

MANFRED SAILER

Distributionsidiosynkrasien: Korpuslinguistische Erfassung und grammatiktheoretische Deutung¹

Abstract

In diesem Aufsatz werden Wortverbindungen aus einer distributionellen Perspektive im Rahmen einer formalen lexikalistischen Grammatiktheorie betrachtet. Ausgehend von unikalischen Elementen (*Tacheles* reden) wird ein Distributionsmodul als Teil des Lexikoneintrags motiviert. Anhand des Verbs *fackeln* wird eine analoge Distributionsanalyse für Polaritätselemente entwickelt. Da Korpora eine zentrale Datenquelle darstellen, werden die Möglichkeiten diskutiert, Distributionsanforderungen lexikalischer Elemente automatisch aus Korpora zu extrahieren. Um dem Distributionsmodul ein klareres Profil zu geben, wird seine Funktion gegenüber der von Selektion und von Konstruktionen abgegrenzt. Abschließende Überlegungen widmen sich einem Versuch, die Rolle von Gebrauchsdaten innerhalb einer formalen Grammatiktheorie zu bestimmen, was zur Skizzierung einer erfahrungsbasierten modelltheoretischen Grammatiktheorie führt.

1. Einführung

Dieser Aufsatz widmet sich dem Thema der „Wortverbindungen“ aus einer distributionellen Sicht. Das bedeutet, ich werde mich auf die Frage konzentrieren, welche Verbindungen Wörter eingehen können und wie dies im Zusammenhang mit dem Rest der Grammatik zu deuten ist.

Formale lexikalistische Grammatiken gehen von der folgenden Arbeitsteilung aus: Eine kleine Menge von allgemeinen Prinzipien kontrolliert die Verwendung von lexikalischen Elementen in wohlgeformten linguistischen Strukturen. Die lexikalischen Elemente ihrerseits haben idiosynkratische Eigenschaften, wie ihre Phonologie, ihre Bedeutung und ihre syntaktische Kategorie und ihre Selektionseigenschaften. Diese Eigenschaften sind nicht aus allgemeinen Prinzipien ableitbar sondern müssen im Lexikon angegeben werden.

¹ Diese Arbeit entstand im Rahmen des Teilprojekts A5 des von der DFG eingerichteten Sonderforschungsbereichs 441 der Universität Tübingen. Wie auch aus den Literaturverweisen ersichtlich, gehen viele Aspekte auf gemeinsame Arbeiten mit Frank Richter zurück. Ich möchte dem IDS für die Einladung zu diesem Tagungsband danken. Für Kommentare und Diskussion danke ich Frank Richter, Jan-Philipp Soehn und Andreas Wagner sowie den Teilnehmern der IDS-Jahrestagung.

Bei der Betrachtung großer Korpora beobachten wir, dass die momentanen Umsetzungen dieser Grammatikarchitektur einen Phänomenbereich außen vor lassen, der traditionell als Kollokationsbeschränkungen von lexikalischen Elementen beschrieben wird. Wie wir sehen werden, können diese Beschränkungen über das hinausgehen, was als statistische Tendenz zu werten ist, und sie sollten deshalb als Teil des sprachlichen Wissens eines Sprechers betrachtet werden. Da diese Beschränkungen sich nicht mit den Mitteln der Selektion beschreiben lassen, argumentiere ich für eine Integration in ein formales Grammatikmodell, bei dem Distributionsbeschränkungen als Teil der lexikalischen Spezifikation ausdrückbar sind. Hinzu tritt ein Distributionsprinzip in der Grammatik, das die Erfüllung dieser Bedingungen in konkreten linguistischen Strukturen erzwingt. Damit wird ein eigenes Distributionsmodul in die Grammatik eingeführt.

Im Anschluss an die Einleitung beginne ich mit der Darstellung von unikalen Elementen, da sie sehr starken und konkreten Auftretensbeschränkungen unterliegen. Dann wende ich mich Elementen mit schwächeren und abstrakteren Beschränkungen zu. Es wird deutlich werden, dass mit dem Maß an Abstraktheit die grammatiktheoretische Relevanz des Phänomens steigt, dass im gleichen Zug jedoch die Möglichkeit einer korpuslinguistischen Erfassung der Daten abnimmt. Danach werde ich die Stellung des eingeführten Distributionsmoduls erläutern, vor allem hinsichtlich seiner Interaktion mit Selektionseigenschaften sprachlicher Elemente auf der einen Seite und mit dem Konstruktionsbegriff auf der anderen.

Bei der Betrachtung von Distributionsdaten wird die Bedeutung der tatsächlichen Spracherfahrung eines Sprechers für sein linguistisches Wissen besonders deutlich. Deshalb möchte ich ausblickend für eine Synthese von zwei Grammatikmodellen plädieren, die jeweils den Aspekt der „Spracherfahrung“ beziehungsweise den des „sprachlichen Wissens“ betonen: die erfahrungsbasierte und die modelltheoretische Grammatik.

2. Unikale Elemente

Unikale Elemente sind sprachliche Zeichen, die nur in einer oder wenigen Verbindungen vorkommen. Diese Elemente werden auch als *phraseologisch gebundene* oder *phraseologisch isolierte* Einheiten oder als *Unikalia* bezeichnet. Oftmals wird ihnen auch terminologisch jegliche Selbständigkeit abgesprochen, und die Verbindungen, in denen sie auftreten, heißen einfach *Phraseologismen mit unikalener Komponente* (Dobrovol'skij 1988, Fleischer 1989, Häcki Buhofer 2000). In der englischsprachigen Literatur finden sich die Bezeichnungen *hapax legomena* und *cranberry words* (Aronoff 1976), oder auch einfach *bound word* (Moon 1998). Typische Vertreter dieser Gruppe sind:

- (1) a. Tacheles reden
 b. essen wie ein Scheunendrescher
 c. klipp und klar

Diese Wendungen werden in Fleischer (1997), Burger (1998) und anderen Einführungstexten als Phraseologismen *par excellence* bezeichnet, da in ihnen das Kriterium der Fixiertheit am deutlichsten ausgeprägt ist. Dennoch zeigen andere Betrachtungen, dass je nach Klassifikationskriterien die Einordnung dieser Wendungen durchaus problematisch ist (Dobrovol'skij/Piirainen 1994).

2.1 Analysealternativen

Prinzipiell ergeben sich zwei Analysemöglichkeiten für Unikalia, die ich hier plakativ als „Konstruktionsanalyse“ und als „Distributionsanalyse“ (oder „Kollokationsanalyse“) bezeichnen möchte. Die Konstruktionsanalyse behandelt die Wendungen in (1) als phraseologische Einheit und kommt somit ohne einen eigenständigen Lexikoneintrag für das auftretende unikale Element aus. Die Distributionsanalyse weist dem unikal Element einen eigenen Lexikoneintrag zu, muss allerdings verhindern, dass das unikale Element frei auftritt.

Ich will zunächst einige Argumente für eine Konstruktionsanalyse der Wendungen in (1) zusammenstellen, um dann jedoch eine Distributionsanalyse zu propagieren.

Erstens tritt das unikale Element außerhalb der Wendung nicht auf, d. h., es hat keine *wendungsexterne Bedeutung*.

Zweitens spart man durch Betrachtung der Wendung als „Ganzes“, d. h. als Phraseologismus, einen (mental) Lexikoneintrag für das unikale Element. Insbesondere Höhle (1978) argumentiert, dass ein eigener Lexikoneintrag für das Wort *Garaus* uns dazu zwingen würde (i) in diesem Lexikoneintrag die Wendung (*jdm den Garaus machen*) zu erwähnen, um freies Auftreten von *Garaus* zu verhindern; und (ii) einen zusätzlichen Lexikoneintrag für die gesamte Wendung einzuführen. Ein solches Vorgehen wird als unerwünscht bewertet.

Drittens fehlt es bislang in vielen formalen Grammatikmodellen (sei es die Chomsky'sche Schule, Tree Adjoining Grammar (Joshi 1987), Head-Driven Phrase Structure Grammar (Pollard/Sag 1994) oder Construction Grammar (Fillmore et al. 1988)) ein ausgearbeiteter Mechanismus, um Distributionsbeschränkungen auszudrücken. Aus diesem Grund hat sich eine Distributionsanalyse als Alternative auch nicht wirklich gestellt.

Nichtsdestotrotz scheint einiges für eine Dekomponierung der obigen Wendungen zu sprechen: Zum einen treten alle anderen an der Wendung beteiligten Elemente in ihrer „normalen“ Bedeutung auf (bzw. in einer auch wendungsextern möglichen Bedeutung). Zum anderen ist die Verbindung des unikal Elements mit den übrigen Elementen der Wendung syntaktisch regulär (bzw. nicht wendungsspezifisch). Darüber hinaus ist es möglich, dem unikal Element eine Bedeutung zuzuweisen, so dass die Gesamtbedeutung der Wendung sich kompositionell ergibt (bzw. ohne die Annahme wendungsspezifischer Interpretationsmechanismen).

Der letzte Punkt verdient einige Erläuterung. Für eine Bedeutungszuweisung an unikale Elemente spricht das Phänomen der Remotivierung (vergl. Häcki Buhofer (2000) und die dort aufgeführte Literatur). In diesem Zusammenhang ist auch die Entunikalisierung zu nennen, d. h. die Ausbreitung der Verwendungskontexte von „remotivierten“ unikalen Elementen (Häcki Buhofer 2000).

Um die Behauptung aufrecht zu erhalten, unikale Elemente haben eine eigene Bedeutung, sollte diese auch angegeben werden. Für die Elemente in (1) ergibt sich eine Annäherung an diese Bedeutung aus den Paraphrasen in (2).

- (2) a. Tacheles: Klartext
- b. Scheunendrescher: „sehr starker Esser“
- c. klipp: klar, deutlich

Für die Wendung *jdm den Garaus machen* wird in der Regel behauptet, dass sie nicht dekomponierbar sei (Höhle 1978, Dobrovolskij 1999). Dabei ist nicht klar, ob die semantische „Leere“ von *Garaus* unabhängig motiviert ist oder lediglich aufgrund der Unikalität induziert. Unter einer Dekompositionsanalyse ergibt sich, dass das Verb *machen* in dieser Wendung in genau derselben Bedeutung und demselben syntaktischen Rahmen auftritt wie in anderen Funktionsverbgefügen:

- (3) *jdm den Hof/ einen Antrag/ eine Szene machen;*
 jdm/etwas ein Ende machen

Damit ergibt sich für *Garaus* eine Bedeutung, die ähnlich der von *Ende* im letzten Beispiel in (3) ist.

