaren Sammlung für das Englische.

Literatur

- Dobrovolskij, D. (1988). *Phraseologie als Objekt der Universallinguistik*. Verlag Enzyklopädie, Leipzig.
- Dobrovolskij, D. (1989). Formal gebundene phraseologische Konstituenten: Klassifikationsgrundlagen und theoretische Analyse. In W. Fleischer, R. Große, and G. Lerchner (Eds.), *Beiträge zur Erforschung der deutschen Sprache*, Volume 9, pp. 57–78. Leipzig, Bibliographisches Institut.
- Dobrovolskij, D. and E. Piirainen (1994a). PGF: Auf dem Präsentierteller oder auf dem Abstellgleis? *Zeitschrift für Germanistik* (NF 4), 65–77.
- Dobrovolskij, D. and E. Piirainen (1994b). Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative. *Folia Linguistica* 27(3–4), 449–473.
- Everaert, M. and K. Kuiper (1996). Theory and Data in Idiom Research. In L. McNair et al. (Ed.), *CLS32: the Parasession on Theory and Data in Linguistics*, pp. 43–57. Chicago Linguistic Society.
- Feyaerts, K. (1994). Zur lexikalisch-semantischen Komplexität der Phraseologismen mit phraseologisch gebundenen Formativen. In *Sprachbilder zwischen Theorie und Praxis*, pp. 133–162. Bochum.
- Fillmore, C., P. Kay, and M. O'Connor (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64, 501–538.
- Kürschner, W. (1983). *Studien zur Negation im Deutschen*. Gunter Narr, Tübingen.
- Ladusaw, W. (1980). *Polarity Sensitivity as Inherent Scope Relations*. Garland Press, New York.
- Lambrecht, K. (1990). “What, me worry? — ‘Mad Magazine sentences revisited. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, pp. 215–228. BLS, Berkeley, USA.
- Nunberg, G., I. A. Sag, and T. Wasow (1994). Idioms. *Language* 70, 491–538.
- O’Grady, W. (1998). The Syntax of Idioms. *Natural Language and Linguistic Theory* 16, 279–312.

- Richter, F. and M. Sailer (2003). Cranberry Words in Formal Grammar. In C. Beyssade, O. Bonami, P. C. Hofherr, and F. Corblin (Eds.), *Empirical Issues in Formal Syntax and Semantics*, Volume 4, pp. 155–171. Paris: Presses Universitaires de Paris-Sorbonne.
- Sailer, M. (2003). Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar. Phil. Dissertation (2000). Arbeitspapiere des SFB 340. 161, Universität Tübingen.
- Sailer, M. (2004). Distributionsidiosynkrasien: Korpuslinguistische Erfassung und grammatiktheoretische Deutung. In K. Steyer (Ed.), *Wortverbindungen — mehr oder weniger fest*, Institut für Deutsche Sprache, Jahrbuch 2003, Berlin, New York, pp. 194–221. de Gruyter.
- Soehn, J.-P. (2003). Von Geisterhand zu Potte gekommen. Eine HPSG-Analyse von PPs mit unikalener Komponente. Magisterarbeit, Universität Tübingen, Seminar für Sprachwissenschaft.
- Soehn, J.-P. and M. Sailer (2003). At First Blush on Tenterhooks. About Selectional Restrictions Imposed by Nonheads. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner (Eds.), *Proceedings of Formal Grammar 2003*, pp. 149–161.
- Steyer, K. (2004). Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikographische Perspektiven. In K. Steyer (Ed.), *Tagungsband der 39. Jahrestagung des IDS*. IDS Mannheim, Mannheim. In Vorbereitung.
- van der Wouden, T. (1997). *Negative Contexts. Collocation, Polarity and Multiple Negation*. London: Routledge.
- Witt, A. (2002). *Multiple Informationsstrukturierung mit Auszeichnungssprachen. XML-basierte Methoden und deren Nutzen für die Sprachtechnologie*. Ph. D. thesis, Universität Bielefeld. PDF-Version über den Schriftenserver der Bibliothek der Universität Bielefeld.
- Zwarts, F. (1997). Three Types of Polarity. In F. Hamm and E. W. Hinrichs (Eds.), *Plurality and Quantification*, pp. 177–237. Kluwer Academic Publishers, Dordrecht.