

Produktivität und Idiomatizität von Präposition-Substantiv-Sequenzen ¹

Tibor Kiss
Sprachwissenschaftliches Institut
Ruhr-Universität Bochum
D-44801 Bochum
tibor@linguistics.ruhr-uni-bochum.de

¹Ich danke Florian Dömges, Katja Keßelmeier, Anneli von Könemann, Antje Müller und Claudia Roch für ihre Hilfe bei der Aufbereitung und Analyse der Daten. Die Fragen und Kommentare von Stefanie Dipper, Joachim Jacobs, Ingrid Kaufmann, Harald Stamm und Doris Schönefeld ermöglichten es mir, die Fragestellung klarer zu formulieren und möglichen Einwänden zu begegnen. Ihnen sei ebenso herzlich gedankt wie zwei anonymen Gutachtern. Miriam Butt danke ich schließlich für die Gelegenheit, diese Überlegungen auf der Konvens 2006 vorstellen zu können.

Abstract The present paper discusses two fundamental properties of so-called determinerless PPs, their purported idiomaticity and their productivity. Combining evidence from German newspaper corpora with methods from statistical natural language processing, we can show that determinerless PPs are no more idiomatic than ordinary phrasal combinations in general. ¿From an empirical point of view determinerless PPs can be classified as productive. Consequently, determinerless PPs cannot be handled by finite lists or by treating them as idiomatic exceptions.

1 Einleitung

Sequenzen aus Präposition und zählbarem Substantiv im Singular treten keineswegs selten auf und sind auch nicht auf einzelne Sprachen beschränkt (vgl. Himmelmann (1998)). Dass die Grammatikschreibung bislang eher einen Bogen um diese Konstruktionen gemacht hat, liegt an einigen Faktoren, die eine Analyse deutlich erschweren. Wenn wir uns hierbei auf das Deutsche konzentrieren, kann man zunächst einmal festhalten, dass Sequenzen aus Präposition und Substantiv die Dudenregel 442 verletzen, die vorsieht, dass „*Substantive mit Merkmalkombination ‚zählbar‘ plus Singular... grundsätzlich immer ein Artikelwort bei sich [haben], und wenn es als letzte Möglichkeit der indefinite Artikel ist.*“

- (1) Die im Untergrund operierenden Widerstandsorganisationen beschleunigten das Ende der schwarzen Selbstverwaltungsorgane, indem sie zum Boykott der Lokalwahlen aufriefen und die gewählten schwarzen Gemeinderatsmitglieder **unter Androhung** physischer Gewalt zum Rücktritt zwangen.

Obwohl das Substantiv *Androhung* zählbar ist und im Singular erscheint, wird es in (1) ohne Artikel realisiert. Dazu kommt, dass die Kombination in (1) von einer entsprechenden Kombination mit Artikel semantisch kaum unterschieden werden kann:

- (2) . . . , indem sie zum Boykott der Lokalwahlen aufriefen und die gewählten schwarzen Gemeinderatsmitglieder **unter der Androhung** physischer Gewalt zum Rücktritt zwangen.

Interessanterweise wird das Fehlen des Artikels häufig weniger als eine Verletzung einer Grammatikregel verstanden, denn als definierendes Charakteristikum dieser Konstruktion, die aber dann – weil sie die Regeln der Grammatik ganz offenkundig verletzt – aus der Grammatikschreibung herausgenommen wird. Deutlich wird dies etwa in der Betrachtung dieser Konstruktionen bei Helbig and Buscha (2001), die sich scheuen, diese Sequenzen *Phrasen* zu nennen und stattdessen von *Wortgruppen* sprechen. Baldwin et al. (2006) folgen dieser Sicht und klassifizieren entsprechende Konstruktionen des Englischen als *multiword expressions*. Wenn überhaupt

über diese Sequenzen gesprochen wird, dann werden Sonderfälle innerhalb dieser schon speziellen Konstruktion behandelt, so etwa bei Trawinski (2003) und Trawinski et al. (2006) im Rahmen der HPSG oder in Burger (2003). Himmelmann (1998) schließt wie einige andere Autoren, dass der Artikel nicht nur fehlen kann, sondern fehlen muss – das Beispiel (2) bleibt damit natürlich außen vor.

Typischerweise werden Präposition-Substantiv-Sequenzen stillschweigend als extragrammatisch verstanden und zugleich der diffusen Klasse ‚idiomatischer Konstruktionen‘ zugeordnet, so etwa bei Fleischer (1982, 300), der kategorisch feststellt: „Die Bildungen sind zum größten Teil idiomatisiert.“ Burger (2003, 36f.), dem allerdings konzidiert werden muss, dass er nur einen Teilbereich der Konstruktionen betrachtet, zweifelt die Produktivität der Konstruktionen ebenso wie ihre Kompositionalität an.

Es kann natürlich nicht bestritten werden, dass Präpositionen-Substantiv-Sequenzen die Dudenregel 442 verletzen, und es ist sicherlich auch nicht verkehrt, einige Präposition-Substantiv-Sequenzen den idiomatischen Ausdrücken zuzuordnen. Fraglich ist allerdings, ob dadurch rechtfertigt werden kann, die Konstruktion gleich der Syntax zu verweisen.

Tatsächlich gibt es hierfür zwei wesentliche Beweggründe. Zum einen lässt sich beobachten, dass Präposition-Substantiv-Sequenzen häufig durch Sprecher einer Sprache nicht kreativ nachvollzogen oder in Isolation als der Sprache zugehörig bewertet werden können.² Zum anderen ist die Kombination von Präposition und Substantiv nicht frei. So kann man die Dudenregel 442 nicht einfach außer Kraft setzen und so behaupten, dass jede Präposition mit jedem artikellosen Substantiv im Singular kombiniert werden kann. Zur Erläuterung können wir hier das Paar *Prämisse-Voraussetzung* betrachten. Die nahezu synonymen Substantive treten im Plural auf, sind also zählbar. Weiterhin können beide innerhalb einer regulären NP als Komplemente der Präposition *unter* erscheinen, aber nur *Voraussetzung* kann auch mit *unter* eine minimale Phrase bilden, während dies für *Prämisse* nicht möglich ist.

- (3) a. Auch Philippe Egli besteht auf einer eigenen Handschrift - **unter Voraussetzung** des Einverständnisses des Ensembles.
b. * Auch Philippe Egli besteht auf einer eigenen Handschrift - **unter Prämisse** des Einverständnisses des Ensembles.
c. Auch Philippe Egli besteht auf einer eigenen Handschrift - **unter der Prämisse** des Einverständnisses des Ensembles.

Die Beispiele in (3) zeigen nicht nur, dass man die Dudenregel 442 nicht einfach dadurch

²Auf die Tatsache, dass es Sprechern nicht leicht fällt, Sequenzen aus Präposition und Substantiv neu zu bilden, machte mich Joachim Jacobs aufmerksam.

retten kann, dass man sie außer Kraft setzt, sondern auch, dass man sie nicht dadurch retten kann, indem man sie für *einzelne* Präpositionen außer Kraft setzt.

Ein weiterer ‚Rettungsversuch‘ für die Dudenregel 442 könnte darin bestehen, die Zählbarkeit der Substantive in Frage zu stellen, die als Komplemente der Präposition auftreten. Hier sind zwei Überlegungen denkbar. Man kann annehmen, dass es sich bei den betreffenden Substantiven formal um Singularia handelt, allerdings entweder gar keine Pluralformen desselben Substantivs existieren oder nur solche, die eine gänzlich andere Interpretation als das Singularium besitzen. Beispiele für die zweite Argumentationslinie wären etwa *Kontrolle* oder *Druck*. Von beiden gibt es Pluralformen, die jedoch jeweils andere Lesarten induzieren, weswegen prominente Verbindungen wie *unter Druck* und *unter Kontrolle* auch hier gar nicht Gegenstand der Untersuchung sind. Im allgemeinen Fall ist diese Argumentation jedoch nicht plausibel, das Beispiele (4) zeigt.

- (4) Milosevic unterschrieb auch **unter Androhungen** von Nato-Bombardementen nicht.

Im Beispiel (4) besitzt *Androhungen* keine wesentlich von *Androhung* in (1) verschiedene Interpretation und es wird auch jeweils ein Komplement realisiert. Man muss wohl zunächst akzeptieren, dass die Realisation eines nicht modifizierten Singulariums als Komplement einer Präposition nicht einfach wegerklärt werden kann.

Der Duden erfasst diese Situation in der Regel 395 dadurch, dass die betreffenden Konstruktionen als implizit aufzählbar bzw. als finite Klasse betrachtet werden. Eine ähnliche Position liegt der Behandlung entsprechender englischer Konstruktionen in Grishman et al. (1998) zugrunde. Im Folgenden wird sich jedoch zeigen, dass die Konstruktion durchaus produktiv ist – wenn auch nicht beliebige Kombinationen gestattet sind – und dass nicht-kompositionelle Kombinationen nur etwa in dem Maße auftreten, in dem wir idiomatische Kombinationen in allen syntaktischen Verbindungen erwarten. Die vorliegende Untersuchung baut auf zwei Zeitungskorpora auf, die zusammen 213 Millionen Wörter umfassen, und zwar der *Neuen Zürcher Zeitung* aus den Jahren 1995 bis 1998 und der *Frankfurter Rundschau* aus den Jahren 1997 bis 1999. Erfreulicherweise besitzen die hier relevanten Konstruktionen einige Charakteristika, die ihre Extraktion aus umfangreichen Korpora deutlich erleichtern: Zum einen handelt es sich jeweils um Bigramme $w_1 w_2$, wobei w_1 die Präposition und w_2 das Substantiv ist. Zur Extraktion der Daten können daher Standardwerkzeuge für die Extraktion von N-grammen verwendet werden, im vorliegenden Fall das Perl-Modul TEXT::NSP. Weiterhin wird das Substantiv nach den orthographischen Regeln des Deutschen mit großem Initial geschrieben, was wiederum die Extraktion der relevanten Kandidatenpaare erleichtert. Die vorliegende Untersuchung konzentriert sich auf Sequenzen, in denen die Präposition *unter* ist und das Substantiv zumindest nach Sicht

der Mehrzahl der Wörterbücher des Deutschen zählbar ist.³

Im Rahmen dieses Aufsatzes werden wir uns auf die Produktivität der Kombination mit ‚fehlendem‘ Artikel und einer möglichen idiomatischen Interpretation von Präposition-Substantiv-Sequenzen konzentrieren. Dies sind jedoch nicht die einzigen Eigenschaften von Präposition-Substantiv-Sequenzen, die im Rahmen einer Grammatik dieser Konstruktion erläuterungsbedürftig sind. Obwohl diese Probleme hier nicht gelöst oder auch nur detailliert diskutiert werden können, sollen diese Fragestellungen zumindest angesprochen werden.