Bei einigen Wendungen mit unikalen Elementen scheinen metaphorisierende Lesarten zwingend:²

- (4) *ins Fettnäpfchen treten; sich ein Armutszeugnis ausstellen; jdn an den Bettelstab bringen, auf Freiersfüßen gehen; jdn am Gängelband führen/ haben; jdm reißt der Geduldsfaden; auf Hochtouren laufen/bringen*

Interessant ist hierbei, dass die Metaphorisierung auch die nicht-unikalen Elemente der Wendung erfasst. Metaphorische Interpretationsstrategien sind nun aber nicht eingeschränkt auf Phraseologismen. Damit liegt also auch hier keine wendungsspezifische Interpretationsstrategie vor. Spricht man andererseits dem unikalen Element jegliche Eigenbedeutung ab, so können die Wendungen nicht mehr als metaphorisch gewertet werden.

In einem Fall scheint mir tatsächlich keine Bedeutungszuweisung möglich:

- (5) *jdn über den Löffel balbieren (jdn in plumper Weise betrügen‘)*

² In der momentanen Sammlung von unikalen Elementen im Projekt A5 sind 33 von 268 (= 12,3 %) Wendungen mit unikalen Elementen als metaphorisch eingestuft.

Hier ist *balbieren* eine Nebenform zu *barbieren*. Bei dieser Wendung lässt sich keine synchrone metaphorische Motiviertheit mehr feststellen, ähnlich wie bei der Wendung *den Löffel abgeben*. Das heißt, wir haben es hier mit einer nichtdekomponierbaren Wendung zu tun, einer Wendung also, in der die Gesamtbedeutung nicht sinnvoll auf die wendungskonstituierenden Elemente verteilt werden kann.

2.2 Distributionsanalyse von unikalen Elementen

Wenn wir nun annehmen, dass unikale Elemente sich wie andere sprachliche Zeichen auch verhalten, d. h. syntaktische und semantische Eigenschaften und insbesondere eine eigene Bedeutung haben, dann ergeben sich die Eigenschaften der Wendungen in (1) auf natürliche Weise. Es bleibt natürlich das Problem, die Unikalität zu erfassen.

Hierzu gibt es bereits etliche Ansätze, wie die Behandlung dekomponierbarer Idiome (der *Idiomatically Combining Expressions* von Wasow et al. (1983)) in der Generalized Phrase Structure Grammar (Gazdar et al. 1985) mittels *partieller Bedeutungsfunktionen*. Ähnlich wurde zur Behandlung von bestimmten Kollokationsphänomenen eine kontext-sensitive Semantik angenommen (Keenan 1974, Partee 1984). In diesen Ansätzen wird die Bedeutung eines Ausdrucks als disjunktive Bedeutungsfunktion aufgefasst, wobei die Disjunkte jeweils mögliche kontextuell unterschiedliche Bedeutungen ausdrücken und im Zweifelsfall die Bedeutungsfunktion in manchen Kontexten nicht definiert ist. Bei unikalen Elementen wäre letzteres in den meisten Kontexten der Fall.³ So werden die beiden scheinbar widersprüchlichen Eigenschaften von unikalen Elementen auf natürliche Weise erfasst: dass Unikalia keine wendungsexterne Semantik haben, innerhalb der Wendung, in der sie auftreten aber sehr wohl eine Semantik zu haben scheinen.

Diese beiden Distributionsanalysen für unikale Elemente formulieren die Auftretensbeschränkungen als Beschränkungen im semantischen Modell. Neben einigen inhärenten Komplikationen dieses Ansatzes (siehe Pulman 1993) gibt es auch Evidenz, dass wir es mit Beschränkungen zu tun haben, die eher mit Bezug auf die linguistische Struktur formuliert werden sollten, als mit Bezug auf die modelltheoretische Interpretation der beteiligten Wörter. Ich will hier kurz Beispiele für drei derartige Beschränkungen geben: Lexembeschränkungen, syntaktische und phonologische Beschränkungen.

Lexembeschränkungen: Etliche unikale Elemente treten in Funktionsverbgefügen auf. Hierbei sind sie auf ein bestimmtes Verb beschränkt. Es ist nicht klar, dass hier auf der Ebene des semantischen Modells ein Unterschied zwischen den Funktionsverben besteht.

- (6) a. jdm den Garaus machen/ *tun/ *treiben
- b. seine Schuldigkeit *machen/ tun/ *treiben
- c. Schindluder *machen/ *tun/ treiben

³ Diese Analyseoption erwähnt auch van der Wouden (1997, S. 46, Fn. 56).

Ähnliches gilt bei der Beschränkung auf ein oder wenige bestimmte Lexeme aus einer Gruppe von Plesionymen:⁴

- (7) a. Tacheles reden/ *sprechen
 b. jdn Mores lehren/ *unterrichten
 c. jdm Angst/ einen Schrecken/ ?Furcht/ *Entsetzen einjagen

Syntaktische Beschränkungen: Die Literatur zu syntaktischen Beschränkungen auf Phraseologismen zeigt (Wasow et al. 1983, Nunberg et al. 1994, Schenk 1995, Dobrovolskij 1999), dass je stärker ein Phraseologismus dekomponierbar ist, er desto freier ist hinsichtlich der syntaktischen Konstruktionen in denen er auftreten kann. Dennoch bleibt hier ein gewisser Rest an idiosynkratischem Verhalten. Dasselbe gilt für Wendungen mit unikalen Elementen. Auch wenn diese Wendungen in den Formen, die auftreten, syntaktisch und semantisch „regulär“ sind, bedeutet das nicht, dass sie in jeder prinzipiell möglichen syntaktischen Konstellation auftreten können. Als Beispiel sei hier die Modifizierbarkeit von unikalen Nomen angeführt, die sich nur mit einer bestimmten Präposition verbinden können.⁵

Sowohl *Haaresbreite* als auch *Daffke* sind unikal. Beiden kann prinzipiell eine Bedeutung zugeordnet werden, wobei keines von beiden referentiell ist. Dennoch kann *Daffke* modifiziert werden, *Haaresbreite* hingegen nicht.

- (8) a. um (*knappe/bloße/...) Haaresbreite
 b. aus (reinem) Daffke

Phonologische Beschränkungen: Im Niederländischen ist die reguläre Form des Possessivpronomens der 3. Person Singular *haar* (ihr-). In Verbindung mit dem Lexem *haar* (*Haar*) tritt jedoch eine Alternativform des Possessivpronomens auf, *heur*. Es ist klar, dass diese Beschränkung primär phonologisch motiviert ist. Allerdings tritt die Form *heur* nicht nur unter Adjazenz zum Nomen *haar* auf:⁶

- (9) a. haar krullerige kapsel
 ihre kräuselige Frisur
 b. heur krullerige haar
 ihr kräuseliges Haar (http://www.spaink.net/groene/net_helden.html)

⁴ Eine Suche auf Google ergab etliche Belege für *Furcht einjagen* und einen einzigen Beleg für *Entsetzen einjagen*.

⁵ Ich danke Jan-Philipp Soehn für Diskussion dieser Daten. Siehe auch Soehn (2003) für eine Auswertung der Daten zu diesem Typ Unikalia in den öffentlich zugänglichen IDS-Korpora geschriebener Sprache.

⁶ *Heur* ist eine südliche Variante zu *haar*. Die im Text besprochene Einschränkung gilt jedoch für die Standardsprache. Ich danke Ton van der Wouden dafür, mich auf diese Besonderheit des Niederländischen hingewiesen zu haben.

Aus dem Genannten ergibt sich die Schlussfolgerung, dass die Distributionsbeschränkungen eines unikaligen Elements mit Bezugnahme auf Teile der linguistischen Struktur, in denen es auftritt, formuliert werden können und müssen, und nicht adäquat als Anforderungen an die Interpretation von semantischen Konstanten gesehen werden kann.

Für einen Lexikoneintrag eines unikaligen Elements ergibt sich damit die in (10) angedeutete Struktur. Wie bereits in der Einleitung erwähnt, hat der Lexikoneintrag ein Feld für die phonologische Repräsentation, die syntaktische Kategorie, die semantische Repräsentation sowie die Selektionseigenschaften.

(10) Beispielllexikoneintrag für *Tacheles*:

Phonologische Repräsentation/	taxeles/
Syntaktische Kategorie	N
Semantische Repräsentation	tacheles'
Selektionsanforderungen	—
Distributionsanforderungen	Themaargument zum Lexem <i>reden</i>

Neu hinzu tritt jedoch das Feld Distributionsanforderungen. Hier werden die Anforderungen an die linguistische Umgebung kodiert, die in keinem der anderen Felder untergebracht werden können. Im Falle von *Tacheles* fordern wir, dass es als Themaargument zum Lexem *reden* auftreten muss.

Es muss noch gewährleistet werden, dass die angegebenen Distributionsanforderungen auch tatsächlich in jeder Äußerung, die ein Auftreten von *Tacheles* hat, erfüllt sind. Hinzu erweitere ich die Grammatik um ein allgemeines Distributionsprinzip:

(11) Distributionsprinzip:

Alle Distributionsanforderungen der lexikalischen Elemente, die in einer Äußerung Σ auftreten, sind in Σ erfüllt.

Die Formulierung des Prinzips in (11) erlaubt es, mit einem einzigen Prinzip die Erfüllungsbedingung für die Distributionsbeschränkungen aller unikaligen Elemente zu erfassen.⁷

Wenn eine solche Grammatikarchitektur anvisiert wird, stellt sich natürlich die Frage, ob Distributionsanforderungen auch für andere sprachliche Elemente als Unikalia zu formulieren sind und auf welche Aspekte der linguistischen Struktur in Distributionsanforderungen Bezug genommen werden kann. Im Folgenden wollen wir anhand eines Wortes, des Verbs *fackeln*, die Notwendigkeit und Nützlichkeit einer Distributionskomponente demonstrieren. Dabei erweist sich die Arbeit mit Korpora und deren Ergänzung um introspektive Urteile als besonders wichtig.

⁷ In Sailer (2000), Richter/Sailer (2003) und Sailer/Richter (2002a, b) wurde ein solches Distributionsmodul für ein konkretes formales Grammatiksystem, die HPSG (Pollard/Sag 1994) ausformuliert.

3. Weitere distributionell beschränkte Elemente

In diesem Abschnitt werde ich das Verb *fackeln* in der Bedeutung von *zögern* diskutieren und dann einen gut untersuchten Typus von distributionell beschränkten Elementen betrachten: Polaritätselemente.

3.1 Das Verb *fackeln*

3.1.1 Daten

In Sailer/Richter (2002a,b) wird das Distributionsverhalten des Verbs *fackeln* auf Basis von 280 Korpusbelegen und weiteren introspektiven Daten untersucht. Die dort berichteten Ergebnisse werden von einer neueren Untersuchung bestätigt, die von der bis Februar 2003 auf 650 gewachsenen Zahl von Belegstellen in den über das COSMAS-I Portal öffentlich zugänglichen Korpora des IDS ausgeht.⁸

Das Verb *fackeln* in der Bedeutung *zögern*, *schwanken* tritt im COSMAS-I Korpus 650 Mal auf. In 84 % (546) der Fälle in Kombination mit *nicht lange*.

