

Parsing mit inkonsistenten Grammatiken

Wolfgang Menzel

Universität Hamburg, Fachbereich Informatik, Vogt-Kölln-Straße 30, D-22527 Hamburg (e-mail: menzel@informatik.uni-hamburg.de)

Parsing with inconsistent grammar models

Summary. A novel approach to the parsing of natural language based on weighted constraints is presented. Well-formedness conditions are exclusively expressed by means of constraints and no generative rule component is available. In general, constraints can be violated and contradictory requirements of a grammar are tolerated. Additional representational levels allow the integration of different linguistic perspectives into a coherent model, thus contributing jointly to the computation of the most plausible structural interpretation. A number of solution procedures is available for such a parsing problem. Their properties are discussed with respect to possible analogies to human language processing, focusing on robustness against deviating utterances and external temporal pressure.

Zusammenfassung. Der Beitrag beschreibt einen neuartigen Ansatz zur Strukturanalyse natürlicher Sprache auf der Basis gewichteter Constraints. Unter völligem Verzicht auf eine generative Regelkomponente werden jegliche Bedingungen an die Wohlgeformtheit einer sprachlichen Struktur mit Hilfe von Constraints ausgedrückt. Constraints sind grundsätzlich verletzbar und daher auch in der Lage, widersprüchliche Forderungen der Grammatik zu tolerieren. Über zusätzliche Repräsentationsebenen können sehr unterschiedliche linguistische Phänomenbereiche in die Modellierung einbezogen werden und so ihren spezifischen Beitrag zur Ermittlung der plausibelsten Interpretation einer Äußerung leisten. Für ein derartig definiertes Parsingproblem existieren verschiedene Lösungsverfahren, deren Eigenschaften insbesondere im Hinblick auf mögliche Parallelen zur menschlichen Sprachverarbeitung diskutiert werden. Im Mittelpunkt steht dabei die Robustheit gegen abweichende Äußerungen, sowie gegen zeitlichen Verarbeitungsdruck.

Praktisch alle heutzutage gebräuchlichen Grammatikformalismen haben eine starke Fundierung in der Begriffswelt

Die hier vorgestellten Forschungen wurden von der Deutschen Forschungsgemeinschaft unter den Aktenzeichen Me 1472/1-1, Ku 811/3-1 und Me 1472/1-2 gefördert.

der klassischen mathematischen Logik. Dies gilt für einfache kontextfreie Regelsysteme und deren Erweiterung zur Unifikationsgrammatik ebenso wie für die stärker von der Logikprogrammierung inspirierten constraint-basierten Formalismen. Von diesem logischen Kern her ererben solche Ansätze zum einen die für eine Computersimulation unverzichtbare Stringenz mathematischer Modellierung sowie die fundamentalen Berechenbarkeitseigenschaften der jeweiligen Modellklasse. Darüber hinaus übertragen sie aber auch zwei eng miteinander zusammenhängende Grundannahmen des Modellierungswerkzeugs auf den Modellierungsgegenstand, die gerade in einem linguistischen Kontext nicht ohne genaue Analyse ihrer Konsequenzen hingenommen werden sollten.

Zum einen ist dies der strikt binäre Wohlgeformtheitsbegriff, der in Anlehnung an die binäre Natur des Wahrheitswertes eine sprachliche Äußerung s nur dann als grammatikalisch akzeptiert, wenn sie aus dem in Form von Grammatik \mathcal{G} und Lexikon \mathcal{D} gegebenen sprachlichen Wissen über die Sprache \mathcal{L} folgt:

$$\exists \mathcal{G} \exists \mathcal{D} \forall s \in \mathcal{L} . \mathcal{G} \cup \mathcal{D} \vdash s$$

Typischerweise setzen jedoch die verfügbaren Ableitungsverfahren als zusätzliche Forderung auch noch globale Konsistenz im Axiomensystem voraus

$$\exists \mathcal{G} \exists \mathcal{D} \forall s \in \mathcal{L} . \mathcal{G} \cup \mathcal{D} \cup \{s\} \not\vdash$$

und schließen damit die Existenz von widersprüchlicher Information im linguistischen Wissen bzw. zwischen linguistischem Wissen und sprachlicher Realität grundsätzlich aus.

Im Gegensatz zu diesen Annahmen erweist sich aber der hier zugrunde liegende binäre Wohlgeformtheitsbegriff als eine sehr starke Idealisierung, denn tatsächlich lassen sich in vielen Fällen verschiedene Grade der grammatischen Wohlgeformtheit unterscheiden, so dass sich etwa für die Anordnungsvarianten im Mittelfeld des deutschen Nebensatzes eine mehr oder weniger deutliche Rangfolge hinsichtlich der Akzeptabilität ergibt. Die verschiedenen Varianten unterscheiden sich dann vor allem durch den kontextuellen bzw. prosodischen Aufwand, der zur jeweils problemlosen Verarbeitung erforderlich ist:

- (1a) ..., weil die Mutter dem Kind vertraut.
 (1b) ..., weil dem Kind die Mutter vertraut.
 (2a) ..., weil sie dem Kind vertraut.
 (2b) ..., weil dem Kind sie vertraut.
 (3a) ..., weil die Mutter ihm vertraut.
 (3b) ..., weil ihm die Mutter vertraut.
 (4a) ..., weil sie ihm vertraut.
 (4b) ..., weil ihm sie vertraut.

Offensichtlich ist hier der Grad der Akzeptabilität nicht nur von der Anordnung, sondern auch von der Oberflächenrealisierung der Konstituenten als Nominalgruppe bzw. Proform abhängig. Insgesamt ergeben sich ungefähr die folgenden Akzeptabilitätsurteile:

$$(1a) > (1b), (2a) \gg (2b), (3a) \sim (3b) \text{ and } (4a) \gg (4b)$$

wobei die Relation \sim für einer ungefähr vergleichbare, $>$ für eine größere und \gg für eine viel größere Akzeptabilität stehen.

Letztendlich können derartige Unterschiede auf das Wirken gegensätzlicher Forderungen innerhalb des Systems der Syntax zurückgeführt werden: Während einerseits die unmarkierte Abfolge das Subjekt in initialer Position präferiert, kann das kommunikative Bedürfnis, andere Satzbestandteile hervorzuheben, eine andere Reihung nahelegen. Orthogonal dazu besteht dann aber auch noch eine Präferenz für Proformen in der initialen Topikposition.

Weitere Beispiele für widersprüchliche Forderungen innerhalb der Grammatik sind

- die Optionalität bestimmter Valenzforderungen, die durch die Anwesenheit adverbialer Ergänzungen oftmals stark an Akzeptabilität gewinnt:

- (5a) Sie spielt Schach.
 (5b) Sie spielt mit ihm.
 (5c) Sie spielt gut.
 (5d) Sie spielt.

Auch hier liegt wieder eine merkliche Abstufung in der Akzeptabilität vor.

- die Extraposition von Relativsätzen, wo die Forderung nach syntaktischer Nähe der Bestandteile des Verbs mit dem Wunsch nach minimaler referentieller Distanz zwischen Relativpronomen und seinem Antezedenten in starkem Konflikt steht:

- (7a) Er lehnt jedes Angebot, das unter diesem Betrag bleibt, rigoros ab.
 (7b) Er lehnt jedes Angebot rigoros ab, das unter diesem Betrag bleibt.

Ein ähnliches Konfliktpotential lässt sich auch bei der pronominalen Referenz beobachten, bei der referentielle Nähe z.B. gegen syntaktische bzw. semantische Parallelitätsforderungen abgewogen werden muss.