Ein weiteres herausstechendes Merkmal von Präposition-Substantiv-Sequenzen ist die Möglichkeit einer Valenzveränderung des Substantivs, die hier anhand des Beispiels (5) erläutert werden kann.

- (5) Raschheit ist nicht ohne eine gewisse Straffung des Verfahrens zu erzielen, eine Einvernahme bei unverfälschtem Zustand der Betroffenen nur unter Verzicht *(auf Transporte) und ein Verzicht nur durch Dezentralisation.

Das Beispiel (5) zeigt zunächst, dass *Verzicht* ohne PP-Komplement erscheinen kann. Das Weglassen des PP-Komplements scheint allerdings unzulässig zu sein, wenn *Verzicht* zusammen mit *unter* realisiert wird.

Diese Annahme wird sowohl qualitativ als auch quantitativ gestützt. So gehen Helbig and Buscha (2001) davon aus, dass Sequenzen aus Präposition und Substantiv *zusammen* ein Komplement besitzen, eine Annahme, die Trawinski (2003) in einem HPSG-Ansatz umsetzt. Im Frankfurter Rundschau-Korpus finden sich insgesamt 2.833 Vorkommen von *Verzicht*, von denen 2.031 zusammen mit einer PP[+auf] erscheinen. Von den insgesamt 57 Vorkommen von *unter Verzicht* findet sich allerdings nur eines, in denen *Verzicht* kein Komplement besitzt.

Eine weitere Fragestellung, die hier ebenfalls nur angesprochen werden kann, betrifft das Vorkommen von Präposition-Substantiv-Sequenzen in Funktionsverbgefügen. Sehr häufig bildet etwa *unter+Substantiv* mit den Verben *stehen* oder *stellen* ein Funktionsverbgefüge. Hierbei geht man in der Literatur gemeinhin davon aus, dass eine Untersuchung jenseits der Feststellung

³Mit der Präposition ist der Kasus des Substantivs bestimmt, so dass eine Lemmatisierung des Substantivs nicht notwendig ist – Lemmatisierungswerkzeuge wurden allerdings verwendet, um Substantive im Singular von solchen im Plural zu trennen. Obwohl es prinzipiell möglich gewesen wäre, POS-annotierte Korpora zu verwenden, um unmittelbar das Suchmuster APPR-NN zu extrahieren (APPR steht hier entsprechend der Notationskonventionen des STTS für eine Präposition, NN für ein Appellativum, vgl. Schiller et al. 1999), haben wir auf die Analyse annotierter Korpora vorerst verzichtet. Die Gründe dafür liegen darin, dass zumindest im vorliegenden Fall einfache reguläre Ausdrücke entsprechende Ergebnisse liefern. Auf den ersten Blick erscheint die Verwendung annotierter Korpora deswegen sinnvoll, weil man etwa nominalisierte Adjektive oder postnominale Eigennamen durch eine Einschränkung des Suchmusters auf APPR-NN ausschließen könnte. Tatsächlich liegt aber gerade in der Unterscheidung zwischen NN und NE (Eigennamen) und in der Unterscheidung von NN und ADJA (flektierte Adjektive) eine häufige Fehlerquelle beim Pos-Tagging, vgl. Kesselmeier and von Kønemann (2005). In der Konsequenz ist die händische Nachkontrolle einer Kandidatenmenge auf der Basis eines POS-annotierten Korpus in etwa so aufwändig wie die Nachkontrolle einer Kandidatenmenge auf der Basis regulärer Suchmuster.

des Funktionsverbgefüges, also etwa eine Aufteilung des Gefüges in das Verb und die Sequenz aus *unter+Substantiv* nicht notwendig oder sinnvoll ist. Dieser Einschränkung liegt die Überlegung zugrunde, dass Präposition-Substantiv-Sequenzen bereits extragrammatisch sind. Wenn dies jedoch fraglich wird, stellt sich auch die Frage danach, ob Funktionsverbgefüge nicht feiner analysiert werden sollten.

Ein letzter Punkt betrifft schließlich die Beobachtung, dass nackte Singularia universell eher im Kontext von Präpositionen als in anderen Kontexten auftreten (vgl. dazu Himmelmann (1998)). Ob diese Beobachtung überhaupt zutreffend ist, vermag ich an dieser Stelle nicht zu bewerten. Es ist sicherlich richtig, dass nackte Singularia als unmittelbare Verboobjekte ungrammatisch sind, aber sie treten im Deutschen auch als Bestandteile von Funktionsverbgefügen ohne Präposition auf.

2 Zur Kompositionalität von Präposition-Substantiv-Sequenzen

Eine kompositionelle Analyse von Präposition-Substantiv-Sequenzen kann auf zwei Weisen in Frage gestellt werden. Zum einen findet sich die Position, dass es sich bei diesen Konstruktionen zum größten Teil um Idiome handelt (vgl. Fleischer (1982, 300)). Dies soll wohl bedeuten, dass die Bedeutung von [? P N] nicht aus der Kombination der Präposition mit dem Substantiv abgeleitet werden kann. Eine Spielart dieser Charakterisierung findet sich in der impliziten Zuordnung dieser Konstruktionen zu den sog. strukturellen Phraseologismen in Burger (2003, 36f.): „Strukturelle Phraseologismen ... haben ‚nur‘ eine Funktion innerhalb der Sprache, nämlich die Funktion, (grammatische) Relationen herzustellen.“ Eine Sequenz aus Präposition und Substantiv ist in diesem Sinne schon deswegen nicht kompositionell zu analysieren, weil sie insgesamt keine Bedeutung aufweist.

Die zweite Position besteht darin, die Kompositionalität von Präposition-Substantiv-Sequenzen zwar nicht prinzipiell in Frage zu stellen, jedoch anzunehmen, dass die Präposition in Präposition-Substantiv-Sequenzen tatsächlich nur eine vermittelnde grammatische Funktion hat (etwa deswegen, weil ein unbegleitetes Substantiv allein nicht adverbiell verwendet werden kann) und die eigentliche semantische Kontribution der gesamten Phrase durch das Substantiv beigesteuert wird. In der Literatur findet sich allerdings nur die exakt umgekehrte Position. Helbig and Buscha (2001, 354) charakterisieren Präposition-Substantiv-Sequenzen als idiomatisch, weil die *Bedeutung des Substantivs* in der Kombination abgeschwächt bzw. ‚sehr verallgemeinert‘ ist, wie anhand des Kontrasts zwischen *an Hand* und *an der Hand* belegt wird.

(6) an Hand des Dokuments \neq an der Hand des Kindes

Diese Position kann allerdings bei einer Berücksichtigung der möglichen Kandidaten für

Substantive in diesen Kombinationen unmittelbar verworfen werden: die meisten Substantive verlieren ihre Bedeutung in diesen Kombinationen keineswegs, wie anhand der Substantive in (7) verdeutlicht werden kann.

- (7) a. Nach dem Verhandlungsstand von Sonntag sollen der linksliberalen so genannten „Dreiergruppe“ **unter Führung** von Mazowiecki neun Ressorts zufallen, darunter das Verteidigungs- und das Außenministerium sowie die wichtigsten Wirtschaftsressorts.
- b. Die nachhaltigste Korrektur der Programmatik fordert ein Antrag, der **unter Federführung** des hessischen Realpolitikers Hubert Kleinert entstanden ist.
- c. Außerdem sei vereinbart worden, wie die Staaten, die am Rubel festhalten, eine gemeinsame Geldpolitik **unter Leitung** der russischen Zentralbank betreiben sollen.

Wäre das Kriterium von Helbig und Buscha paradigmatisch, so sollten Präposition-Substantiv-Sequenzen auch keine Alternanten zulassen, in denen ein Komplement des Substantivs in das Substantiv inkorporiert wird, d.h. eine Sequenz aus Präposition und Kompositum vorliegt. Dies folgt aus der Annahme, dass die Semantik des Kompositums in einem solchen Fall nicht fundiert wäre, weil der Kopf des Kompositums keinen semantischen Beitrag leisten könnte. Die folgenden Beispiele verdeutlichen, dass diese Position nicht korrekt sein kann.

- (8) a. Dieser Auffassung schlossen sich das Sanitätsdepartement und das Verwaltungsgericht des Kantons Solothurn an und verboten den Verkauf des «TMK-Krebsrisikotests» in Apotheken und Drogerien **unter Strafandrohung**.
- b. Dies widerruft er später, er habe es **unter der Folterandrohung** bloss erfunden
- c. Und da man das Gesetz **unter Sanktionsandrohung** zu befolgen habe, sei ein Handeln in Freiheit und somit echtes moralisches Verhalten gar nicht möglich.
- d. Die im Untergrund operierenden Widerstandsorganisationen beschleunigten das Ende der schwarzen Selbstverwaltungsorgane, indem sie zum Boykott der Lokalwahlen aufriefen und die gewählten schwarzen Gemeinderatsmitglieder **unter Androhung** physischer Gewalt zum Rücktritt zwangen.

Die Grundbedeutung des Substantivs bleibt erhalten, egal ob das semantische Argument als Komplement oder als Bestandteil des Kompositums realisiert wird. Die Kombinationen von *unter* und einem Kompositum treten mit und ohne Artikel auf, ohne dass sich ein Bedeutungsunterschied ergibt, wie (8a, b) zeigen. In (8b) ist beispielsweise der Artikel weglassbar, ohne dass

sich eine Bedeutungsveränderung ergibt. Wie die Beispiele weiterhin zeigen, ist es auch nicht sinnvoll, dies dadurch zu erklären, dass die Kombination in (8d) regelhaft kompositionell ist, während das Substantiv in (8a) eigentlich nicht mehr als Kompositum analysiert werden sollte. Wenn man auf diese Weise die Produktivität der Bildung in Frage stellt, steht man vor dem Problem, die Beispiele in (8b, c) zu erklären.⁴ Im Abschnitt 2.1 werden wir zunächst verdeutlichen, dass es gute Gründe gibt, auch die Annahme zu verwerfen, dass die Präposition keinen semantischen Beitrag zur Kombination liefert. Eine Diskussion der ersten Position (generelle Idiomatizität der Kombination) erfolgt in Abs. 3.1.