Tabelle 1:

Auftretensmuster von *fackeln*

	<i>nicht</i>		<i>nicht mehr</i>		<i>ohne</i>		<i>nie, etc</i>		kein Neg		Σ	
	#	%	#	%	#	%	#	%	#	%	#	%
<i>lange</i>	546	84.129	11	1.694	47	7.242	12	1.848	3	0.462	619	95.378
anderer DM	1	0.154	-	0	-	0	2	0.308	-	0	3	0.308
<i>viel</i>	-	0	-	0	1	0.154	-	0	-	0	1	0.154
kein DM	16	2.465	1	0.154	5	0.770	1	0.154	4	0.616	27	4.160
Σ	563	86.661	12	1.849	53	8.166	15	2.310	7	1.077	650	100

(12) Der neue Geschäftsführer fackelte nicht lange.

Dennoch ist *fackeln* nicht auf diese beiden Elemente eingeschränkt. Tatsächlich kann *fackeln* in einem großen Spektrum „negativer“ Kontexte auftreten und auch der durative Modifikator kann variieren. Das Korpus beinhaltet beispielsweise die folgenden Belege:

- (13) a. Es gebe zwei Punkte, in denen die Politesse keine Minute fackle.
b. Keine Sekunde tät ich da fackeln.

Die Befunde sind in Tabelle 1 zusammengefasst. Insgesamt tritt in 95,9% (623) der Korpusbelege ein durativer Modifikator auf, ein negativer Kontext liegt sogar in 99,2% (645) der Fälle vor.

⁸ COSMAS-I war bis zum 7.3.2003 erreichbar über <http://corpora.ids-mannheim.de/~cosmas/>. Die Korpusergebnisse und die Kollokationsauswertung, von denen hier berichtet wird, beziehen sich auf den damaligen Stand der IDS-Korpora. Seit März 2003 sind die IDS-Korpora über COSMAS-II, das Nachfolgeportal zu COSMAS-I, erreichbar. Zum Zeitpunkt der Endfassung dieses Aufsatzes ist COSMAS-II nur über kommerzielle Betriebssysteme aufrufbar, so dass es mir nicht möglich war, die Aussagen in Bezug auf neuere Korpusbelege zu überprüfen

Aufgrund der Natur von Korpora tritt natürlich nicht jeder mögliche Negationsträger mit jedem möglichen durativen Modifikator auf. Dies zeigt sich an den freien Feldern in Tabelle 1. An dieser Stelle kann nur mit introspektiven Daten weitergearbeitet werden. Aus den introspektiven Daten er-

Tabelle 2:

Introspektive Bewertung der Auftretensmuster von *fackeln*

	<i>nicht</i>	<i>nicht mehr</i>	<i>ohne</i>	<i>nie, etc</i>	schwache Neg	kein Neg
<i>lange</i>	!	!	!	!	!	?*
anderer DM	!	!	!	!	!	?*
<i>viel</i>	!	!	!	!	?*	*
kein DM	?*	?	?*	?	*	*

gibt sich das Bewertungsmuster in Tabelle 2.⁹

Ein Beispiel für eine nicht im Korpus auftretende Konstellation, die aber dennoch vollkommen grammatisch ist, ist in (14) gegeben:

(14) Unter solchen Umständen fackelt eine Politesse nicht eine Minute.

Fehlt sowohl Negation als auch ein durativer Modifikator, so ist ein Auftreten von *fackeln* ausgeschlossen:

(15) a. ?* Unter solchen Umständen fackelt eine Politesse beim Ausstellen von Strafzetteln.

b. * Als die ersten Schwierigkeiten auftraten, fackelten die Organisatoren, statt gleich nach einem Ausweichtermin zu suchen.

3.1.2 Analyse

In Sailer/Richter (2002a, b) wurde gezeigt, dass keine der beiden Kollokationsanforderungen von *fackeln* auf bereits standardmäßig vorhandene Information zum Verb oder auf unabhängige allgemeine Prinzipien zurückgeführt werden kann.

Aus der semantischen Nähe zu *zögern* ergibt sich, dass der durative Modifikator nicht als Komplement von *fackeln* betrachtet werden kann. Die Möglichkeit, dass ein solcher Modifikator auftritt, folgt aus der Aktionsart von *fackeln*. In negierten Sätzen ist darüber hinaus ein durativer Modifikator sowieso immer möglich. Allerdings folgt aus der Möglichkeit noch nicht, dass ein Modifikator auch tatsächlich auftreten muss, wie es jedoch sowohl die introspektiven als auch die Korpusdaten zeigen.

Es ist außerdem nicht möglich die obligatorische Präsenz des Modifikators auf unabhängige pragmatische Prinzipien zurückzuführen. Ein solches Verfahren schlägt Steinitz (1988) für Verben mit sehr schwachem semantischem Eigenbeitrag vor. Das Verb *fackeln* kann jedoch nicht als semantisch schwach aufgefasst werden, womit auch diese Option entfällt.

⁹ Hierbei steht „!“ für vollständig grammatisch, „*“ für ungrammatisch. „?“ und „?*“ sind zwei dazwischen liegende Werte.

Das gerade Gesagte lässt sich auch auf die Negationsanforderung von *fackeln* übertragen. Hier ist die Form und auch die syntaktische Funktion des Negationsträgers im Satz frei. Damit lässt sich die Negationsanforderung nicht als Selektionseigenschaft kodieren, davon abgesehen, dass der Negationsträger auch in nicht selezierter Position, zum Beispiel als Adjunkt (*nicht, nie, keinen Moment*), auftreten kann.

Es ist auch eher zweifelhaft, ob sich die Negationsanforderung auf pragmatische Prinzipien zurückführen lässt. Damit scheidet also auch für die Negationsanforderung eine alternative Erklärung aus.

Bleibt noch der Versuch einer Konstruktionsanalyse für die Verwendungen von *fackeln*. Erschien ein solches Vorgehen im Falle von unikalen Elementen noch auf der Hand liegend, ist es hier, wenn überhaupt zweite Wahl. Nach dem Konstruktionsbegriff der Construction Grammar (Fillmore et al. 1988, Kay 1998, Kay/Fillmore 1999) ist eine Konstruktion eine idiosynkratische Verbindung von Form und Bedeutung. Die Verbindung des Verbs *fackeln* mit dem durativen Modifikator weist jedoch, wie oben gezeigt keinerlei syntaktische oder semantische Idiosynkrasien auf.

Wenn wir nun alle anderen Analysealternativen verwerfen müssen, bleibt nur noch eine Distributionsanalyse analog zur vorgeschlagenen Behandlung von unikalen Elementen. Eine solche Analyse reduziert sich auf die Formulierung eines Lexikoneintrags, in dem die Distributionsbeschränkungen explizit erfasst sind. In Interaktion mit dem Distributionsprinzip ergibt sich die richtige Verteilung von *fackeln*. Ein Lexikoneintrag von *fackeln* kann dann wie folgt aussehen:

(16) Lexikoneintrag für *fackeln*:

Phonologische Repr.	/fak-/
Syntaktische Kat.	V
Semantische Repr.	fackeln' (ähnliche Bedeutung wie <i>zögern</i>)
Selektion	NP.subj
Distribution	logische Form der Äußerung: ... NEG [... Dur.MOD [... fackeln' ...] ...] ...

Fackeln ist ein intransitives Verb, dessen Bedeutung ähnlich der von *zögern* ist. Im Distributionsfeld ist kodiert, dass die semantische Konstante, die *fackeln* einführt, in der logischen Form der Äußerung im Skopus eines durativen Modifikators und eines Negationsoperators stehen muss.

3.2 Polaritätselemente

Das Beispiel *fackeln* hat gezeigt, dass lexikalische Elemente in Bezug auf bestimmte Operatoren in der semantischen Repräsentation distributionell beschränkt sein können. Ein sehr prominenter derartiger Operator ist die Negation. Negationssensitive Elemente, d. h. *negativ poläre Elemente* (NPEs)

und *positiv poläre Elemente* (PPEs) müssen, beziehungsweise dürfen nicht, im Skopus eines negativen Operators stehen.

- (17) a. Niemand/ *Peter versteht auch nur irgendwas von Filmen. (NPE)
 b. *Niemand/ Peter ist schon gekommen. (PPE)

Diese Elemente werden seit Klima (1964) in großem Umfang untersucht. Eine Sammlung von Polaritätselementen im Deutschen findet sich in Kürschner (1983).

Die Vielzahl von Arbeiten zu Polaritätselementen beinhaltet auch einige Ansätze, bei denen Polaritätselemente explizit als kollokationell beschränkt angesehen werden. Am deutlichsten ist dies in van der Wouden (1997) ausgearbeitet. In dieser Arbeit wird von einer kontext-sensitiven Semantik für lexikalische Elemente ausgegangen. Die Interpretation von Polaritätselementen ist dann nur unter bestimmten (semantischen) Eigenschaften des Kontexts definiert.

Die frappierende Analogie zwischen Negationssensitivität und anderen Distributionsanforderungen ergibt sich jedoch vor allen Dingen bei einer genaueren Betrachtung der Auftretenskontexte einzelner Polaritätselemente. Bereits im Falle von *fackeln* haben wir gesehen, dass die Negationssensitivität nur eine von zwei Distributionsbeschränkungen des Verbs ist.

Eine aus der deutschen phraseologischen Literatur (Dobrovolskij 1988, Fleischer 1997, Dobrovolskij/Piirainen 1994) zusammengetragene Liste von unikalen Elementen beinhaltet etliche negationssensitive Elemente. In (18) finden sich einige Beispiele.

- (18) kein(en) Hehl aus etwas machen; aus seinem Herzen keine Mördergrube machen; etwas nicht wahrhaben wollen; jdn nicht ausstehen können; sich nicht lumpen lassen; sein Lebtag nicht; nicht viel Federlesens machen

Nehmen wir aus dieser Sammlung das unikale Element *Hehl* heraus. Dieses Wort ist ein Nomen, dessen Bedeutung in etwa die von *Geheimnis* ist. Es ist distributionell eingeschränkt auf die Verbindung mit dem Funktionsverb *machen*, das auch in der freien Verbindung *aus etwas ein Geheimnis machen* auftritt. Darüber hinaus muss *Hehl* in negativen Kontexten stehen.