Besonders offensichtlich ist die Allgegenwart von Inkonsistenz auf der semantischen Ebene. Selektionale Restriktionen etwa sind besonders konfliktträchtig, da semantische Kategorien durch Typverschiebung und metaphorische Uminterpretation starken Modifikationen unterworfen sind. Ein noch stärkeres Spannungsverhältnis ergibt sich

in vielen Fällen zwischen dem propositionalen Gehalt einer Äußerung und dem jeweils verfügbaren Hintergrundwissen. Wollte man hier auf der Forderung nach globaler Konsistenz beharren, würde die Kommunikation neuer und ungewöhnlicher Inhalte völlig unmöglich:

- (8) Schaf erschoss Schäfer im Schlaf (Yahoo, 9.3.2001)

Es sollte deutlich werden, dass Inkonsistenz keinesfalls nur als unerwünschte Abweichung vom widerspruchsfreien Kern der natürlichen Sprache betrachtet werden kann. Sie stellt vielmehr einen grundlegenden Mechanismus bereit, der maßgeblich zum Erfolg zwischenmenschlicher Kommunikation beiträgt. Sprecher nutzen Techniken wie Topikalisierung und Metaphorik systematisch, um ihre Redebeiträge möglichst interessant zu gestalten und so die Aufmerksamkeit des Hörers auf sich zu ziehen. Das Konfliktpotential eines unvollständigen Satzes mit seinen offenen Valenzforderungen etwa kann auch als Einladung zum Wechsel der Dialoginitiative verstanden werden und erfüllt somit eine ganz konkrete kommunikative Funktion. Zudem spielen Inkonsistenzen eine wichtige Rolle bei der Produktion humoristischer Äußerungen, weil sie besonders gut geeignet sind, den Kontrast zwischen konkurrierenden Interpretationen aufzuzeigen und daher gern verwendet werden, einen absurden Sachverhalt zu verdeutlichen.

Aus der Sicht der Logik versucht man, sich dem Problem inkonsistenter Axiomenmengen über die verschiedenen Kalküle zum nichtmonotonen Schließen zu nähern. So werden etwa beim Default-Schließen ausgewählte Axiome z.B. für die Kongruenzforderung in einer deutschen Nominalgruppe durch die Angabe entsprechender Normalitätsforderungen unter Vorbehalt gestellt:

$$\text{nom}(X) \wedge \text{det}(Y) \wedge \text{part_of}(X,Z) \wedge \text{part_of}(Y,Z) \wedge \neg \text{abnormal}(\text{agree}(X,Y,\text{case})) \rightarrow \text{case}(X)=\text{case}(Y)$$

Durch zusätzliche Annahmen über ein irgendwie geartetes "unnormales Verhalten", z.B.

$$\text{abnormal}(\text{agree}(X,Y,\text{case}))$$

werden Extensionen der Theorie konstruiert, in denen die betreffenden Bedingungen dann bei Bedarf, (z.B. beim Vorliegen einer agrammatischen Konstruktion) de facto außer Kraft gesetzt sind. Damit wird die Widersprüchlichkeit der zu modellierenden Domäne in den abgeleiteten Theoremen direkt reflektiert und steht z.B. für die Zwecke einer Fehlerdiagnose zur Verfügung.

Leider greifen solche Ansätze für die Modellierung natürlicher Sprache immer noch zu kurz, da sie zum einen die Behandlung von Inkonsistenzen über den Begriff der Extension letztendlich auch wieder nur auf die Wahrung der Konsistenz in möglichen Welten zurückführen, weshalb spätestens dann, wenn Annahmen über "abnormes Verhalten" gehäuft auftreten, eine Vielzahl von gleichberechtigten Interpretationsmöglichkeiten generiert und ausgewertet werden muss. Zum anderen neutralisiert jede Abnormitätsannahme die zugehörigen Axiome des Modellwissens vollständig, wodurch unter Umständen wertvolle, für die Disambiguierung unverzichtbare Information über Häufigkeit oder Stringenz verloren gehen kann.

Den Erfordernissen der Sprachverarbeitung eher angemessen wäre demgegenüber ein Verarbeitungsmodell,

das widersprüchliche Information innerhalb einer Axiomenmenge nicht nur tolerieren kann, sondern aufgrund von Bewertungen über die relative Wichtung von Constraintverletzungen in die Lage versetzt wird, zwischen widersprüchlichen Befunden abzuwägen. Dazu wird es jedoch erforderlich, die Sicht auf Inkonsistenz als reparables Phänomen ganz aufzugeben und die Behandlung von Widersprüchen zum Grundprinzip der Verarbeitung zu erheben.

Einen Schritt in diese Richtung geht die Optimalitätstheorie (Prince & Smolensky, 1993). Wohlgeformtheitsbedingungen (Prinzipien) über den Basisstrukturen der zu modellierenden Domäne (Phonemsequenzen in der Phonologie bzw. Phrasenstrukturbäume in der Syntax) können einander widersprechen und werden grundsätzlich als verletzbar angesehen. Überdies sind die Prinzipien in eine Hierarchie der Verletzbarkeit eingeordnet, so dass im Falle einer Inkonsistenz aus der Vielzahl möglicher Strukturrepräsentationen diejenige ausgewählt werden kann, die sich bezüglich der Hierarchie nur durch minimale Constraintverletzungen, d.h. maximale Harmonie, auszeichnet.

Einen radikaleren Ansatz bietet hier die Constraint Dependency Grammar (CDG, Heinecke et al., 1998), die über die Optimalitätstheorie in dreierlei Hinsicht hinausgeht:

- Wohlgeformtheitsbedingungen werden nicht nur hierarchisiert, sondern mit Straffaktoren gewichtet, die zu einer Gesamtbewertung für komplette Strukturbeschreibungen akkumuliert werden können. Dadurch kann insbesondere bei massiven Abweichungen von der sprachlichen Norm auch unterhalb der jeweils "gravierendsten" Constraintverletzung noch eine Wichtung der Interpretationsvarianten erfolgen, so dass letztendlich auch eine als relativ schwerwiegend eingestufte Constraintverletzung noch durch die Kombination mehrerer geringfügiger Verletzungen neutralisiert werden kann. Auf der anderen Seite lassen sich (im Rahmen der numerischen Genauigkeit endlicher Repräsentationen) die Straffaktoren der Constraint Dependency Grammar stets so vergeben, dass sie die strikte Hierarchisierung der Optimalitätstheorie simulieren.
- Auf ein kontextfreies Grundgerüst als Basis für die Constraintanwendungen wird vollständig verzichtet, da sich der überwiegende Teil der Forderungen an die Struktur wohlgeformter Äußerungen als Constraints selbst formulieren lässt. Normalerweise werden dabei zahlreiche dieser Constraints als verletzbar deklariert sein, so dass für den Parser sehr große Freiheitsgrade bei der Festlegung der jeweils plausibelsten Strukturbeschreibung bestehen, auch wenn die Struktur aus der Perspektive einer klassischen Sprachmodellierung in sich selbst widersprüchlich sein sollte.
- Wegen des relationalen Charakters der Dependenzgrammatik ist die Modellierung nicht auf die traditionellen Subordinationsstrukturen der Oberflächensyntax beschränkt, sondern kann bestimmte Aspekte anderer sprachlicher Ebenen einschließen, solange diese sich als Relationen zwischen den Wortformen der sprachlichen Äußerung darstellen lassen. Hierzu gehört etwa die Beziehung der Vorfeldposition zur Informationsstruktur und die Abbildung der Verbvalenzen auf die Vergabe der thematischen Rollen.

Die Constraint Dependency Grammar stellt damit ein recht leistungsfähiges Paradigma für Experimente zum Zusammenspiel unterschiedlicher Wissenskomponenten bei der Sprachverarbeitung dar. Von besonderem Interesse ist dabei das Verhalten beim Vorliegen gestörter bzw. unvollständiger sprachlicher Äußerungen, sowie bei Ressourcenknappheit.

1 Parsing mit Constraints

Grundsätzlich lässt sich jedes nebeneffektfreie Logikprogramm, und somit letztendlich auch jede formale Grammatik, als Menge von Constraints über dem Raum möglicher Lösungen interpretieren. Von Constraints im eigentlichen Sinne spricht man jedoch eher dann, wenn die Datenstruktur, über der die Constraints formuliert sind, relativ allgemein gehalten ist. Als typisches Beispiel hierfür kann die Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag, 1994) dienen, in der das kontextfreie Grundgerüst auf einen rudimentären Bestand an sehr generellen Strukturschemata zurückgedrängt worden ist, während Constraints in Form von Implikationen über typisierten Merkmalsstrukturen den zentralen Kern des Grammatikwissens bilden.