2.1 Semantischer Beitrag der Präposition

In Ausnahmefällen (z.B. bei *anhand*) mag die Bedeutung des Substantivs in der Kombination Präposition+Substantiv tatsächlich sehr schwach ausgeprägt sein, der Regelfall ist es jedoch nicht. Entsprechende Zweifel entstehen, wenn man stattdessen annimmt, dass nicht das Substantiv, sondern die Präposition in ihrer Bedeutung abgeschwächt bzw. als reines Synsemantikon verwendet wird. Nach dieser Hypothese, die wir im folgenden *Hypothese der synsemantischen Präposition (HSP)* nennen wollen, würde etwa die Interpretation der Kombination in (9) gänzlich durch das Substantiv bestimmt.

- (9) Irak ist gemäß der Waffenstillstandsresolution zur Beendigung des Golf-Krieges verpflichtet, seine Massenvernichtungswaffen und deren Produktionsanlagen **unter Aufsicht** der UN zu zerstören.

Dies steht etwa im Widerspruch zur Bedeutungsbestimmung der Präposition *unter* im Duden-Wörterbuch Deutsch als Fremdsprache (DDaF). Das DDaF zählt die in (9) aufgeführte Verwendung von *unter* zur Kernbedeutung von *unter* (DDaF, 2002, 948f.): „(mit Dativ und Akkusativ) kennzeichnet eine Abhängigkeit, Unterordnung o.Ä.: *unter Aufsicht, unter jemandes Leitung; unter ärztlicher Kontrolle; unter jemandem arbeiten, unter jemandem stehen, jemand oder etwas unter sich haben ...*“. Hierbei wird auch nicht unterschieden zwischen diesen Verwendungen von *unter* im Kontext eines Artikels oder einer ähnlichen Konstruktion, die ein NP-Komplement induziert (*unter jemandes Leitung*) und den hier einschlägigen Kontexten. Auch aus reinen Plausibilitätserwägungen sollte die Hypothese verworfen werden, dass die Präposition in die-

⁴Eine interessante Fragestellung ergibt sich in Hinsicht auf die Frage, welche Artikel in den Beispielen in (8) hinzufügar bzw. weglassbar sind. Der definite Artikel in (8b) kann sicherlich als weglassbar gelten, allerdings ist in (8c) wohl nur ein *indefinit* Artikel hinzufügar, während in anderen Konstruktionen definite und indefinite Artikel frei austauschbar sind.

(i) Ursprünglich war der Artikel als Verbot der Beleidigung jeder Religion eingeführt worden, der Diktator engte ihn jedoch nach 1980 auf die Beleidigung des Islam ein, und er verschärfte ihn **durch** ~~o/~~**eine/die Androhung** der Todesstrafe oder lebenslänglicher Haft.
Ich danke Ingrid Kaufmann für diesen Hinweis.

ser Konstruktion synsemantisch verwendet wird. Wenn *unter jemandes Leitung* kompositionell analysiert werden soll, dann sicherlich auch *unter Leitung von*. Aber auch über reine Plausibilitätserwägungen hinausgehende Überlegungen verdeutlichen, dass die HSP in Präposition-Substantiv-Kombinationen zumindest nicht den Regelfall beschreibt.

2.2 Konditionale Interpretationen von *unter*

Der DDaF listet unter den Bedeutungen für *unter* u.a. auch eine konditionale Bedeutung und eine Bedeutung des Begleitumstands auf.

- (10) mit Dativ) **a)** kennzeichnet einen Begleitumstand: *unter Tränen, unter Schmerzen, unter dem Beifall der Menge zogen sie durch die Stadt. ... c)* kennzeichnet eine Bedingung o.Ä.: *unter der Voraussetzung, unter der Bedingung; sie akzeptierte es nur unter Vorbehalt.* (DDaF, 2002, 948f.)

Für die Präposition *mit* gilt wiederum, dass sie zwar einen Begleitumstand denotieren kann, nicht jedoch eine Bedingung im o.g. Sinne.⁵ Wenn es also nicht um die konditionale Lesart geht, dann sollten *mit* und *unter* frei austauschbar sein, was durch (11) auch bestätigt wird.

- (11) Die Saarbergwerke hingegen rechnen **mit/unter Berufung** auf „ernstzunehmende Energieprognosen“ mit einem Exportbedarf beim Strom.

Wenn die Konstruktion aber intentional konditional interpretiert wird, dann sollte ein solcher Austausch dann nicht möglich sein, wenn die Präposition tatsächlich bedeutungskonstitutiv wirkt.

- (12) a.* Die Arbeitsgemeinschaft Berliner Mieterberatungsgesellschaften hatte bereits Ende Dezember die betroffenen MieterInnen aufgefordert, die Mieterhöhung im Januar nur **mit Vorbehalt** zu zahlen.
b. Die Arbeitsgemeinschaft Berliner Mieterberatungsgesellschaften ... hatte bereits Ende Dezember die betroffenen MieterInnen aufgefordert, die Mieterhöhung im Januar nur **unter Vorbehalt** zu zahlen.

In Hinsicht auf die Beispiele in (11) und (12) könnte man einwenden, dass sie tatsächlich nur zeigen, dass die Substantive *Berufung* und *Vorbehalt* eben entweder eine konditionale Interpretation besitzen oder eben nicht. Dies kann im Falle von *Vorbehalt* zumindest nicht ausgeschlos-

⁵Die Präposition *mit* besitzt – wie Helbig and Buscha (2001) zurecht anmerken – eine konditionale Bedeutung: *Mit Glück kann er das Spiel gewinnen., mit etwas mehr Zeit, mit einiger Übung.* Der Unterschied zwischen *mit* und *unter* scheint darin zu liegen, dass konditionale *mit*-PPen eine Bedingung einführen, die außerhalb eines intentionalen Handelnden liegen, während *unter*-PPen eine Bedingung einführen, die intentional im Handelnden verankert werden kann. Um gerade diese Lesart geht es hier.

sen werden.⁶ Man muss jedoch anerkennen, dass eine solche konditionale Interpretation nicht auf ein Substantiv wie *Belastung* angewandt werden kann. Dennoch zeigen Kombinationen mit *unter Belastung* die einschlägige konditionale Interpretation und verbieten eine Ersetzung von *unter* durch *mit*:

- (13) a. Im Gegenteil, oft steigern sie sich **unter Belastung** zu neuen Bestleistungen, wie der Wettkampf in Mosnang bewies.
b. **Unter Belastung** braucht das Herz für die grössere Belastung auch mehr Sauerstoff, der durch die roten Blutkörperchen dem Herzen über drei grosse Kranzgefässe zugeleitet wird.
- (14) a.* Im Gegenteil, oft steigern sie sich **mit Belastung** zu neuen Bestleistungen, wie der Wettkampf in Mosnang bewies.
b.* **Mit Belastung** braucht das Herz für die grössere Belastung auch mehr Sauerstoff, der durch die roten Blutkörperchen dem Herzen über drei grosse Kranzgefässe zugeleitet wird.
- (15) a. Ab Neujahr wird er wieder **mit Belastung** trainieren können.
b. „Er hat mir geraten, nach jedem Dauerlauf noch eine 56er-Runde zu laufen. Oder 400m in einer Minute **mit Belastung**, dann ohne. Man muß es probieren!“
- (16) a. Ab Neujahr wird er wieder **unter Belastung** trainieren können.
b. „Er hat mir geraten, nach jedem Dauerlauf noch eine 56er-Runde zu laufen. Oder 400m in einer Minute **unter Belastung**, dann ohne. Man muß es probieren!“

Die Beispiele in (13) zeigen zunächst typische konditionale Interpretationen von *unter Belastung*. Die Beispiele in (14) zeigen, dass *unter* hier nicht durch *mit* ersetzt werden kann. Die Beispiele in (15) verdeutlichen, dass eine Kombination von *mit* und *Belastung* keineswegs aus unabhängigen Gründen verboten ist. Die Beispiele in (16) wurden aus (15) abgeleitet, indem *mit* in (15) durch *unter* in (16) ersetzt wurde. Da hier eine Interpretation des Begleitumstands (bzw. der Art und Weise) vorliegt, erwarten wir die Möglichkeit einer Kombination mit *unter* ebenso wie mit *mit*.

Insgesamt zeigt die Diskussion, dass die Präposition ebenso wie das Substantiv in der Kombination bedeutungskonstitutiv ist, somit also insgesamt bezweifelt werden muss, dass im Regelfall das Substantiv oder die Präposition synsemantisch zu analysieren sind.

2.3 Valenzmuster

Die Diskussion im vorhergehenden Abschnitt sollte bereits verdeutlicht haben, dass die *Hypothese der synsemantischen Präposition (HSP)* nicht aufrechterhalten werden kann. Aber auch im

⁶Ich danke Stefanie Dipper, die diesen Einwand formulierte.

Zusammenspiel zwischen Syntax und Semantik lassen sich Argumente gegen diese Hypothese finden, wenn man die Valenz und die mögliche Realisation des Substantivs im Kontext unterschiedlicher Präpositionen betrachtet. Hierzu muss man sich zunächst klar machen, dass es vor dem Hintergrund der HSP merkwürdig erscheint, dass ein gegebenes Substantiv mit mehr als einer Präposition auftritt. Wenn die Funktion der Präposition allein darin besteht, eine ansonsten unzulässige adverbelle Verwendung des Substantivs sicherzustellen, dann müsste eigentlich für beliebige Interpretationen ein- und dieselbe Präposition ausreichen. Gegen diese Überlegung könnte man einwenden, dass selbst nach der HSP das Auftreten des Substantivs deswegen mit mehr als einer Präposition zulässig ist, weil das Substantiv mit der Kernbedeutung der Präposition (die im vorliegenden Kontext wegen der rein syntaktischen Funktion der Präposition natürlich ausgeschaltet ist) kompatibel sein muss. Dies vorausgesetzt würden wir eigentlich nur dann das Vorkommen eines Substantivs mit mehr als einer Präposition erwarten, wenn die semantische Kompatibilität der Gesamtphrase mit Präposition₁ nicht jedoch mit Präposition₂ gegeben ist. Anhand von (11), (15) und (16) haben wir allerdings gezeigt, dass selbst die einschränkende Prämisse keine korrekten Vorhersagen macht, so dass die Frage erneut gestellt werden muss: Warum kann ein Substantiv im Kontext mehrerer Präpositionen auftreten, wenn die Präposition selbst keinen semantischen Beitrag zur Phrase liefert?