Umgekehrt stellt sich bei Elementen, die normalerweise nicht in phraseologischen Sammlungen als Unikalia genannt werden, sondern eher in Listen von Polaritätselementen auftreten, in detaillierten Einzelstudien ihrer Auftretensmuster heraus, dass sie oft nicht nur negationssensitiv sind, sondern weiteren distributionellen Beschränkungen unterliegen. Besonders deutlich ist das in den Arbeiten von Jack Hoeksema zum Niederländischen formuliert (Hoeksema 1996, 2001a, 2002a). Fürs Deutsche ist mit den Studien zum Verb *fackeln* ein solcher Fall relativ robust etabliert (Sailer/Richter 2002a, b). Vergleichbare Studien zu anderen deutschen Polaritätselementen fehlen. Allerdings legt eine Betrachtung der Beobachtungen zum Niederländischen nahe, dass im Deutschen ähnliche Verhältnisse herrschen.

Eine relativ homogene Gruppe von Polaritätselementen bilden die sogenannten *Minimizers*.

- (19) a. Er hat kein Sterbenswörtchen gesagt.
 b. Zur Demo ist kein Schwein gekommen.
 c. Van Gogh hat mit seinen Bildern keine müde Mark verdient.
- (20) a. Er hat kein Sterbenswörtchen/ *Schwein gesagt.
 b. Zur Demo ist kein *Sterbenswörtchen/ Schwein gekommen.

Semantisch kann man die Minimizers alle gleich behandeln, nämlich als Indefinita. In den Beispielen in (19) ist jedoch schon intuitiv naheliegend, dass sie sich hinsichtlich der Prädikate, mit denen sie sich verbinden stark unterscheiden.

Untersuchungen für das Niederländische, das besonders reich an Minimizers ist, zeigen, dass in dieser Gruppe, eine starke Spezialisierung auftritt (Hoeksema 2001b, 2002b). Auch in einer diachronen Studie (Hoeksema 2001b), finden sich analoge Ergebnisse: Minimizers tendieren dazu, sich auf Verbindungen mit Prädikaten mit bestimmten lexikalisch-semantischen Klassen zu spezialisieren.

Aus den Beobachtungen zu Polaritätselementen ergibt sich, dass es sich hierbei um lexikalische Elemente handelt, die zweifelsohne einen eigenen Lexikoneintrag und eine inhärente Bedeutung haben. Diese Elemente sind dennoch von ihrem Auftretenspotenzial her stark beschränkt; in erster Linie auf negative Kontexte, aber darüber hinaus auch auf weitere Eigenschaften der linguistischen Umgebung.

Daraus ergibt sich, dass Polaritätselemente und unikale Elemente sich nur graduell unterscheiden, indem im Falle der Unikalia die Distributionsbeschränkungen in der Regel lokaler und weniger abstrakt sind. Die Existenz von Polaritätselementen unter den unikalen Elementen, wie *Hehl*, zeigt deutlich, dass hier kein prinzipieller Unterschied vorliegt.

4. Distributionsidiosynkrasien in Korpora

Zur Etablierung der Distributionsbeschränkungen von unikalen Elementen und vor allem auch bei Polaritätselementen ist es von zentraler Bedeutung, Korpusdaten und introspektive Daten miteinander zu verbinden. Durch die Korpusdaten wird die Sensitivität für bestimmte Auftretensbeschränkungen geweckt. Durch die introspektiven Daten können diese Beschränkungen dann genauer geprüft werden, so dass sich am Ende robuste und aussagekräftige Generalisierungen über die Auftretensmöglichkeiten eines Elements ergeben. Dies wurde am Beispiel des Verbs *fackeln* illustriert.

Dabei stellen sich zwei zentrale Fragen im Hinblick auf die Nutzbarkeit von Korpora: Erstens, wie exakt können Distributionsidiosynkrasien aus dem Korpus extrahiert werden? Zweitens, können weitere distributionell beschränkte Elemente automatisch erkannt werden?

Betrachten wir zunächst unser bisheriges Vorgehen. Sobald ein bestimmtes Lexem, wie das Verb *fackeln*, als potenziell interessant erkannt ist, kann zur Datenerhebung relativ problemlos das COSMAS-I Korpus oder ein beliebiges anderes herangezogen werden. Das in COSMAS integrierte Kollokationswerkzeug arbeitet auf Wortformen- und Lexembasis, so dass lediglich die beiden Kernfälle *nicht lange* und *ohne lange zu* erkannt werden. Ein Blick in die Ergebnisübersicht der Kollokationsermittlung von COSMAS-I, wiedergegeben in Abbildung 1, macht das deutlich. In der ersten Spalte finden sich die Belegnummern, in der zweiten das statistische Maß, nach dem die Stärke der Verbindung von *fackeln* und seinen Kollokatoren angegeben ist. Die dritte Spalte gibt die Kollokatoren an und die vierte schließlich die Anzahl der entsprechenden Korpusbelege.¹⁰

Das bedeutet im Hinblick auf unsere Leitfragen, dass wir bei der Verwendung der meisten Korpora und Kollokationswerkzeuge lediglich auf Wortformen und Lexeme beschränkt sind, sowohl was die Erkennung von Distributionskontexten anbelangt, als auch was die Erkennung weiterer distributionell beschränkter Elemente angeht.

Es hat sich jedoch gezeigt, dass Auftretensbeschränkungen in Form von Sensitivität auf eine bestimmte Wortform oder ein bestimmtes Lemma nur einen Teil dessen ausmachen, was an Auftretensbeschränkungen von lexikalischen Elementen bislang bereits etabliert wurde. Dies habe ich am Beispiel *fackeln* erläutert. Zur empirischen Bestätigung oder auch Falsifizierung des hier vertretenen Distributionsansatzes ist es notwendig, systematisch nach weiteren Arten von Distributionsbeschränkungen zu suchen. Als natürliche Hypothese aus den betrachteten Daten ergibt sich die Annahme, dass es noch deutlich mehr Gruppen von distributionell beschränkten Elementen geben muss als die bisher diskutierten.

Im Fall von *fackeln* hat sich eine bestimmte Konstellation von Konstanten in der semantischen Repräsentation des Satzes als relevant erwiesen. Dabei treten zwei Arten von Generalisierungen über die konkreten Sätze auf: Generalisierungen über die semantische Klasse bestimmter Ausdrücke und Generalisierungen über bestimmte Operatorkonstellationen in der semantischen Repräsentation.

Für Generalisierungen des ersten Typs müsste im Falle von *fackeln* über die Arten des Modifikators abstrahiert werden: *Moment*, *Minute*, *Augenblick*, ... Im Falle von *Sterbenswörtchen* entsprechend über die Art des Prädikats. In lexikalisch-semantischen Netzen wie dem GermaNet ist eine solche Generalisierung möglich, da semantische Beziehungen zwischen einzelnen Lexikoneinträgen entsprechend kodiert sind. Vielsprechend ist in diesem Zusammenhang auch die Arbeit des Projekts SALSA (*Saarbrücken Lexical Semantics Annotation and Analysis*, Erk et al. 2003). Im Rahmen dieses Pro-

¹⁰ Bei den Daten sind auch Verwendungen von *(ab)fackeln* mitberücksichtigt, die wir in der vorhergehenden Diskussion ausgeschlossen haben.

Abbildung 1:
Ergebnis der Kollokationsauswertung zu *fackeln* in COSMAS-I

BelegNr	LLR	Kollokatoren	Häufigkeit
1+1:	4387	lang nicht sondern Der und	1
2+4:		lang nicht sondern und	4
6+13:		lang nicht sondern	13
19+26:		lang nicht Der und	26
45+25:		lang nicht Der	25
70+200:		lang nicht und	200
270+253:		lang nicht	253
523+10:		lang ohne und	10
533+35:		lang ohne	35
568+1:		lang sondern	1
569+2:		lang Der	2
571+3:		lang und	3
574+12:		lang	12
586+1:	1344	nicht herum	1
587+2:		nicht Der	2
589+7:		nicht und	7
596+24:		nicht	24
620+12:	130	ohne	12
632+2:	107	herum	2
634+39:	54	ab	39
673+2:	38	Ball	2
675+1:	27	jährig	1
676+1:	17	sondern	1
677+2:	15	Handball	2
679+1:	14	Feuerwerk	1
680+1:	9	Flamme	1
681+3:	7	da	3
684+3:	7	Feuer	3
687+1:	7	Minute	1
688+8:	6	und	8
696+1:	6	also	1
697+1:	6	garen	1
698+43:		statistisch unspezifisch	43

jekts wird ein bereits syntaktisch annotiertes Korpus um eine lexikalisch-semantische Annotation angereichert, die sich auf die Klassifikation in FrameNet (Baker et al. 1998) bezieht.

Eine Extraktion von Generalisierungen über lexikalisch-semantische Information findet sich bereits in Ansätzen zur automatischen Extraktion von semantischen Subkategorisierungseigenschaften (Resnik 1993, Wauschkuhn 1999, Wagner 2000, 2002). Exemplarisch für solche Ansätze sei hier kurz Wagner (2002) dargestellt. Das englische Verb *eat* hat eine starke semantische Selektionspräferenz für sein Thema-Arguments auf Lebensmittel. Die Begriffe in (Euro)WordNet sind hierarchisch angeordnet, und dort gibt es auch den für unser Beispiel relevanten Oberbegriff, FOOD. Wenn nun im

Korpus Belege gefunden werden für *eat+potatoes* und *eat+milk chocolate*, dann soll daraus mit Hilfe eines informationstheoretisch motivierten statistischen Verfahrens der geeignetste Knoten im semantischen Netz gefunden werden, der mit den Korpusbelegen kompatibel ist. Würde statt FOOD ein allgemeinerer Knoten gewählt, z. B. ENTITY, so wären zwar die Korpusbelege auch erfasst, allerdings wäre der Knoten zu generell. Das gewählte informationstheoretische Verfahren soll die notwendige Einschränkung gewährleisten.

Es sei in Erinnerung gerufen, dass wir bei der Evaluation der *fackeln*-Belegstellen ganz analog vorgegangen sind: Durch konstruierte Beispiele wurde getestet, ob die angenommenen Generalisierungen zu viel zulassen, oder auf der richtigen Ebene im semantischen Netz ansetzen.

Das Problem der automatischen Extraktion von lexikalisch-semantischen Beschränkungen, das sich für Distributionsdaten stellt, ist damit nur eine Verallgemeinerung eines bereits gut untersuchten Falls. Die automatische Erkennung der richtigen Selektionsgeneralisierung liegt im Falle von subkategorisierten Elementen bei etwa 77%, gemessen an einem aus EuroWordNet extrahierten *Gold Standard*. Dabei sind im Bereich der Auswertung und Bewertung der Ergebnisse noch viele Fragen offen.

