Regelabstraktionen und -erweiterungen für eine kognitiv inspirierte Kategorialgrammatik

Tilman Vierhuff und Thurid Spieß

Zusammenfassung

Im Rahmen des Sonderforschungsbereichs 360 „Situierte künstliche Kommunikatoren“ wurde eine Variante von Steedmans *Combinatory Categorial Grammar* entwickelt, die den ursprünglichen kategorialgrammatischen Ansatz um Regeln für spezifische Eigenschaften der deutschen Sprache ergänzt.

Dieser Bericht zeigt Redundanzen auf, die sich aus dem Zusammenspiel von Steedmans Definitionen und den Erweiterungen ergeben und entwickelt Regeln, die diese Redundanzen vermeiden. Umgekehrt wird der Regelmechanismus in anderer Hinsicht stark modifiziert und erweitert, um die Verarbeitung verschiedener Phänomene im Zusammenhang mit Konjunktionen zu ermöglichen, ohne auf eine inkrementelle Verarbeitung der Konjunkte verzichten zu müssen.

1 Einleitung

Im Sonderforschungsbereich 360 „Situierte Künstliche Kommunikatoren“ hat sich eine Variante von Steedmans *Combinatory Categorial Grammar* [Ste93] als Formalismus für das Sprachverstehenssystem etabliert [HKS+98]. Ihre Performanz wurde in verschiedenen Bereichen wie der Anbindung von Präpositionalphrasen [HR97], der freien deutschen Konstituentenstellung [HE99] und der elliptischen Koordination [VHH01] bewiesen. Besonders hervorzuheben sind dabei die frühe Integration von Syntax und Semantik und die Unterstützung inkrementeller Verarbeitung [HER+99], die ein gutes Antwortverhalten bei kognitiv inspirierter Verarbeitung erlauben.

Wie jeder Grammatikformalismus lässt dabei auch die CCG Spielraum für Erweiterungen. In diesem Paper wird die in [VHH01] angesprochene Verarbeitung von Koordinationen und Ellipsen weiter ausgebaut, wobei ein besonderer Schwerpunkt auf größerer Inkrementalität und Effizienz liegt. Auch im Bereich der Modalverben und Nebensätze wird die Inkrementalität verbessert oder überhaupt erst ermöglicht, indem das Konzept optionaler Argumente und variabler Argumentlisten („Dollarkonvention“, [AS82]) auf *multiple Argumente* ausgedehnt wird.

Auf der anderen Seite entdeckt man bei näherer Betrachtung des bislang verwendeten Regelsatzes Redundanzen. Wir zeigen, wie sich unter Ausnutzung der Gemeinsamkeiten die bisherigen Regeln der CCG unter zwei allgemeine Reduktionsregeln subsumieren lassen.

Nachdem in Abschnitt 2 die im SFB 360 verwendete Variante der Kategorialgrammatik auf Basis der Steedmanschen CCG eingeführt wird, bespricht Abschnitt 3 die naheliegende Generalisierung der Regeln. Ausgehend von dem so gekürzten Regelsatz wird in Abschnitt 4 besprochen, wie mit Erweiterungen der Kategorien eine bessere Inkrementalität in Bezug auf Modalverben und Nebensätze erreicht werden kann, und eine Gruppe von Regeln diskutiert, die eine inkrementelle Verarbeitung zahlreicher elliptischer Koordinationen erlaubt.

2 Die zugrundeliegende Variante der Combinatory Categorial Grammar

2.1 Kategorialgrammatik

Die Kategorialgrammatik wurde erstmals 1929 von Lesniewski formuliert [Les29], von Ajduciewicz [Ajd35] weiterentwickelt und dann von Bar-Hillel verstärkt auf natürliche Sprache angewendet [Bar64]. Seitdem sind zahlreiche Varianten der Kategorialgrammatik entwickelt worden, die bezüglich ihrer Mächtigkeit, Repräsentationsformalismen und Effizienz zum Teil erheblich voneinander unterscheiden. Einen guten Überblick findet man z. B. in [W93].

Definition 1 Eine Kategorialgrammatik besteht aus

1. einer endlichen Menge von atomaren Kategorien
2. einem oder mehreren Konnektoren, mittels denen aus den atomaren Kategorien rekursiv komplexe Kategorien gebildet werden können
3. syntaktischen Regeln, die die Kombination von Kategorien definieren
4. einem Lexikon, das Wörter auf Kategorien abbildet

Im Gegensatz zu den in der zeitgenössischen Computerlinguistik vorherrschenden Phrasenstrukturgrammatiken (z. B. HPSG, vgl. [PS94]) definieren Kategorialgrammatiken keine großen Mengen spezieller Regeln für bestimmte Arten von Wörtern oder Konstituenten, sondern beschränken sich auf einen kleinen generischen Regelsatz, dessen Regeln erst im Zusammenhang mit den komplexen Kategorien ihre Mächtigkeit erhalten. Auf diese Weise verlagern Kategorialgrammatiken den Großteil ihrer Komplexität in das Lexikon.

Wählt man für Definition 1.2 die Konnektoren '/' und '\' und das Bildungsprinzip

wenn X und Y Kategorien sind, sind auch X/Y und $X\backslash Y$ Kategorien

so sind zwei übliche syntaktische Regeln die Vorwärts- und Rückwärts-Funktionsanwendung:

(1) $X/Y + Y \Rightarrow X$ Vorwärts-Funktionsanwendung
(2) $Y + X\backslash Y \Rightarrow X$ Rückwärts-Funktionsanwendung

Mit diesen einfachen Regeln, die der mathematischen Anwendung von Präfix- bzw. Postfix-Funktionen auf ihre Argumente entsprechen, ist bereits die Verarbeitung einfacher Sätze möglich (Abbildung 1).

Man beachte, dass die Konnektoren ihre Bedeutung erst durch den Kontext der Regeln wie z. B. (1) gegenüber (2) erhalten. Gemäß der von [AS82] eingeführten Konvention wird in diesem Paper die Kategorie links vom Schrägstrich stets als Ergebnis und die rechts vom Schrägstrich als Argument der funktionalen Kategorie betrachtet. Dies erhöht die Übersichtlichkeit komplexer Kategorien mit mehreren Argumenten.

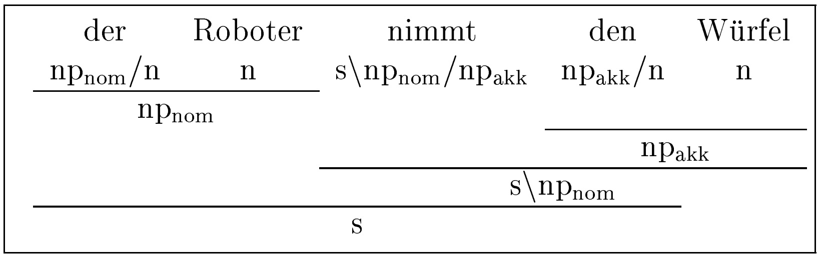

Abbildung 1: Einfache Satzverarbeitung bei ausschließlicher Verwendung der Funktionsanwendungsregeln

2.2 Kombinatorische Erweiterungen der Kategorialgrammatik

Der Formalismus der reinen Kategorialgrammatik kann durch Erweiterungen des Regelsatzes mächtiger gemacht werden. Bei der von Steedman formulierten Variante *Combinatory Categorial Grammar* (CCG) werden die in Abbildung 2 dargestellten Kombinatoren aus Currys kombinatorischer Logik [CF58] in die Kategorialgrammatik aufgenommen. Kombinatoren sind spezielle Funktionen höherer Ordnung in Form geschlossener Terme, also Terme ohne freie Variablen.

Für die CCG wird die gewöhnliche Kategorialgrammatik somit um drei Klassen von kombinatorischen Regeln erweitert, von denen jede einem von

Curry definierten Kombinator entspricht (Abbildung 3). Diese kombinatorischen Regeln erheben die Ausdrucksfähigkeit der Kategorialgrammatik von den kontextfreien zu den mild kontextsensitiven Sprachen [JVW91].

