

LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung (GLDV)

LDV-Forum 15 (1998) 2

Zeitschrift für Computerlinguistik und Sprachtechnologie

GLDV-Journal for Computational Linguistics and Language Technology

Offizielles Organ der GLDV

Herausgeber

Prof. Dr. Gerhard Knorz;
Gesellschaft für Linguistische
Datenverarbeitung

Anschrift: Fachhochschule
Darmstadt, Fachbereich Infor-
mation und Dokumentation,
Haardtring 100, D-64289
Darmstadt; Tel: (06151)
16-8499; Fax: (06151)
16-8980; e-mail: knorz@
www.iud.fh-darmstadt.de

Redaktion

Gerhard Knorz

Wissenschaftlicher

Beirat

Prof. Dr. W. Hoepfner
(hoepfner@unidui.
uni-duisburg.de);
Prof. Dr. Gerhard Knorz; Prof.
Dr. Winfried Lenders
(lenders@uni-bonn.de);

Editorial

Es gibt Fehler und Vorkommnisse, deren Wahrnehmung man nicht dadurch entschärfen kann, dass man sie zu erklären versucht. Freud'sche Versprecher gehören dazu und gegenwärtig auch das Erscheinen des LDV-Forum, das man gar nicht mehr adäquat mit „verspätet“ attributieren kann. Immerhin liegt dieses Heft 1998/2 zu Beginn des Jahres 2000 (!!!) nun endlich vor, mit insgesamt 124 Seiten, einem respektablem fachlichen Teil und einem interessanten Beitrag aus Koblenz in der Rubrik „Aus der Lehre für die Lehre“. Kaum jemand wird es für vernünftig halten, eine Verspätung von mehr als 12 Monaten zu investieren, um ein Heft so überrandvoll mit Inhalten zu füllen, aber die im Sommer von mir erteilte Druckfreigabe ist aufgrund eines Mißverständnisses „hängengeblieben“.... (jetzt bin ich doch beinahe schon beim Erklären!)

Um den zeitlichen Verzug nicht in ähnlicher Weise als Dauerproblem durch die Folgejahre zu schleppen, wie etwa die Bundesrepublik ihr Haushaltsdefizit (an dem sie aber nun ja auch offensichtlich entschlossen ansetzen will), wird das Jahr 1999 durch ein Doppelheft abgedeckt werden, und zwar mit dem inhaltlichen Schwerpunkt „Maschinelle Übersetzung“. Dadurch, dass der gleichnamige GLDV-Arbeitskreis die inhaltliche Ausgestaltung des Themas übernommen hat, sollte es ein so gravierendes Problem wie mit dem

Prof. Dr. Ulrich Schmitz (e-mail: ulrich.schmitz@uni-essen.de)

Erscheinungsweise

2 Hefte im Jahr, halbjährlich zum 31. Mai und 31. Oktober. Preprints und redaktionelle Planungen sind laufend und aktuell unter der Adresse <http://www.iud.fh-darmstadt.de/iud/wwmeth/publ/ldvforum/menu1.htm> einsehbar.

Bezugsbedingungen

Für Mitglieder der GLDV ist der Bezugspreis des LDV-Forum im Jahresbeitrag mit eingeschlossen.

Jahresabonnements können zum Preis von DM 40,- (incl. Versand), Einzel Exemplare zum Preis von DM 20,- (zuzügl. Versandkosten) bestellt werden:

LDV-Forum, c/o IKS, Poppeldorfer Allee 47, 53115 Bonn

Fachbeiträge

Unaufgefordert eingesandte Fachbeiträge werden vor Veröffentlichung von mindestens 2 ReferentInnen begutachtet. Manuskripte sollten deshalb möglichst frühzeitig eingereicht werden und bei Annahme zur Veröffentlichung in jedem Fall elektronisch und zusätzlich auf Papier übermittelt werden. Artikel werden am besten in Microsoft Word für Windows® oder Word Perfect® für Windows erstellt. Eine Dokumentvorlage für Word für Windows kann unter der Adresse <ftp://www.iud.fh-darmstadt.de/>

Heft 98/2 nicht geben. Mit dem Erscheinen ist im Frühjahr 2000 zu rechnen.

Der Gedanke, dass die Arbeitskreise der GLDV eine aktive Rolle im Hinblick auf das LDV-Forum übernehmen, ist keineswegs neu, aber gegenwärtig durchaus vielversprechend. Die erfolgreiche Jahrestagung 1999 zum Thema „Multilinguale Corpora: Kodierung, Strukturierung, Analyse“ kann in diesem Sinn als eine aus einem Arbeitskreis hervorgegangene Aktivität angesehen werden. Und der Erfolg der Tagung, gerade was die Anzahl der Teilnehmer und der eingereichten Vorträge betrifft, kann ebenfalls sicher zumindest teilweise der Strategie zugeschrieben werden, die Arbeitskreise substantziell in das Verfahren einzubinden.

Keineswegs sollen allerdings die Arbeitskreise Sie, liebe Leserin und lieber Leser, von der Anforderung befreien, selbst aktiv zu werden. Die Aufforderung, sich zu beteiligen, lässt keine der gebotenen Möglichkeiten aus: Bieten Sie an, eine Rezension zu schreiben! Berichten Sie über eine Tagung! Verfassen Sie eine Nachricht! Reichen Sie einen Fachbeitrag ein! Oder äußern sie sich konstruktiv kritisch zu aktuellen Entwicklungen und zu den Inhalten des LDV-Forum. Gerade zu Letzterem werden sie von mir und den zwei Autoren Roland Hausser und Bernhard Schröder explizit aufgefordert, die in kontroversen Beiträgen die Rolle der Logik für eine Definition der Semantik natürlicher Sprachen diskutieren. Lassen Sie sich also anregen und zur Meinungsäußerung provozieren!

Das wünsche ich Ihnen und mir

Ihr

Gerhard Knorz

itud/wwmeth/publ/ldvforum/
ldvforum.dot heruntergeladen
werden. Sie enthält die wich-
tigsten Styles.

Rubriken

Die namentlich gezeichneten
Beiträge geben ausschließlich die
Meinung der AutorInnen wieder.
Einreichungen sind – wie bei
Fachbeiträgen – an den
Herausgeber zu übermitteln.

Redaktionsschluß

Für Heft **15** (1998) 2:
15. August 1998

Druck und Vertrieb

IKS, Poppelsdorfer Allee 47,
53115 Bonn

Satz

Kurt Thomas, IKP, Universität
Bonn

Auflage

400 Exemplare

Anzeigen

Preisliste und Informationen:
IKS e. V.

Bankverbindung

IKS e. V.: PGA Köln (BLZ
370 100 50), Konto 385647-505

GLDV-Anschrift

Prof. Dr. R. Hausser
Universität Erlangen-Nürnberg
Abteilung für Computerlinguistik
Bismarkstraße 12
D-91054 Erlangen
e-mail: rrrh@linguistik.uni-
erlangen.de

PS: Jetzt habe ich quasi mit einer versteckten
Entschuldigung begonnen. Lassen Sie mich mit
einer offenen Entschuldigung schließen: Im letzten
Heft (LDV-Forum 98/1) ist eine Rezension er-
schienen, die sich auf zwei Monografien in
unterschiedlichen Reihen (*Lexicographica Series
Maior* und *Sprache und Information*) bei
Niemeyer bezieht. Die Kürzel LSM und SI, mittels
derer auf die beiden unterschiedlichen Werke ver-
wiesen wurde, waren aus unerfindlichen Gründen
systematisch vertauscht. Sie finden den Fehler auf
Seite 111 und 114 korrigiert, und ich bitte um
Entschuldigung!