Was die direkte Übertragbarkeit der Methoden und Ergebnisse zur Extraktion von Selektionspräferenzen auf allgemeinere Distributionsphänomene zusätzlich erschwert, ist das Folgende: Bei der Bestimmung semantischer Selektionseigenschaften ist die syntaktische Konstellation, innerhalb derer nach Generalisierungen gesucht wird, von vornherein klar. Ausgehend von einem Verb werden seine syntaktischen Komplemente auf ihre semantischen Eigenschaften hin geprüft. In unserem Fall ist es a priori nicht entschieden, welche syntaktische Funktion im Satz untersucht werden muss. Im Falle von *fackeln* war es eine Subklasse von Modifikatoren. Bei *Sterbenswörtchen* handelte es sich um das Prädikat.¹¹

Im Falle der Negationsanforderung von *fackeln* und anderen Polaritätselementen ist die syntaktische Realisierung des geforderten Kontextelements gänzlich frei. Hier hilft eine syntaktische Annotation nur begrenzt. Vielmehr geht es um „logische Form“-Annotation, da es sich um Sensitivität bezüglich „satzsemantischer“ Eigenschaften handelt, und nicht „lexikalisch-semantischer“ Eigenschaften. So unterscheiden sich beispielsweise die Wörter *niemand*, *jeder* und *jemand* nicht hinsichtlich ihrer lexikalischen Semantik, die in jedem Falle so etwas wie *Person* bezeichnet. Alle drei können in Argumentstellen von Prädikaten auftreten, die eine lexikalisch-semanti-

¹¹ Untersuchungen zu Unikalien deuten darauf hin, dass prinzipiell alle direkten Dependenten des zu betrachtenden Elements in Frage kommen, ebenso wie alle Wörter, von denen das zu betrachtende Element über eine rekursive Dependenzbeziehung abhängt (Soehn 2003, Soehn/Sailer 2003). Zu ähnlichen Überlegungen vergleiche auch Kuiper/Everaert (2000).

sche Beschränkung auf Menschen haben. Nichtsdestotrotz unterscheiden sich die drei genannten Wörter was ihre satzsemantischen Eigenschaften betrifft. Dies zeigt sich gerade bei ihrer Kombinierbarkeit mit Polaritätselementen.

(21) a. Niemand/ Jeder/ *Jemand, der auch nur irgendwas von Filmen versteht, mag „Lola rennt“.

b. Niemand/ *Jeder/ *Jemand hier versteht auch nur irgendwas von Filmen.

So erlaubt *niemand* das Auftreten eines NPEs sowohl in seinem Restriktor als auch in seinem Skopus. *Jeder* kann nur in seinem Restriktor ein NPE lizensieren, und *jemand* überhaupt nicht (Ladusaw 1980).

Zur automatischen Bestimmung von Negationsbeschränkungen lexikalischer Elemente müsste ein Korpus hinsichtlich der Negativität einzelner Wörter annotiert sein. Da eine große Zahl von verschiedenen negativen Kontexten bekannt ist, und auch bestimmte Hierarchisierungen hier bereits etabliert sind, ist ein solches Unternehmen durchaus Erfolg versprechend. Als Grundlage für eine solche Annotation kann zunächst die Klassifikation der Grade von Negativität dienen, wie sie von Frans Zwarts ausgearbeitet wurde (Zwarts 1981, 1986, 1995, 1997). Hier werden vier Grade unterschieden: *Antimorphizität*, die durch die Negationspartikel *nicht* hervorgerufen wird. *Antiadditivität* liegt bei sogenannten N-Wörtern vor (*niemand*, *niemals*). Fallende Monotonie schließlich ist eine schwache Form von Negativität. Sie kann durch Elemente wie *wenige* oder *höchstens* ausgedrückt werden. Für manche Elemente scheint eine andere schwache Form der „Negativität“ zu genügen: *Nonveridikalität* (Giannakidou 1998).

Eine solche Negationsannotation ist prinzipiell machbar, allerdings bislang noch in keinem Korpus durchgeführt. Durch sie wäre es möglich, eine Liste von polaritätssensitiven Elementen aus dem Korpus zu extrahieren und die Stärke der Negationsanforderung gemäß der angedeuteten Hierarchien zu bestimmen.

Wenn der hier verfolgte Ansatz sich empirisch bewähren soll, ist es wichtig, zu zeigen, dass auch andere satzsemantische Operatoren für Distributionsbeschränkungen konstitutiv sein können. Kandidaten hierfür sind Quantoren ebenso wie Tempus-, Modus- und Frageoperatoren.

Wie bei der Negation ist auch hier eine Annotation möglich. Bei Quantoren lässt sich ausgehend von syntaktischen Strukturen eine feinere Annotation vornehmen. Dabei kann jedoch aus praktischen Gründen kein Quantorenskopus annotiert werden, da dieser in der Regel nicht eindeutig disambiguiert werden kann. Dieses Problem ist in der reichen Literatur zur Unterspezifikation in der Semantik diskutiert (Reyle 1993, Pinkal 1996) und in den Beispielen in (22) in Anlehnung an Pafel (1998) illustriert.

(22) a. Zwei von diesen Sonaten kennt jeder Pianist. (2 Lesarten)

b. Auch nur eine von diesen Sonaten kennt kein Pianist. (1 Lesart)

Satz (22a) hat zwei Lesarten: in der einen Lesart gibt es zwei bestimmte Sonaten, die jeder Pianist kennt. Hier hat also der Quantor im Subjekt engen Skopus. In der anderen ist es für jeden Pianisten so, dass er zwei Sonaten kennt, wobei diese nicht für alle Pianisten dieselben sein müssen. Bei dieser Lesart hat der Quantor an Subjektposition weiten Skopus. Im Gegensatz dazu ist der (b)-Satz skopuseindeutig. Hier gibt es nur eine Lesart, bei der das Subjekt weiten Skopus hat. Dies liegt daran, dass das Polaritätselement *auch nur* im Skopus der durch *kein* eingeführten Negation stehen muss.

Die Sätze illustrieren, dass zwar im allgemeinen Sätze mit mehreren semantischen Operatoren, wie Quantoren und Negation, skopusambig sind, dass es aber im Einzelfall sein kann, dass durch die besonderen Distributionseigenschaften einzelner Elemente bestimmte Lesarten ausgeschlossen sind. Für die Erkennung eines Polaritätselementes wäre es also prinzipiell sehr wohl notwendig, alle möglichen Lesarten anzugeben; sonst könnte beispielsweise das NPE in (22b) als nicht-polär interpretiert werden, d. h. mit weitem Skopus über die Negation.

Trotz solcher Details, ist einzig die Untersuchung der Ko-Vorkommen einzelner lexikalischer Elemente mit Ausdrücken, die bestimmte logische Konstanten beitragen, praktikabel und scheint einiges Potenzial zu besitzen, um die hier angerissenen Fragestellungen anzugehen.

Zusammenfassend ergibt sich, dass Korpora auch für abstraktere Distributionsbeschränkungen die primäre Datenquellen sein können. Hierbei sind zur Extraktion von Distributionsbeschränkungen zwei Arten von Generalisierung über konkrete Belegstellen notwendig, die eine entsprechende Kodierung in den Korpora voraussetzen: Erstens die Identifizierung einer bestimmten syntaktischen oder satzsemantischen Konstellation und zweitens die genaue Ausformulierung der Wortgruppen oder lexikalisch-semantischen Klassen, die in dieser Konstellation auftreten können. Diese beiden Schritte sind nicht unabhängig von einander: so musste von allen möglichen und angetroffenen Ergänzungen des Verbs *fackeln* eine bestimmte Klasse isoliert werden, nämlich die der durativen Modifikatoren. Die Bestimmung genau dieser Klasse kann jedoch letztendlich nur aufgrund der lexikalischen Semantik ihrer Elemente erfolgen. Die Tatsache, dass immer mehr Korpora entstehen, die über reiche linguistische Annotation verfügen, erhöht die Durchführbarkeit der notwendigen Untersuchungen.

5. Distributionsidiosynkrasien als Teil der Grammatikarchitektur

Was bedeuten die hier gemachten Beobachtungen für eine formale Analyse natürlicher Sprache? Wir haben gesehen, dass lexikalische Elemente idiosynkratische distributionelle Beschränkungen aufweisen können. Eine solche Betrachtung ist für Polaritätselemente plausibel und erhöht auch die

Plausibilität einer Dekompositionsanalyse von Wendungen mit unikalen Elementen. Zur Integration von Distributionsbeschränkungen haben wir in den vorangegangenen Abschnitten ein zusätzliches Feld in Lexikoneinträge eingeführt und die Grammatik um ein Prinzip angereichert, das die Erfüllung aller Distributionsanforderungen in einer Äußerung erzwingt.

Die Information im neuen Feld des Lexikoneintrags beinhaltet eine Beschreibung der möglichen Auftretenskontexte des Lexems. Damit erhalten wir genau die Art von Analyse, die in Höhle (1978) als abwegig erachtet wurde. Am Beispiel von *fackeln* und anderer Polaritätselemente haben wir jedoch gesehen, dass diese Form der Kodierung durchaus sinnvoll ist. In Sailer/Richter (2002a, b) wurde insbesondere argumentiert, dass die Distributionsbeschränkung von *fackeln* auf durative Modifikatoren weder als Selektionseigenschaft aufgefasst werden sollte, noch aus allgemeinen Prinzipien folgt, noch über eine Konstruktionsanalyse behandelt werden sollte.

Da jede formale Grammatik mit den Begriffen der Selektion und der Konstruktion arbeitet, ist es wichtig, zu klären, in welcher Hinsicht Distributionsidiosynkrasien sich von ihnen unterscheiden.

5.1 Aufgabenteilung: Selektion – Kollokation

Folgt man dem hier vertretenen Dekompositionsansatz, dann ergibt sich die Frage, ob bei einem so mächtigen Distributionsmechanismus, ein getrennter Mechanismus für normale „Selektion“ in der Grammatik überhaupt notwendig ist. Ist Selektion dann nur ein Spezialfall von „Kollokation“?

Eine solche Auffassung ist vermutlich kompatibel mit vielen Arbeiten zu Kollokationen, wie Sinclair (1991). Auch Korpusmethoden können prinzipiell nicht diskriminieren zwischen obligatorischer Kookurrenz von Elementen, zwischen denen Selektion besteht, und solcher wo keine Selektion vorliegt: in beiden Fällen wird nur erkannt, dass das eine Element nahezu ausschließlich in Verbindung mit dem anderen auftritt.

Viele formale Grammatiksysteme schränken jedoch die Möglichkeiten der Selektion massiv ein:

(23) Beschränkungen der Selektion:

Syntaktische Beschränkungen (kategoriale, oder C-Selektion):

1. Selektion ist lokal: d. h. ein selegiertes Element und sein Selektor stehen in einer bestimmten syntaktischen Nähe. In der Regel sind sie Schwestern.
2. Selektion nimmt nur Bezug auf „lokale“ Eigenschaften des selegierten Elementes. Dies bedeutet, dass die syntaktische Kategorie des selegierten Elementes beschränkt werden kann, die syntaktische Struktur der selegierten Konstituente aber keine Rolle spielen darf, ebensowenig wie die konkrete phonologische Form.