S f g x = f x (g x)	Substitution
B f g x = f (g x)	Komposition
T f g = g (f)	Type Raising

Abbildung 2: Kombinatoren von Curry, die in der CCG verwendet werden

Weiterhin führte Steedman die Dollar-Konvention ein, nach der eine Variable für eine beliebig lange Folge von Argumenten stehen kann [AS82]. In Abbildung 3 betrifft dies jeweils die Variable $Z. Durch diese Konvention werden auch Argumentkategorien für die Verarbeitung zugänglich, die in implizit linksassoziativ geklammerten Kategorien höherer Ordnung tiefer verschachtelt sind[1].

Eine jüngere Erweiterung der CCG betrifft die Verarbeitung von Koordinationen und Ellipsen. Hier definiert Steedman zwei Koordinationsregeln (Abbildung 4). Sie verknüpfen in zwei Schritten zwei Konjunkte gleicher Art miteinander; Koordinationen unterschiedlicher Kategorien (z. B. *„Ich bin Demokrat und stolz darauf“*) sind damit nicht möglich, aber dafür bereits eine Vielzahl von einfachen elliptischen Strukturen (z. B. *„Nimm einen grünen und einen roten Würfel“*).

Für detaillierte Beschreibungen der Steedmanschen CCG sei auf seine Ausführungen in [Ste87, Ste93, Ste00] verwiesen.

2.3 Die Variante der CCG im SFB 360

Das Sprachverstehensmodell im SFB 360 „Situierte Künstliche Kommunikatoren“ strebt die Verwirklichung folgender Ziele an (z. B. [HKS+98], [HR97]):

[1]Die Notation der Dollarkonvention ist durch Steedmans Arbeiten nicht durchgehend konsistent. Für die gleiche Kategorie sind die Notationen , $X\$/Z$ [AS82], $X/..Z$ [Ste90] und $X/\$Z$ [Ste96] zu finden

Substitution (entsprechend Currys Kombinator S)	
$(X/Y)/\$Z + Y/\$Z \Rightarrow X/\$Z$	Vorwärts-Substitution
$(X/Y)\backslash\$Z + Y\backslash\$Z \Rightarrow X\backslash\$Z$	vorwärts kreuzende Substitution
$Y\backslash\$Z + (X\backslash Y)\backslash\$Z \Rightarrow X\backslash\$Z$	Rückwärts-Substitution
$Y/\$Z + (X\backslash Y)/\$Z \Rightarrow X/\$Z$	rückwärts kreuzende Substitution
Komposition (entsprechend Currys Kombinator B)	
$X/Y + Y/\$Z \Rightarrow X/\$Z$	Vorwärts-Komposition
$X/Y + Y\backslash\$Z \Rightarrow X\backslash\$Z$	vorwärts kreuzende Komposition
$Y\backslash\$Z + X\backslash Y \Rightarrow X\backslash\$Z$	Rückwärts-Komposition
$Y/\$Z + X\backslash Y \Rightarrow X/\$Z$	rückwärts kreuzende Komposition
Type Raising (entsprechend Currys Kombinator T)	
$X \Rightarrow T/(T\backslash X)$	Vorwärts-Type-Raising
$X \Rightarrow T\backslash(T/X)$	Type Raising

Abbildung 3: Die den Curryschen Kombinatoren entsprechenden Regeln in der Kategorialgrammatik

- Die Modellierung soll kognitiven Prinzipien folgen, die aus psycholinguistischen Experimenten gewonnen wurden, und ihrerseits Anstöße für neue Fragestellungen liefern

- Die Verarbeitung soll inkrementell erfolgen (entsprechend der Unmittelbarkeitshypothese von Just & Carpenter [JC87]), wobei noch zu untersuchen ist, welche Inkremente menschlicher Sprachverarbeitung zugrunde liegen

- Zwischen verschiedenen Ebenen der Sprachverarbeitung (Syntax, Semantik, Pragmatik) soll möglichst früh eine enge Interaktion stattfinden; neu gewonnene Information soll dabei unmittelbar in bereits erzeugte Teilrepräsentationen integriert werden

- Neben allen theoretischen Ansprüchen muss die Simulation auch pragmatischen Anforderungen genügen, d. h., der Grammatikformalismus muss effizient verarbeitbar sein

Coordination rule A	$conj + X \Rightarrow [X]_{\&}$
Coordination rule B	$X' + [X]_{\&} \Rightarrow X''$

Abbildung 4: Steedmans Koordinationsregeln

Diese Ziele lassen sich mit einer leicht variierten Grammatik nach dem Vorbild der CCG gut umsetzen [HE99]. Zum einen wurde der Regelsatz um weitere Regeln ergänzt, die Komposition mit Type-Raising verbinden. Mit diesen Regeln kann die freie deutsche Satzstellung effizient behandelt werden, wobei ein Konstituentenzähler in der Repräsentation der Verben sicherstellt, dass nur grammatikalisch korrekte Sätze akzeptiert werden. Andererseits konnte auf drei Varianten der Kompositions-, sämtliche Substitutions- und implizite Type-Raising-Regeln verzichtet werden.

Eine höhere Flexibilität der Verbkategorien wird erreicht, indem man zwischen zwingend notwendigen und optionalen Komplementen der Verben unterscheidet. Optionale Argumentpositionen, die noch nicht gefüllt sind (z. B. zusätzliche beschreibende Präpositionalphrasen), können am Ende der Sprachanalyse ignoriert werden.

Der dabei entstandenen Regelsatz der CCG, den das Sprachverstehenssystem des SFBs momentan verwendet, enthält sechs Grundregeln (Abbildung 5), die jeweils durch Constraints auf bestimmte Arten von Argumenten beschränkt wurden, um Übergeneralisierungen zu vermeiden und statt einer wortbasiert inkrementellen eine an den Konstituenten orientierte Verarbeitung zu erreichen [HER+99]. Eine genaue Beschreibung dieser Regeln findet man in [HE99]. Im folgenden werden die Regeln dieses Grundregelsatzes als Regel (1) bis Regel (6) bezeichnet.

Für die Verarbeitung von Modalverbkonstruktionen wurden neue Kategorien mit variablen Argumentlisten definiert, die die Form (s/\$Z)/($\mathrm{s_{inf}}$/\$Z) besitzen[3]. Folglich akzeptieren sie ein Verb im Infinitiv, das eine nicht näher

[2]nur anwendbar, wenn der sogenannte Hauptfunktor [HE99], d.h. die innerste Ergebniskategorie von X, die Kategorie s hat und damit einen Konstituentenzähler i besitzt

[3]In [HE99] werden die variablen Argumentlisten mit Args bezeichnet, es ergeben sich also Kategorien der Form (s/Args)/($\mathrm{s_{inf}}$/Args). Da im Grunde diese Argumentlisten dieselbe Bedeutung haben wie die ebenfalls von [HE99] benutzte Steedmansche Dollar-Konvention, vereinheitlichen wir an dieser Stelle die Notation.

(1)	$X/Y + Y \Rightarrow X$	Functional Application, Forward
(2)	$X/Y + Y/\$Z \Rightarrow X/\$Z$	Functional Composition, Forward
(3)	$(X/Y)/\$Z + Y \Rightarrow X/\$Z$	Functional Composition, Crossed Backward
(4)	$Y + X\backslash Y \Rightarrow X$	Functional Application, Backward
(5)	$Y + X_i/Y \Rightarrow X_{\mathrm{i}-1}$	Functional Application, Crossed Forward[2]
(6)	$Y + (X_i/Y)/\$Z \Rightarrow X_{\mathrm{i}-1}/\$Z$	Functional Composition, Crossed Forward[2]

Abbildung 5: Regelsatz des Sprachverstehenssystems im SFB 360

spefizierte Menge von Argumenten erwartet, und ergeben ein semantisch modifiziertes konjugiertes Verb mit denselben Argumenten. Dieser Mechanismus der zweifachen variablen Argumentlisten ist elegant und ermöglicht, wie man in Abbildung 6 sieht, die Verarbeitung von Modalverben, genügt aber offenbar nicht dem Anspruch an inkrementelle Verarbeitung.