Deutscher Wortschatz im Internet

U. Quasthoff

Institut für Informatik

Universität Leipzig

04109 Leipzig

quasthoff@informatik.uni-leipzig.de

Entstehung der Sammlung

Die mittlerweile vorliegende Datensammlung, die momentan sicher eine der umfangreichsten frei zugänglichen Datensammlungen zur deutschen Sprache ist, entstand Anfang der 90er Jahre mit einer Wortliste mit sporadisch vorhandenen Angaben zu Grammatik und Sachgebiet. Diese war angelegt worden, da damals ein allgemeiner Mangel an frei verfügbaren maschinenlesbaren Daten zum deutschen Wortschatz bestand. Anliegen der damals begonnenen Sammlung war und ist, verfügbare Daten zunächst zu sammeln und sie (sobald wie möglich) zu nutzen, um fehlende Angaben zu ergänzen und eventuelle Fehler zu beseitigen. Dazu bietet sich speziell die Redundanz an, die in einer so großen Sammlung zu finden ist. Hier erweist sich auch die Sammlung von Vollformen als günstig, da bei einer späteren Reduktion auf Grundformen die Menge der flektierten Formen zu einer Grundform die korrekte Erkennung dieser Grundform erleichtert. Ebenso läßt sich dann das korrekte Flexionsschema leichter erkennen bzw. überprüfen. Nachdem lange mit verschiedenen Wortlisten gearbeitet worden war, wurde 1994 auf ein relationales Datenbanksystem umgestellt, um eine einfachere Datenverwaltung und einheitliche Zugriffsmöglichkeiten zu bekommen.

Schnell stellte sich heraus, daß bei einigen so gesammelten Wörtern für den Betrachter im nachhinein nicht festgestellt werden kann, ob es sich um ein fehlerhaft geschriebenes Wort, einen Eigennamen oder vielleicht einen ihm unbekanntem Fachbegriff handelt. Um solche Fragen wenigstens in der überwiegenden Mehrzahl der Fälle klären zu können, wurde ab ca. 1996 zusätzlich für jede neue Wortform ein Belegsatz gesammelt. Dazu wurde ein eigener Satzsegmentierer entwickelt (s. u.).

Das Vorhandensein der Beispielsätze wiederum ermöglicht Untersuchungen von Kollokationen. Erste Versuche sahen sehr erfolgversprechend aus, allerdings war das verwendete Material an Beispielsätzen nicht repräsentativ, da immer nur Sätze mit neuen Wörtern gesammelt wurden. Deshalb wurde 1998 dazu überge-

gangen, möglichst alle Sätze zu sammeln. Ein Volltext-Index ermöglicht effektives Suchen in diesen Sätzen.

Seit März 1998 ist die Datensammlung des Projekts Deutscher Wortschatz zu großen Teilen im Internet unter <http://wortschatz.uni-leipzig.de> zugänglich. Die Nutzerzahlen zeigen wachsendes Interesse, Anfang 1999 liegen die Zugriffszahlen bei 300–400 pro Tag.

Entsprechend ihrer Entstehung ist es nicht Anliegen der Sammlung, auf irgendeine Weise normierend auf den Sprachgebrauch einzuwirken. Zwar werden offensichtliche Fehler (z. B. orthographische Fehler oder Segmentierungsfehler, d. h. hauptsächlich Leerzeichen mitten im Wort oder fehlende Leerzeichen zwischen den Wörtern), wenn sie erkannt werden, als solche gekennzeichnet und die entsprechenden Wörter sind dann nicht mehr sichtbar, aber sonst wird in der Sammlung die deutsche Sprache dargestellt, wie sie in der Gegenwart maschinenlesbar zugänglich ist.

Vorliegendes Datenmaterial

Wortformen

Momentan liegen in der Datenbank ca. 5 Millionen Wortformen vor. Neben der absoluten Häufigkeit (d. h. der Anzahl, wie oft eine Wortform im Text gezählt wurde) ist eine Häufigkeitsklasse angegeben, welche diese Häufigkeit relativ zur absoluten Häufigkeit des häufigsten Wortes *der* mißt. Diese Häufigkeitsklasse bleibt bei der Auswertung weiterer Quellen (theoretisch, d. h. bei gleicher Zusammensetzung des Gesamtkorpus) unverändert, während sich die absoluten Häufigkeiten natürlich ändern.

Die Häufigkeitsklasse eines Wortes berechnet sich folgendermaßen: Sei $n(\textit{wort})$ die absolute Häufigkeit des Wortes *wort*, $n(\textit{'der'})$ die absolute Häufigkeit des Wortes *der*. Die Häufigkeitsklasse von *wort* ist dann definiert als

$$h(\textit{wort}) = \log_2 (n(\textit{'der'})/n(\textit{wort}))$$

und wird auf die nächstgelegene ganze Zahl gerundet.

Die folgende Tabelle enthält die häufigsten 20 Wörter und ihre Häufigkeitsklassen. Die absoluten Zahlen stammen vom Juli 1998.

Wort	Anzahl	Häufigkeitsklasse
der	7507541	0
die	6836196	0
und	4965268	1
in	3850947	1
den	2758375	1
von	2232689	2
zu	2081975	2
das	1889278	2
mit	1843993	2
sich	1751629	2
nicht	1680592	2
des	1625102	2
ist	1603975	2
auf	1595642	2
für	1584354	2
im	1535076	2
dem	1463015	2
ein	1323984	2
eine	1229663	3
als	1081635	3

Niederfrequente Wörter haben folgende Häufigkeitsklassen:

Anzahl	Häufigkeitsklasse
12– ...	18
6–11	19
3–5	20
1–2	21

Grammatikangaben

Die Grammatikangaben unterscheiden sich für Grundformen und für flektierte Formen. Für flektierte Formen erfolgt nur ein Verweis auf die Grundform, während für die Grundform Wortklasse und Flexionstyp angegeben sind. Damit lassen sich wieder sämtliche flektierten Formen generieren, und für eine bestimmte flektierte Form lassen sich alle Ableitungen ermitteln. Weiterhin sind Eigennamen als solche markiert. Beispiele für die Verwendung als Eigennamen werden wie weiter unten beschrieben explizit angegeben.

Für ein Wort können mehrere Grammatikangaben vorhanden sein, z.B. verschiedenen Flexionsmuster für *Bank*.

Weiterhin enthält die Datenbank eine morphologische Analyse des Wortes. Diese baut auf dem Vorgehen von [WOT] auf und ermittelt Morphemgrenzen auf der Grundlage von Übergangswahrscheinlichkeiten, die aufbauend auf einer Ausgangsmenge von uns ermittelt wurden.

Sachgebiete

Die Klassifikation der Sachgebiete folgt im wesentlichen der Klassifikation der Schlagwort-Norm-Datei (SWD) der Deutschen Bibliothek, Frankfurt ([NOR]). Diese Sachgebietstitel sind hierarchisch angeordnet, so daß der *Bereich 27. Medizin* weiter untergliedert ist und z.B. die feineren Bereiche *27.9 Innere Medizin* und *27.9b Pulmonologie* enthält. Diese Struktur ermöglicht auch die Beibehaltung des Sachgebiets *Medizin* für Begriffe, deren genauere Einordnung uns momentan unbekannt ist.

Sachgebietsangaben aus anderen Quellen wurden eingearbeitet, so daß das gesamte Klassifikationsschema momentan ca. 1700 Sachgebietstitel enthält.