Verglichen mit anderen Problemstellungen aus dem Bereich des Constraint-Satisfaction handelt es sich beim HPSG-Parsing um verhältnismäßig komplexe Bedingungen, die über einer einzigen Systemvariablen (und zwar für die Strukturbeschreibung des komplexen sprachlichen Zeichens "Satz") definiert sind. Wegen der rekursiven Natur des hierfür erforderlichen Datentyps (gerichtete azyklische Graphen) ist zudem die Domäne der Variablen potentiell unbeschränkt.

Aus der Perspektive der Lösungsverfahren für Constraint-Satisfaction-Probleme (Tsang, 1993) erscheint eine derartige Problemstellung eher ungewöhnlich. Betrachtet werden in diesem Bereich vielmehr Aufgaben mit zahlreichen Variablen, jedoch sehr einfach strukturierten Datentypen als Wertebelagungen. Für solche Problemstellungen existiert dann ein ganzes Spektrum von Lösungsverfahren mit unterschiedlichen und auch aus kognitiver Sicht sehr interessanten prozeduralen Eigenschaften, wie 'fail-soft'-Robustheit und Unterbrechbarkeit.

Voraussetzung für die Anwendbarkeit derartiger Techniken ist allerdings eine geeignete Zerlegung des Parsingproblems, so dass sich die gesuchte Lösung als Wertebelagung für mehrere Systemvariablen ergibt. Eine solche Darstellung des Parsings als Constraint-Satisfaction-Problem wurde erstmals von Maruyama (1990) für annotierte Dependenzstrukturen angegeben.¹ Dabei entsprechen die Variablen den unterzuordnenden Wortformen und die Werte stellen Paare aus der Identität des Unterordnungspartners (im Falle des Spitzenknotens ein reservierter Bezeichner) und der jeweils vorliegenden Unterordnungsbeziehung (Label) dar. Jede vollständige Auswahl aus diesem Wertevorrat entspricht dann einer Dependenzstruktur. Diese muss jedoch nicht notwendigerweise azyklisch und zusammenhängend sein. Werden diese Eigenschaften ebenfalls gefordert, sind zusätzlich (globale) Constraints erforderlich, die dann jedoch

¹ Ein vergleichbarer Ansatz, jedoch für unterspezifizierte Dependenzstrukturen findet sich in (Karlsson et al., 1995)

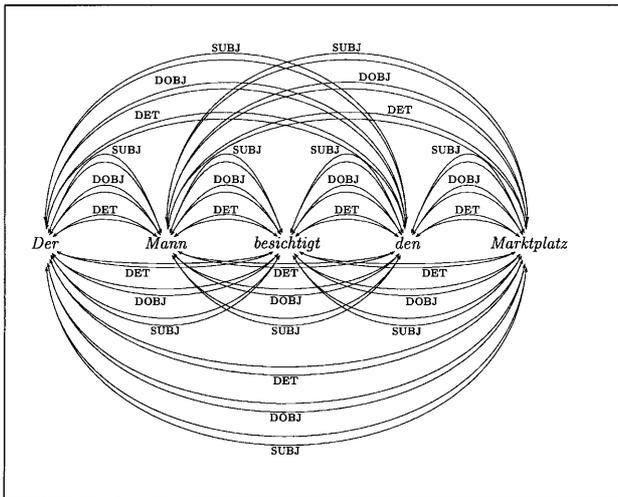


Abb. 1. Der vollständige Hypothesenraum eines Parsingproblems: Jede Wortform modifiziert jede andere Wortform mit jeder möglichen Unterordnungsrelation.

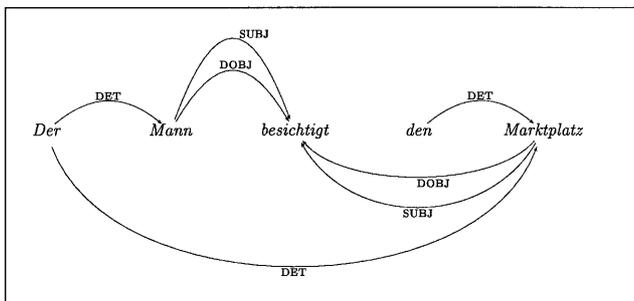


Abb. 2. Der Hypothesenraum nach der Anwendung von drei unären Constraints.

zweckmäßigerweise vom Lösungsverfahren selbst überprüft werden sollten.

Abbildung 1 zeigt den vollständigen Hypothesenraum für ein stark vereinfachtes Parsingproblem: Betrachtet werden nur die drei Unterordnungsrelationen SUBJ, DOBJ und DET. Durch ggf. auch wiederholte Constraintanwendung kann der Raum möglicher Lösungen sukzessiv eingeschränkt werden, weshalb sich das Parsing auch als Disambiguierung einer hinsichtlich struktureller Anbindung und Dependenzrelation maximal mehrdeutigen, andererseits aber auch hochgradig unterspezifizierten initialen Struktur verstehen lässt.

Abbildung 2 zeigt den Hypothesenraum nach der Anwendung von drei unären Constraints, die alle potentiellen Unterordnungsbeziehungen isoliert von ihrer strukturellen Umgebung betrachten², und zwar

- Ein Artikel modifiziert ein Nomen im rechten Kontext mit dem Label DET.
- Ein Nomen kann das finite Verb entweder als Subjekt oder direktes Objekt modifizieren.
- Das finite Verb modifiziert nichts, ist also der Spitzenknoten der Dependenzstruktur.

Obwohl durch diese Bedingungen bereits ein Großteil der potentiellen Unterordnungsmöglichkeiten ausgeschlos-

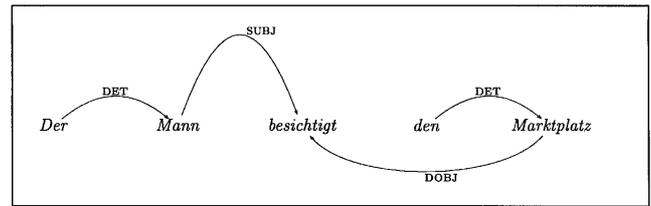


Abb. 3. Eine vollständig disambiguierte Dependenzstruktur.

sen werden kann, ist eine vollständige Disambiguierung dieses Beispiels allein mit unären Constraints nicht möglich. Hierzu müssen binäre Constraints herangezogen werden, welche die wechselseitige Verträglichkeit von zwei Unterordnungsbeziehungen beurteilen. Im betrachteten Beispiel sind hierfür die beiden Constraints

- Eine Wortform kann nicht zwei Mal mit dem gleichen Label modifiziert werden.
- Das Subjekt steht im Nominativ.

ausreichend. Die resultierende Struktur ist in Abb. 3 angegeben.

Typische Problemfälle, die beim Lösen derartiger Constraint-Satisfaction-Aufgaben auftreten können, sind mit der Existenz unter- bzw. überbeschränkter Constraint-Satisfaction-Probleme verbunden. Während im ersten Fall offenbar das vorhandene grammatische Wissen zur vollständigen Disambiguierung nicht ausreicht (vgl. Harper et al., 1994), verhindert im zweiten Fall widersprüchliche Information im zugrunde liegenden Axiomensystem die Ermittlung überhaupt eines Berechnungsergebnisses. In der Tat stellt sich die "punktgenaue" Disambiguierung einer sprachlichen Äußerung als eine sehr schwierige Gratwanderung zwischen diesen beiden Extremen heraus, da für die Disambiguierung verschiedener Sätze oftmals ganz unterschiedliche Constraints benötigt werden. Leider zeigt es sich immer wieder, dass dieses Problem allein durch eine gezielte Manipulation am grammatischen Wissen kaum gelöst werden kann. Erforderlich ist vielmehr die gleichzeitige Erfüllung zweier scheinbar unvereinbarer Teilziele:

- eine Anreicherung der Grammatik mit möglichst vielen, für die Disambiguierung notwendigen Constraints und
- die gezielte Rücknahme von Constraints zur Vermeidung von logischen Inkonsistenzen.

Letztendlich handelt es sich hier um das gleiche Dilemma, in dem sich jede Grammatikentwicklung befindet, solange sie sich dem Ideal einer hohen Akzeptanzschärfe zwischen korrekten und inkorrekten Äußerungen verpflichtet fühlt. Auch aus dieser Sicht scheint also ein systematischer Ansatz zur Behandlung von Inkonsistenz unumgänglich.