Ähnliches gilt in Nebensätzen, deren konjugiertes Verb im Deutschen am Ende steht. Ohne Informationen darüber, welche Komplemente das Vollverb erwartet, können bereits reduzierte Konstituenten nicht inkrementell in eine Satzsemantik integriert werden (Abbildung 7).

der Roboter	soll	die Schraube	mit der Zange	nehmen
np	(s/$Z)/(s/$Z)	np	pp	s/np/np/pp
			s/np/np	
			s/np	
		s/np		
	s			

Abbildung 6: Verarbeitung von Modalverben mit der SFB-Grammatik

(weil)	der Roboter	die Schraube	mit der Zange	nimmt
	np	np	pp	s/np/np/pp
			s/np/np	
		s/np		
	s			

Abbildung 7: Verarbeitung von Nebensätzen mit der SFB-Grammatik

Weitere Regeln zur Verarbeitung einfacher und bestimmter elliptischer Koordinationen, die von Steedmans Koordinationsregeln inspiriert sind (vgl. Abschnitt 4) wurden in jüngerer Zeit hinzugenommen [Vie00]. Bedingt durch die Eigenschaften der Steedmanschen Vorbilder konkurrieren hier allerdings die Ziele der Inkrementalität und der Effizienz: Wird im Satz *„Nimm eine Schraube und einen Würfel“* der Teilsatz aus den ersten drei Wörtern inkrementell verarbeitet, steht die Nominalphrase *„eine Schraube“* für eine Koordination nicht mehr zur Verfügung. Um den Satz mit über Steedmans Arbeit motivierten Regeln verarbeiten zu lassen, muss man auch hier auf inkrementelle Verarbeitung verzichten (Abbildung 8).

3 Generalisierungen

Wie in [HE99] dargelegt, ist der Regelsatz des SFB 360 geeignet, zahlreiche Phänomene der deutschen Sprache beschreiben, wenngleich mitunter Abstriche bei Inkrementalität oder Effizienz gemacht werden müssen. Jedoch fallen bei näherer Betrachtung der Regeln in ihrem Zusammenspiel vermeidbare Redundanzen auf.

In diesem Abschnitt wird zunächst gezeigt, wie sich die Regeln (1) bis (3) und (5) bis (6) jeweils zu einer Regel zusammenfassen lassen. Anschließend wird ein Weg dargestellt, wie auf den rückwärts gerichteten Konnektor und damit auf Regel (4) verzichteten werden kann.

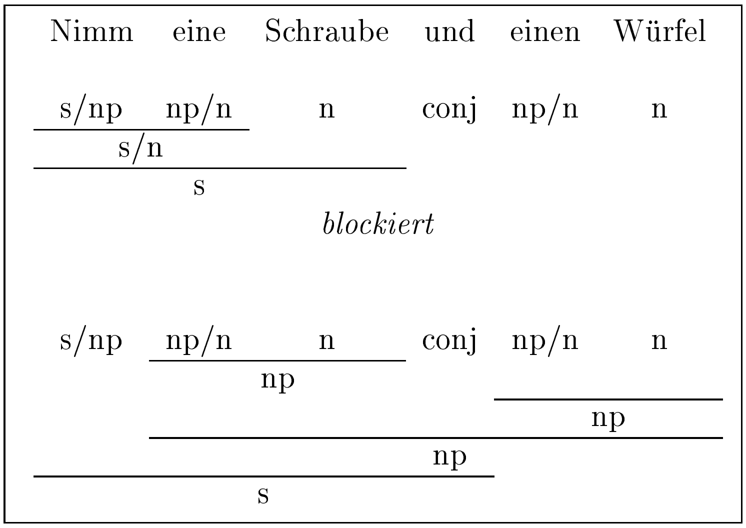

Abbildung 8: Die Steedmanschen Koordinationsregeln werden von inkrementeller Verarbeitung blockiert

3.1 Subsumption von Regeln

Interpretiert man \$Z entsprechend der Dollarkonvention als eine *beliebig* lange Folge von Argumenten, schließt dies auch die leere Folge mit ein. Unter dieser Interpretation wird die Funktionsanwendung zum Spezialfall der Komposition mit einer „nullstelligen Funktion“, wodurch eine explizite Regel für diesen Fall – die Funktionsanwendung (Regel 1) – entbehrlich wird. Das gleiche gilt für Regel 3, die rückwärts gekreuzte Komposition, die ebenso wie die Kompositionsregel als Verallgemeinerung der Funktionsanwendung aufgefasst werden kann. Auf diese Weise lassen sich die ersten drei Regeln zur folgenden Regel zusammenfassen:

(7) $W/X/\$Y + X/\$Z \Rightarrow W/\$Y/\$Z$ verallgemeinerte Functional Composition

Die Regeln (5) und (6) operieren beide ausschließlich auf Kategorien, deren Ergebnis letztlich ein Satz ist, da sie den erwähnten Konstituentenzähler als Index enthalten und dieser nur von der Kategorie s geführt wird. Der

Zähler zeigt gemäß dem von Hildebrandt eingeführten System die Anzahl der Argumente an, die vorgezogen werden dürfen, und wird bei der Verarbeitung eines solchen Arguments um eins reduziert [HE99]. Mit dem gleichen Argument wie bei den Regeln 1 und 2 können auch diese beiden Regeln zu einer zusammengefasst werden:

(8) $Y + X_i/Y/\$Z \Rightarrow X_{\mathrm{i}-1}/\$Z$ verallgemeinerte vorwärtsgekreuzte Funktionskomposition

Im Unterschied zu den vorwärtsgerichteten Regeln kommt die Grammatik hier ohne die Komposition von Funktionen aus, da bei einer inkrementellen Verarbeitung weiter links stehende Konstituenten bereits vollständig reduziert vorliegen.

Die Regeln (7) und (8) allein ermöglichen folglich dieselbe Funktionalität wie die bisherigen fünf Regeln, die den vorwärtsgerichteten Konnektor '/' benutzen. Zu beachten dabei ist, dass X und Y ihrerseits jeweils beliebig komplexe funktionale Kategorien sein können. In Abschnitt 3.3 wird beschrieben, wie man mit einem Konnektor und damit den beiden soeben beschriebenen Regeln ohne die zusätzliche Regel (4) auskommen kann.

3.2 Constraints und Argumentreihenfolgen

Bislang konnten die Variablen in den Regeln jeweils mit Constraints belegt werden, die ihre Anwendungsmöglichkeiten auf bestimmte Argumente festlegten oder bestimmte Argumente von der Verarbeitung ausschlossen. So wurden Sätze wie in Abbildung 9 durch eine geschickte Wahl dieser Constraints vermieden, indem insbesondere die Anwendung der Funktionskomposition stark eingeschränkt wurde [HE99]. Zwischen verschiedenen Ausformungen der Regeln (7) und (8) kann nun weniger leicht unterschieden werden, trotzdem ist eine Beschränkung der Variablen durch Constraints natürlich weiterhin möglich.

Bei der Komposition funktionaler Kategorien muss nun auf anderem Wege darüber hinaus gesichert werden, dass bislang noch nicht gefüllte Argumentpositionen in der Reihenfolge ihres Auftretens gesättigt werden. Das kann über eine verzeigerte Stack-Struktur erreicht werden, in der die jeweils zugefügten Argumente verkettet werden, so dass sich eine Folge von Gruppen jeweils gemeinsam zu reduzierender Argumente ergibt. Dieses Vorgehen übersteigt allerdings den gewöhnlichen Unifikationsmechanismus, mit dem derzeit im SFB sytaktische und semantische Informationen integriert werden, und muss deshalb in konkreten Implementationen prozedural behandelt werden.

Der Roboter	steckt	die	in den Würfel	Schraube
np_{nom}	$s_1/np_{nom}/np_{akk}/pp_{in}$	np_{akk}/n_{akk}	pp_{in}	n_{akk}
	$s_0/np_{akk}/pp_{in}$			
	$s_0/n_{akk}/pp_{in}$			
		s_0/n_{akk}		
			s_0	

Abbildung 9: Übergeneralisierung durch Benutzung der zusammengefassten Regeln ohne Berücksichtigung der Argumentreihenfolgen

Interessantere Probleme ergeben sich, wenn mehrere optionale Argumente auftreten, die sich auf verschiedene Entitäten beziehen. Im Falle mehrfacher Präpositionalphrasen unterschiedlicher Art können unerwünschte Resultate entstehen, weil Bezüge, obwohl sie semantisch nicht einwandfrei sind, trotzdem an entsprechende Objekte gebunden werden (Abbildung 10). Die semantische Repräsentation ist die gleiche wie bei dem Satz „*Nimm mit der linken Hand den Würfel mit den Löchern*“.