Beschreibung

Für einige Begriffe liegen kurze Beschreibungen vor, die keiner einheitlichen Form folgen.

Thesaurusrelationen

Häufig sind außer Sachgebietsangaben noch mehr inhaltliche Zusammenhänge zwischen Begriffen angegeben. Typischerweise findet man in Wörterbüchern zum einen Synonyme und gelegentlich Antonyme, zum anderen Hyponyme und Hyperonyme. In vielen Quellen ist für verschiedene Zwecke der Hinweis *vgl.* verwendet. Momentan sind diese Daten noch nicht befriedigend aufgearbeitet, vorgesehen ist aber eine schrittweise Erweiterung der Menge der Relationen. Zunächst sollen (ähnlich wie in WordNet, [FEL]) verschiedene Meronym-Relationen hinzugenommen werden, anschließend weitere Relationen, wie sie z.B. in der Meaning-Text-Theory ([STE]) vorgeschlagen werden. Kandidaten für Begriffspaare in solchen Relationen werden mit den weiter unten beschriebenen Kollokationen ermittelt.

Sätze

Die Beispielsätze erfüllen einen mehrfachen Zweck. Zunächst dienen sie im klassischen Sinn als Belegstellen. Weiterhin sind sie in vielen Fällen geeignet, zu entscheiden, ob es sich bei einem bestimmten Wort, welches nicht zum passiven Wortschatz des Benutzers gehört, um einen Fachbegriff, einen Eigennamen oder nur einen Schreibfehler handelt. Interessant ist, daß bei wirklich seltenen Wörtern der Autor damit rechnet, daß der Leser sie nicht kennt und deshalb ein Beispielsatz eine definitonische Erklärung enthält. Hier sei ein solcher Beispielsatz für Hunni angegeben: *Ein Hunderter ist immerhin noch ein Hunderter, wird aber, um ihn klein zu machen, seit ein paar Jahren Hunni genannt. (Quelle: Frankf. Rundschau 1992).*

Um nur möglichst wenig fehlerbehaftete Einträge zu erhalten, wurden von vornherein nur solche Quellen benutzt, die bereits selbst einen hohen Qualitätsstandard besitzen. Im wesentlichen wurden Materialien ausgewertet, die auf CD-ROM vorliegen und uns von den entsprechenden Verlagen für das Projekt zur Verfügung gestellt wurden. Durch die Zerlegung in einzelne Sätze und die unzusammenhängende Speicherung in der Datenbank konnte sichergestellt werden, daß die Originaltexte aus der Datenbank nicht rekonstruierbar sind.

Momentan stehen ca. 10 Millionen Beispielsätze zur Verfügung, die über einen Volltextindex erschlossen sind.

Kollokationen

Als Kollokationen wollen wir hier Paare von Wortformen bezeichnen, die signifikant häufig gemeinsam in einem Satz auftreten, bezogen auf die Menge der vorliegenden Beispielsätze. Unabhängig vom konkreten Verfahren werden solche Signifikanzen statistisch berechnet. Ein solches Verfahren gilt als gut, wenn es zu einem gegebenen Begriff möglichst viele Begriffe aus dem „semantischen Umfeld“ liefert und nur wenige andere, „unpassende“ Begriffe.

Das hier angewendete Signifikanz-Maß berechnet sich folgendermaßen:

Sei a die Anzahl der Sätze mit Wort A , b die Anzahl der Sätze mit Wort B , n die Gesamtanzahl aller Sätze und k die Anzahl der Sätze, welche die Wörter A und B enthalten. Wir setzen $x=ab/n$, und definieren:

Gilt $(k+1)/x < 0.1$ (dies ist der typische Fall), so setzen wir

$$\text{sig}(A,B) = x - k \cdot \log x + \log k!$$

Diese Formel leitet sich aus einem Ansatz her, der eine Poissonsche Verteilung nutzt und damit die Wahrscheinlichkeit des Ereignisses bestimmt, daß bei statistischer Unabhängigkeit die Wörter A und B insgesamt (mindestens) k -mal gemeinsam auftreten. In [LEM] sind einige weitere Verfahren zur Ermittlung von Signifikanzen vorgestellt.

Im Gegensatz zu anderen Arbeiten, bei denen Kollokationen eher exemplarisch für einige wenige Wörter ermittelt werden, sind hier die Kollokationen für alle Wörter ab einer gewissen Mindestfrequenz berechnet und abrufbar.

Außer Kollokationen auf Satzebene wurden weiterhin alle signifikanten rechten und linken Nachbarn ermittelt. Beispielsweise treten als signifikante rechte Nachbarn zum Wort *Insel* fast sämtliche Namen von Inseln auf, geordnet in einer Reihenfolge, die sich als Mischung von Bekanntheit und Nähe interpretieren läßt: *Sachalin, Rügen, Okinawa, Mindanao, Honshu, Hainan, Usedom, Shikoku, Java, Borneo, Texel, Kyushu, Lesbos, Sylt, Hokkaido, Skye, Mainau, Luzon, Sumatra, Borkum, Batam, Chios, Korfu, Elba, Tasmanien, Amrum, Rhodos, Herrenchiemsee, Cheju, Helgoland, Euböa, Flores, Sulawesi, Juist, Negros, Brac, Taipa, Mauritius, Oahu.*

Umfang der Daten

Die folgende Tabelle gibt einen Überblick über die Anzahl der verschiedenen Angaben.

Art der Angabe	Anzahl der Einträge
Wortliste	4872716
Beispiele	7441775
Grammatik	2772369
Pragmatik	33948
Beschreibung	135914
Morphologie	3461877
Sachgebiet	1255813
Relationen	449619
Kollokationen (im Satz)	3154745
Kollokationen (Nachbarn)	709076
Volltext-Index	103730292

Möglichkeiten zur Suche über das Internet

Um die Nutzung der Datenbank von verschiedenen Plattformen aus und unabhängig vom Standort zu ermöglichen, wurde ein Zugang über das Internet ermöglicht. Verschiedene Abfragen ermöglichen unterschiedlich komplexe Aktionen.

Einfache Abfrage

Die einfache Anfrage ermöglicht das alphabetische Nachschlagen nach einzelnen Wörtern. Weiterhin können die üblichen Wildcards ? und * als Platzhalter für einen oder beliebig viele Zeichen benutzt werden. Die Benutzung dieser Wildcards ist an jeder Position möglich. Beispielsweise liefert die Suchanfrage nach *?i?i?i?i* den Eintrag *Minibikini*. Es ist allerdings zu beachten, daß eine Suche mit einer Wildcard am Anfang des Suchbegriffs eine sequentielle Suche durch die Wortliste erfordert, die mehrere Minuten dauern kann. Wird ein gesuchtes Wort (auf eine Anfrage ohne

wortschatz lexikon [Zum Inhaltsverzeichnis](#)

Das Wortschatz-Lexikon stellt sich zum Ziel, den deutschen Wortschatz so vollständig wie möglich zur Verfügung zu stellen. Grundlage ist der gegenwärtig in maschinenlesbaren Quellen zur Verfügung stehende Text. Momentan enthält die Datenbank fast 5 Millionen Wortformen.

Gesuchtes Wort: **nachschiagen**

Zu vielen Wörtern erhalten Sie zusätzliche Angaben:

- Häufigkeitsklasse,
- Grammatikangaben (Verweis auf die entsprechende Grundform bzw. Flexionsklasse),
- Beispielsatz (Der erste Satz, in dem uns dieses Wort begegnet ist),
- Sachgebiets-, Pragmatik- und Morphologieangaben.