Semantische Beschränkungen (Selektionsbeschränkungen oder S-Selektion)

3. „Linking“: Selektion ermöglicht es, die Realisierung von semantischen Argumentstellen in der Syntax zu erzwingen.
4. Selektion nimmt nur Bezug auf lexikalisch-semantische Eigenschaften des selegierten Elements. Selektionsanforderungen wie „nimmt ein universell quantifiziertes Subjekt“ sollen ausgeschlossen werden.
5. Es werden lexikalisch-semantische Konzepte selegiert, nicht aber einzelne Lexeme.

Es besteht weitgehender Konsens hinsichtlich dieser Annahmen zur Selektion. In manchen Grammatiktheorien, wie beispielsweise in der HPSG, wird versucht, zumindest die syntaktischen Beschränkungen fest in die Grammatik zu integrieren.

Damit ergibt sich die Frage, wie Fälle behandelt werden sollten, in denen klar Selektion zwischen einem distributionell beschränkten Element und seinem Kollokator vorliegt, jedoch die Beschränkungen des Selektors auf das selegierte Element deutlich stärker sind, als dies mittels Selektion möglich ist. Beispiele für diese Konstellation sind *Angst einjagen* oder *sattsam bekannt*.¹²

Nun mag man einwenden, dass die Beschränkungen in (23) willkürlich sind, oder aus anderer Perspektive, zum Beispiel der eines Lexikographen, zum Teil zu strikt. Beispielsweise wird in Soehn (2003) und Soehn/Sailer (2003) gezeigt, dass viele der hier distributionell behandelten Daten in der HPSG mittels „Selektion“ behandelt werden können, wenn man (i) auf die semantischen Beschränkungen in (23) verzichtet und (ii) wechselseitige Selektion von Kopf und Komplement zulässt. Ähnlich führen Kuiper/Everaert (2000) eine neue Art Selektion ein, lexikalische (oder L-)Selektion. L-Selektion unterliegt nicht den strikten Lokalitätsbedingungen von C- und S-Selektion und kann auch einzelne Lexeme selegieren.

Es ist nicht klar, ob solche selektionsähnlichen Mechanismen das angesprochene Phänomen adäquat angehen. Edmonds/Hirst (2002) zeigen beispielsweise, dass es auch in Fällen wie *Angst einjagen* und *sattsam bekannt* sinnvoll sein kann, die strengen Beschränkungen für Selektion aufrecht zu erhalten und ein zusätzliches Kollokationsmodul anzunehmen. Im Rahmen eines Sprachgenerierungssystems unterscheiden sie zwei Schritte: Im ersten Schritt werden die an Selektion orientierten Entscheidungen vorgenommen. Dies beinhaltet syntaktische Einschränkungen und Einschränkungen hinsichtlich semantischer Konzepte. Im zweiten Schritt findet die Auswahl konkreter lexikalischer Elemente statt. Hier wird dann zwischen Plesionymen

¹² Diese Fälle sind in Richter/Sailer (2003) diskutiert.

unterschieden, und feinere, kollokationelle Bedingungen spielen eine Rolle. Durch eine solche Architektur wird zum einen eine größere Effizienz erreicht, da jeweils der Entscheidungsraum kleiner ist, zum anderen wird der Beobachtung Rechnung getragen, dass die Entscheidungskriterien jeweils gänzlich unterschiedlicher Natur sind.

Der Ansatz von Edmonds/Hirst (2002) scheint uns vielversprechend für eine Verbindung des traditionellen formalgrammatischen Selektionsbegriffs mit den hochidiosynkratischen Distributionsbeschränkungen lexikalischer Elemente. Statt einer Ignorierung der tatsächlichen Sprachgegebenheiten (wie bislang in formalen Grammatiken) oder einer Reduktion von allem auf Kollokation, entsteht eine Aufgabenteilung zwischen einem Selektionsmechanismus und einem Kollokationsmechanismus. Die Kriterien in (23) geben einen Anhaltspunkt, entlang welcher Linie die Aufgabenverteilung stattfindet.

5.2 Aufgabenteilung: Kollokation – Konstruktion

Im Zusammenhang mit unikalen Elementen und dem Verb *fackeln* haben wir bereits die Möglichkeiten von Konstruktionsanalysen angesprochen. In der Konstruktionsgrammatik werden komplexe sprachliche Ausdrücke als Einheiten betrachtet. Es gibt keine prinzipiellen Beschränkungen auf die Anzahl oder die Beziehung oder die Struktur der Elemente, die in einer Konstruktion zusammengefasst sind. Das unterscheidet Konstruktionen von den Beschränkungen auf Selektion in (23).

In der *Construction Grammar* von Fillmore und Kay wird eine möglichst reduzierte Kodierung der grammatischen Strukturen einer Sprache angestrebt. Daraus ergibt sich, dass nur dort Konstruktionen angenommen werden, wo tatsächlich Idiosynkrasien vorliegen. In Kay/Fillmore (1999) wird deshalb ein eigener Selektionsmechanismus entworfen, der die Realisierung von Argumenten einer sehr generellen Argument-Realisierungs-Konstruktion überlässt. Bei der Diskussion von unikalen Elementen und *fackeln* bin ich von dieser Art der Konstruktionsgrammatik ausgegangen und habe aufgrund der kombinatorischen Regularität der betrachteten Elemente gegen eine Konstruktionsanalyse argumentiert.

Die Grundannahmen von Fillmore und Kollegen zu Konstruktionen sind in (24) zusammengefasst:

(24) Eine Konstruktion ist ein syntaktischer Komplex, der mindestens eines der folgenden Kriterien erfüllt:

1. Seine Teile sind syntaktisch auf idiosynkratische Weise zusammengefügt.
2. Die Interpretation erfordert eine eigene Strategie, d. h. Bedeutungsaspekte der Teile werden ignoriert, Bedeutungsaspekte kommen hinzu, o. ä.

Wir haben also gezeigt, dass dieser Begriff von Konstruktion nicht den der Distribution ersetzen kann. Natürlich gilt das Umgekehrte ebenso wenig. Es gibt Phrasen mit idiosynkratischem Aufbau, sei dies syntaktisch oder semantisch.

Fillmore et al. (1988) und Nunberg et al. (1994) listen etliche Beispiele für das Englische auf, und auch das Deutsche ist reich an Beispielen. Auf der Ebene der „lexikalischen Konstruktionen“ finden wir Beispiele wie in (25); auf der Ebene der „grammatischen Konstruktionen“ u. a. den Typus der Phraseoschablone (vergl. (26)).

(25) mir nichts dir nichts

(26) a. Der und ein guter Schwimmer? (Lambrecht 1990)
b. Urlaub ist Urlaub.

Bei diesen Beispielen kann eine Distributionsbeschränkung für die auftretenden lexikalische Elemente nicht verwendet werden, um die Eigenschaften der Konstruktionen vollständig zu erfassen.

Die Interaktion von Kollokation und Konstruktion lässt sich an folgendem Beispiel gut illustrieren. In (27) liegt ein phraseologischer Teilsatz vor.

(27) Peter weiß, wo Barthel den Most holt.

Die Eigenschaften dieses Teilsatzes sind in Sailer (2000) im Detail besprochen. Dort wird gezeigt, dass der Teilsatz als „Konstruktion“ zu behandeln ist, da er (i) syntaktisch fest ist (siehe (28) aus Fleischer 1997, S. 103) und (ii) seine Bedeutung sich nicht aus der seiner Teile ergibt; v. a., da *Barthel* und *Most* in diesem Fall überhaupt keine Bedeutung zu haben scheinen.

(28) Der wusste schon, wo (der) Barthel den Most holt/ *holte

Andererseits ist die Verbindung des Teilsatzes mit dem Matrixverb *wissen* sowohl syntaktisch als auch semantisch völlig regelhaft. So folgt aus (27), dass Peter etwas weiß und der semantische Gehalt des Teilsatzes kann sogar pronominal wieder aufgenommen werden (siehe (29)).

(29) Peter weiß, [wo Barthel den Most holt]_i, und Hans ahnt es_i.

Der unikale Teilsatz ist nicht nur intern irregulär sondern auch distributionell auf Kombination mit dem Lexem *wissen* beschränkt:

(30) Er ahnt/ fragt sich, *[wo Barthel den Most holt]/[wo Hans das Bier holt].

In der Analyse müssen wir also einen phrasalen Lexikoneintrag für die Konstruktion *wo Barthel den Most holt* annehmen. Auch phrasale Lexikoneinträge haben ein Feld für Distributionsbeschränkungen. Im Falle des phrase-

ologisierten Teilsatzes steht dort die Bedingung, dass der Satz nur als Komplementsatz zum Verb *wissen* auftreten kann.

Die Spezifikation der Distributionsanforderungen in Lexikoneinträgen übernimmt demnach eine Aufgabe, die sich nicht mit den etablierten Domänen von Selektion und Konstruktion deckt, sondern sie ergänzt und mit ihnen interagiert. Alle lexikalischen Zeichen, auch phrasale, können Distributionsanforderungen aufweisen. Ein wichtiger Unterschied zu Konstruktionen und auch zum Selektionsmechanismus ist, dass (i) Distributionsbeschränkungen keinen strikten Lokalisierungsbedingungen gehorchen müssen und v. a. dass (ii) Distributionsbeschränkungen keine „generative“ Funktion haben, d. h. sie lizenzieren keine anderweitig ausgeschlossenen Verbindungen (wie Konstruktionen) oder legen die Beziehungen zwischen verschiedenen Elementen im Satz fest (wie Selektion). Ihre einzige Aufgabe ist die der Strukturbeschränkung.¹³ Wir haben es also mit einem unabhängigen Mechanismus zu tun.

6. Zusammenfassung und Ausblick: Erfahrungsbasierte modelltheoretische Syntax

In diesem Aufsatz habe ich versucht, den Begriff der Distributionsidiosynkrasie empirisch und grammatiktheoretisch anzufüllen. Dabei habe ich Beispiele für lexikalische Elemente mit Distributionsidiosynkrasien angeführt und gezeigt, dass Distributionsbeschränkungen bereits etablierte Techniken wie Selektion und Konstruktionsanalyse ergänzen. Als empirischer Bereich für Distributionsanalysen zeichnen sich unikale Elemente und Polaritätselemente ab, sowie allgemein Kollokationsphänomene. Das vorgeschlagene Distributionsmodul ermöglicht eine Integration von Kollokationsdaten in den Abdeckungsbereich formaler Grammatiksysteme.