Ein Ausweg aus diesem Problem ist, äußere optionale Argumente zu verwerfen, sobald ein weiter innen liegendes Argument gefüllt wird. Kann eine Kategorie ein zuerst erwartetes optionales Argument nicht reduzieren, passt aber zu einem weiter innen im Stack stehenden obligatorischen Argument, ist das optionale Argument als in diesem Kontext unerfüllt zu verwerfen, um die erfolgreiche Reduktion zu ermöglichen.

Andererseits existieren andere Argumente wie Nebensätze, die aus dem normalen Satzgefüge ausgegliedert werden können. Der Satz „*Nimm den Würfel mit der linken Hand, der Löcher hat*“ muss verarbeitet werden können, auch wenn der optionale Relativsatz erst nach der früher verlangten optionalen Präpositionalphrase folgt.

Von den bisherigen Regeln wären optionale Argumente in der falschen Reihenfolge ebenso wie nachgestellte Relativsätze nicht akzeptiert worden. Hier sind weitere Forschungsarbeiten nötig um festzustellen, welche Reihen-

nimm	den Würfel	mit der linken Hand	mit den Löchern
$s_{imp}\ /np_{akk}/pp_{opt}$	np_{akk}/pp'_{opt}	pp_1	pp_2
$s_{imp}/pp_{opt}/pp'_{opt}$			
	s_{imp}/pp'_{opt}		
		s_{imp}	

Abbildung 10: Auflösung von Bezügen optionaler Präpositionalphrasen in einem Satz, der nicht akzeptiert werden sollte

folgen unter welchen Umständen verarbeitet werden müssen, und diese Bedingungen in das Modell zu integrieren.

Bei den Regeln (5) und (6) wird über den Konstituentenzähler gesteuert, ob Argumente vorangehen oder nachfolgen, und der Konnektor wird somit im Prinzip richtungsfrei gemacht. Entsprechend der historischen Entwicklung der CCG (vgl. [AS82]) benutzen dafür wir vorläufig den vorwärtsgerichteten Schrägstrich '/'. Eine andere Variante, die den ungerichteten Charakter des Konnektors betont, wird in Abschnitt 4.2 im Zusammenhang mit weiteren Variationen motiviert und eingeführt.

3.3 Beschränkung auf einen Konnektor

Regel (4) als einzige Regel mit dem rückwärts gerichteten Konnektor \ kam bisher nur bei der Verarbeitung von optionalen präpositionalen Ergänzungen für Nominalphrasen zum Einsatz wie z. B. bei dem Satz in Abbildung 11. Dabei wurde jede Präposition im Lexikon durch zwei konkurrierende Kategorien, nämlich pp/np und np\np/np, repräsentiert, von denen eine im Verlauf der Sprachverarbeitung durch Regelanwendungen und den Unifikationsmechanismus ausgeschlossen wurde.

Diese Redundanz der Lexikoneinträge kann mit den vorhandenen Ausdrucksmitteln bereits vermieden werden, indem das Konzept der optionalen Argumente von den Verbkategorien auf die Nominalphrasen ausgedehnt wird und diesen eine Kategorie der Form np/pp_{opt} zugeordnet wird. Mithin ist nur noch eine Kategorie für jede np notwendig, und die Regel (4) wird entbehrlich.

Nimm	die Schraube	mit	dem roten Kopf
s_{imp}/np_{akk}	np_{akk}	$np_X/np_{dat}\backslash np_X$	np_{dat}
	np_{akk}/np_{dat}		
		np_{akk}	
	s_{imp}		

Abbildung 11: Bisherige Verarbeitung optionaler Präpositionalphrasen

3.4 Bewertung der Inkrementalität

Es kann der Einwand erhoben werden, dass die Verarbeitung in Abbildung 12 nicht den Eigenschaften der konstituentenbasierten Inkrementalität entspricht, die erst vollständige Konstituenten als Funktionsargumente akzeptiert. Begründet wird diese Herangehensweise damit, dass dabei linguistisch nicht zu rechtfertigen Kategorie entstehen. Jedoch ist derzeit noch nicht erforscht, welches die Inkremente bei menschlicher Sprachverarbeitung sind; es ist daher nicht plausibel, bestimmte Zwischenzustände von vornherein auszuschließen. Die Kategorie pp/n z. B. ist bereits für einige Wörter (*am*, *zum*, *im*…) in der deutschen Sprache enthalten, obwohl sie durch konstituentenbasierte Inkrementalität nicht aus ihren Ursprungswörtern (Präposition pp/np und Artikel np/n) erzeugt werden dürfte, da die np erst als vollständige Konstituente vorliegen muss, bevor sie als Argument der pp dienen kann. Dagegen kann die Reduktion eines Artikels (np/n) und eines Adjektivs (n/n) zu einem semantisch angereicherten Artikel (np/n) vom formal linguistischen Standpunkt aus als zweifelhaft betrachtet werden; dennoch wird sie von der konstituentenbasierten Inkrementalität akzeptiert [HER+99].

Verschiedene Arbeiten fordern ferner eine wortbasierte inkrementelle Verarbeitung, die der hoch integrativen menschlichen Sprachverarbeitung am ehesten entspreche (maßgebend z. B. [JC87]). Nicht zuletzt ist der hohe mögliche Grad an Inkrementalität eine Stärke der Kategorialgrammatik, die sie jedoch großenteils einbüßt, wenn man ihre Reduktionen auf Phrasen beschränkt und sie so zu einer Spielart einer Phrasenstrukturgrammatik macht. Weitere psycholinguistische Versuche werden Indizien für potenzielle Inkremente liefern, die durch Constraints leicht in den Rahmen der CCG integriert werden können.

Nimm	die Schraube	mit	dem	roten	Kopf
s_{imp}/np_{akk}	np_{akk}/pp_{opt}	pp_{mit}/np_{dat}	np_{dat}/n_{dat}	n_{dat}/n_{dat}	n_{dat}/pp_{opt}
s_{imp}/pp_{opt}					
	s_{imp}/np_{dat}				
		s_{imp}/n_{dat}			
			s_{imp}/n_{dat}		
				s_{imp}/pp_{opt}	
				s	

Abbildung 12: Verarbeitung mit neuen Regeln und optionalem pp-Argument der nps

4 Erweiterungen

4.1 Inkrementalität bei Modalverben und Nebensätzen

Die vorgestellte Verarbeitung von Modalverben und Nebensätzen mit den Argumentlisten variabler Länge ist elegant und flexibel. Andererseits ist sie jedoch nicht inkrementell. Abbildung 13 zeigt, dass abgesehen von der Verarbeitung auf Konstituentenebene alle Reduktionen rechtsassoziativ verlaufen, d.h. erst wenn das letzte Wort vorliegt, können die Konstituenten integriert werden.

Auch für Sätze mit Verbletztstellung ist jedoch eine inkrementelle Verarbeitung wünschenswert. So ist es für Menschen, die den Teilsatz *„Ich muss vor Monatsende sechshundert Euro an meinen Vermieter...“* hören, nicht nur offensichtlich, dass ein transitives Vollverb im Infinitiv folgen muss, das Kopf aller inzwischen aufgebauten Konstituenten ist, sondern sie können auch dessen Semantik deduzieren und wahrscheinliche Hypothesen angeben.

Die Ableitungsstruktur in Abbildung 13 zeigt jedoch zu diesem Zeitpunkt nur, dass *irgendein* Verb zu erwarten ist, ohne jegliche Anhaltspunkte für die davon geforderten Konstituenten oder gar Ansätze für eine semantische Interpretation zu liefern. Um die Verarbeitung inkrementell zu machen, ist eine Erweiterung der DAGs und ihrer Unifikation notwendig, die das sukzessive

Erweitern der Verbstruktur erlaubt.