Verwenden Sie das Symbol * am Ende, um alle Wörter mit einem bestimmten Wortbeginn zu erhalten (zum Beispiel liefert hex* als Ergebnis u. a. Hexadezimalcode, hexenhaft, Hexylamin). Umlaute können auf englischen Tastaturen auch durch vorangestellte Anführungszeichen (´) eingegeben werden. (z. Bsp. ´a für ä, ´s für ß oder ´e für é). Groß- / Kleinschreibung wird bei der Eingabe nicht beachtet.

Da viele Angaben automatisch erzeugt wurden und manche Quellen fehlerbehaftet sind, sind auch in unseren Daten noch viele Fehler. Wir arbeiten dran, aber auch Ihre Hilfe ist uns willkommen.

e-mail an das Wortschatz-Team

Wildcards hin) nicht gefunden, so startet eine Suchhilfe und sucht nach Wörtern, die sich von der Eingabe auf genau eine der folgenden Arten unterscheiden darf:

- Austauschen eines Buchstabens gegen einen anderen,
- Hinzufügen eines Buchstabens,
- Weglassen eines Buchstabens,
- Vertauschen zweier benachbarter Buchstaben,
- Ersetzungen aus der folgenden Liste: ä↔ae, ö↔oe, ü↔ue, ß↔ss, f↔ph.

Besteht das Suchergebnis aus mehreren Wörtern, so wird zunächst eine Auswahlliste präsentiert, aus der ein bestimmtes Stichwort ausgewählt werden kann.

Gibt man als Suchbegriff beispielsweise *Gärtens* ein, so erhält man folgende Korrekturvorschläge: *Gärten*, *Gärens*, *Gättens*, *Gartens* und *Märtens*. Als Ergebnis werden die vorhandenen Angaben zu dem ausgewählten Wort angegeben, beispielhaft seien hier zwei Einträge für einen Eigennamen und ein flektiertes Verb angegeben:

Wort: Chemnitz:

Häufigkeitsklasse: 12 (Anzahl: 1207)

Beschreibung:

Stadt in Deutschland (über 250000 E)

Stadt in Sachsen

Sachgebiet:

Nachname

STADT

Grammatikangaben:

Wortart: Eigenname

Beispiel:

Gegen die Konsumgenossenschaft Chemnitz ist auf Antrag der Gewerkschaft HBV das Konkursverfahren eröffnet worden. (Quelle: *FAZ 1994*)

Karin Chemnitz (Quelle: *Telefonbuch*)

Kollokationen im Satz: Zwickau, Dresden, Mosel, Leipzig, Sachsen, FC, sächsischen, Berlin, Riedel, Weißwasser, Plauen, Drechsler

Wort: ermüdet:

Häufigkeitsklasse: 16 (Anzahl: 113)

Morphologie: er|müd|et (+er=müd%et)

Grammatikangaben: Wortart: Verb

Stammform: ermüden

Relationen zu anderen Wörtern:

Vergleiche: matt, schlaff, schlapp

Synonyme: kaputt, kraftlos, erschöpft

für Synonyme siehe: müde, bettreif,

ruhebedürftig, schlafbedürftig,

schläfrig, schlaftrunken, übermüde

Beispiel:

Der Sachbuchautor Miloslav Stingl, Mitarbeiter des Mitteldeutschen Rundfunks, ist ein Kenner seiner Heimatstadt, schreibt aber einen korrekt-farblosen Stil, der rasch ermüdet.

(Quelle: *FAZ 1994*)

Graphische Darstellung der Kollokationsmengen

Die zu einem Wort angegebenen typischen Kollokationen geben typischerweise verschiedene Zusammenhänge wieder. Unter den semantischen Zusammenhängen lassen sich folgende identifizieren:

- inhaltlich ähnliche Begriffe, sog. Kohyponyme (Schafe und Rinder zu Schweine);
- Kombinationen von Namen (Bill zu Clinton, Romeo zu Julia);
- Verb-Nomen-Kombinationen (beißen – Hund – Herrchen).

In einigen Fällen liefern die Kollokationsmengen auch Wörter in einer sprachlich ähnlichen Funktion, Beispiele dafür sind weiter unten angegeben.

Um die Struktur zwischen den einzelnen Wörtern in einer Kollokationsmenge darzustellen, wurde eine Visualisierung als Graph gewählt, die auf dem Prinzip des Simulated Annealing beruht (JEAD, DAV). Dazu werden die Begriffe als Knoten des Graphen dargestellt. Neben dem Ausgangsbegriff werden all die Begriffe aus der Kollokationsmenge mit aufgenommen, zwischen denen untereinander mindestens noch eine Verbindung besteht. D.h. zwei Begriffe A und B sind durch eine Kante verbunden, falls $\text{sig}(A,B)$ einen Schwellwert übersteigt. Die Stärke der Verbindungslinie entspricht dabei der Größe dieses Signifikanzwertes, zu schwache Verbindungen werden der Übersichtlichkeit wegen nicht eingezeichnet.

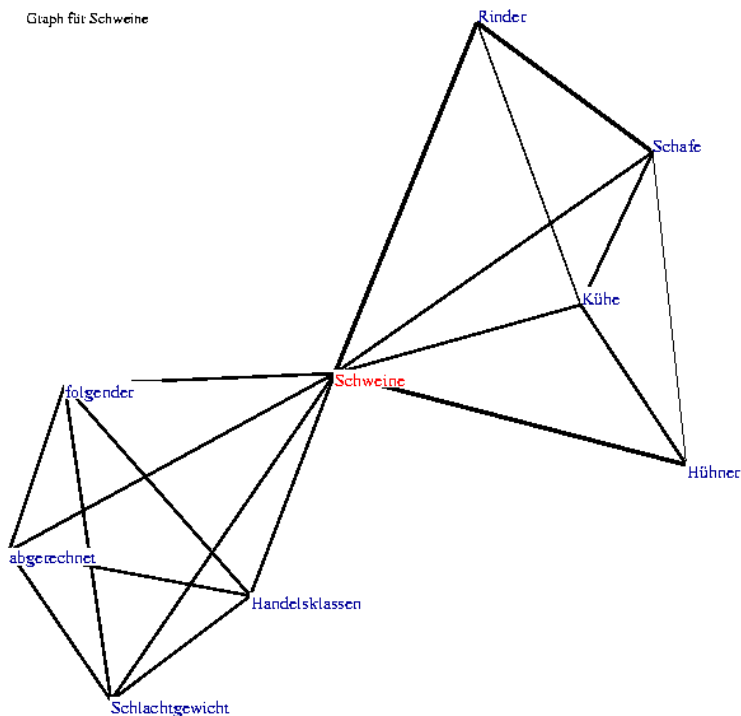
Die Anordnung der Punkte erfolgt nach dem folgenden Schema: Zunächst werden die Punkte zufällig verteilt. Zwischen je zwei Punkten herrscht eine konstante Abstoßungskraft, zusätzlich herrscht zwischen zwei Punkten eine Anziehungskraft proportional $\text{sig}(A,B)$. Ermöglicht man nun vorsichtig Bewegungen der Punkte entsprechend dieser Kräfte, so ergibt sich nach einiger Zeit eine stabile Anordnung, welche die inhaltlichen Zusammenhänge relativ gut zeigt.