(17) Valenzdistribution bei *Vorbehalt*

| | unter | mit |
|--|-----------|-----------|
| <i>F</i> (Präposition) | 1.321.466 | 8.926.887 |
| P Vorbehalt | 615 | 266 |
| [<i>PP</i> P [<i>NP</i> ... Vorbehalt]] | 584 | 195 |
| P Vorbehalt gegen | 0 | 0 |
| [<i>PP</i> P [<i>NP</i> ... Vorbehalt gegen]] | 0 | 3 |
| P Vorbehalt [<i>NP</i> ...] | 96 | 4 |
| [<i>PP</i> P [<i>NP</i> ... Vorbehalt [<i>NP</i> ...]]] | 260 | 34 |

Umgekehrt finden wir natürlich auch Substantive, die nur mit bestimmten Präpositionen kombiniert werden können. Dies wurde im vorhergehenden Abschnitt bereits thematisiert, erfährt aber unter der HSP keine Erklärung. Schließlich stellt sich speziell die Frage, wie diese Hypothese mit der Beobachtung umgeht, dass die Valenz eines Substantivs bzw. eines substantivischen Kopfes in nicht unwesentlicher Weise durch die Präposition determiniert wird.⁷ Wir haben hierbei alle Vorkommen von *unter* und *mit* mit entweder folgendem Substantiv oder folgender NP aus dem IDS-Archiv *W-öffentlich* extrahiert. Die Sammlung *W-öffentlich* enthält deutsche, österreichische und schweizerische Zeitungskorpora aus den Jahren 1990-2005. Der Umfang aller Korpora beträgt zusammen ca. 1 Milliarde Wörter.

Die zweite Zeile der Tabelle (17) verdeutlicht zunächst, dass *Vorbehalt* im Kontext der Präposition *unter* wesentlich häufiger auftritt als im Kontext der Präposition *mit*, obwohl die Präposition *mit* in den Korpora wesentlich häufiger auftritt als die Präposition *unter*. Von *unter* finden sich ca. 1,3 Millionen Vorkommen, von *mit* fast 9 Millionen. Ein ähnliches Verhältnis findet sich auch im Kontext einer vollständigen Präpositionalphrase (Zeile 3), denn auch vollständige Präpositionalphrasen mit *unter* als Kopf, deren NP-Komplement wiederum *Vorbehalt* als Kopf besitzt, sind häufiger also solche mit *mit* und *Vorbehalt* in den entsprechenden syntaktischen Positionen.

Die Zeilen 4 und 5 der Tabelle (17) zeigen, dass das Auftreten eines PP[+gegen]-Komplements von *Vorbehalt* ganz entschieden eingeschränkt ist, wenn das Substantiv entweder im Kontext einer Präposition (*unter*, *mit*) (Zeile 4) oder einer PP (Zeile 5) erscheint. Diese Beobachtung ist auch insofern bemerkenswert als die Blockade des PP[+gegen]-Komplements nicht als

⁷Trawinski (2003) diskutiert nur solche APPR-NN-Sequenzen, in denen ein obligatorisches Komplement von N vorliegt. Es wäre dennoch interessant zu überlegen, ob ihr Ansatz auf die in (17) vorgestellte Distribution übertragen werden kann. Trawinski verwendet zur Analyse den Mechanismus der Argumentanhebung. Dieser sieht aber nicht vor, dass eine abhängige Valenz nicht nur von einem höheren Regenten übernommen, sondern durch die Übernahme auch verändert wird.

syntaktische Anomalie der Konstruktion P *Vorbehalt* abgetan werden kann: Das Verhalten im Kontext [_{PP} P [_{NP} ... [_{PP} gegen ...]] ist ja nahezu gleich.

Die Beobachtungen der Zeilen 2 bis 5 der Tabelle (17) finden vor dem Hintergrund der Überlegungen des Abschnitts 2.2 eine einfache Erklärung: Die Präposition *unter* bietet insgesamt ein breiteres Spektrum von Kernbedeutungen als die Präposition *mit*, die insbesondere eine konditionale Bedeutung nicht einschließt. Wenn man nun davon ausgeht, dass die Kombination der Präposition mit dem Substantiv einerseits und der NP andererseits jeweils kompositionell erfolgt, dann erwartet man geradezu, dass die Eigenschaften der einen Kombination auch als Eigenschaften der anderen Kombination auftauchen. Was auch immer der Grund sein mag, der im Kontext der Präposition *unter* die Realisation des PP-Komplements des Substantivs mit *gegen* verbietet, dieser Grund ist in jedem Fall von der regulären Kombination [_{PP} P NP] auf die Kombination aus P und Substantiv übertragbar. Daraus folgt natürlich, dass ein Sonderstatus der einen Kombination zumindest vor dem Hintergrund der Valenzwahl nicht behauptet werden kann.

Die hier bereits beobachteten Korrelationen zwischen regulären PPen einerseits und P-Substantiv-Kombinationen andererseits setzen sich auch bei der Wahl eines postnominalen Genitivkomplements in den Zeilen 6 und 7 fort.

Diese Überlegungen zeigen, dass die Valenzrealisationen an die regierende Präposition gebunden sind, ganz unabhängig davon, ob es sich um eine reguläre Präpositionalphrase oder um die hier in Frage stehenden Kombinationen handelt. Wenn dies für beide Konstruktionen gilt, dann muss man in der Konsequenz auch behaupten, dass eine synsemantische Präposition sowohl in den Kombinationen aus Präposition und Substantiv als auch in den Präpositionalphrasen aktiv ist, was wiederum nicht in Einklang zu bringen wäre mit den semantischen Abweichungen, die in Abschnitt 2.2 diskutiert wurden. Dass die Kombinationen mit *unter* wesentlich häufiger sind als die Kombinationen mit *mit*, erklärt sich unmittelbar, wenn man annimmt, dass *unter* nicht nur eine von *mit* verschiedene, sondern eben auch reichhaltigere Bedeutung besitzt, was erneut im Widerspruch zur HSP steht.

3 Die empirische Bestimmung von Idiomatizität und Produktivität

Die Überlegungen des letzten Abschnitts haben gezeigt, dass eine vollständige nicht-kompositionelle Analyse von Präposition-Substantiv-Sequenzen unplausibel ist. In diesem Abschnitt werden wir zwei computerlinguistische Verfahren anwenden, die eine empirische Bestimmung der Idiomatizität und der Produktivität der betreffenden Konstruktionen gestatten. Das von Dunning (1993) vorgeschlagene Verfahren der Bestimmung der *likelihood ratio* als Assoziationsmaß für die strukturelle wechselseitige Abhängigkeit zweier Worte kann verwendet werden, um zu be-

stimmen, wie viele der relevanten Instanzen tatsächlich idiomatisch sind. Die Berechnung der Produktivität einer Konstruktion basiert auf dem von Baayen (2001) und Evert (2004) vorgeschlagenen Verfahren. Hierbei interessieren uns sowohl Substantive, die zählbar sind und im Singular mit der Präposition *unter* kombiniert werden, als auch Substantive, die als Objekt von *unter* im Plural realisiert worden sind. Die Produktivität und Kompositionalität der letzten Gruppe steht außer Frage. Gerade deswegen kann sie für die verwendeten Verfahren auch als Kontrollgruppe verwendet werden.

3.1 Kompositionalität und likelihood ratio

Ein Maß für die Idiomatizität eines Ausdrucks kann im allgemeinen Fall schon deswegen nicht bestimmt werden, weil das Konzept der Idiomatizität nicht unbedingt klar umrissen ist. Dazu kommt, dass eine allgemeine Bestimmung der Idiomatizität eines Ausdrucks berücksichtigen müsste, dass Idiome unterschiedlich strukturiert sind und damit im Regelfall aus unterschiedlich vielen Elementen bestehen. Der vorliegende Fall ist allerdings ein Spezialfall, der eine allgemeine quantitative Bestimmung der Idiomatizität deswegen gestattet, weil die hier fraglichen Sequenzen immer aus genau zwei Elementen bestehen. Somit kann man Sequenzen aus Präposition und Substantiv auch als Präposition-Substantiv-Kollokationen charakterisieren. Eine Kombination aus Präposition und Substantiv muss, wenn sie ein Idiom ist, eine feste Fügung sein. Natürlich ist nicht jede Kollokation ein Idiom, aber jedes Idiom, das aus zwei adjazenten Elementen besteht, muss eine Kollokation sein. Im Gegensatz zur Bemessung allgemeiner Idiome ist die Computerlinguistik in der quantitativen Identifikation von Kollokationen recht erfolgreich. Die Grundüberlegung ist nun, zu bestimmen, ob es sich bei den jeweiligen Kandidatenpaaren um Kollokationen handelt. Wenn Sequenzen aus Präposition und Substantiv im Regelfall idiomatisch sind, dann sollte eine sehr hohe Anzahl dieser Sequenzen auch als Kollokation bestimmt werden können. Wenn umgekehrt nur eine relativ geringe Anzahl dieser Sequenzen als Kollokation bestimmt werden kann, dann muss die Idiomatizität der Konstruktion deutlich in Frage gestellt werden.