Nach diesen Betrachtungen zum kompetenzgrammatischen Status von Distributionsbeschränkungen möchte ich mich zum Schluss der Frage nach dem Zusammenhang von Gebrauch und Grammatik zuwenden. Da Gebrauchsdaten die primäre Datenquelle meiner Untersuchungen darstellen und der Gebrauch letztendlich die einzige Motivation für die Distributionsidiosynkrasien zu sein scheint, stellt sich diese Frage zwangsläufig. Hinzu kommt, dass ich mich bislang auf grammatikalisierte Gebrauchsbeobachtungen beschränkt habe, und sich damit die Frage ergibt, welchen Status andere, weniger absolute Auftretensmuster haben. Ich werde zunächst kurz auf erfahrungsbasierte linguistische Modelle eingehen, mich dann dem Paradigma der modelltheoretischen Syntax zuwenden, um schließlich die Möglichkeiten einer Synthese dieser beiden Ansätze zu skizzieren.

Sowohl in der Computerlinguistik als auch in verschiedenen theoretischen Richtungen werden in den letzten Jahren verstärkt Modelle diskutiert, die

¹³ Dies kann natürlich zur Klärung von sonst möglichen Ambiguitäten führen.

den Sprachgebrauch ins Zentrum der Untersuchung stellen. Beispiele für solche Ansätze sind das *Analogical Modeling* (Skousen 1989, 1992), die *Exemplar-based Grammar* (Chandler 1993), die *Experience-based Grammar* (DOP, Bod 1998), das *Usage-based Model* (Croft 2001, Croft/Cruse 2003) oder das *Memory-based Processing* (Daelemans 1999).

Gemeinsam ist diesen Vorschlägen, dass Eigenschaften der tatsächlichen sprachlichen Erfahrung eines Sprechers für den Sprachaufbau als zentral erachtet werden, wohingegen die Rolle von abstrakten, allgemeinen Regeln als recht gering eingeschätzt wird. Im „reinsten“ Fall eines solchen Ansatzes, gibt es gar keine abstrakten Regeln, sondern lediglich eine Menge von „Beispielsätzen“. Neue Sätze werden dann entweder aus Bestandteilen dieser Beispielsätze zusammengesetzt, oder in Analogie zu ihnen gebildet. Diese Ansätze haben sich in praktischen computerlinguistischen Aufgaben als sehr erfolgreich erwiesen. Ihre besondere Stärke liegt darin, dass sie vollständig automatisiert sind und extrem robust sind, d. h. auch mit Unregelmäßigkeiten und unbekanntem Input keine Probleme haben. Ein weiterer konzeptueller Vorteil dieser Ansätze liegt in ihrer Nähe zu kognitionspsychologischen Modellen.

Das Kennzeichnende dieses Typus von Sprachmodell ist, dass linguistische Beschränkungen nicht explizit kodiert sind, sondern sich lediglich implizit ergeben, durch Häufigkeiten bestimmter Merkmalskombinationen im Belegpool, oder durch gänzliche Abwesenheit anderer Kombinationen.

Eine interessante Möglichkeit der Verbindung von verwendungsbasierten Ansätzen mit formalen Grammatiktheorien ergibt sich im Rahmen der modelltheoretischen Syntax (MTS). Bereits Riehemann (2001) fordert im Zusammenhang mit der Betrachtung der syntaktischen Flexibilität von Phraseologismen eine Verbindung von formaler und erfahrungsbasierter Grammatikforschung. Simov (2001, 2002) zeigt, wie eine solche Verbindung von HPSG und DOP formal aussehen könnte. Die folgende knappe Darstellung orientiert sich an Sailer (2003).

Nach Pullum/Scholz (2001) ist die Grundidee der MTS, dass eine Grammatik als eine endliche Menge von Axiomen einer formalen Beschreibungssprache aufgefasst wird, zu der es eine modelltheoretische Interpretation gibt. Unter einer solchen Annahme ist eine Grammatik des Deutschen beispielsweise eine Menge von Beschränkungen, die von allen (möglichen/potenziellen) Sätzen des Deutschen erfüllt wird und genau von diesen. Modelltheoretische Grammatikformalismen sind beispielsweise die HPSG (Pollard/Sag 1994), ebenso wie modelltheoretische Reformulierungen der LFG, der GPSG und der GB (vergleiche die Übersicht in Pullum/Scholz (2001)).

Der entscheidende Vorteil der modelltheoretischen Syntax gegenüber dem generativ-aufzählenden Grammatiktyp ist, dass die linguistischen Strukturen, d. h. das *Modell*, im Mittelpunkt steht. Für jedes Modell gibt es unendlich viele Grammatiken. Wir können also je nach Bedürfnissen eine eher datennahe Beschreibung wählen, oder sehr abstrakte, generelle Beschrei-

bungen suchen, die beispielsweise auch auf Modelle anderer Sprachen zu treffen. Damit bietet dieser Ansatz eine hervorragende Basis, für eine Zusammenführung von eher empirischen, computerlinguistisch-praktischen und stark theoretischen Arbeiten.

Unter einer solchen Perspektive ist ein „Belegstellenpool“, wie ihn die erfahrungsbasierten Ansätze annehmen, nichts anderes als eine echte Teilmenge des Modells der Grammatik. Vermutlich ist es eine besonders ausgezeichnete Teilmenge, da sie ausreicht, um die Gesamtgrammatik daraus zu extrapolieren.

Eine *erfahrungsbasierte modelltheoretische Grammatik* (EMG) besteht aus einer *Erfahrungsmenge* E , d. h., dem Belegpool der erfahrungsbasierten Ansätze, und einer *Theorie* θ , d. h., der Theorie der modelltheoretischen Ansätze. Die Erfahrungsmenge E muss ein Modell der Theorie θ sein, jedoch kein vollständiges („exhaustives“, King 1999) Modell. Unter erfahrungsbasierter Perspektive ergibt sich aus E zusammen mit dem sprachunabhängigen Extrapolationsmechanismus automatisch eine Theorie, deren exhaustives Modell dann die zu beschreibende Sprache ist. Unter modelltheoretischer Perspektive ist die Form der Theorie θ das einzig interessante, Eigenschaften von E bleiben weitgehend unberücksichtigt. In einem theoretischen Rahmen, der beide Sichtweisen vereint, können nun Phänomene angemessener einzelnen Betrachtungsweise zugeordnet werden.

Für die hier untersuchten Distributionsdaten ergibt sich eine differenzierte Einordnung. Die Bedeutung der Arbeit mit Korpusdaten zur Erkennung von Distributionsmustern wurde in den vorangegangenen Abschnitten hinreichend ausgeführt. Es wurde jedoch auch gezeigt, dass in manchen Fällen die Distributionsmuster derart strikt sind, dass Muttersprachler sogar in der Lage sind, Grammatikalitätsurteile zu fällen, die davon abhängen, ob ein bestimmtes Muster vorliegt oder nicht. Für solche Fälle haben wir das Distributionsmodul in die Grammatik integriert. Grammatikalisierung von Distributionskontexten setzt natürlich voraus, dass alle Auftreten eines lexikalischen Elements im Erfahrungspool E ebenfalls diesem Muster entsprechen.

In vielen Fällen sind Distributionsanforderungen jedoch nicht so absolut. Es finden sich Produktions- und Interpretationspräferenzen für bestimmte Kontexte. Allerdings werden Abweichungen von diesem Muster von Sprechern nicht als ungrammatisch gewertet. In solchen Fällen ist das Distributionsmuster lediglich in E sehr prominent, es ist jedoch noch nicht grammatikalisiert. Selbstverständlich können immer ausgeprägtere Produktionspräferenzen dazu führen, dass ein bestimmtes Auftretensmuster im Laufe der Zeit freiere Varianten im Erfahrungspool verdrängt. In einer solchen Situation lässt sich dann eine zunehmende Grammatikalisierung beobachten.

Schließlich besteht im Rahmen einer EMG auch die Möglichkeit, zwischen reinen Ko-Vorkommen (Kookkurrenzen) von lexikalischen Einheiten

und bestimmten sprachlichen Mustern und grammatisch relevanten Kombinationen (Kollokationen) zu unterscheiden: Kookkurrenzen treten auf der Ebene des Erfahrungsmenge E auf, Kollokationen treten in jedem Modell der Grammatik auf, werden also von der Theorie θ gefordert.

Die in diesem Aufsatz diskutierten Distributionsdaten zeigen, dass eine Verbindung von erfahrungsbasierten und modelltheoretischen Grammatiksystemen notwendig und im Rahmen einer EMG auch möglich und potenziell fruchtbar ist. Distributionsdaten sind auf der Wortebene bereits relativ gut im Rahmen verschiedener empirischer und korpuslinguistischer Ansätze untersucht; andererseits erfordern die abstrakteren Distributionsbeschränkungen, wie sie beispielsweise bei Polaritätselementen relevant sind, eine stärkere Berücksichtigung formaler Grammatiksysteme.

Literatur

- Aronoff, Mark (1976): *Word Formation in Generative Grammar*. MIT Press, Cambridge, Massachusetts and London, England.
- Baker, Collin F./Fillmore, Charles J./Lowe, John B. (1998): The Berkeley FrameNet Project. In: *Proceedings of COLING-ACL*, Montreal, Canada.
- Bod, Rens (1998): *Beyond Grammar. An Experience-Based Theory of Language*. CSLI Publications, Stanford.
- Burger, Harald (1998): *Phraseologie: Eine Einführung am Beispiel des Deutschen*. Erich Schmidt Verlag, Berlin.
- Chandler, Steve (1993): Are Rules and Modules Really Necessary for Explaining Language? In: *Journal of Psycholinguistic Research* 22 (6), S. 593–606.
- Croft, William (2001): *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford.
- Croft, William/Cruse, Alan D. (2003): *Cognitive Linguistics*. Cambridge University Press, Cambridge.
- Daelemans, Walter (1999): Memory-based Language Processing. Introduction to the Special Issue. In: *Journal for Experimental and Theoretical Artificial Intelligence* 11 (3), S. 287–292.
- Dobrovol'skij, Dmitrij (1988): *Phraseologie als Objekt der Universallinguistik*. Verlag Enzyklopädie, Leipzig.
- Dobrovol'skij, Dmitrij (1999): Haben transformationelle Defekte der Idiomstruktur semantische Ursachen? In: Fernandez Bravo, N./Behr, I./Rozier, C. (Hg.): *Phraseme und typisierende Rede*. S. 25–37. Stauffenburg, Tübingen.
- Dobrovol'skij, Dmitrij/Piirainen, Elisabeth (1994): Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative. In: *Folia Linguistica* 27 (3–4), S. 449–473.
- Edmonds, Philip/Hirst, Graeme (2002): Near-synonymy and Lexical Choice. In: *Computational Linguistics* 28 (2), S. 105–144.
- Erk, Katrin/Kowalski, Andrea/Pinkal, Manfred (2003): A Corpus Resource for Lexical Semantics. In: *Proceedings of IWCS5, Tilburg 2003*. Version vom 12.12.2002.
- Fillmore, Charles/Kay, Paul/O'Connor, M. (1988): Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. In: *Language* 64, S. 501–538.
- Fleischer, Wolfgang (1989): Deutsche Phraseologismen mit unikalischer Komponente – Struktur und Funktion. In: Gréciano, G. (Hg.): *Europhras* 88. S. 117–126. Univ. des Sciences Humaines, Strasbourg.