Diese Kollokationsgraphen sind relativ aussagekräftig, wenn die absolute Häufigkeit des Ausgangsbegriffes eine Mindestgröße von ca. 500 übersteigt. Dies trifft gegenwärtig für ca. 26.000 Begriffe zu.

Im folgenden sind vier solche Graphen kurz beschrieben.

Kollokationen für *Schweine*

Graph für Schweine

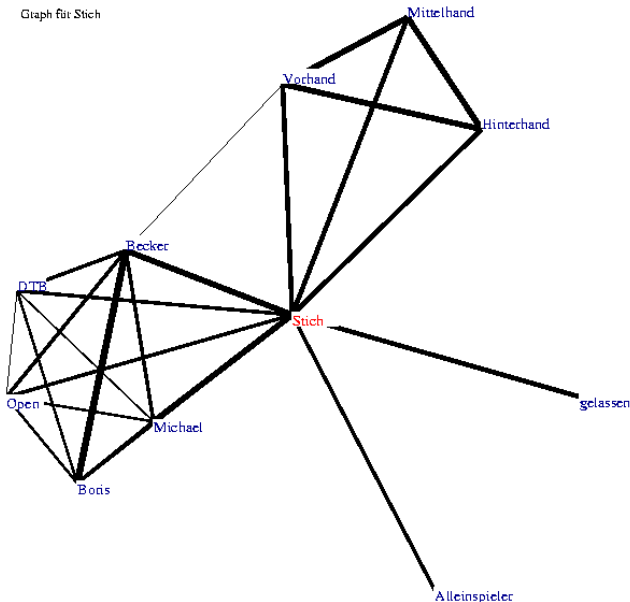


Der Graph für *Schweine* zerfällt sehr schön in zwei Teile. Im rechten Teil finden wir eine Zusammenstellung von anderen Nutztieren (ebenfalls im Plural), die untereinander stark zusammenhängen: Rinder, Hühner, Kühe, Schafe. Im linken Teil finden wir Wörter, die typische Sachverhalte zum Thema Schweine kennzeichnen, z.B.: Schweine werden nach Schlachtgewicht in Handelsklassen eingeteilt und abgerechnet.

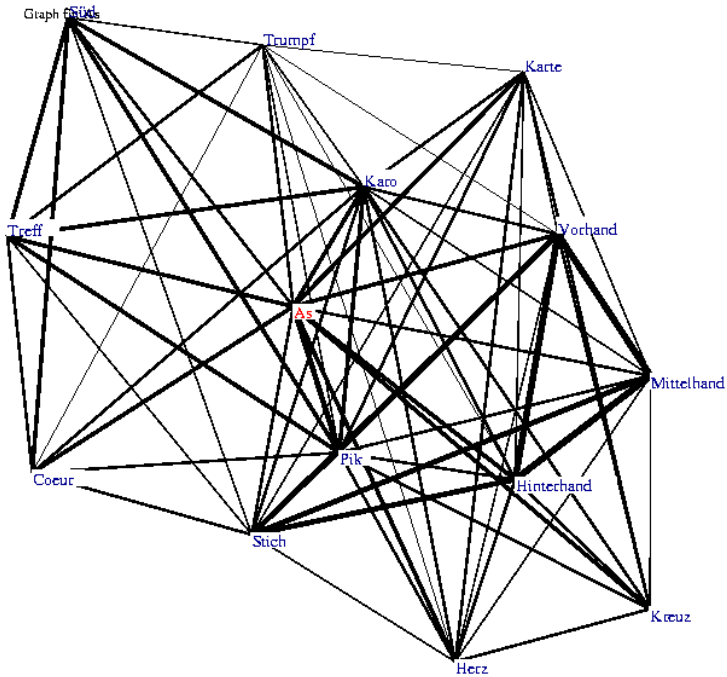
Kollokationen für *Stich*

Beim Graphen von *Stich* fallen uns zwei große Gruppen auf, die untereinander stark zusammenhängen: Einmal geht es um Tennis (Michael Stich usw.), einmal um das Skatspiel, bei dem die drei Mitspieler Vorhand, Mittelhand und Hinterhand heißen. Die zwei Teile entsprechen also verschiedenen Bedeutungen des Wortes *Stich* (einmal als Eigenname, speziell für den Tennisspieler Michael Stich, einmal um den Stich beim Kartenspiel). Interessant ist, daß diese beiden Gruppen auch noch über eine dünne Verbindung zwischen *Becker* und *Vorhand* zusammenhängen; diese Verbindung entsteht durch das starke Vorhandspiel von Boris Becker beim Tennis! Wir haben also auch bei *Vorhand* zwei Bedeutungen!

Die isolierte Verbindung von *Stich* und *gelassen* resultiert aus der Redewendung *im Stich gelassen*.

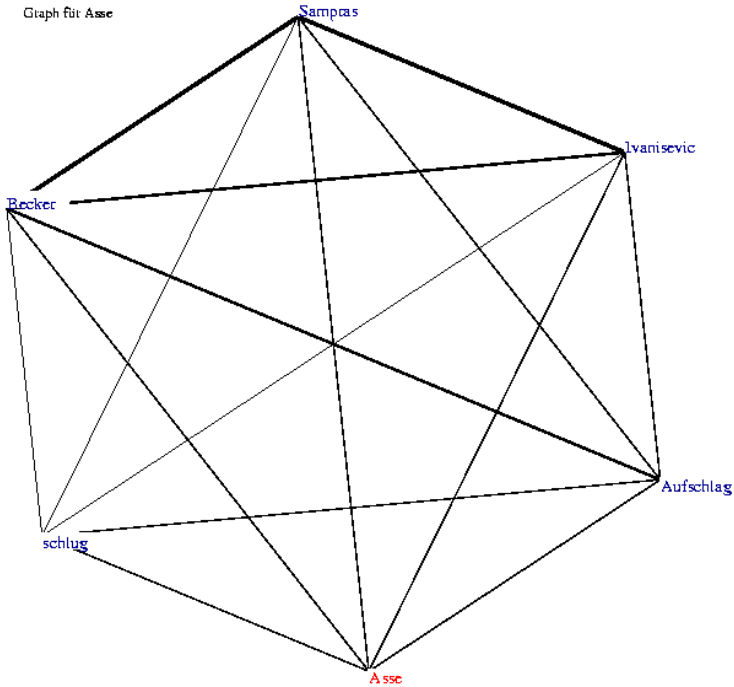


Kollokationen für *As*



Der Graph für *As* zeigt ausschließlich Begriffe aus dem Bereich Kartenspiel, und zwar auch nur Bezeichnungen für die Farben und die Mitspieler.

Kollokationen für *Asse*



Hier finden wir die Begriffe aus dem Bereich Tennis, die wir bei *As* vergeblich gesucht haben. Die Erklärung besteht einfach darin, daß in einem Bericht über Tennis offensichtlich nur eine große Anzahl von Assen erwähnenswert ist.

Diese sehr verschiedenen semantischen Umfelder für verschiedene flektierte Formen einer Grundform würden sehr verwischen, wenn wir sofort bei der Auswertung der Texte jedes Wort auf seine Grundform zurückführen würden. Erst

die Häufigkeitsangaben für die verschiedenen flektierten Formen ermöglichen das Erkennen der obigen Strukturen.

Spezielle Anfragen

Außer Anfragen, die wie oben von einem bestimmten Wort ausgehen, sind auch andere Anfragen realisierbar. Um den Nutzer nicht mit der Datenbanksprache SQL zu konfrontieren, wurden einige Anfragen vorbereitet, für die spezielle Eingabemasken vorliegen.