Das erfolgreichste Maß zur Identifikation einer Kollokation ist Dunning's (1993) *log likelihood ratio*, das, wie (18) zeigt, auf dem Verhältnis zwischen einer Unabhängigkeitshypothese und einer Abhängigkeitshypothese basiert und angibt, wie viel wahrscheinlicher das Auftreten eines Paares $Wort_1$ $Wort_2$ unter der Abhängigkeitshypothese in einem Korpus wird.

$$(18) \quad \log \lambda = -2 \cdot P(w_2 \text{ ist von } w_1 \text{ abhängig}) / P(w_2 \text{ ist nicht von } w_1 \text{ abhängig})$$

Die Bestimmung einer *log likelihood ratio* ist unter anderem deswegen den anderen Ver-

fahren zur Identifikation von Kollokationen vorzuziehen, weil sie bereits auf relativ geringe Datenmengen angewandt werden kann.⁸

Das Verfahren gibt einen Schwellenwert an, den eine feste Fügung überschreiten muss, um überhaupt als Kollokationskandidat zu gelten. Es hat allerdings auch eine klare Interpretation zum Vergleich der Werte: Wenn ein Bigramm $Wort_1 Wort_2$ einen *log likelihood*-Wert von n besitzt und dieser Wert über dem Schwellenwert liegt, dann wird das tatsächliche Auftreten des Paares mit einer $e^{n/2}$ -mal höheren Wahrscheinlichkeit durch die Hypothese vorhergesagt, dass eine Abhängigkeit zwischen $Wort_1$ und $Wort_2$ besteht, als durch die Hypothese, dass $Wort_1$ und $Wort_2$ „zufällig“ nebeneinander stehen (vgl. Manning and Schütze (1999, 173ff.)). Als Schwellenwert für eine Restfehlerrate von 0,5 % setzen wir hier 7,88. Wenn wir nun als Illustration einen Wert für $n = 8$ verwenden, der nur geringfügig höher liegt als dieser Schwellenwert, so wird das tatsächliche Auftreten durch die Annahme einer Abhängigkeit zwischen $Wort_1$ und $Wort_2$ knapp 55-mal wahrscheinlicher. Bei der Beobachtung eines *log likelihood*-Wertes von $n = 35$ kommt man aufgrund der exponentiellen Steigerungsrate schon auf einen Wert von $3,9 \cdot 10^7$!

Der Leser wird sicherlich mit mir darin übereinstimmen, dass die Hypothese einer vollkommenen Unabhängigkeit von $Wort_1$ und $Wort_2$ im Kontext einer syntaktischen Realisation außerordentlich unwahrscheinlich ist und somit erst ein relativ hoher *log likelihood*-Wert überhaupt die Schlussfolgerung gestattet, dass hier tatsächlich eine Abhängigkeit, somit eine feste Fügung und damit eine idiomatische Verbindung zwischen $Wort_1$ und $Wort_2$ besteht.

In der folgenden Untersuchung haben wir für den Jahrgang 1995 des NZZ-Korpus die *log likelihood ratio* für Paare der Form *unter+Substantiv_{sg}* und *unter+Substantiv_{pl}* bestimmt. Der Umfang des Jahrgangs beträgt ca. 26 Millionen Wörter.⁹ Substantive im Plural können im Deutschen regelhaft ohne Artikel erscheinen. Die Konstruktion sollte somit in etwa so viele idiomatische Instanzen besitzen wie jede regelhafte syntaktische Konstruktion des Deutschen.

Die Extraktion der Daten aus dem Korpus erfolgte mit dem Perl-Modul TEXT::NSP. Die resultierenden Daten wurden zunächst nach elementaren orthographischen Gesichtspunkten gefiltert (Substantive beginnen mit Großbuchstaben), automatisch nach Singular- und Pluralvorkommen klassifiziert und anschließend händisch aufbereitet, um bei den Singularvorkommen nur zählbare Gattungsbezeichnungen zu betrachten. Die Klassifikation in Singular- und Pluralvorkommen basiert auf einem von Jan Strunk aufbereiteten Lexikon, das aus der Tübinger Baumbank des Deutschen/Schriftsprache (TüBa-D/Z) in der Version mit Koreferenzannotation (vom 14.07.2006) extrahiert wurde, vgl. http://www.sfs.uni-tuebingen.de/de_tuebadz.shtml.

⁸Eine detaillierte Darstellung der Interpretation von (20) findet sich – ebenso wie die mathematischen Details – in Manning and Schütze (1999).

⁹Zur Bestimmung von Kollokationen sind Korpora in einer Größe von 10-25 Millionen Wörtern vollkommen ausreichend und man kann ggf. auch mit wesentlich kleineren Kandidatenmengen arbeiten (vgl. Manning and Schütze (1999, 163-175), Kiss and Strunk (2006))

Für die Bestimmung der Zählbarkeit wurden die folgenden Wörterbücher konsultiert, wobei ein Substantiv, das nur in einem der Wörterbücher als nicht zählbar beschrieben wurde, bereits ausgeschlossen wurde: Mackensen (1977), Wahrig (2006), Bertelsmann *Wörterbuch der deutschen Sprache*, Wahrig *Rechtschreibung* (die letzteren beiden über *wissen.de*), das *Deutsche morphologische Wörterbuch* von Canoo.net, das digitale Wörterbuch der Deutschen Sprache des 20. Jahrhunderts (DWDS).

Eine Vielzahl von Substantiven besitzen entweder homonyme oder polyseme Instanzen, die sich in der Zählbarkeit voneinander unterscheiden. Als Beispiel mag hierzu *Druck* dienen, das ja unter anderem auch das Erzeugnis eines Druckvorgangs benennen kann und als solches natürlich zählbar ist. Aber diese Interpretation von *Druck* ist natürlich im Kontext *unter* irrelevant, so dass dieses Paar nicht betrachtet wurde. Im Einzelfall war somit jeweils zu entscheiden, ob die hier relevante Lesart auch diejenige ist, die tatsächlich im Kontext von *unter* erscheint. Ein heuristisches Verfahren hierzu ist die Kombination des betreffenden Substantivs mit *viel/viele* bzw. *einig/einige*. Ist ein Substantiv nicht zählbar, dann ist eine Kombination mit *viel+Singular* erforderlich, so wie etwa *unter Druck* \Rightarrow *unter viel Druck/*unter vielen Drucken*, während bei zählbaren Substantiven im Regelfall eine Kombination mit *viele+Plural* erforderlich bzw. zumindest möglich ist.¹⁰

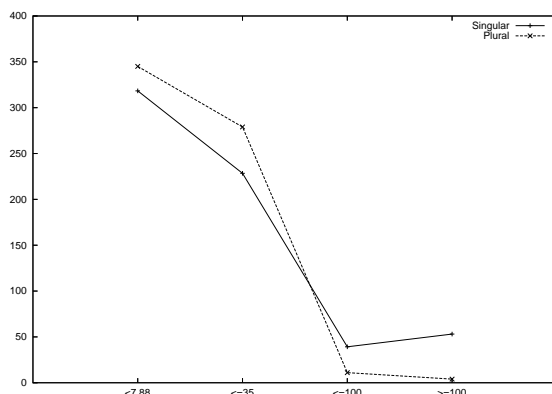
Wenn das Substantiv zählbar ist, ist eine Kombination mit *viele+Plural* erforderlich, ist es nicht zählbar, ist eine Kombination mit *viel+Singular* erforderlich, so wie etwa *unter Druck* \Rightarrow *unter viel Druck/*unter vielen Drucken*. Das Ergebnis dieser Untersuchung ist in der Graphik (19) zusammengefasst.

Das Diagramm zeigt die Verteilung der Kandidaten mit zählbaren Singularia und Pluralia. Wenn wir uns zunächst nur auf die Werte für die Kombinationen mit Singularformen konzentrieren, stellen wir fest, dass nahezu 50 % aller Kandidatenpaare aus der Stichprobe schon dann durch den Test fallen, wenn man die sowieso wenig plausible Hypothese vertritt, dass adjazente Wörter in einem Korpus vollkommen unabhängig voneinander sein können. Ein sehr geringer Teil der Kandidaten bei den Singularformen erreicht einen *log likelihood*-Wert von 100 und mehr und lässt somit den Schluss zu, dass es sich hierbei um eine Kollokation handelt. Dies zeigt, dass es zumindest keine empirischen Argumente dafür gibt, die Konstruktion *unter+Substantiv_{sg}* insgesamt als Kollokation zu klassifizieren. Wenn sie aber keine Kollokation ist, kann sie aufgrund ihrer syntaktischen Struktur – Adjazenz von zwei Wörtern w_1 und w_2 – auch nicht als

¹⁰Das Verfahren ist heuristisch und es folgt aus der Kombinierbarkeit mit *viel+Singular* noch nicht, dass das betreffende Substantiv nicht zählbar ist, wie anhand des folgenden Beispiels mit *Vorbehalt* verdeutlicht werden kann:

- (i) Paracelsus, von den Deutschen beansprucht als grosser Deutscher, von den Franzosen als grosser Franzose (Bürgerrecht Strassburg), von den Österreichern als grosser Österreicher und von den Schweizern als grosser Schweizer kann also nur *mit sehr viel Vorbehalt* zu einem grossen Appenzeller gemacht werden. (A97/OKT.30484 St. Galler Tagblatt, 20.10.1997; «Das Weltbild der Heilkunst verändert»)

(19) Verteilung der *log likelihood* bei Singular- und Pluralvorkommen im Kontext von *unter*



festen Fügung und somit auch nicht als Idiom analysiert werden.