2 Parsing mit gewichteten Constraints

Grundsätzlich ist mit der Abbildung der Parsingaufgabe auf ein Constraint-Satisfaction-Problem ein sehr flexibler Ansatz gegeben, der Erweiterungen und Ergänzungen in ganz verschiedene Richtungen zulässt. Für praktische Zwecke unumgänglich und auch relativ nahe liegend ist die Einbezie-

² Alle hier verwendeten Constraints stellen starke Vereinfachungen dar.

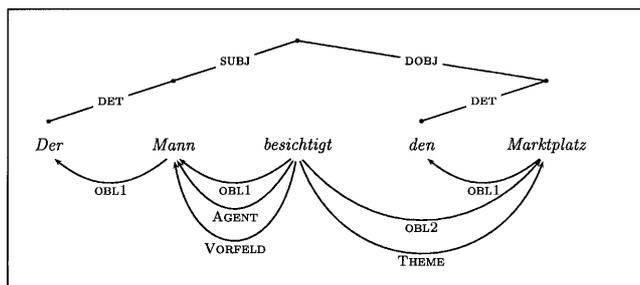


Abb. 4. Eine Mehrebenenrepräsentation unter Einbeziehung der Vorfeldbesetzung, sowie der Funktor-Argument-Struktur.

hung lexikalischer Mehrdeutigkeit, die auch als ein Spezialfall von Hypothesenunsicherheit bei der Erkennung gesprochener Sprache behandelt werden kann (Harper & Helzermann 1995). Das Parsing erfolgt dann nicht mehr über einer einfachen Sequenz von lexikalischen Elementen, sondern in unterschiedlich komplex strukturierten Hypothesen, die eine Vielzahl von Kombinationsmöglichkeiten beschreiben. Durch einen zusätzlichen Mechanismus muss dann natürlich dafür Sorge getragen werden, dass alternative Lesarten bzw. zeitlich überlappende Hypothesen nicht mehr in ein und dieselbe Strukturbeschreibung integriert werden können.

Im Interesse der Erschließung zusätzlicher Information für die strukturelle Disambiguation erscheint die Hinzunahme weiterer linguistischer Repräsentationsebenen nützlich. Diese Erweiterung ist vor allem schon deshalb nahe liegend, weil für die Behandlung von Obligatoritätsforderungen ohnehin mehrere Hilfeebenen erforderlich sind. Die Notwendigkeit hierfür ergibt sich aus der Asymmetrie des Lösungsverfahrens, das zwar für jede Wortform die eindeutige Auswahl eines Partners zur strukturellen Anbindung sicherstellen kann, nicht jedoch die eindeutige Auswahl einer unterzuordnenden Wortform z.B. zur Sättigung obligatorischer Valenzforderungen. Hierfür sieht Maruyama (1990) inverse Dependenzbeziehungen (sog. needs-Rollen) vor. In Abb. 4 sind diese als OBL1 bzw. OBL2 markiert.

Während inverse Dependenzbeziehungen jedoch nur einen rein technischen Charakter besitzen, lässt sich die Grundidee der parallelen Disambiguation auf mehreren Ebenen leicht auf die Einbeziehung zusätzlicher struktureller Phänomene übertragen. Offensichtliche Kandidaten hierfür sind Aspekte der semantischen Struktur sowie der Informationsstruktur eines Satzes. Hierzu zählen etwa die Vergabe der thematischen Rollen durch das Verb (z.B. AGENT bzw. THEME in Abb. 4) und die Vorfeldbesetzung im Deutschen³. Ebenfalls recht gut geeignet erscheinen die innersprachlichen Referenzbeziehungen. In speziellen Anwendungen können selbst Teile des Hintergrundwissens — etwa die räumlichen Relationen zwischen Referenzobjekten — in die Modellierung einbezogen werden, um so auch außersprachliche Wissensquellen für die Disambiguation zu mobilisieren (Menzel, 1998).

Teilstrukturen auf den verschiedenen Repräsentationsebenen lassen sich ebenfalls mit Hilfe von Constraints aufeinander abbilden. Derartige Mapping-Constraints bewirken

³ Ein systematischer Ansatz zur Modellierung der topologischen Struktur mit Hilfe von Constraints findet sich in (Duchier & Debusmann (2001))

dann einen bidirektionalen Informationsfluss zwischen den beteiligten Ebenen und damit ein im Idealfall synergetisches Zusammenwirken der verschiedenen Wissensquellen bei der Disambiguation.

Von zentraler Bedeutung für die Behandlung inkonsistenter Information und damit für die Leistungsfähigkeit des gesamten Ansatzes ist jedoch die prinzipielle Möglichkeit zur Berücksichtigung von Constraintverletzungen. Um die dabei unvermeidlich auftretende drastische Ausdehnung des Hypothesenraums zumindest teilweise wieder zu kompensieren, werden Constraints mit Straffaktoren p_i annotiert, die mögliche Constraintverletzungen im Hinblick auf die Akzeptabilität der daraus resultierenden Strukturbeschreibung bewerten. Straffaktoren werden multiplikativ verknüpft und lassen sich so zu Gesamtbewertungen für komplette Strukturbeschreibungen akkumulieren.

Aufgrund ihrer Straffaktoren lassen sich Constraints grob in drei Klassen einordnen, die ihre Rolle im Disambiguationprozess charakterisieren:

- Constraints mit einem Straffaktor $p_i = 0$ entsprechen den klassischen (harten) Constraints und dürfen unter keinen Umständen verletzt werden. Derartige Constraints werden benötigt, um den Raum möglicher Strukturbeschreibungen sinnvoll einzugrenzen und dürften etwa die Rolle des kontextfreien Rückgrats der Optimalitätstheorie spielen.
- Constraints mit einem Straffaktor nahe Null ($0 < p_i \ll 1$) stellen verletzbare Bedingungen für die Wohlgeformtheit einer Äußerung dar. Beispiele hierfür sind Valenz-, Rektions- und Kongruenzforderungen, sowie die strikten Anordnungsrestriktionen für das finite Verb im deutschen Haupt- und Nebensatz.
- Constraints mit einem Straffaktor nahe Eins ($0 \ll p_i < 1$) sind Standardforderungen bzw. Präferenzen, die beim Fehlen von gegenteiliger Information eine vollständige Disambiguation auf die unmarkierte Lesart erzwingen können. Hierzu zählen z.B. die eingangs angesprochenen Anordnungspräferenzen im Mittelfeld des deutschen Nebensatzes, aber auch Distanzphänomene bei der strukturellen Anbindung bzw. bei der Referenzauflösung. Auch selektionale Restriktionen haben diesen Präferenzcharakter. Präferenzen erschließen damit eine recht große Klasse sprachlichen Wissens für die Disambiguation, das bei einer klassischen binären Modellierung zwangsläufig ungenutzt bleiben müsste.

In vielen Fällen ermöglicht die Verwendung der mit Straffaktoren versehenen Constraints eine effiziente Modellierung. Versucht man z.B. die Akzeptabilitätsurteile für die Äußerungen (1a) bis (4b) nachzubilden, so reichen hierfür drei annähernd gleich stark gewichtete Constraints aus

- Das Subjekt sollte vor dem Objekt stehen.
- Nur die Konstituente in der Topikposition sollte pronominalisiert werden.
- Ein pronominalisiertes Subjekt sollte vor einem pronominalisiertem Objekt stehen.

Diese erzeugen dann für die vier Satzpaare (1a/b) bis (4a/b) Bewertungen, die dem ein-, zwei-, null- bzw. zweifachen des jeweils gewählten Straffaktors entsprechen und

damit recht gut mit den oben angegebenen intuitiven Akzeptabilitätsurteilen übereinstimmen.

Aufgrund der Einbeziehung gewichteter Constraints stellt sich die strukturelle Analyse einer sprachlichen Äußerung nunmehr als Optimierungsaufgabe dar, die aus der Menge möglicher Strukturhypothesen diejenige mit der maximalen Gesamtbewertung ermittelt. Dieser Optimierungsprozess führt somit zu einem globalen Abgleich von möglicherweise widersprüchlicher Evidenz auf bzw. zwischen den verschiedenen Repräsentationsebenen im Rahmen eines einheitlichen Berechnungsverfahrens. Constraints modellieren zwar nach wie vor Wohlgeformtheitsbedingungen, können aber für den Fall, dass keine geeignetere Interpretation zur Verfügung steht, auch als verletzt betrachtet werden.