Folgende Anfragen werden momentan unterstützt:

- Suche nach Wörtern bestimmter, fester Länge;
- Suche nach Wörtern aus einem bestimmten Sachgebiet;
- Suche nach Wörtern mit mehreren Sachgebietsangaben;
- Suche nach frei wählbaren Zeichenketten in den Beispielsätzen;
- Suche nach frei wählbaren Zeichenketten in den Beschreibungen;
- Angabe der häufigsten Sachgebietsangaben, Grammatikangaben usw.;
- Durchschnitt zweier Kollokationsmengen.

Das letzte Beispiel ist folgendermaßen anwendbar: Beispielsweise enthält die Kollokationsmenge zu *Australien* neben anderen Ländern und australischen Persönlichkeiten auch die wichtigen australischen Städte. Weiterhin enthält die Kollokationsliste zu *Hauptstadt* die Namen vieler Hauptstädte. Der Durchschnitt beider Mengen (geordnet nach Signifikanz) enthält als erste Stadt tatsächlich *Canberra*, die australische Hauptstadt. Ähnlich kann man als Durchschnitt der Kollokationsmengen von *Rinder* und *Schweine* die Kohyponyme *Schafe*, *Kühe*, *Ziegen*, *Pferde*, *Kälber* in dieser Reihenfolge ermitteln, unterbrochen nur vom Begriff BSE.

Besteht die Ergebnismenge aus mehreren Wörtern, so sind verschiedene Anzeigemöglichkeiten auswählbar:

- alphabetisch sortiert
- rückläufig alphabetisch sortiert
- sortiert nach Häufigkeit
- sortiert nach Wortlänge
- Anzeige nur von Grundformen

Weitere Anfragen können bei Bedarf hinzugefügt werden.

Tools

Der Satzsegmentierer

Um Texte in einzelne Sätze zerlegen zu können, wurde ein eigener Satzsegmentierer entwickelt. Er hat folgende Eigenschaften:

- Eine sehr umfangreiche Liste der Abkürzungen, die mit Punkt beendet werden (als *Prof.*, aber nicht *BRD*), um solche Punkte nicht mit Satzenden zu verwechseln.
- Zusätzliche Zeichen sind als Satzende definierbar, z.B. kann eingestellt werden, ob CRLF als Satzende gewertet werden soll.
- Globale Ersetzungen für Zeichen sind möglich, dies ist nötig bei Sonderzeichen aus manchen Quellen.
- Die Zeichen, aus denen ein Wort bestehen darf, sind frei wählbar. Voreingestellt sind die Buchstaben a–z, ä, ö, ü, ß, é und das Zeichen - in der Wortmitte. Nicht als Wörter akzeptiert werden so *50%ig* und *Privileg(ium)*.
- Ebenso frei wählbar ist die Liste der Whitespace-Zeichen, also der Zeichen, die Wortzwischenräume kennzeichnen. Dazu kann man außer Leerzeichen, Tabulator, CRLF, und den Satzzeichen beispielsweise noch Klammern wählen. Dies sollte aber für jeden Text einzeln entschieden werden, da beispielsweise die Hinzunahme von Klammern aus *Privileg(ium)* fälschlicherweise das Wort *ium* separiert.

Weiterhin kann man dem Satzsegmentierer eine Liste bekannter Wörter mitgeben, dann erhält man als Ergebnis eine Liste der neuen Wörter mit jeweils einem Beispielsatz.

Der Satzsegmentierer steht frei zur Verfügung unter <http://wortschatz.informatik.uni-leipzig.de/Download.html>.

Umgang mit fehlerhaften Einträgen

Das Vorgehen, alle Einträge in der Datenbank per Hand zu kontrollieren und die falschen Einträge zu löschen, ist aus mindestens zwei Gründen nicht gangbar: Zunächst ist es aus Kapazitätsgründen unmöglich, alle Einträge per Hand zu bearbeiten. Weiterhin wird ein fehlerhaft geschriebenes Wort wieder automatisch aufgenommen, sobald es nach dem Löschen wieder in einer neuen Quelle gefun-

den wird. Deshalb werden in der Datenbank keine Einträge gelöscht, sondern mit Qualitätsangaben bewertet.

Die Wortformen können mit positiven und negativen Qualitätsangaben versehen werden. Positive Qualitätsangaben sind:

- Manuelle Klassifikation als korrekt.
- Autorisierung durch eine oder besser mehrere zuverlässige Quellen. Solche Quellen können sein:
- Wortlisten aus anderen Computerlexikographie-Projekten,
- Wortlisten aus Wörterbüchern,
- Wörter mit sehr großer Häufigkeit.
- Konsistenz zusätzlicher Angaben, z. B.:
- Eine vorhandene flektierte Form entspricht dem Flexionsmuster der entsprechenden Grundform.
- Das Flexionsmuster einer vorhandenen flektierten Form entspricht der morphologischen Zerlegung.

Negative Qualitätsangaben sind:

- Manuelle Klassifikation als fehlerhaft.
- Inkonsistenz zusätzlicher Angaben, z. B.
- Nicht passend zum angegebenen Flexionsmuster
- Schwierigkeiten bei der morphologischen Analyse
- Phonologisch inakzeptabel: Das Wort ist nicht aussprechbar.

Diese Kriterien ermöglichen eine teilweise automatische Überprüfung von Lexikoneinträgen.

Angabengewinnung und Fehlerbearbeitung

Angaben aus Kollokationslisten

Kollokationslisten bilden häufig einen sehr guten Ausgangspunkt, um eine Liste von Kandidaten für Wörter mit einer bestimmten Eigenschaft zu bekommen. Beispielsweise enthalten die ausgewerteten Texte gelegentlich englische Zitate, Titel von Musikstücken usw. Die darin enthaltenen englischen Wörter werden automatisch in die Datenbank aufgenommen und müssen von Hand als englisch-

sprachig markiert werden. Die Suche danach unter ca. 5 Millionen Wörtern gleicht der Suche nach der Nadel im Heuhaufen. Kollokationsmengen helfen hier, Kandidaten zu sammeln.

Beispielsweise liefert die Kollokationsliste für *the*:

of, and, to, The, on, for, is, from, you, with, that, it, world, are, be, not, We, at, World, we, have, this, by, they, when, You, can, When, into, what, your, or, But, time, And, like, over, Breaking, only, one, but, shall, which, has, What, road, as, On, same, people, out, our, This, It, way, best, who, no, my, more, his, up, their, ...

Ähnlich lassen sich Ausdrücke in Mundart identifizieren.

Die Kollokationsliste für *ick* liefert:

det, nich, Ick, Det, hab, is, ne, ooch, keene, wat, wees, uff, de, ma, nu, keen, dat, aba, och, jing, jut, Nee, meen, Jöre, een, mach, inne, watt, wa, jenuch, kieke, janze, kumm, janz, tau, Mutta, janzen, hätt, sag, wieda, kleene, ha, hör, imma, un, habense, kriejen, ejal, zwee, nischt, nee, Wetta, jedacht, hebb, heff, ...

Die Kollokationsliste für *ned* liefert:

Des, is, woas, ma, hod, mi, wos, amoi, Bua, scho, aa, kumma, wia, woin, kriegst, hätt, gwesn, ka, koa, derf, ko, oda, muaß, san, Erd, spuin, de, gesund, liaba, woas, kan, machn, frog, skandiert, warn, nu, fei, ausm, wär, Naa, mar, koan, falschrum, sogn, Hinterlechner, hudeln, oiß, ...