Die Verteilung der Kombinationen mit Pluralformen zeigt eine sehr große Ähnlichkeit mit der Distribution der Singularformen, obwohl die prinzipielle Regelmäßigkeit und damit auch Kompositionalität der Kombinationen mit Pluralformen außer Frage steht. Bei den Pluralvorkommen findet sich ein noch größerer Prozentsatz mit einem Wert von weniger als 7,88 und Kandidaten mit einem Wert jenseits der 35 sind extrem selten. Aus letzterem folgt, dass es nur sehr wenige Kombinationen des Typs *unter+Substantiv_{pl}* gibt, die extrem feste Fügungen bilden. Aus ersterem folgt, dass die Kombinationen mit Pluralvorkommen in etwa so viele freie Wortverbindungen umfasst wie die Kombination mit Singularvorkommen. Insgesamt verhalten sich also beide Gruppen sehr ähnlich, was in der Konsequenz bedeutet, dass man auch die Realisationen mit Pluralformen als eher idiomatisch bezeichnen müsste, wollte man Fleischer (1982) folgen und die Realisationen mit Singularformen als überwiegend idiomatisch bezeichnen.¹¹

Es erscheint mir vor dem Hintergrund dieser Ergebnisse allerdings sinnvoller zu behaupten, dass sowohl in der Gruppe der Kombinationen mit Singularformen als auch in der Gruppe der Kombinationen mit Pluralformen idiomatische Verbindungen auftreten, obwohl die letzteren syntaktisch natürlich in einem noch zu bestimmenden Sinne regelhafter sind als die ersteren. Das Auftreten idiomatischer Verbindungen in einer Kombination erlaubt allerdings keineswegs den Schluss, dass die Kombination selbst idiomatisch ist. Richtig ist es vielmehr zu akzeptieren, dass es für jede Kombination eine Gruppe von Instanzen gibt, deren Verbindung idiomatisch ist. Wenn wir annehmen (ohne es allerdings hier beweisen zu können), dass die Kombination

¹¹Ein Beispiel für einen sehr hohen *log likelihood*-Wert ist *unter Druck* mit 5.422,6. Aber selbst *unter Druck* – das nicht zu den hier relevanten Kandidaten zählt, weil *Druck* in der hier einschlägigen Lesart nicht zählbar ist – ist zumindest dekomponierbar im Sinne von Nunberg et al. (1994).

unter+Substantiv_{pl} in diesem Sinne regelhaft ist, dann bedeutet die o.g. Distribution, dass auch in einer regelhaften Kombination festere Fügungen mit einem hohen Prozentsatz auftreten und – da die Verteilungen bis zum Wert 35 doch sehr ähnlich sind – dass entsprechend auch die Kombination *unter+Substantiv_{sg}* zu den in diesem Sinne semantisch regelhaften Kombinationen zu zählen ist, also insgesamt nicht als feste Fügung oder „zum größten Teil idiomatisiert“ bezeichnet werden darf.

3.2 Die Bestimmung der Produktivität: Korpusaufbereitung

Baayen (2001) beschreibt einen statistischen Ansatz zur Bestimmung morphologischer Produktivität, der in einem Produktivitätsmaß mündet, das eine natürliche Interpretation für die Produktivität einer bestimmten morphologischen Klasse besitzt. Dieses Produktivitätsmaß, bezeichnet durch $P(N)$, gibt an, wie produktiv eine morphologische Operation in einem Korpus mit N Wortvorkommen ist. Die natürliche Interpretation dieses Maßes ist die Wahrscheinlichkeit, mit der man einen neuen Ausdruck einer morphologischen Klasse findet, nachdem man schon N Worte betrachtet hat.

Es bietet sich an, dieses Produktivitätsmaß auf Sequenzen von *unter* und zählbarem Substantiv im Singular zu übertragen und die Produktivität dieser Kombination zu bestimmen. Baayens Ansatz entwirft ein Modell, auf dessen Basis für Korpora die erwartete Produktivität einer morphologischen Klasse bestimmt werden kann. Evert (2004) stellt zwei Modelle vor, die die Übertragung dieses Modells auch auf syntaktische Kombinationen gestattet. In Dömges et al. (2007) diskutieren wir die mathematischen Aspekte dieses Verfahrens, insbesondere die Implikationen, die sich daraus ergeben, dass ein infinites Modell bessere Näherungswerte liefert als ein Modell, das von einem endlichen Vokabular ausgeht. Der vorliegende Ansatz basiert somit auf den Arbeiten von Evert (2004) und Dömges et al. (2007).

Zur Bestimmung des Produktivitätsmaßes verwenden wir zwei Korpora: einerseits vier Jahrgänge der *Neuen Zürcher Zeitung* (NZZ 1995-1998) und andererseits drei Jahrgänge des *Frankfurter Rundschau* (FRR 1997-1999). Aus diesen beiden Korpora wurden alle Vorkommen von *unter* mit folgendem Wort beginnend mit Großbuchstaben extrahiert und nach dem Vorkommen sortiert. Die Korpora umfassen knapp 213 Millionen Wörter. Die Datenextraktion erfolgte auf Stichproben mit 6, 12, 26, 52, 78, 107 und 213 Millionen Wörtern. Die Ausgabe der Kandidatenpaare erfolgte auf der Basis ihrer *log likelihood*-Werte. Dieses Vorgehen liefert zum einen eine sehr gut lesbare Strukturierung der Daten und erlaubt zum anderen die leichte Extraktion solcher Kandidaten, die tatsächlich als idiomatische Kombinationen gewertet werden können.

Auf der Basis der Stichproben erfolgte eine erste automatisierte Filterung der Daten, an die

sich eine händische Filterung angeschlossen hat.¹² Die automatisierte Filterung umfasste einerseits die Lemmatisierung auf der Basis des oben bereits angesprochenen Lexikons und andererseits die Identifikation von *unter+Substantiv* im Kontext von Funktionsverbgefügen. Die Filterung wurde schrittweise auf den Stichproben durchgeführt, so dass die Ergebnisse der händischen Filterung bei den jeweils folgenden Stichproben als Basis der Filterung diente, die dann wiederum händisch komplettiert wurde. Insgesamt erfolgte die Filterung der Daten in den folgenden Schritten:

1. Extraktion von *unter+Eigennamen, Genitiv, Adjektiv*: beispielsweise *unter Sun-Yat-Sen, unter Berns* oder *unter Diversem*.¹³
2. Extraktion von *unter+Titel*: beispielsweise *unter Premierminister, unter PLO-Chef*.
3. Extraktion von *unter+Zahl* oder *Maßeinheit*: beispielsweise *unter Null, unter Prozent*.
4. Extraktion von *unter+Substantiv* als Resultat einer Konversion, die in einem nicht zählbaren Substantiv resultiert: beispielsweise *unter Nutzen, unter Kopfschütteln*.
5. Extraktion von *unter+Substantiv im Plural*.
6. Extraktion aller als sicher idiomatisch geltenden Kandidatenpaare: *unter Dach* (aus *unter Dach und Fach*).
7. Extraktion aller Kombinationen *unter+Substantiv im Singular*, in denen das Substantiv nicht zählbar ist (auf Basis des in Abs. 3.1 beschriebenen manuellen Verfahrens), etwa *unter Druck, unter Mitwirkung*.
8. Extraktion aller Kombinationen *unter+Substantiv im Singular*, die vornehmlich in Funktionsverbgefügen auftreten: beispielsweise *unter Kontrolle, unter Beweis, unter Vertrag*.

Bei der Extraktion der Kombinationen, die vornehmlich in Funktionsverbgefügen auftreten, erfolgte die Analyse auf Basis der Kookkurenzanalyse, die als Bestandteil von COSMAS

¹²Für die Hilfe bei dieser äußerst mühseligen Arbeit danke ich Antje Müller und Claudia Roch nochmals ganz ausdrücklich.

¹³Zur Extraktion von Eigennamen und nominalisierten Adjektiven hätte man sicherlich auch eine automatisierte Filterung durch POS-Tagging verwenden können. Allerdings liefern POS-Tagger gerade in diesen Kontexten im Deutschen häufig Fehler, so dass auch in diesem Fall eine händische Kontrolle notwendig gewesen wäre. Für zukünftige Arbeiten wird man allerdings sicherlich direkt auf getaggtten Korpora aufsetzen.

zur Verfügung gestellt wird.¹⁴ Ein Kandidatenpaar wurde schon dann ausgeschlossen, wenn es entsprechend der Kookkurenzanalyse mit einem oder mehreren Verben ein Funktionsverbgefüge bildet. Es wurde nicht überprüft, ob das entsprechende Kandidatenpaar auch außerhalb dieses Funktionsverbgefüges auftritt. Bei zukünftigen Untersuchungen könnte die Kandidatenmenge zur Bestimmung des Produktivitätsmaßes also noch um solche Kandidatenpaare erweitert werden, die zwar in Funktionsverbgefügen auftreten, aber nicht nur in Funktionsverbgefügen. Eine interessante Untergruppe dieser (ausgeschlossenen) Kandidaten sind Kombination von *unter+N+leiden*, wobei *N* etwa *Fäulnis*, *Rheuma*, *Schwindsucht*, *Langeweile*. Die verbleibenden Kandidatenpaare wurden in die Berechnung des Produktivitätsmaßes einbezogen.

3.3 Die Bestimmung der Produktivität: Produktivitätsmaß

Das Produktivitätsmaß kann am besten anhand eines Beispiels eines nicht produktiven Prozesses erläutert werden. Baayen (2001) bietet hierzu die Produktivität des Würfels an. Wie kann dies sichtbar gemacht werden? Wir betrachten hierzu $V(N)$, die Anzahl der Vokabulartypen in einem Korpus von N Vorkommen und $V(1, N)$, die Anzahl derjenigen Typen des Vokabulars, die nur einmal in einem Korpus von N Wörtern vorkommen, d.h. die *Hapax legomena*. Beim Würfeln wird sich $V(N)$ so entwickeln, dass irgendwann einmal die Seitenzahl des Würfels erreicht ist – das ist das nicht-produktive Vokabular des Würfels. Gleichzeitig wird $V(1, N)$ zunächst willkürlich ansteigen, um nach weiteren Würfelversuchen schließlich auf 0 zurückzufallen, weil es keine Zahl mehr geben wird, die nur einmal gewürfelt worden ist. Das Produktivitätsmaß kann nun bestimmt werden als das Verhältnis zwischen $E[V(1, N)]$, d.h. dem Erwartungswert für Typen, die in einer Stichprobe von N Kandidaten nur einmal auftreten, und N selbst, so wie dies in (20) dargestellt ist.

$$(20) \quad P(N) = E[V(1, N)]/N$$

Im vorliegenden Kontext werden wir also fragen: Wie ist der Anteil des einmal vorkommenden Vokabulars eines bestimmten Typs in einem Gesamtvorkommen von N Wörtern? Das resultierende Verhältnis $P(N)$ ist auch zu deuten als die Wahrscheinlichkeit, dass wir nach N Wörtern auf einen neuen Typ des betreffenden Vokabulars treffen.

Ist eine Kombination nicht produktiv, so erwarten wir über eine lange Anzahl von Versuchen (d.h. bei einem immer umfangreicher werdenden Korpus), dass $V(N)$ – die Anzahl der Kandi-

¹⁴Cyril Belica: Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analysemethoden. © 1995 Institut für Deutsche Sprache, Mannheim. Ein anonymer Gutachter merkte zu Recht an, dass es wünschenswert wäre zu charakterisieren, wie die Kookkurenzanalyse arbeitet. Bedauerlicherweise wird das Verfahren jedoch nicht beschrieben, andererseits ergab die manuelle Kontrolle, dass nur sehr wenige der verbleibenden Kandidaten als Bestandteil eines Funktionsverbgefüges bestimmt werden konnten. Diese wurden dann entsprechend manuell aussortiert.

datenpaare *unter+Substantiv_{sg}* – einen asymptotischen Verlauf nimmt, also irgendwann einmal nicht mehr ansteigt. Schließlich sollte $V(1, N)$ in einem solchen Fall 0 werden und entsprechend $P(N)$ dann auch auf 0 fallen.