Letztendlich steht dank der Kopplung von Mehrebenenrepräsentation und gewichteten Constraints erstmalig eine Systemarchitektur zur Verfügung, in die sich sehr unterschiedliche linguistische Komponenten in modularer Weise integrieren lassen, ohne dabei der Interaktion zwischen den Komponenten von vornherein Beschränkungen aufzuerlegen. Sie trägt damit der von Trueswell, Tanenhaus und Garnsey (1994) erhobenen Forderung nach Modularität auf der Ebene der Repräsentation, aber Interaktivität auf der Ebene des Berechnungsverfahrens in geradezu idealer Weise Rechnung, erlaubt jedoch im Gegensatz zu konnektionistischen Ansätzen wegen der durchgängig symbolischen Realisierungsgrundlage ganz gezielte experimentelle Untersuchungen zum Zusammenwirken der unterschiedlichen Ebenen in einem komplexen sprachverarbeitenden System.

Entscheidende Voraussetzung für die Planung und Durchführung derartiger Computersimulationen ist jedoch die Klärung der beiden folgenden Fragestellungen:

- Für welche Modellklassen kann man noch Lösungsverfahren mit akzeptablem Ressourcenbedarf angeben?
- Lassen sich natürlichsprachliche Phänomene mit den jeweils noch praktikablen Modellklassen in der erforderlichen Breite modellieren?

Da der Rechenzeitaufwand für die Lösung von Constraint Satisfaction Problemen mit der Stelligkeit der verwendeten Constraints extrem stark ansteigt, wurden bisher ausschließlich Grammatiken mit ein- und zweistelligen Constraints untersucht. Die damit spezifizierbaren Teilstrukturen enthalten dann maximal zwei Unterordnungsbeziehungen bzw. bis zu vier daran beteiligte Wortformen. Insbesondere bei der Modellierung der ebenenübergreifenden Beziehungen macht sich eine derartige Einschränkung jedoch bereits recht unangenehm bemerkbar. Dennoch ist es trotz dieser vereinfachenden Annahmen gelungen, unter gezielter Verwendung von Approximationstechniken weite Bereiche der Grammatik natürlicher Sprache mit ausreichender Genauigkeit zu modellieren (Schröder et al., 2000).

Auf der Seite der Lösungsverfahren steht eine ganze Reihe von Algorithmen zur Verfügung, die sich durch unterschiedliche, teils komplementäre Eigenschaften auszeichnen (Menzel & Schröder, 1998b). Hierzu zählen:

- Verfahren zum Wegstreichen (Pruning) von individuellen Wertebelegungen aufgrund lokaler heuristischer Kriterien. Sie zeichnen sich zumindest in der Pruningphase durch einen geringen Zeitbedarf und eine gute Vorher-

sagbarkeit des Terminierungszeitpunkts aus. Dem steht jedoch ein unakzeptabel hoher Aufwand zur Initialisierung des Hypothesenraums und eine sehr geringe Zuverlässigkeit des Verfahrens gegenüber.

- Verfahren zur Bestensuche, die bestehende Teillösungen durch Hinzunahme erfolgversprechender Abhängigkeitsbeziehungen sukzessiv zu einer kompletten Strukturbeschreibung erweitern, wobei die Auswahl der Kandidaten durch die Constraintgewichte gesteuert wird. Derartige Verfahren sind bei hinreichender Ressourcenallokation tatsächlich in der Lage, das gesuchte Optimum zuverlässig zu ermitteln. Allerdings ist der Zeitpunkt, zu dem dieses Ergebnis vorliegen wird, praktisch nicht prognostizierbar.
- Transformationsbasierte Verfahren, die durch partielle Modifikation einer bereits bestehenden Komplettlösung deren schrittweise Verbesserung anstreben. Dadurch dass hier zu jedem Zeitpunkt eine vollständige Lösung bekannt ist, sind derartige Algorithmen praktisch beliebig unterbrechbar und bieten sich daher vor allem zur Modellierung der zeitlichen Adaption bei der Sprachverarbeitung an. Zu unterscheiden ist hierbei zwischen Verfahren, die aufgrund eines perfekten Gedächtnisses Mehrfachanalysen vermeiden können (Foth, Menzel & Schröder, 2000) und solchen, die den diesbezüglichen Mehraufwand zugunsten eines einfachen parametrischen Suchraummodells in Kauf nehmen (Schulz, 2000). Nicht ganz unproblematisch ist in beiden Fällen die Wahl eines geeigneten Abbruchkriteriums zur Terminierung der Lösungssuche.

3 Robustheit gegen Normabweichungen

Auffälliges Merkmal des Parsings mit gewichteten Constraints ist sicherlich seine erstaunliche Robustheit gegenüber nicht-normgerechten sprachlichen Äußerungen. Typischerweise ist das Verfahren auch bei schwerwiegenden Constraint-Verletzungen noch in der Lage, plausible strukturelle Interpretationen für die sprachlichen Eingaben zu erzeugen. Zu diesem hohen Grad an Robustheit tragen vor allem zwei Faktoren maßgeblich bei:

- die Existenz gewichteter Constraints, durch die der Prozess des Abwägens zwischen konkurrierenden Interpretationsvarianten wirksam gesteuert werden kann, sowie
- die gezielte Nutzung von struktureller Redundanz zwischen den verschiedenen Repräsentationsebenen, die im Zusammenspiel mit den Constraintgewichten eine effiziente Nutzung der jeweils spezifischen Informationsbeiträge ermöglicht, ohne den Gesamtprozess der Strukturanalyse von der Korrektheit dieser Beiträge abhängig zu machen.

Durch Experimente mit gezielt gestörten sprachlichen Daten konnte gezeigt werden, dass kompensatorische Effekte zwischen den verschiedenen Repräsentationsebenen wirksam zur Steigerung der Robustheit des Gesamtsystems beitragen (Menzel & Schröder, 1998a). So ermöglicht in vielen Fällen erst eine massive semantische Unterstützung die sinnvolle Interpretation syntaktisch abweichender Daten, während andererseits starke syntaktische Indikatoren einen

wichtigen Beitrag zur Analyse semantisch fragwürdiger Inhalte leisten können. Eine optimale Performanz ergibt sich allerdings erst im synergetischen Zusammenwirken der beiden Ebenen.

Aus kognitionswissenschaftlicher Perspektive erscheint besonders interessant, dass hierbei die Behandlung agrammatischer Äußerungen nicht als Sonderfall der Verarbeitung auftritt, der in einen nachgeordneten Verarbeitungsschritt nach dem Scheitern der eigentlichen Analyse ausgelagert werden kann (vgl. z.B. Mellish (1989)), sondern als integraler Bestandteil des normalen Verarbeitungsablaufs angesehen werden muss. Interpretiert man eine deutlich vom Normalniveau abweichende Gesamtbewertung für eine Äußerung als Äquivalent für das bekannte, aber noch unspezifische Gefühl "Hier stimmt etwas nicht!", werden durch das CDG-Parsing auch die für die menschliche Sprachverarbeitung charakteristischen Kausalbeziehungen korrekt wiedergegeben: Agrammatikalität kann auch unbemerkt bleiben und erfolgreiches Sprachverstehen hat keinesfalls auch das Detektieren von Fehlern zur Voraussetzung. Vor allem beim Vorhandensein starker semantischer Unterstützung aus dem sprachlichen bzw. außersprachlichen Kontext einer Äußerung bleiben Formabweichungen eher unbemerkt.

Damit liegt hier erstmals eine Systemarchitektur vor, mit der wesentliche Aspekte der erstaunlich großen Robustheit des menschlichen Sprachverstehens im Simulationsexperiment nachvollzogen werden können. Sie rückt daher auch die Frage nach den kognitiven Grundlagen robuster Sprachverarbeitung verstärkt in den Mittelpunkt, eine Frage die bislang aufgrund einseitig idealisierender Annahmen über die Qualität sprachlicher Daten weitgehend ignoriert worden ist.