Ich	muss	sechshundert Euro	an meinen Vermieter	...
np	(s/\$Z)/(s/\$Z)	np	pp	

Abbildung 13: Fehlende Inkrementalität bei der Verarbeitung von Modalverben mit der bisherigen SFB-Grammatik aus [HE99]

Multiple Argumente

In Abschnitt 2.3 wurde erläutert, wie optionale Argumente verarbeitet werden können, die gegebenenfalls auftreten dürfen, aber nicht zwingend notwendig sind. Für erweiterbare Argumentlisten liegt der Fall ähnlich, statt einfach auftretender optionaler Argumente steht hier jedoch ein variables Argument für eine mehrfach unifizierbare und dabei immer komplexer werdende Kategorie.

Wir schreiben $(s/(s/\$X))/{*X}$ für eine Kategorie mit *multiplen Argumenten*, bei der $*X$ beliebig oft mit neuen Verbargumenten gefüllt werden kann, die es an die innere Argumentliste $\$X$ weitergibt. Sobald keine weitere passende Kategorie mehr vorliegt, wird das multiple Argument wie ein optionales behandelt und kann weggelassen werden. Der Stern in $*X$ steht somit für 0 bis potentiell unbeschränkt viele optionale Argumente, die die variable Argumentliste $\$X$ erweitern. Mit dieser Kategorie wird eine vorausschauende, inkrementelle Ableitung von Modalverbsätzen (Abbildung 14) ebenso wie von Nebensätzen (Abbildung 15) mit deutscher Verbletztstellung ermöglicht.

Zu berücksichtigen ist dabei, dass in der konventionellen Kategorialgrammatik, wie sie bisher besprochen wurde, die Reihenfolge der Argumente dabei genau mit der definierten Reihenfolge in der Verbkategorie übereinstimmen muss. Wie man die Verbkategorien für die realitv freie deutsche Verbstellung verallgemeinern kann, zeigt der Abschnitt 4.2.

[4] Aus Gründen der Übersichtlichkeit wurde in diesen und manchen der folgenden Ableitungsstrukturen darauf verzichtet, die Reduktion der Konstituenten darzustellen

Ich	muss	600 Euro	an meinen Vermieter	überweisen
np	(s/(s/\$X))/∗X	np	pp	s/np/np/pp

(s/(s/np/\$X))/∗X

(s/(s/np/np/\$X))/∗X

(s/(s/np/np/pp/\$X))/∗X

(s/(s/np/np/pp))

s

Abbildung 14: Verarbeitung von Modalverben mit multiplen Argumenten[4]

weil	der Roboter	die Schraube	mit der Zange	nimmt
(s/(s/\$X))/∗X	np	np	pp	s/np/np/pp

(s/(s/np/\$X))/∗X

(s/(s/np/np/\$X))/∗X

(s/(s/np/np/pp/\$X))/∗X

(s/(s/np/np/pp))

s

Abbildung 15: Verarbeitung von Nebensätzen mit multiplen Argumenten

4.2 Aufspaltung der Regelmenge

Bei der Verwendung und Betrachtung der Regeln nach [HE99] fällt auf, daß es offensichtlich zwei Klassen von Regeln gibt: Solche, die den Konstituentenzähler verändern und somit nur auf Kategorien des Typs s, und solche, die auf Kategorien ohne diesen Zähler angewendet werden können. Obwohl die Regeln grundsätzlich als allgemeine Regeln mit Variablen X, Y, Z formuliert werden, schränkt der Konstituentenzähler die Anwendung auf eine von genau zwei Argumentmengen ein.

Weiterhin bestehen bei den Regeln mit Veränderung des Konstituentenzählers noch mehr gemeinsame Eigenschaften, die als Nebeneffekte aus der Einführung dieses Zählers entstehen. Insbesondere betrifft dies den Konnektor, dessen Richtung ignoriert wird und werden muss – anderenfalls wäre der Zähler offensichtlich redundant.

Diese versteckte Aufteilung lässt sich explizit machen, indem man bewusst zwei Mengen von Regeln angibt. Menge A enhält nur ein einziges Element, nämlich die inkrementelle Standardregel der Kategorialgrammatik, die nur auf andere Kategorien als s angewendet werden kann und Functional Application sowie Functional Composition in sich vereinigt[5] (vgl. auch Abschnitt 3):

$$X/Y + Y/\$Z \Rightarrow X/\$Z$$

Menge B enthält Regeln, die zur Verarbeitung der deutschen Konstituentenstellung am Verb notwendig sind, mit den dafür nötigen Variationen des üblichen Formalismus (Konstituentenzählung, Ignorieren der Schrägstrichrichtung). Im folgenden werden sie zur besseren Unterscheidung von Regelsatz A bewusst anders dargestellt. Der ungerichtete Konnektor '|' verdeutlicht, dass die Argumente vor oder hinter dem Verb auftreten dürfen.

$$\begin{array}{ccccc} s|X|\ldots|Y|\ldots|Z & + & Y|\$W & \Rightarrow & s|X|\ldots|Z|\$W \\ Y & + & s_i|X|\ldots|Y|\ldots|Z & \Rightarrow & s_{\mathrm{i}-1}|X|\ldots|Z \end{array}$$

Zusammen mit der Einführung des ungerichteten Konnektors lässt sich auch die freie Konstituentenstellung in der Kategorieschreibweise umsetzen. Was die Dollarkonvention erlaubt, kann in den Regeln auf folgende Weise veranschaulicht werden, indem die scheinbare Reihenfolge der Argumente durch vertikale Anordnung aufgehoben wird.

$$s\left|\begin{array}{l} X \\ \vdots \\ Y \\ \vdots \\ Z \end{array}\right. \quad + \quad Y'\left|\begin{array}{l} W_1 \\ \vdots \\ W_n \end{array}\right. \quad \Rightarrow \quad \left(s\left|\begin{array}{l} X \\ \vdots \\ Z \end{array}\right. \right)\left|\begin{array}{l} W_1 \\ \vdots \\ W_n \end{array}\right.$$

Das Vorziehen des Subjekts eines Hauptsatzes oder aller Konstituenten eines Nebensatzes vor das Verb ist keine fehlerhafte Ausnahmeerscheinung, sondern die erwünschte Regel und sollte auch entsprechend dargestellt werden. Die verwendete Schreibweise spiegelt wider, dass der Konnektor in den

[5] Die gekreuzte Alternative mit der Ausgangskategorie X/Y/$Z ist hier entbehrlich, da sie nur in Verbindung mit dem Konstituentenzähler auftritt

erwähnten Regeln eben nicht vorwärtsgerichtet ist und seine Richtung verletzt wird, wie die Beschreibung in [HE99] nahelegt. Stattdessen wechselt der Zähler bei dieser Betrachtung von der Anzahl der Richtungsverstöße zur Anzahl der vor das Verb gezogenen Kategorien, was weniger auf eine Wertung als richtig oder falsch hinausläuft und somit der deutschen Grammatik näherkommt.

Die Aufteilung macht deutlich, wie die verwendete Grammatik arbeitet: Mit der Regel A werden Konstituenten erzeugt, die mit Regeln der Menge B an das Verb gebunden werden. Dieselbe Aufgabenteilung existierte vorher schon implizit und wird nun durch die veränderte Schreibweise unterstrichen.

4.3 Verarbeitung von Koordinationen

Der folgende Abschnitt verfolgt einen neuen Weg, elliptische Koordinationen zu verarbeiten, der noch weiter von der klassischen Kategorialgrammatik abweicht. Wie in Abschnitt 2.2 besprochen, stellen die Steedmanschen Konjunktionsregeln ein Effizienzproblem dar. Eine inkrementelle Verarbeitung ist nicht möglich und führt zu nicht weiter reduzierbaren Zwischenzuständen. Der hier besprochene Weg kann zwar auch keine wortweise Inkrementalität leisten, erlaubt aber immerhin eine inkrementelle Verarbeitung der zu koordinierenden Teilsätze und akzeptiert spezielle elliptische Phänomene, die von der CCG bisher nicht verstanden werden konnten.

Um verschiedene koordinative Strukturen elegant verarbeiten zu können, führen wir zunächst weitere Änderungen der Regelmenge B ein. Die Argumente des Satzes werden, wenn sie mit Werten belegt werden, nicht vollständig gelöscht, sondern in der Verbkategorie als belegt markiert. Eine Kategorie für den vollständig verarbeiteten Satz *„Der Roboter nimmt die Schraube“* würde damit wie in Abbildung 16 dargestellt. Genau genommen stehen *Roboter* und *Schraube* für die ganzen Nominalphrasen, die den Argumenten zugewiesen wurden, während ihre Semantik z. B. mit *structure sharing*, λ-Ausdrücken oder ähnlichen Mechanismen mit Einträgen in der Satzsemantik verbunden wird. Ein belegtes Argument kann nicht noch einmal mit einem anderen Wert belegt werden.