Die Kollokationsliste für *een* liefert:

de, un, dat, as, op, anner, se, vun, ok, speelt, för, düsse, is, wat, nich, mol, ick, det, na, ward, Dat, sick, deit, Se, Tied, he, hett, ut, giff, dor, ne, to, wa, As, aba, nu, jedeen, Keerl, snackt, Nootebooms, tein, sitt, eegens, grote, uff, ...

Solche Kollokationslisten enthalten natürlich nicht nur Wörter der gesuchten Art, sind aber zur manuellen Bearbeitung viel besser geeignet als die gesamte Wortliste.

Ableitung von Angaben aus Wortenden

Bei einem hinreichend großen Grundwortschatz sind neue Wörter häufig Komposita, deren letzter Bestandteil bereits bekannt ist. Damit lassen sich aus dem Wortende häufig sowohl grammatische als auch semantische Angaben ableiten. Das soll am folgenden Beispiel demonstriert werden: Zunächst hat es den Anschein, als ob folgende Regel gilt:

Alle Nomina auf –ung sind weiblich sind und gehorchen dem gleichen Flexionsmuster.

Aber es gibt Ausnahmen, nämlich *Dung* und *Schwung* und die dazugehörigen Komposita. Also wird unsere Regel verfeinert zu:

Alle Nomina auf –ung, außer –dung und -wung sind weiblich sind und gehorchen dem gleichen Flexionsmuster. Die Fälle –dung und –wung müssen einzeln weiter untersucht werden.

Um dieses Vorgehen zu automatisieren, wurde die Lernstrategie von Liang ([LIA]) beim TEX-Silbentrenner so verändert, daß sie beliebige Angaben aus Wortenden lernt.

Literatur

- [EAD] Eades, Peter, 1984. A Heuristic for Graph Drawing, *Congressus Numerantium*, vol. 42, pp. 149–160.
- [DAV] Davidson R. and D. Harel: Drawing Graphs Nicely Using Simulated Annealing, *ACM Transaction on Graphics*, Vol. 15, No. 4, 1996, Seiten 301–331.
- [FEL] Fellbaum, Chr. (ed.) *WordNet: An Electronic Lexical Database*, MIT Press 1998
- [LEM] Lemnitzer, L.: Komplexe lexikalische Einheiten in Text und Lexikon; in: Heyer, G., Wolff, Ch.: *Linguistik und neue Medien*, Wiesbaden, Dt. Universitätsverlag, 1998.
- [LIA] Liang, F. M.: *Word hy-phen-a-tion by com-pu-ter*. Technical Report STAN-CS-83-977, Stanford University, August 1983.
- [NOR] *Normdaten-CD-ROM*, Die Deutsche Bibliothek Frankfurt / Main, 1998.

-
- [QUA] Quasthoff, Uwe: Tools for Automatic Lexicon Maintenance: Acquisition, Error Correction, and the Generation of Missing Values. In: Proceedings of the first International Conference on Language Resources & Evaluation, ELRA 1998, S. 853–856.
- [STE] Steele, J.: Meaning-text theory: linguistics, lexicography and implications, University of Ottawa Press; Ottawa, London, Paris, 1990
- [WOT] Wothke, K. (1993). Morphologically based automatic phonetic transcription, IBM Systems Journal 32, S. 485–511.

Die vier Elementar-Ontologien semantischer Interpretation und ihre empirischen Folgen

Roland Hausser

Universität Erlangen-Nürnberg
Abteilung Computerlinguistik (CLUE)
rrh@linguistik.uni-erlangen.de

Diese Arbeit vergleicht die Interpretation der Logiksprachen, der Programmiersprachen und der natürlichen Sprachen. Sie zeigt, daß die Semantiken dieser Sprachen auf unterschiedlichen Ontologien beruhen und untersucht, wie sich diese auf die empirische Analyse der Wahrheit, des Epimenides-Paradoxons, propositionaler Einstellungen und der Vagheit auswirken.

1 Metasprachliche Semantik

In der Logik wird eine semantische Beziehung zwischen der formalen Sprache und der Welt etabliert, indem diese beiden Ebenen und die systematische Beziehung zwischen ihnen *definiert* werden. Die Theorie hinter dieser Methode wurde in ihrer noch heute gültigen Form von ALFRED TARSKI (1902–1983) präsentiert.

Die Interpretation einer Logiksprache soll bestimmen, unter welchen Umständen die Ausdrücke der Objektsprache wahr sind. Die Objektsprache ist die Sprache, die semantisch interpretiert wird (z. B. zitierte Ausdrücke wie ' ϕ & ψ '), während die Definitionen der semantischen Interpretation in einer Metasprache formuliert werden. Tarskis metasprachliches Grundschema zur Charakterisierung der Wahrheit ist die sogenannte T-Bedingung.

1.1 Schema von Tarskis T-Bedingung

T: x ist ein wahrer Satz dann und nur dann, wenn p.

Die T-Bedingung als Ganzes ist ein Satz der Metasprache, der den Satz x der Objektsprache zitiert und als p übersetzt. Tarski verwendet das folgende Beispiel:

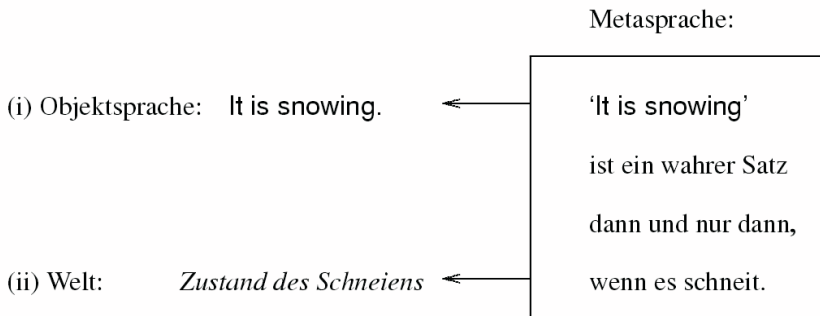
1.2 Instantiierung von Tarskis T-Bedingung

'It is snowing' ist ein wahrer Satz dann – und nur dann – wenn es schneit.

Dieses Beispiel ist von trügerischer Einfachheit, die zu Mißverständnissen geführt hat.¹ Was 1.1 und 1.2 in ihrer provokativen Simplizität für sich allein betrachtet nicht zum Ausdruck bringen, ist der genaue Charakter der *Zwei-Ebenen-Struktur*, die allen Formen der semantischen Interpretation zugrunde liegt.

Ein genaueres Studium von Tarskis Text zeigt, daß sich der Sinn der T-Bedingung keineswegs auf eine formelhafte Wiederholung des objektsprachlichen Ausdrucks in der metasprachlichen Übersetzung beschränkt. Die T-Bedingung hat vielmehr eine zweifache Funktion. Erstens wird mit Hilfe der Metasprache eine systematische Verbindung zwischen der Objektsprache und der Welt hergestellt; die Metasprache ist also das Mittel, um den *Zuordnungsalgorithmus* in der logischen Semantik zu realisieren. Zweitens wird mit Hilfe der Metasprache Wahrheit charakterisiert: der Wahrheitswert von x in der Objektsprache soll über die Interpretation von p in der Metasprache bestimmbar sein.

Beide Funktionen setzen voraus, daß die Metasprache sowohl auf (i) die Objektsprache als auch (ii) auf den Zustand der Welt (Modell) direkt Bezug nehmen kann. Diese von der Metasprache zwischen den beiden Ebenen der Objektsprache und der Welt hergestellte Verbindung ist in 1.3 schematisch dargestellt.