Tatsächlich gehen die Parallelitäten zur menschlichen Sprachverarbeitung deutlich über das reine Tolerieren bzw. Detektieren von Constraintverletzungen hinaus. Sie schließen gleichermaßen die Fähigkeit zur Diagnose der Fehlerursache mit ein. Entscheidend dafür ist die Tatsache, dass die Ermittlung der optimalen Strukturinterpretation automatisch auch die Identifikation der durch sie verletzten Constraints zur Folge hat. Da außerdem die Constraints der CDG in den weitaus meisten Fällen ganz elementare Wohlgeformtheitsbedingungen betreffen (Kongruenz wird gefordert, eine Valenz muss gesättigt sein usw.) und sie somit auch über eine elementar charakterisierbare Negation verfügen (die Kongruenz ist verletzt, die Valenz ungesättigt usw.), können Constraintverletzungen ohne größeren Aufwand auch als Fehlererklärungen interpretiert werden. Dies eröffnet insbesondere für Anwendungen in Sprachlehrsystemen neue Perspektiven, da wegen der Verfügbarkeit der Mehrebenenrepräsentation syntaktische Abweichungen nunmehr nicht mehr isoliert, sondern eingebettet in den zu kommunizierenden propositionalen Gehalt diagnostiziert werden können [Menzel und Schröder 1998b, Menzel und Schröder 1998c].

Auch hier ergibt sich wieder eine weitgehend "naturanalogue" Staffelung, welche die mit der jeweiligen Teilaufgabe verbundenen kognitiven Anforderungen widerspiegelt: Während das Tolerieren einer Constraintverletzung offenbar die geringsten kognitiven Anforderungen stellt, sind für das Detektieren, das Lokalisieren und schließlich auch das Diagnostizieren jeweils aufwendigere Verarbeitungsleistungen zu erbringen.

4 Robustheit gegenüber Ressourcenbeschränkungen

Nicht nur hinsichtlich der Robustheit gegen Normabweichungen, sondern auch bezüglich der zeitlichen Einflussfaktoren verfügen menschliche kognitive Prozesse über einen enormen Spielraum zur Anpassung an wechselnde interne oder externe Vorgaben. Herausragend ist hierbei insbesondere die Fähigkeit, den Zeitaufwand bzw. die Verarbeitungskapazität für die Lösung einer Aufgabe zu Lasten der Ergebnisqualität zu reduzieren. Beispiele für diese Flexibilität finden sich natürlich auch in der Sprachverarbeitung, so etwa beim zumeist selbst vorgegebenen Lesetempo, oder aber bei der Anpassung an das vom jeweiligen Sprecher abhängige Sprechtempo in der mündlichen Kommunikation.

Berechnungsverfahren, die über die fundamentale Fähigkeit des Abwägens zwischen dem Nutzen eines Resultats und dem für seine Erzeugung jeweils benötigten Aufwand verfügen, werden in der Literatur auch als Anytime-Algorithmen bezeichnet (Boddy & Dean, 1994). Anytime-Eigenschaften können am einfachsten realisiert werden, wenn es gelingt, unterbrechbare Algorithmen zu entwerfen, die zu jedem Zeitpunkt über eine vollständige Lösung verfügen und kontinuierlich an deren qualitativer Verbesserung arbeiten. Genau auf diesem Prinzip basieren die oben genannten Transformationsverfahren. Da sie mit der Gesamtbewertung über ein quantitatives Maß für den Analysefortschritt verfügen, können sie dieses jederzeit mit der Nützlichkeit des zu erwartenden Berechnungsergebnisses in Beziehung setzen. Dabei ist die Nützlichkeit eines Resultats eng mit den jeweils noch verfügbaren Zeitreserven verbunden und fällt daher jenseits eines bestimmten Zeitpunkts deutlich ab. Es sollte daher möglich sein, die Berechnung abzubrechen, wenn der noch zu erwartende Analysefortschritt den damit verbundenen Verlust an Nützlichkeit nicht mehr aufwiegen kann.

Eine zweite wichtige Klasse von Anytime-Algorithmen bilden die Vertragsalgorithmen. Diese müssen in der Lage sein, die zur Lösung einer Aufgabe erforderlichen Mittel in Abhängigkeit von der noch verfügbaren Zeit so gezielt einzusetzen, dass nach Ablauf der vorgegebenen Zeit ein Resultat mit der geforderten Qualität tatsächlich zur Verfügung steht. Voraussetzung dafür ist

- die Kenntnis eines Performanzprofils, das die durchschnittlich erreichbare Qualität der Resultate in Abhängigkeit vom investierten Zeitaufwand beschreibt, und
- die grundsätzliche Fähigkeit des Verfahrens, bei Bedarf zu einer beschleunigten und entsprechend weniger gründlichen Arbeitsweise überzugehen.

Erste Ansätze für ein solches Verhalten bietet das CDG-Parsing dank der Constraintgewichte. Bereits mit einer sehr einfachen Vorschrift, die alle Constraintgewichte unterhalb eines vorgegebenen Schwellwerts θ auf Null setzt, lässt sich der gewünschte Beschleunigungseffekt erzielen. Abbildung 5 zeigt den mittleren Zeitbedarf sowie das erreichte Qualitätsniveau in Abhängigkeit von der Wahl des Schwellwerts für die Bestensuche. Deutlich erkennbar ist der erhebliche Geschwindigkeitszuwachs bei moderatem Qualitätsabfall im Bereich $0 < \theta < 0.2$.

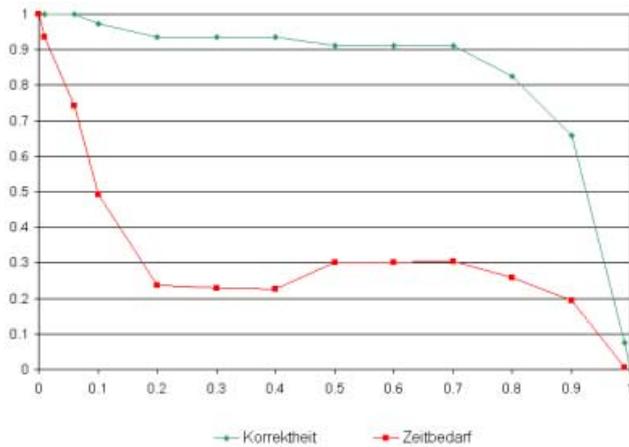


Abb. 5. Reduktion von Zeitbedarf und Analysequalität in Abhängigkeit von der Wahl des Schwellwerts θ

Ein ganz ähnliches Verhalten ergibt sich auch für den transformationsbasierten Ansatz, wenn der Schwellwert für die maximal in Kauf zu nehmende Bewertungsverschlechterung während eines Transformationsschritts variiert wird.

5 Inkrementelle Verarbeitung

Bislang wurden all die genannten Experimente zum qualitativen Zeitverhalten nur für die Verarbeitung kompletter Sätze durchgeführt. Tragfähige Aussagen über Analogien zum menschlichen Vorbild sind aber erst dann möglich, wenn auch das maschinelle Verfahren der temporalen Struktur der sprachlichen Äußerung Rechnung trägt und die Verarbeitung bereits parallel zum Empfang der Eingabedaten erfolgen kann.

Einen wichtigen Ansatzpunkt für die Realisierung einer solchen inkrementellen Analyse bieten wiederum die grundlegenden Robustheitseigenschaften des Parsers, die in jedem Fall gewährleisten, dass auch unvollständige Äußerungen eine optimale Strukturbeschreibung zugewiesen bekommen. Auf diese Weise lassen sich zum Beispiel beliebige Satzpräfixe mit schrittweise anwachsender Länge verarbeiten, wobei die dabei auftretenden Constraintverletzungen den noch offenen Erwartungen an den rechten Kontext entsprechen und die bisher akkumulierte Gesamtbewertung als Gradmesser für die Vollständigkeit der Eingabedaten interpretiert werden kann:

Satzpräfix	Diagnose	Bewertung
Der ...	Fehlendes finites Verb	0.01
Der Mann ...	Fehlendes finites Verb	0.01
Der Mann besichtigt ...	Fehlendes zweites Komplement	0.10
Der Mann besichtigt den ...	—	1.00

Dabei ergibt sich die relativ hohe Bewertung im ersten und letzten Fall über eine Interpretation des Artikels als Demonstrativpronomen in Verbindung mit der Tatsache, dass Hinweise aus dem außersprachlichen Kontext, wie das Vorliegen bzw. die Abwesenheit einer Zeigegeste, in der Modellierung noch nicht berücksichtigt wurden.