Die Regeln der Menge B werden damit insgesamt so erweitert, dass sie die in Abbildung 17 gezeigte Form annehmen.

Ferner wird in der Kategorie s gespeichert, welches die zuletzt reduzierte Kategorie war, durch die der Konstituentenzähler vermindert wurde. Dies ist

$$s_{\text{nehmen}} \left| \begin{array}{l} np_{\text{nom}} \\ np_{\text{akk}} \end{array} \right.$$

$$s_{\text{nehmen}} \left| \begin{array}{l} np_{\text{nom}}=\mathit{Roboter} \\ np_{\text{akk}}=\mathit{Schraube} \end{array} \right.$$

Abbildung 16: Satzkategorie mit leeren und mit gefüllten Argumentpositionen

die zuletzt topikalisierte Konstituente, der bei der Koordination elliptischer Teilsätze besondere Bedeutung zukommen wird.

Zur Verarbeitung von Konjunktionen definieren wir nun eine neue Regelmenge C. Wir beginnen mit den zwei Grundregeln C1 und C2 (Abbildung 18, die im wesentlichen mit Steedmans Conjunction Rules identisch sind [Ste93]. Anschließend behandeln wir die spezialisierten Regeln C3 bis C5 (Abbildung 19, 21 und 23) zur Verarbeitung vielfältiger Koordinationsphänomene.

Die Basisregeln C1 und C2 erlauben die in Abschnitt 2.2 eingeführte Koordination gleicher Kategorien. Die zweite Regel muss Sonderfälle wie die unterschiedliche Semantik der Koordination von Adjektiven und Nominalphrasen berücksichtigen und gegebenenfalls entsprechend instantiiert werden. Syntaktisch entspricht $\left[\frac{X}{X'}\right]_{\text{CONJ}}$ einer Kategorie vom Typ X und kann als solche weiterverarbeitet werden, die Schreibweise soll lediglich andeuten, dass die Semantik beider Ursprungskategorien in ihr enthalten ist.

Auf Basis dieser Grundregeln definieren wir nun spezielle Kordinationsregeln, die die inkrementelle Verarbeitung vielfältiger deutscher Ellipsen erlauben. Vergleiche dazu die bisherige Verarbeitung von Ellipsen in der CCG in [VHH01]).

Regel (C3) (Abbildung 19) erlaubt nachträgliche Addition von Argumenten an bereits belegte Einträge. Ein typischer Satz, dessen inkrementelle Verarbeitung diese Regel erforderlich macht, wäre *„Der Roboter nimmt die Schraube und den Würfel“* (Abbildung 20)[6]. Mit den Steedmanschen Regeln

[6] Die semantische Verknüpfung des Substantives mit der Nominalphrase wurde in Abbildung 20 exemplarisch durch *structure sharing* wie in Unifikationsgrammatiken dargestellt. Grundsätzlich ist dafür natürlich auch jeder andere Formalismus (z. B. λ-Ausdrücke) geeignet

$$\text{(B2)} \quad s\left|\begin{array}{l}X\\ \vdots\\ Y\\ \vdots\\ Z\end{array}\right. \quad + \quad Y'\left|\begin{array}{l}W_1\\ \vdots\\ W_n\end{array}\right. \quad \Rightarrow \quad \left(s\left|\begin{array}{l}X\\ \vdots\\ Y{=}Y'\\ \vdots\\ Z\end{array}\right. \right)\left|\begin{array}{l}W_1\\ \vdots\\ W_n\end{array}\right.$$

$$\text{(B2)} \quad Y' \quad + \quad s_i\left|\begin{array}{l}X\\ \vdots\\ Y\\ \vdots\\ \mathrm{Z}\end{array}\right. \quad \Rightarrow \quad s_{\mathrm{i}-1}\left|\begin{array}{l}X\\ \vdots\\ Y{=}Y'\\ \vdots\\ Z\end{array}\right.$$

Abbildung 17: Erweiterte Regeln aus Menge B für die effiziente Verarbeitung von Koordinationen

führt eine inkrementelle Ableitung zunächst zu einem Garden Path, da die freie *np* nach der Konjunktion nicht mehr unifiziert werden kann, sobald der Satz davor abgeschlossen ist (vgl. Abbildung 8 in Abschnitt 2.2). Die neue Regel erlaubt das nachträgliche Einfügen des Konjunkts in die Objektposition des vorher inkrementell geparsten Satzes.

Beachte, daß diese Regel auch in Sätzen mit Nachträgen wie *„Die Schraube nimmt der Roboter, und den Würfel“* von Nutzen ist. Hier wird die Objektposition zunächst mit der Schraube gefüllt, worauf sich bis zum Komma ein vollständiger Satz ergibt, und schließlich um ein weiteres Objekt ergänzt. Hier geht die neue Regel in Hinsicht auf Robustheit weit über die konventionelle Koordination hinaus.

Regel C4 (Abbildung 21) erlaubt es einem elliptischen, nachgestellten Teilsatz, ein topikalisiertes Argument oder das Subjekt zu übernehmen. Dies erlaubt die Verarbeitung von Sätzen wie *„In den Wald ging der Jäger und fing einen Hasen“* (aus [Kle93]) oder *„Den Würfel sehe ich und nimmt der Roboter“*, da die fraglichen Konstituenten jeweils in Topic-Position bzw. in der Funktion des Subjektes auftreten (Abbildung 22). Andererseits unterbindet sie Sätze wie *„Ich nehme den Würfel und sieht der Roboter“* oder *„Ich gehe in den Wald und läuft der Hase“*, in denen das topikalisierte Subjekt

$$
\begin{array}{ll}
\text{(C1)} & conj + X \Rightarrow [X]_{\&} \\
\text{(C2)} & X + [X']_{\&} \Rightarrow \left[\frac{X}{X'}\right]_{\mathrm{CONJ}}
\end{array}
$$

Abbildung 18: Basis–Koordinationsregeln

$$
\text{C3} \quad S\left|\begin{array}{l} X \\ \vdots \\ Y{=}Y' \\ \vdots \\ Z \end{array}\right. \quad + \quad [Y'']_{\&} \quad \Rightarrow \quad S\left|\begin{array}{l} X \\ \vdots \\ Y = \left[\frac{Y'}{Y''}\right]_{\mathrm{CONJ}} \\ \vdots \\ Z \end{array}\right.
$$

Abbildung 19: Koordinationsregel C3 zur Verknüpfung von Konstituenten mit bereits reduzierten Argumenten

nicht mit einer entsprechenden Konstituente im elliptischen Teilsatz unifiziert werden kann.

Weiterhin ist es sinnvoll zu verlangen, daß nach Anwendung der Regel $i = j$ gilt, d.h. beide Sätze gleich viele vorgezogene Konstituenten besitzen müssen, einschließlich der gemeinsamen. Durch dieses einfach umzusetzende Constraint wird der pathologische Satz *„In den Wald ging der Jäger und einen Hasen fing“* vermieden, in dem der zweite Teilsatz im Gegensatz zum ersten zwei vorgezogene Konstituenten besitzt, nämlich den Jäger und den Hasen.

Regel C5 (Abbildung 23) schließlich dient zur Verknüpfung zweier elliptischer Teilsätze mit beliebig gefüllten Konstituentenpositionen, die sich gegenseitig ergänzen. Ein Beispiel ist *„Der Roboter nimmt und montiert den Würfel mit der Zange“*, bei dem *Roboter* in *montiert* übernommen wird, während *Würfel* und *Zange* in *nimmt* übergehen (Abbildung 24). Diese Regel kann nahezu beliebig gesättigte elliptische Teilsätze verknüpfen und ist damit überaus mächtig.