Anstatt $P(N)$ nun wie Baayen (2001) als Wahrscheinlichkeit des Auftretens eines neuen Worttyps zu bestimmen, nachdem bereits N Wörter betrachtet wurden, bestimmen wir die Wahrscheinlichkeit des Auftretens von $Wort_2$ nach einem festgelegten $Wort_1 = unter$, nachdem bereits N einschlägige Konstruktionen, d.h. *unter+Substantiv_{sg}*, beobachtet wurden. Wir messen also die Wahrscheinlichkeit, dass eine neue Instanz einer Regel gebildet wurde, die $Wort_1 = unter$ und $Wort_2$ verbindet. Die zur Bestimmung der Produktivität im vorliegenden Fall notwendigen Werte sind in Tabelle (21) zusammengefasst, aus der hervorgeht, dass im Korpus von 213 Millionen Wörtern insgesamt 16.427 Kombinationen von *unter+Substantiv_{sg}* auftreten.

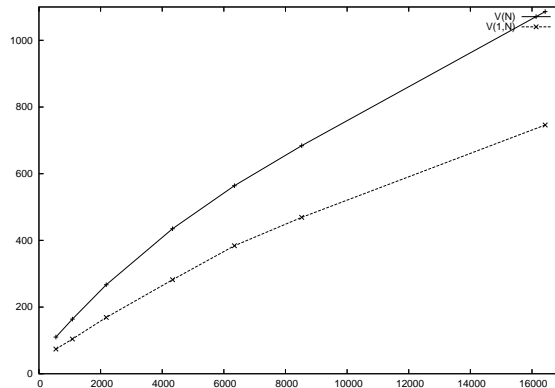
(21) Vokabularentwicklung, *Hapax Legomena* und Produktivität für *unter+Substantiv_{sg}*

| N | V(N) | V(1, N) | E[V(1, N)] | P(N) |
|--------|-------|---------|------------|-------|
| 551 | 110 | 74 | 96,66 | 0,175 |
| 1.086 | 164 | 104 | 123,47 | 0,114 |
| 2.185 | 267 | 169 | 166,41 | 0,076 |
| 4.336 | 435 | 282 | 249,93 | 0,058 |
| 6.343 | 564 | 384 | 332,19 | 0,052 |
| 8.526 | 684 | 469 | 400,43 | 0,047 |
| 16.427 | 1.086 | 746 | 748,81 | 0,046 |

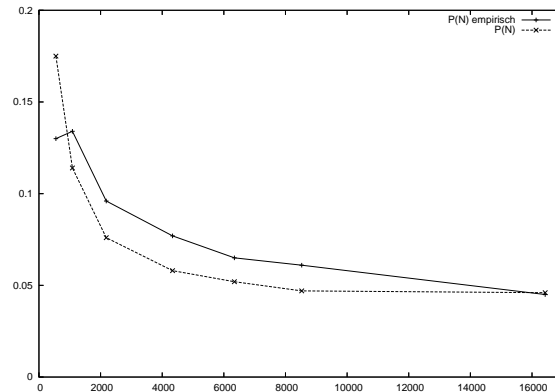
Wie wurde $E[V(1, N)]$ bestimmt? Evert (2004) bietet hierzu zwei Modelle an, die sich darin unterscheiden, dass das eine Modell – ein sog. finites Zipf-Mandelbrot-Modell – von einer endlichen Vokabulargröße ausgeht, während das andere Modell – ein generalisiertes Zipf-Mandelbrot-Modell – die Annahme unendlich vieler Typen zulässt. Dömges et al. (2007) zeigen, wie basierend auf den empirischen Verteilungen $E[V(1, N)]$ geschätzt werden kann, wobei das generalisierte Modell die bessere Schätzung liefert. Anhand der Tabelle (21) ist nun zunächst erkennbar, dass $E[V(1, N)]$ zunächst oberhalb des empirisch beobachtbaren $V(1, N)$ liegt, dann unterhalb des empirisch beobachtbaren und sich bei $N = 16.427$ sehr deutlich annähert. Die folgende Graphik verdeutlicht zunächst, dass weder $V(N)$ noch $V(1, N)$ einen asymptotischen Verlauf nehmen.

Der Verlauf der Kurven in Abbildung (22) ist bereits ein Indikator dafür, dass die Kombination von *unter+Substantiv_{sg}* ein produktiver Prozess ist. Dies wird auch durch den Verlauf von $P(N)$ bestätigt. In der folgenden Abbildung (23) ist sowohl der empirische Verlauf von $P(N)$, d.h. basierend auf $V(1, N)$, als auch der auf der Schätzung basierende Verlauf dargestellt.

(22) Vokabularentwicklung für *unter+Substantiv_{sg}*



(23) Produktivität für *unter+Substantiv_{sg}*



Die Abbildung (23) zeigt, dass die Produktivität bei kleinem N zunächst sehr hoch ist und bei größer werdendem N abfällt. Wichtig ist hierbei insbesondere, dass die Produktivität sich auf einem niedrigen Niveau stabilisiert, was wiederum bedeutet, dass eine relativ gleich bleibende Wahrscheinlichkeit dafür gegeben werden kann, dass bei steigendem N auch neue Vokabulartypen beobachtet werden können. Die Abbildung (23) verdeutlicht weiterhin das Verhältnis zwischen geschätzter und empirischer Produktivität. Der Verlauf der beiden Kurven verdeutlicht die weiter oben getroffene Behauptung, dass die Schätzung bei größer werdender Kandidatenmenge immer klarer dem empirischen Verlauf entspricht. Dennoch wird man sich vielleicht insgesamt fragen, wie produktiv denn nun die Kombination von *unter+Substantiv_{sg}* ist. Um dieser Frage nachzugehen, haben wir $P(N)$ für die Kombination mit Substantiven im Singular mit der Produktivität für die Kombination mit Substantiven im Plural verglichen. Dass die letztere Kombination produktiv ist, steht ja außer Frage. Die wesentlichen Werte sind nun in Tabelle (24)

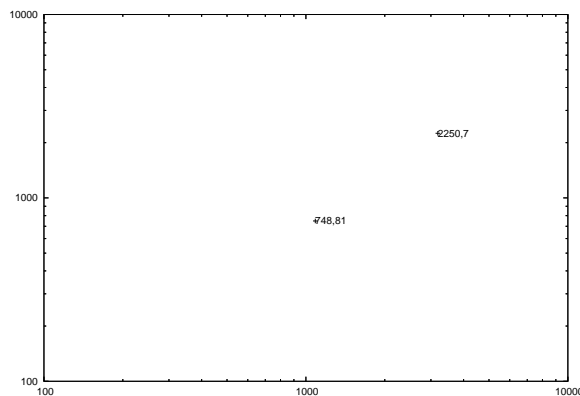
zusammengefasst.

(24) Produktivität von Substantiven im Plural in Kombination mit *unter*

| N | V(N) | E[V(1, N)] | P(N) |
|--------|-------|------------|-------|
| 348 | 191 | 157,00 | 0,451 |
| 681 | 361 | 298,00 | 0,438 |
| 1.431 | 639 | 497,00 | 0,347 |
| 2.910 | 1.117 | 839,00 | 0,288 |
| 4.416 | 1.480 | 1.066,00 | 0,241 |
| 6.056 | 1.855 | 1.314,00 | 0,217 |
| 11.421 | 3.188 | 2.250,70 | 0,197 |

Wir vergleichen nun die Produktivität von *unter+Substantiv_{sg}* und *unter+Substantiv_{pl}* anhand zweier Verhältnisse: Zum einen betrachten wir das Verhältnis zwischen der Vokabularentwicklung überhaupt und der Entwicklung der *Hapax legomena*. Zum anderen das Verhältnis zwischen der Vokabularentwicklung und der Produktivität.

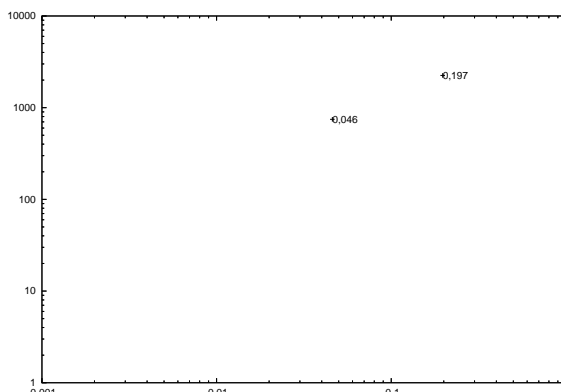
(25) $V(N)/V(1, N)$



Die Abbildung (25) zeigt einerseits, dass die Kombination von *unter* mit einem Substantiv im Plural ein deutlich umfangreicheres Vokabular besitzt, aber – und dies ist hier entscheidend – die beiden Kombinationen teilen in etwa das Verhältnis zwischen Vokabularumfang überhaupt und der *Hapax legomena*. Ein unproduktiver Prozess könnte dennoch eine Vielzahl unterschiedlicher Instanzen besitzen, aber die Anzahl von Typen, die nur einmal auftreten, wäre dann relativ gering, sogar mit Werten unter 100.