- Fleischer, Wolfgang (1997): *Phraseologie der deutschen Gegenwartssprache*. 2. überarbeitete Auflage. Niemeyer, Tübingen.
- Gazdar, Gerald/Klein, Ewan/Pullum, Geoffrey/Sag, Ivan (1985): *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge, Mass.
- Giannakidou, Anastasia (1998): *Polarity Sensitivity as Nonveridical Dependency*. John Benjamins, Amsterdam/Philadelphia.
- Häcki Buhofer, Annelies (2000): *Phraseologisch isolierte Wörter und Wortformen*. Im Druck. Manuskript unter <http://www.germa.unibas.ch/seminar/whoiswho/publikationen/ahaeckiwortformen.pdf>.
- Hoeksema, Jack (1996): *Zeg het met rozen: een vergelijking van drie idiomatische uitdrukkingen*. In: *TABU Bulletin voor Taalwetenschap* 26 (3), S. 129–149.
- Hoeksema, Jack (2001a): *Polarity Items in Context: Primary and Secondary Licensing*. Handout zu einem Vortrag auf dem Workshop on Negation, Paris, 2.10.2001.
- Hoeksema, Jacob (2001b): *Rapid change among expletive polarity items*. In: Brinton, L. J. (Hg.): *Historical Linguistics 99. Selected Papers from the 14th International Conference on Historical Linguistics*. S. 175–186. Benjamins, Amsterdam/Philadelphia.
- Hoeksema, Jacob (2002a): *De negatief-polaire uitdrukkingen van het Nederlands. Inleiding en lexicon*. Manuskript, Rijksuniversiteit Groningen.
- Hoeksema, Jack (2002b): *Meaning and Use of Polarity Items*. Handout zu einem Vortrag an der Universität Tübingen, 18.11.2002.
- Höhle, Tilmann N. (1978): *Lexikalistische Syntax. Die Aktiv-Passiv-Relation und andere Infinitivkonstruktionen im Deutschen*. Niemeyer, Tübingen.
- Joshi, Aravind K. (1987): *An Introduction to Tree Adjoining Grammars*. In: Manaster-Ramer, A. (Hg.): *Mathematics of Language*. S. 87–114. Benjamins, Amsterdam/Philadelphia.
- Kay, Paul (1998): *An Informal Sketch of a Formal Architecture for Construction Grammar*. In: Bouma, G./Kruijff, G.-J./Oehrle, R. T. (Hg.): *Proceedings of the FHCG-98*. S. 175–183.
- Kay, Paul/Fillmore, Charles J. (1999): *Grammatical Constructions and Linguistic Generalizations: the „What’s X doing Y?“ Construction*. In: *Language* 75, 1, S. 1–33.
- Keenan, Edward L. (1974): *The Functional Principle: Generalizing the Notion of „Subject-of“*. In: *Chicago Linguistic Society* 10, S. 298–310. Chicago Linguistic Society, Chicago.
- King, Paul J. (1999): *Towards Truth in HPSG*. In: Kordoni, V. (Hg.): *Tübingen Studies in HPSG, Arbeitspapiere des SFB 340, Nr. 132, Band 2*, S. 301–352. Universität Stuttgart und Universität Tübingen.
- Klima, Edward (1964): *Negation in English*. In: Fodor, J. A./Katz, J. (Hg.): *The Structure of Language*. S. 246–323. Prentice Hall, Englewood Cliffs, New Jersey.
- Kuiper, Koenraad/Everaert, Martin (2000): *Constraints on the Phrase Structural Properties of English Phrasal Lexical Items*. In: *Proceedings of the 8th Annual Conference of the Polish Association for the Study of English*. S. 151–170.
- Kürschner, Wilfried (1983): *Studien zur Negation im Deutschen*. Gunter Narr, Tübingen.
- Ladusaw, William (1980): *Polarity Sensitivity as Inherent Scope Relations*. Garland Press, New York.
- Lambrecht, Knud (1990): *„What, me worry? – „Mad Magazine“ sentences revisited*. In: *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, S. 215–228. BLS, Berkeley, USA.
- Moon, Rosamund (1998): *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Clarendon, Oxford.

- Nunberg, Geoffrey/Sag, Ivan A./Wasow, Thomas (1994): Idioms. In: *Language* 70, S. 491–538.
- Pafel, Jürgen (1998): Skopus und logische Struktur. Studien zum Quantorenskopos im Deutschen. Arbeitspapiere des SFB 340, Nr. 129. Universität Stuttgart und Universität Tübingen.
- Partee, Barbara H. (1984): ‚Compositionality‘. In: Landman, F./Veltman, F. (Hg.): *Varieties of Formal Semantics. Proceedings of the 4th Amsterdam Colloquium, September 1982*. S. 281–311. Foris, Dordrecht.
- Pinkal, Manfred (1996): Radical Underspecification. In: Dekker, P./Stokhof, M. (Hg.): *Proceedings of the 10th Amsterdam Colloquium*. S. 587–606. ILLC, Amsterdam.
- Pollard, Carl/Sag, Ivan A. (1994): *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago/London.
- Pullum, Geoffrey K./Scholz, Barbara C. (2001): On the Distinction between Model-Theoretic and Generative-Enumerative Syntactic Frameworks. In: de Groote, P./Morrill, G./Retoré, C. (Hg.): *LACL 2001*. pp. 17–43. Springer, Berlin/Heidelberg.
- Pulman, Stephen G. (1993): The Recognition and Interpretation of Idioms. In: Cacciari, C./Tabossi, P. (Hg.): *Idioms: Processing, Structure, and Interpretation*, Kapitel 11, S. 249–270. Lawrence Erlbaum Associates.
- Resnik, Philip Stuart (1993): *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph. D. thesis, University of Pennsylvania.
- Reyle, Uwe (1993): Dealing with Ambiguities by Underspecification: Construction, Representation and Deduction. In: *Journal of Semantics* 10 (2), S. 123–179.
- Richter, Frank/Sailer, Manfred (2003): Cranberry Words in Formal Grammar. In: Beyssade, C./Bonami, O./Cabredo Hofherr, P./Corblin, F. (Hg.): *Empirical Issues in Formal Syntax and Semantics, Volume 4*. Presses Universitaires de Paris-Sorbonne, Paris. S. 155–171.
- Riehmann, Susanne Z. (2001): *A Constructional Approach to Idioms and Word Formation*. Ph. D. thesis, Stanford University.
- Sailer, Manfred (2000): *Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar*. Dissertation, Universität Tübingen. Erschienen als: *Arbeitsberichte des SFB 340, Nr. 161*. Universität Stuttgart und Universität Tübingen. 2003.
- Sailer, Manfred (2003): *Experience-Based Model-Theoretic Syntax*. Manuskript, Sonderforschungsbereich 441, Universität Tübingen.
- Sailer, Manfred/Richter, Frank (2002a): Collocations and the Representation of Polarity. In: Alberti, G./Balogh, K./Dekker, P. (Hg.): *Proceedings of the 7th Symposium on Logic and Language*. Pécs, S. 129–138.
- Sailer, Manfred/Richter, Frank (2002b): Not for Love or Money: Collocations! In: Jäger, G./Monachesi, P./Penn, G./Wintner, S. (Hg.): *Proceedings of Formal Grammar 2002*. S. 149–160.
- Schenk, André (1995): The Syntactic Behavior of Idioms. In: Everaert, M./v. d. Linden, E.-J./Schenk, A./Schreuder, R. (Hg.): *Idioms. Structural and Psychological Perspectives*. S. 253–271. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Simov, Kiril (2001): Grammar Extraction from an HPSG Corpus. In: *RANLP 2001 Conference*, S. 285–287.
- Simov, Kiril (2002): Grammar Extraction and Refinement from an HPSG Corpus. In: *Proceedings of the ESSLI Workshop on Machine Learning Approaches in Computational Linguistics*, S. 38–55.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford Univ. Press, Oxford.

- Skousen, Royal (1989): *Analogical Modelling of Language*. Kluwer, Dordrecht/Boston/London.
- Skousen, Royal (1992): *Analogy and Structure*. Kluwer, Dordrecht/Boston/London.
- Soehn, Jan-Philipp (2003): *Von Geisterhand zu Potte gekommen. Eine HPSG-Analyse von PPs mit unikalener Komponente*. Magisterarbeit, Universität Tübingen, Seminar für Sprachwissenschaft.
- Soehn, Jan-Philipp/Sailer, Manfred (2003): *At First Blush on Tenterhooks. About Selectional Restrictions Imposed by Nonheads*. In: Jäger, Gerhard/Monachesi, Paola/Penn, Gerald/Wintner, Shuly (Hg.): *Proceedings of FGVienna: The 8th Conference on 'Formal Grammar'*. CSLI Publications. Im Erscheinen.
- Steinitz, Renate (1988): *Adverbial-Syntax*. Akademie-Verlag, Berlin.
- van der Wouden, Ton (1997): *Negative Contexts. Collocation, Polarity and Multiple Negation*. Routledge, London.
- Wagner, Andreas (2000): *Enriching a Lexical Semantic Net with Selectional Preferences by Means of Statistical Corpus Analysis*. In: *Proceedings of the ECAI-2000 Workshop on Ontology Learning*, Berlin, August 2000.
- Wagner, Andreas (2002): *Learning Semantic Roles for Wordnets*. In: *Proceedings of the ESSLLI-2002 Workshop on Machine Learning Approaches in Computational Linguistics*. Trento, August 2002.
- Wasow, Thomas/Sag, Ivan A./Nunberg, Geoffrey (1983): *Idioms: An Interim Report*. In: Hattori, S./Inoue, K. (Hg.): *Proceedings of the XIIIth International Congress of Linguists*. S. 102–115.
- Wauschkuhn, Oliver (1999): *Automatische Extraktion von Verbvalenzen aus deutschen Textkorpora*. Shaker, Aachen.
- Zwarts, Frans (1981). *Negatief polaire uitdrukkingen*. In: *GLOT 4* (1), S. 35–132.
- Zwarts, Frans (1986): *Categoriale grammatica en algebraïsche semantiek. Een onderzoek naar negatie en polariteit in het Nederlands*. Dissertation, Rijksuniversiteit Groningen.
- Zwarts, Frans (1995): *Nonveridical Contexts*. In: *Linguistic Analysis* 25, S. 286–312.
- Zwarts, Frans (1997): *Three Types of Polarity*. In: Hamm, F./Hinrichs, E. W. (Hg.): *Plurality and Quantification*. S. 177–237. Kluwer, Dordrecht/Boston/London.