Einerseits sind nicht alle Möglichkeiten, Argumente elliptischer Teilsätze zu ergänzen, sinnvoll, andererseits bestehen zum Teil mehrere Möglichkeiten, welche Argumente übernommen werden sollen. Das erwähnte Beispiel *„Der Roboter nimmt und montiert den Würfel mit der Zange"* ist ambig, da nicht aus der Satzstruktur hervorgeht, ob beide Aktionen oder nur die zweite mit der Zange vorgenommen werden. Allgemeiner formuliert, ist die Anbindung optionaler Argumente nicht eindeutig und führt zu Mehrdeutigkeiten, die im Kontext aufgelöst werden müssen.

Welche Kombinationen von Konstituenten im Umgang mit der Regel C5 sinnvoll sind, wird in entsprechenden Experimenten zu prüfen sein. In ihrem gegenwärtigen Zustand ist sie zu allgemein, da z. B. Zeitangaben auf den einen Teilsatz beschränkt sein können und dementsprechend nicht in den anderen übernommen werden sollten (z. B. *„Ich kaufe und koche morgen ein Huhn"*).

	Der	Roboter	nimmt	die	Schraube	und	den	Würfel
	np/n	n	$s\left\vert\begin{array}{l}np_{nom}\\np_{akk}\end{array}\right.$	np/n	n	$conj$	np/n	n
A	np_{nom}							
B2	$s\left\vert\begin{array}{l}np_{nom}\text{=Roboter}\\np_{akk}\end{array}\right.$							
B1	$\left(s\left\vert\begin{array}{l}np_{nom}\text{=Roboter}\\np_{akk}=\boxed{1}\end{array}\right.\right)/n_{\boxed{1}}$							
A			$s\left\vert\begin{array}{l}np_{nom}\text{=Roboter}\\np_{akk}\text{=Schraube}\end{array}\right.$					
A							np_{akk}	
C1						$[np_{akk}]_{\&}$		
C3			$s\left\vert\begin{array}{l}np_{nom}\text{=Roboter}\\np_{akk}=\begin{bmatrix}\text{Schraube}\\\text{Würfel}\end{bmatrix}\end{array}\right.$					

Abbildung 20: Anwendung der der Regel C3 zur Ergänzung bereits reduzierter Argumente um ihre Konjunkte

$$
S_i \left| \begin{array}{l} X \\ \vdots \\ Y{=}Y'_{(\text{topic}\vee\text{subject})} \\ \vdots \\ Z \end{array} \right. \quad + \quad \left[S_j \left| \begin{array}{l} \text{V} \\ \vdots \\ Y \\ \vdots \\ \text{W} \end{array} \right. \right]_{\&} \quad \Rightarrow \quad \left[\begin{array}{c} {\scriptstyle S_i} \left| \begin{array}{l} \text{X} \\ \vdots \\ \text{Y=Y'} \\ \vdots \\ \text{Z} \end{array} \right. \\ \hline {\scriptstyle S_{\text{j}-1}} \left| \begin{array}{l} \text{V} \\ \vdots \\ \text{Y=Y'} \\ \vdots \\ \text{W} \end{array} \right. \end{array} \right]_{\text{CONJ}}
$$

Abbildung 21: Koordinationsregel C4, die die Übernahme einer topikalisierten Konstituente oder des Subjekts in einen elliptischen Teilsatz erlaubt

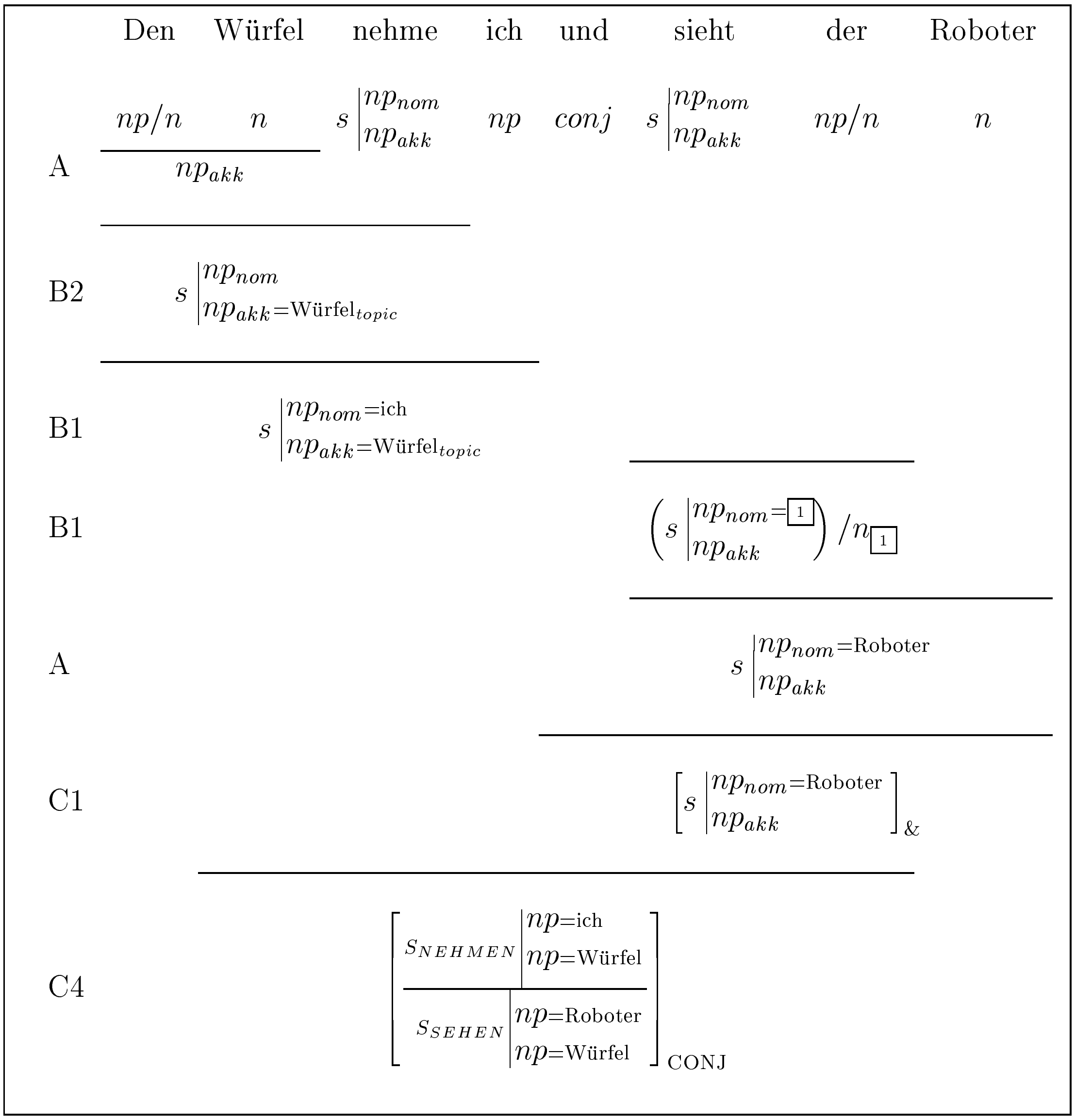

Abbildung 22: Anwendung der Regel C4 zur Übernahme einer topikalisierten Konstituente in einen elliptischen Teilsatz

$$
S\left|\begin{array}{l}
X_1\\
\vdots\\
X_h = X'_i\\
\vdots\\
X_j = X'_j\\
\vdots\\
X_m
\end{array}\right.
\quad + \quad
\left[S'\left|\begin{array}{l}
X_1\\
\vdots\\
X_i{=}X'_i\\
\vdots\\
X_k{=}X'_k\\
\vdots\\
X_n
\end{array}\right. \right]_{\&}
\quad \Rightarrow \quad
\left[\begin{array}{c}
{\scriptstyle S}\left|\begin{array}{l}
X_1\\
\vdots\\
X_h = X'_h\\
\vdots\\
X_k = X'_k\\
\vdots\\
X_m
\end{array}\right.\\
\hline
{\scriptstyle S'}\left|\begin{array}{l}
X_1\\
\vdots\\
X_h = X'_h\\
\vdots\\
X_k{=}\ X'_k\\
\vdots\\
X_n
\end{array}\right.
\end{array} \right]_{\text{CONJ}}
$$

Abbildung 23: Koordinationsregel C5, die zwei elliptische Teilsätze verknüpft; die Übernahme von Konstituenten muss ggf. eingeschränkt werden