Von einer tatsächlich inkrementellen Verarbeitung kann natürlich erst dann gesprochen werden, wenn in jedem einzelnen Analyseschritt wieder von den zuvor bereits ermittelten Zwischenergebnissen ausgegangen wird. Obwohl durch den Einsatz entsprechender Heuristiken der Zeitbedarf für eine derartige Analysestrategie schon deutlich reduziert werden konnte (Foth, Menzel & Schröder, 2000), so ist jedoch noch kein Verfahren bekannt, das den über größere Zeiträume etwa linearen Zeitaufwand bei der inkrementellen Verarbeitung einer Äußerung durch den Menschen auch nur annähernd nachbilden könnte.

6 Zukünftige Arbeiten und offene Fragestellungen

Verfahren zum Constraint-Satisfaction bringen von Natur aus ein verhältnismäßig hohes Potenzial zur Realisierung als nebenläufige Prozedur mit, wobei die Parallelisierung der Berechnung sowohl durch Verteilung der Variablen des Problems auf mehrere Prozessoren als auch durch parallele Auswertung der Constraints erreicht werden kann. Helzerman & Harper (1992) konnten zeigen, dass für CDGs mit harten Constraints auf einer massiv parallelen Architektur ein logarithmischer Rechenzeitaufwand in Abhängigkeit von der Satzlänge n erreichbar ist, wenn hierfür $\mathcal{O}(n^4)$ Prozessoren zur Verfügung stehen. Allerdings basiert der dabei verfolgte Ansatz auf einer Compilation des vorliegenden Parsingproblems in das parallele Programm und ist daher aus kognitiver Sicht weniger interessant. Gefragt sind letztendlich Lösungen zur dynamischen Lastverteilung, die den temporalen Verlauf des Analyseprozesses auch aktiv beeinflussen können.

Ein alternativer Ansatz zur nebenläufigen Realisierung des Parsings besteht in einer relativ grobkörnigen Parallelisierung auf der Ebene kompletter, aber unterschiedlich parametrisierter Lösungsverfahren. Dabei kann man sich die Tatsache zu Nutze machen, dass sich die gesteuerte lokale Suche sehr gut auf verschiedene Satzstrukturen hin optimieren lässt. Besonders aussichtsreich sind hier Architekturen mit konkurrierenden Instanzen des Lösungsverfahrens, die sich wechselseitig über den Analysefortschritt informieren. Erste Simulationsexperimente haben ergeben, dass sich auf dieser Grundlage sogar ein Qualitätszuwachs gegenüber den jeweiligen Einzelkomponenten erreichen lässt (Schulz, 2000). Sollte sich herausstellen, dass auf dieser Basis auch ein Geschwindigkeitszuwachs erzielt werden kann, dann deutet sich damit u.U. auch ein Erklärungsansatz für die Race-Phänomene an, die dazu führen, dass Menschen global mehrdeutige Sätze schneller verarbeiten können als die jeweils zugrunde liegenden eindeutigen Varianten (Traxler, Pickering & Clifton, 1998).

Eine weitere interessante Fragestellung betrifft das ebenfalls sehr hohe Potenzial des Ansatzes zum partiellen Parsing. Weicht man die Forderung nach einem finiten Verb als Spitzenknoten einer syntaktischen Oberflächenstruktur graduell auf, eröffnet sich die Möglichkeit, auch Satzfragmente erfolgreich zu analysieren. In der Tat wird von dieser Eigenschaft sowohl auf den nicht-syntaktischen Ebenen als auch beim inkrementellen Parsing bereits massiv Gebrauch gemacht, da in diesen Fällen vollständige und zusam-

menhängende Analyseebäume (zumindest temporär) ja gar nicht gefordert werden können.

Baut man dieses Potenzial gezielt aus, indem man spezialisierte, jedoch semantisch stark vorgeprägte Grammatiken auf einen unbekanntem Text anwendet, so sollte es möglich sein, diejenigen Konstrukte aus dem Text herauszufiltern und strukturell zu analysieren, die für die gegebene Grammatik relevant erscheinen. Dies läuft dann auf die Simulation einer überwiegend kursorischen Textrezeption hinaus, bei der sich die Aufmerksamkeit des Lesers nur auf ganz bestimmte Informationsbeiträge richtet, während der überwiegende Teil der sprachlichen Information als irrelevant ausgesondert werden kann. Anwendungsperspektiven hierfür ergeben sich vor allem im Bereich der Inhaltsextraktion aus Texten.

Der fundamentale Unterschied zu anderen etablierten Ansätzen im Bereich des partiellen Parsings (Hobbs et al., 1996; Abney, 1996) besteht hierbei vor allem darin, dass die angestrebte "Tiefe" der Analyse nicht a priori vorgegeben ist, sondern sich erst aus dem dynamischen Zusammenspiel von Grammatik und aktueller Datenlage ergibt.

Ein anderer Aspekt, der im Zusammenhang mit dem partiellen Parsing Aufmerksamkeit verdient, ist die Frage nach den grundsätzlichen Mechanismen der Sprachanalyse bei beschränkter Sprachkompetenz. Schließlich ist dies die typische Situation, mit der ein Fremdsprachlerner bei der Rezeption eines fremdsprachlichen Textes konfrontiert wird: Große Teile der Lexik und der Grammatik sind noch unbekannt und dennoch muss der Inhalt wenigstens grob erfasst werden. Dank einer starken semantischen Unterstützung wird der Leser jedoch in der Lage sein, zahlreiche der noch fehlenden Informationsbruchstücke aus dem Kontext zu erschließen, auch ohne die hierfür üblichen Nachschlagewerke hinzuziehen zu müssen. Da dieser Prozess zudem wiederholt ablaufen kann, sind geeignete Berechnungsmodelle gefragt, die hier – und über den engen Bereich des Fremdsprachenerwerbs hinaus – erklären können, warum sich ein Text nach mehrfachem Lesen besser verarbeiten lässt als noch beim ersten Versuch.

In letzter Konsequenz stellen sich analoge Fragen natürlich auch für den Erstsprachenerwerb, obschon hier der Weg hin zu ersten Simulationsexperimenten noch weitgehend im Dunkeln liegt. Grundsätzlich eröffnet ein Constraint-basierter Ansatz auch in dieser Hinsicht einige interessante Perspektiven auf das Problem des Erstsprachenerwerbs, insbesondere hinsichtlich der Frage nach dem Ursprung des sprachlichen Regelsystems.

Im Prinzip lassen sich auf der Grundlage der CDG Äußerungen ja auch mit einer leeren Grammatik verarbeiten, nur ist in einem derartigen Extremfall das Resultat völlig nutzlos, weil maximal ambig. Verständnisprobleme entstehen in einem Constraint-basierten Modell also nicht dadurch, dass der Parser scheitert, sondern vielmehr dadurch, dass die vorhandenen Constraints noch zu schwach sind, um Struktur in die vorliegende Äußerung zu bringen. In einem solchen Fall hinterlässt die Analyse eine verwirrende Vielzahl konkurrierender Interpretationen. Die Hinzunahme einzelner gewichteter Constraints kann diese übergroße Mehrdeutigkeit praktisch schlagartig auf eine einzige präferierte Lösung reduzieren, allerdings wird dies in aller Regel nur punktuell geschehen. Damit sollten vor allem die sehr einfach strukturier-

ten Äußerungen bereits verständlich werden, während komplizierte Relationengefüge noch Interpretationsschwierigkeiten bereiten. In ständigem Wechselspiel von Constraintverschärfung und -abschwächung kann sich so allmählich ein optimales Grammatiksystem herausbilden.