Diese Überlegungen werden auch durch den Vergleich in Abbildung (26) bestätigt. Unproduktiv sind Prozesse, wenn sie eine hohe Anzahl von *Hapax legomena* erreichen, die *Hapax legomena* aber mit N nicht ‚mithalten‘ können, wenn es also zwar viele Typen gibt, diese je-

$$(26) \quad V(1, N)/P(N)$$



doch in keinem Verhältnis zu den Kandidaten überhaupt stehen. In einem solchen Fall würde ein Wert in der Nähe der y-Achse beobachtbar sein. Unproduktiv sind Prozesse aber natürlich auch, wenn entweder die Zahl der Hapaxtypen oder die Vokabulartypen überhaupt sehr niedrig sind, wie Baayen (2001) anhand des niederländischen Suffix *-se* verdeutlicht. Dieser besitzt deswegen einen sehr hohen Produktivitätswert, weil praktisch das gesamte Vokabular nur einmal auftritt! Da das Vokabular selbst allerdings außerordentlich klein ist, ergibt sich bei hoher Produktivität zugleich ein sehr niedriger Wert für $V(1, N)$. Die Abbildungen (25) und (26) zeigen somit, dass die Kombination von *unter+Substantiv_{sg}* zwar nicht so produktiv ist wie die Kombination von *unter+Substantiv_{pl}*. Die erstere Kombination zählt aber dennoch klar zu den produktiven Kombinationen. Der in der Duden-Regel 395 implizierte Ansatz einer Möglichkeit der Auflistung der fraglichen Konstruktionen muss somit verworfen werden.

4 Implikationen für eine Analyse

Da die Regelmäßigkeit der Konstruktion Präposition+Substantiv typischerweise in Frage gestellt wird, gibt es nur sehr wenige Analysen dieses Phänomens. Trawinski (2003) und Trawinski et al. (2006) behandeln eine Teilklasse, die sich dadurch auszeichnet, dass Präposition+Substantiv ein obligatorisches PP-Komplement besitzen, also etwa Kombinationen wie *in Verbindung mit*.¹⁵

¹⁵Trawinski (2003) und Trawinski et al. (2006) behaupten, dass *in Verbindung mit* keine interne Modifikation gestattet. Diese Annahme ist sicherlich falsch, wie etwa der folgende Beleg verdeutlicht:

- (i) ... die Korpuskulartheorie entstand also auf der Grenze zwischen Philosophie und Physik, in enger Verbindung mit dem Problem der mathematischen Punkte und des Diskontinuums ... (LIM/LI1.00200, Heller, B.: Grundbegriffe der Physik im Wandel der Zeit; S.82-86)

Trawinski (2003) schlägt hierzu eine Analyse vor, in der die Präposition die Argumente des regierten Substantivs erbt, somit also das Substantiv und das Präpositionalkomplement des Substantivs als Komplemente der Präposition realisiert werden. Die Analyse trifft keine Implikationen für die im vorliegenden Aufsatz behandelten Konstruktionen, ist aber sicherlich nicht auf diese Klasse übertragbar. Eine Anhebungsanalyse würde nämlich implizieren, dass die Valenz des Substantivs ohne Änderungen von der Präposition übernommen wird. Betrachtet man nun als Beispiel das Substantiv *Vorbehalt*, so kann man feststellen, dass dieses ein sententiales Komplement besitzt, aber nicht, wenn es zusammen mit *unter* realisiert wird. Bei einer Anhebungsanalyse müsste man somit Konstruktionen wie **unter Vorbehalt, dass niemand kommt* blockieren.

Eine zweite Analyse findet sich für verwandte Konstruktionen des Englischen in Baldwin et al. (2006). Baldwin et al. (2006) unterteilen Präposition+Substantiv-Kombinationen in drei Klassen, die jeweils unterschiedlich analysiert werden.

Nicht produktive, nicht modifizierbare und nicht-kompositionelle Kombinationen sollten nach Baldwin et al. (2006) schlicht im Lexikon aufgelistet werden. In allen Kombinationen des Typs *unter+Substantiv* gibt es hierfür außerordentlich wenige Kandidaten, einer davon dürfte aber *unter Dach und Fach* sein, das den Kriterien in Gänze genügt, also insbesondere auch keine interne Modifikation zulässt.

Die zweite Klasse umfasst Kombinationen von Präposition und zählbarem Substantiv, in denen ein Substantiv realisiert wird, das auch als unmittelbares Argument eines Verbs ohne Artikel realisiert werden kann. Beispiele aus dem Englischen sind *school* oder *church*, verallgemeinert setzen Baldwin et al. (2006, 166) die semantische Klasse *institution* bzw. *social/geographic spaces* an. Hierbei muss man allerdings berücksichtigen, dass der Artikelgebrauch im Englischen in diesen Kontexten vom Gebrauch des Artikels im Deutschen deutlich abweicht. Substantive dieses Typs finden sich im Deutschen außerordentlich selten, z.B. *Essen ist fertig.* vs. *Das Essen ist fertig.*

Die dritte Gruppe ist schließlich diejenige Gruppe, der die hier diskutierten Kandidaten vollständig angehören, d.h. kompositionelle, modifizierbare, produktive Kombinationen von Präposition und Substantiv. Obwohl Baldwin et al. (2006, 173) zunächst eine Analyse verwerfen, in der eine NP auch als N' realisiert werden kann – weil diese ganz offensichtlich übergenerieren wurde, vgl. die Diskussion in Abs. 1, – schlagen sie schließlich eine Analyse vor, in der die Präposition eine artikellose Nominalprojektion selektiert (Baldwin et al., 2006, 175f.). Sie erkennen jedoch an, dass die von ihnen vorgestellte Regel weitere Einschränkungen benötigt, skizzieren diese Einschränkungen aber nur sehr knapp: „*These restrictions could be the result of selection for specific semantic classes of nouns by the preposition or they could alternatively be interpretations entirely contributed by the preposition on top of the nominal semantics.*“

(Baldwin et al., 2006, 176).

Der Verweis auf semantische Klassen ist allerdings, wie die Diskussion der Beispiele in (3) gezeigt hat, keinesfalls ausreichend. Wenn man die dort diskutierten Substantive *Prämisse* und *Voraussetzung* betrachtet, so scheinen andere Eigenschaften eine Rolle zu spielen, etwa die Tatsache, dass letzteres Substantiv eine *ung*-Ableitung ist, oder, dass ersteres Substantiv nicht nativ ist. Insgesamt kann man für das Deutsche beobachten, dass bei Kombinationen mit *unter* eine Vielzahl von Abstrakta als Substantiv auftreten, die jeweils alle mit *-ung* gebildet wurden, so dass also anstelle von semantischen Einschränkungen sicherlich auch formale Kriterien aufgeführt werden könnten.

Schließlich bleibt in der Analyse von Baldwin et al. (2006) vollkommen unklar, warum es eine Diskrepanz zwischen empirischer und kognitiver Produktivität gibt. Eine mögliche Erklärung ergäbe sich, wenn man annehmen würde, dass der in Baldwin et al. (2006) vorgestellte Lexikoneintrag für die produktive Selektion von artikellosen Substantiven zwar als Bestandteil der Grammatik vorliegt, aber eine Anwendung der Regel blockiert wird, weil in der Grammatik die bereits angesprochenen Voraussetzungen für die Anwendung (spezielle semantische Klassen, formale Kriterien) nicht gegeben sind. Somit würde dem Sprecher eine notwendige Voraussetzung für die Anwendung des Lexikoneintrags fehlen und die Unsicherheit der Sprecher in Bezug auf die Konstruktionen wäre erklärt. Eine solche Vorstellung ist allerdings weder mit der HPSG, in der Baldwin et al. (2006) ihre Analyse entwickeln, noch mit anderen Grammatikmodellen kompatibel. Wir müssen anerkennen, dass wir hier noch am Anfang der Forschung stehen.

Literatur

Harald Baayen. *Word Frequency Distributions*. Kluwer, Dordrecht, 2001.

Timothy Baldwin, John Beavers, Leonoor van der Beek, Francis Bond, Dan Flickinger, and Ivan A. Sag. In search of a systematic treatment of determinerless pps. In Patrick Saint-Dizier, editor, *Syntax and Semantics of Prepositions*, pages 163–179. Springer, 2006.

H. Burger. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Erich Schmidt Verlag, Berlin, 2003.

DDaF. Duden Deutsch als Fremdsprache Standardwörterbuch. Bibliographisches Institut und F.A. Brockhaus AG, 2002.

F. Dömges, T. Kiss, A. Müller, and C. Roch. Measuring the Productivity of Determinerless PPs. In F. Costello, J. Kelleher, and M. Volk, editors, *Erscheint in: Proceedings of the ACL 2007 Workshop on Prepositions*, Prag, Tschechische Republik, 2007.

- T. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19:61–74, 1993.
- DWDS. Das Digitale Wörterbuch der Deutschen Sprache des 20. Jh. (DWDS). URL <http://www.dwds.de/woerterbuch>.
- Stefan Evert. A simple Inre model for random character sequences. In *Proceedings of the 7mes Journees Internationales d'Analyse Statistique des Donnees Textuelles*, pages 411–422, 2004.
- W. Fleischer. *Phraseologie der deutschen Gegenwartssprache*. VEB, Leipzig, 1982.
- R. Grishman, C. MacLeod, and A. Myers. *Complex Syntax Reference Manual*. NYU: Proteus Project, 1998.
- G. Helbig and J. Buscha. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Langenscheidt, Leipzig, 2001.
- Nikolaus Himmelmann. Regularity in irregularity: Article use in adpositional phrases. *Linguistic Typology*, 2:315–353, 1998.
- K. Kesselmeier and A. von Könnemann. Kategorisierungsprobleme bei der Wortarten-Annotation von Textkorpora. B.A.-Arbeit. Sprachwissenschaftliches Institut, Ruhr-Universität Bochum. 2005.
- T. Kiss and J. Strunk. Unsupervised Multilingual Sentence Boundary Detection. *Computational Linguistics*, (32):485–525, 2006.
- L. Mackensen. Geschichte des deutschen Wortschatzes. In *Deutsches Wörterbuch : Rechtschreibung, Grammatik, Stil, Worterklärung, Fremdwörterbuch*. Südwest-Verlag, München, 9. aufl. edition, 1977.
- C.D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge/London, 1999.
- G. Nunberg, I.A. Sag, and T. Wasow. Idioms. *Language*, (70):491–538, 1994.
- Beata Trawinski. The syntax of „Complex Prepositions“ in German: An HPSG Approach. In *Proceedings of GLIP*, volume 5, pages 155–166, 2003.
- Beata Trawinski, Manfred Sailer, and Jan-Philipp Soehn. Combinatorial Aspects of Collocational Prepositional Phrases. In Patrick Saint-Dizier, editor, *Syntax and Semantics of Prepositions*, pages 181–196. Springer, 2006.

Wahrig. *Wahrig, Die Deutsche Rechtschreibung.* Wissen Media Verlag GmbH, Gtersloh/Mnchen, 2006.