Der Roboter — nimmt — und — montiert — den Würfel — mit der Zange

$$np \qquad s\left|\begin{array}{l}np_{nom}\\ np_{akk}\\ pp_{instr}\end{array}\right. \qquad conj \qquad s\left|\begin{array}{l}np_{nom}\\ np_{akk}\\ pp_{instr}\end{array}\right. \qquad np \qquad pp_{instr}$$

B2

$$s\left|\begin{array}{l}np_{nom}\text{=Roboter}\\ np_{akk}\\ pp_{instr}\end{array}\right.$$

B1

$$s\left|\begin{array}{l}np_{nom}\\ np_{akk}\text{=Würfel}\\ pp_{instr}\end{array}\right.$$

B1

$$s\left|\begin{array}{l}np_{nom}\\ np_{akk}\text{=Würfel}\\ pp_{instr}\text{=Zange}\end{array}\right.$$

C1

$$\left[s\left|\begin{array}{l}np_{nom}\\ np_{akk}\text{=Würfel}\\ pp_{instr}\text{=Zange}\end{array}\right.\right]_{\&}$$

C5

$$\left[\begin{array}{c} s_{NEHMEN}\left|\begin{array}{l}np_{nom}\text{=Roboter}\\ np_{akk}\text{=Würfel}\\ pp_{instr}\text{=Zange}\end{array}\right. \\ \hline s_{MONTIEREN}\left|\begin{array}{l}np_{nom}\text{=Roboter}\\ np_{akk}\text{=Würfel}\\ pp_{instr}\text{=Zange}\end{array}\right.\end{array}\right]_{\text{CONJ}}$$

Abbildung 24: Anwendung der Regel C5 zur Kombination zweier elliptischer Teilsätze

5 Zusammenfassung

Im SFB 360 wird eine Variante der Combinatory Categorial Grammar eingesetzt, die durch ihre Mächtigkeit vielfältige deutsche Phänomene mehr oder weniger inkrementell und kognitiv adäquat verarbeiten kann.

In diesem Report haben wir gezeigt, dass der Regelsatz dieser Grammatik (ebenso wie der mancher anderen Kategorialgrammatik) bedingt durch Eigenschaften der dafür verwendeten Dollar-Konvention und der Type-Raising-Regeln Redundanzen enthält, die durch Umformulierung und Vereinigung der Regeln vermieden werden können. Des weiteren lässt sich auf den zweiten, rückwärts gerichteten Konnektor verzichten, wobei gleichzeitig eine inkrementelle Verarbeitung von Nebensätzen und Modalverbkonstruktionen möglich wird, die bisher nicht zugänglich war.

Auf der anderen Seite zeigen sich die Steedmanschen Konjunktionsregeln nicht in der Lage, verschiedene elliptische deutsche Sätze inkrementell und effizient zu verarbeiten. Hier sind entsprechend verschiedenen Phänomenen bestimmte Spezialregeln vonnöten, die elliptische Teilsätze und fehlende Konstituenten integrieren können. Die notwendigen Erweiterungen der Kategorialgrammatik werden durch leichte Variationen der kategorialen Schreibweise wiedergegeben, wobei die Dollarkonvention durch eine intuitivere Notation ersetzt wird.

Literatur

[AS82] Ades, A.E. and Steedman, M. On the Order of Words. *Linguistics and Philosophy, 4*, 517-558, 1982.

[Ajd35] Ajdukiewicz, K. Die syntaktische Konnexität. *Studia Philosophica, 1*, 1-27, 1935.

[Bar64] Bar-Hillel, Y. *Language and Information: Selected Essays on their Theory and Application*, Reading, MA: Addison-Wesley, 1964.

[CF58] Curry, H. and Feys, R. *Combinatory Logic, Vol. 1.* Amsterdam: North-Holland, 1953.

[GKS+93] Günther, Kindt, Schade, Sichelschmidt, und Strohner. *Elliptische Koordination. Strukturen und Prozesse lokaler Textkohärenz.* Linguistische Berichte 146: 311-328. Westdeutscher Verlag, 1993.

[HE99] Hildebrandt, B. und Eikmeyer, H.-J.: *Sprachverarbeitung mit Combinatory Categorial Grammar: Inkrementalität und Effizienz.* Report 1999/05 – Situierte Künstliche Kommunikatoren, SFB 360. Universität Bielefeld, 1999.

[HER+99] Hildebrandt, B., Eikmeyer, H.-J., Rickheit, G. und Weiß, P. Inkrementelle Sprachrezeption. *4. Fachtagung der Gesellschaft für Kognitionswissenschaft, 28.9.-1.10.1999, Universität Bielefeld,*, 1999.

[HKS+98] Hildebrandt, B., Knoll, A., Scheering, C. und Zhang, J.: *Koginitive Aspekte bei der Realisierung eines robusten Robotersystems für Konstruktionsaufgaben.* Report 1998/01 – Situierte Künstliche Kommunikatoren, SFB 360. Universität Bielefeld, 1998.

[HR97] Hildebrandt, B. und Rickheit, G.: *Verarbeitung von Präpositionalphrasen in der Combinatory Categorial Grammar.* Report 1997/06 – Situierte Künstliche Kommunikatoren, SFB 360. Universität Bielefeld, 1997.

[JVW91] Joshi, A., Vijay-Shanker, K. and Weir, D. *The Convergence of Mildly Context-Sensitive Formalisms*. In P. Sells, S. Shieber and T. Wasow (Herausgeber): *Processing of Linguistic Structure*, 31–81, MIT Press, Cambridge MA, 1991.

[JC87] Just, M. and Carpenter, P. *The psychology of reading and language comprehension.* London: Allyn and Bacon, 1987.

[Kle93] Klein, W. *Ellipse*. In J. Jakobs und A. von Steckow: *Syntax. Ein internationales Handbuch zeitgenössischer Forschung.* 1. Halbband, Vol.1. 763–799, 1993.

[Les29] Lesniewski, S. Grundzüge eines neuen Systems der Grundlagen der Mathematik. *Fundamentae Mathematicae*, 14, 1929.

[PS94] Pollard, C. and Sag, I. A. *Head–Driven Phrase Structure Grammar*. Chicago: University of Chicago Press, 1994.

[Shi86] Shieber, S. M. *An Introduction to Unification Based Grammar*. Center for the Study of Linguistics and Information, Stanford, 1986.

[Ste87] Steedman, M. *Combinatory Grammars And Human Language Processing*. In J. L. Garfield (Herausgeber): *Modularity In Knowledge Representation And Natural–Language Understanding*, 197–205, Cambridge, MA: MIT Press, 1987.

[Ste90] Steedman, M. *Gapping As Constituent Coordination*. Linguistics and Philosophy 13: 207-263. Kluwer Academic Publishers, 1990.

[Ste91] Steedman, M. *29th Annual Meeting of the Association for Computational Linguistics, Proceeding of the Conference, 18-21 June 1991, Berkley, California, USA*. Type-Raising and directionality in combinatory grammar. Morristown, NJ: ACL, 71-78, 1991.

[Ste93] Steedman, M. *Categorial Grammar*. Lingua, 90, 221-258, 1990.

[Ste96] Steedman, M. *Surface Structure and Interpretation*, Cambridge, MA: MIT Press, 1996.

[Ste00] Steedman, M. *The Syntactic Process*. The MIT Press, Cambridge Mass, 2000.

[Vie00] Vierhuff, T. *Ein sprachverarbeitendes System zur Verarbeitung koordinativer Strukturen in der Combinatory Categorial Grammar*. Unveröffentlichte Diplomarbeit. Universität Bielefeld. 2000.

[VE01] Vierhuff, T. und Eikmeyer, H.-J. *Directional unification for processing elliptic coordinations in CCG*. 5. Fachtagung der Gesellschaft für Kognitionswissenschaft, 25.9.-28.9.2001, Max-Planck-Institut für Neuropsychologie, Leipzig, 2001.

[VHH01] Vierhuff, T., Heidenreich, S. und Hildebrandt, B. *Integrating elliptic conjunctions into a CCG based parser*. Report 2001/7 – Situierte Künstliche Kommunikatoren, SFB 360. Universität Bielefeld, 2001.

[W93] Wood, M. *Categorial Grammar*. London: Routledge, 1993.