Eine der zentralen Fragen in diesem Zusammenhang ist aber sicherlich die nach der Erlernbarkeit der Constraintgewichte auf der Grundlage strukturell annotierter Korpusdaten. Erste Experimente auf der Basis eines evolutionären Ansatzes haben gezeigt,

- dass es auf diese Weise durchaus möglich ist, das Performanzniveau einer manuell optimierten Grammatik signifikant zu verbessern,
- dass auch bei einer zufälligen Wahl der initialen Gewichte für die verletzbaren Constraints das ursprüngliche Performanzniveau der manuell optimierten Grammatik schon nach knapp 200 Generationen wieder erreicht werden kann, sowie
- dass bei geeigneter Wahl des Selektionskriteriums auch die Verarbeitungsgeschwindigkeit der Analyse deutlich erhöht werden kann (Schröder et al., 2001).

Als ernstzunehmendes Problem bei der Modellierung sprachlicher Phänomene hat sich wiederholt die eingeschränkte Ausdruckskraft des Formalismus herausgestellt. Auch hier sind Weiterentwicklungen denkbar, die sich vor allem auf drei Problembereiche konzentrieren müssen:

- die punktuelle Einbeziehung höherstelliger Constraints in die Disambiguierung,
- die Schaffung zumindest eingeschränkter Möglichkeiten zur Repräsentation von Konstituenteigenschaften, die für die Behandlung von Koordinations- und Skopusphänomenen unverzichtbar erscheinen, sowie
- die Kopplung mit Wissensrepräsentationsformalissen, um so eine klare Trennung zwischen sprachlicher Form und bezeichnetem Referenzobjekt vornehmen zu können.

7 Zusammenfassung

Ausgehend von dem Problem eines sinnvollen Umgangs mit Inkonsistenz bei der strukturellen Analyse natürlicher Sprache wurde eine Systemarchitektur entwickelt, die auf der Idee der Disambiguierung von Dependenzstrukturen mit Hilfe von gewichteten Constraints beruht. Constraints können somit als verletzbar betrachtet werden, weshalb die Aufgabe der Analyse in der Ermittlung der jeweils optimalen Strukturbeschreibung unter Einbeziehung der gesamten verfügbaren Evidenz besteht. Da die Disambiguierung auf mehreren, miteinander graduell gekoppelten Repräsentationsebenen parallel abläuft, können vielfältige Informationsbeiträge in den Berechnungsprozess einfließen, wobei wegen des verletzbaren Charakters der Constraints auch relativ schwache Präferenzen und unsichere Information einen entscheidenden Beitrag zum Erfolg der Verarbeitung leisten können.

Dank des Zusammenspiels von Mehrebenenrepräsentation und verletzbaren Constraints besitzt das System einige aus kognitiver Sicht interessante prozedurale Eigenschaften. Sie betreffen vor allem das Verhalten der Sprachverarbeitung in den Grenzbereichen

- Analyse und Diagnose stark abweichender Eingabedaten,
- Sprachverstehen mit stark beschränkter Sprachkompetenz, sowie
- Sprachverstehen unter Zeitdruck

und rücken damit eine Reihe neuartiger Fragestellungen in den Aufmerksamkeitsfokus kognitiver Modellierung.

Danksagung. Mein besonderer Dank gilt Ingo Schröder, dessen Initiativen und Arbeiten alle Aspekte dieser Forschung entscheidend geprägt haben. Ich danke auch Kilian Foth, Johannes Heinecke, Jürgen Kunze, Bernhard Lahres, Andreas Nolda, Horia F. Pop, Michael Schulz und Niels Stockfleth für ihre wertvollen Beiträge zum besseren Verständnis des Constraint-basierten Parsings.

Literatur

- Abney, S. (1996) Partial parsing via finite-state cascades. In: *Proc. ESSLLI Workshop on Robust Parsing*
- Boddy, M. & Dean, T.L. (1994) Deliberation scheduling for problem solving in time-constrained environments. *Artificial Intelligence*, 67(2), 245–285
- Duchier, D. & Debusmann, R. (2001) Topological dependency trees: A constraint-based account of linear precedence. In: *Proceedings 39th/10th Annual Meeting of the ACL/EACL*, pp. 180–187, Toulouse, France
- Foth, K., Menzel, W. & Schröder, I. (2000) A transformation-based parsing technique with anytime properties. In: *Proc. 4th Int. Workshop on Parsing Technologies*, pp. 89–100, Trento, Italy
- Harper, M.P., Jamieson, L.H., Mitchell, C.D., Ying, G., Potisuk, S., Srinivasan, P.N., Chen, R., Zoltowski, C.B., McPheters, L.L., Pellom, B. & Helzerman, R.A. (1994) Integrating language models with speech recognition. In: *Proceedings of the AAIL-94 Workshop on the Integration of Natural Language and Speech Processing*, pp. 139–146
- Harper, M.P. & Helzermann, R.A. (1995) Extensions to constraint dependency parsing for spoken language processing. *Computer Speech and Language*, 9(3), 187–234
- Heinecke, J., Kunze, J., Menzel, W. & Schröder, I. (1998) Eliminative parsing with graded constraints. In *Proceedings 17th International Conference on Computational Linguistics, 36th Annual Meeting of the ACL, Coling-ACL '98*, pp. 526–530, Montreal, Canada
- Helzerman, R.A. & Harper, M.P. (1992) Log Time Parsing the MasPar MP-1. In: *Proceedings of the 6th International Conference on Parallel Processing*, pp. 209–217
- Hobbs, J.R., Appelt, D., Bear, J., Israel, D., Kameyama, M., Stickel, W. & Tyson, M. (1996) Fastus: A cascaded finite-state transducer for extracting information from natural-language text. In: Roche and Schabes, Eds., *Finite State Devices for Natural Language Processing*. Cambridge MA: MIT Press
- Karlisson, F., Voutilainen, A., Heikkilä, J. & Anttila, A., Eds. (1995) *Constraint Grammar – A Language-Independent System for Parsing Unrestricted Text*. Berlin, New York: Mouton de Gruyter
- Maruyama, H. (1990) Structural disambiguation with constraint propagation. In: *Proceedings 28th Annual Meeting of the ACL*, pp. 31–38
- Mellish, C. (1989) Some chart based techniques for parsing ill-formed input. In: *Proceedings 27th Annual Meeting of the ACL*, pp. 102–109, Vancouver
- Menzel, W. (1998) Sehen und Verstehen: Der Beitrag bildlicher Information zur robusten Sprachverarbeitung. In: *Tagungsband 28. Jahrestagung der Gesellschaft für Informatik, Magdeburg*, pp. 161–170, Berlin: Springer-Verlag
- Menzel, W. & Schröder, I. (1998a) Constraint-based diagnosis for intelligent language tutoring systems. In: *Proceedings IT & KNOWS, XV. IFIP World Computer Congress*, pp. 484–497, Vienna und Budapest
- Menzel, W. & Schröder, I. (1998b) Decision procedures for dependency parsing using graded constraints. In: S. Kahane & Alain Polguère, Eds., *Proc. Coling-ACL Workshop on Processing of Dependency-based Grammars*, pp. 78–87, Montreal, Canada
- Menzel, W. & Schröder, I. (1998c) Model-based diagnosis under structural uncertainty. In: *Proceedings 13th European Conference on Artificial Intelligence*, pp. 284–288, Brighton, UK
- Pollard, C. & Sag, I.A. (1994) *Head-Driven Phrase Structure Grammar*. Chicago: The University of Chicago Press
- Prince, A. & Smolensky, P. (1993) Optimality theory: Constraint interaction in generative grammar. Technical Report 2, Center for Cognitive Science, Rutgers University
- Schröder, I., Menzel, W., Foth, K. & Schulz, M. (2000) Modeling dependency grammar with restricted constraints. *Traitement Automatique des Langues (T.A.L.)*, 41(1), 97–126
- Schröder, I., Pop, H.F., Menzel, W. & Foth, K. (2001) Learning grammar weights using genetic algorithms. In: *Proceedings Euroconference Recent Advances in Natural Language Processing*, pp. 235–239, Tsigov Chark, Bulgaria
- Schulz, M. (2000) Parsen natürlicher Sprache mit gesteuerter lokaler Suche. Diplomarbeit, Fachbereich Informatik, Universität Hamburg
- Traxler, M.J., Pickering, M.J. & Clifton, C. (1998) Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558–592
- Trueswell, J.C., Tanenhaus, M.K. & Garnsey, S.M. (1994) Semantic influences on parsing: Use of thematic roles information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318
- Tsang, E. (1993) *Foundations of Constraint Satisfaction*. London: Academic Press