1.3 Verhältnis von Objekt- und Metasprache

Der direkte Bezug der Metasprache auf die Welt wird *Verifikation* genannt. Die Verifikation von T besteht in der praktischen Möglichkeit, tatsächlich festzustellen, ob p zutrifft oder nicht. Um beispielsweise feststellen zu können, ob It is snowing wahr ist oder nicht, muß es möglich sein, festzustellen, ob es tatsächlich schneit oder nicht. Solange die Möglichkeit der Verifikation nicht grundsätzlich garantiert ist, ist die T-Bedingung (i) für die Charakterisierung der Wahrheit inhaltsleer (siehe auch 2.1) und (ii) als Methode der Zuordnung semantischer Objekte disfunktional.

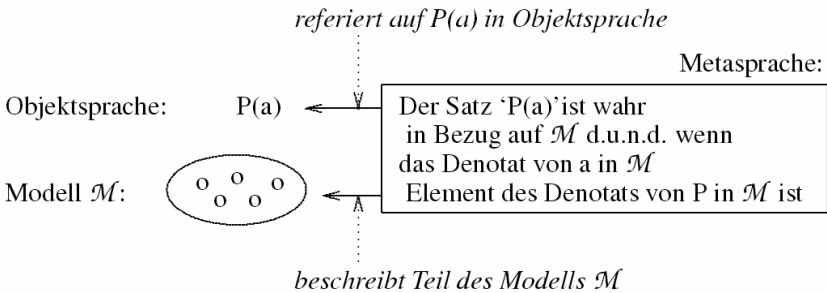
Laut Tarski erfordert diese Konstruktion, daß (i) alle Zeichen und Ausdrücke der Metasprache explizit aufgezählt werden und (ii) jedes Zeichen und jeder Ausdruck der Metasprache eine klare Bedeutung hat (*has a clear meaning*, Tarski 1935., S. 172). Diese gewissenhafte Behandlung der Metasprache illustriert Tarskis Beispiel des Klassenkalküls, an dem er seine Methode detailliert vorführt. Die einzigen Ausdrücke, die Tarski hierbei benutzt, sind Begriffe wie nicht, und, ist enthalten in, ist Element von, Individuum, Klasse und Relation. Die Bedeutung dieser Ausdrücke ist insofern unmittelbar klar und verständlich, als es sich ausschließlich um mathematische Objekte und mengentheoretische Operationen handelt.

Das Gleiche gilt für die semantischen Regeln im Standard-Prädikatenkalkül, weshalb es sich auch dort um eine wohldefinierte Semantik im Tarskischen Sinn handelt. Eine solche semantische Definition wird in 1.4 (gegenüberliegende Seite) analog zu 1.3 als T-Bedingung dargestellt.

1.4 T-Bedingung bei logischer Definition

Die Möglichkeit, diese T-Bedingung zu verifizieren, wird durch nicht mehr und nicht weniger als die Tatsache garantiert, daß bei jedem beliebigen Modell M jeder Sprecher des Deutschen mit einer minimalen Kenntnis der Mengentheorie in der Lage ist, zu *sehen* (im Sinne der unmittelbaren Anschauung), ob die im Übersetzungsteil von T angegebene Relation von M erfüllt wird oder nicht.

Die Berufung auf die unmittelbare Anschauung ist in der Geschichte der Mathematik schon immer als die letzte Instanz herangezogen worden:



En l'un les principes sont palpables mais éloignés de l'usage commun de sorte qu'on a peine à tourner la tête de ce côté-là, manque d'habitude : mais pour peu qu'on l'y tourne, on voit les principes à peine; et il faudrait avoir tout à fait l'esprit faux pour mal raisonner sur des principes si gros qu'il est presque impossible qu'ils échappent.

[Die Prinzipien {des mathematischen Geistes} sind greifbar, aber fern des allgemeinen Gebrauchs, weshalb es Mühe macht, sich ihnen zuzuwenden, da die Gewohnheit fehlt ; aber kaum hat man sich ihnen zugewendet, übersieht man die Prinzipien vollständig; und es bedürfte eines völlig verkehrten Verstandes, um falsch auf Grund von Prinzipien zu schließen, die so offensichtlich sind, daß sie einem kaum entgehen können.]

B. PASCAL (1623–1662), *Pensées*, 1951:340

Die Tarskische Methode ist insofern auf die Bereiche von Mathematik, Logik und Naturwissenschaft beschränkt, als nur dort ausreichend sichere Methoden der Verifikation zur Verfügung stehen.

2 Prozedurale Semantik

Im Gegensatz zu einer semantischen Definition wie 1.4 verletzt die folgende Instantiierung der T-Bedingung die Verifikations-Voraussetzung.

2.1 Beispiel einer inhaltsleeren T-Bedingung

'A ist rot' ist ein wahrer Satz dann – und nur dann – wenn A rot ist.

Diese Instantiierung der T-Bedingung ist formal korrekt, aber inhaltsleer, weil in ihr keine Rückführung der objektsprachlichen Bedeutung auf verifizierbare Begriffe der Metasprache stattfindet. Stattdessen wird der Ausdruck der Objektsprache in der metasprachlichen Übersetzung nur wiederholt.

Im Rahmen ihres mengentheoretischen Fundamentes hat eine modelltheoretische Semantik keine Möglichkeit, ein Inhaltswort wie rot wahrheitsfunktional so zu analysieren, daß seine Bedeutung z. B. im Unterschied zu blau adäquat charakterisiert wird. Hier bleibt nur die Möglichkeit, die Metatheorie durch das Hinzuziehen zusätzlicher Wissenschaften, z. B. der Physik, zu erweitern.

Dafür wird aus der hinzugezogenen Wissenschaft eine möglichst kleine Anzahl von zusätzlichen Grundbegriffen zur Erweiterung der Metasprache verwendet. Das Ergebnis erfüllt seinen Zweck, wenn die Bedeutungen der zusätzlichen Ausdrücke verifizierbar sind. In diesem Sinn könnte man z. B. 2.1 etwa wie folgt verbessern.

2.2 Verbesserte T-Bedingung für rot

'A ist rot' ist ein wahrer Satz dann – und nur dann – wenn A Licht im Frequenzbereich zwischen α und β refraktiert.

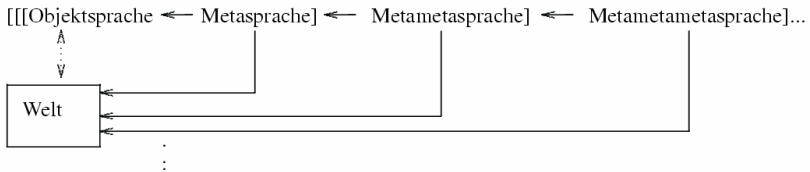
Hier wird in der metasprachlichen Übersetzung eine Rückführung auf elementarere Begriffe (insbesondere die Zahlen a und b auf einer empirisch etablierten Frequenzskala und den physikalisch wohlverstandenen Begriff von refraktiertem Licht) geleistet. Damit wird der objektsprachliche Ausdruck in der T-Bedingung in einer nichttrivialen Weise charakterisiert, die physikalisch verifizierbar ist.

Beispiele wie 2.1 zeigen, daß die Objektsprache Sätze enthalten kann, für die es in einer gegebenen Metasprache nur inhaltsleere Übersetzungen gibt. Das heißt nicht, daß ein Satz wie x ist rot nicht sinnvoll ist oder keinen Wahrheitswert hat. Es heißt nur, daß die Metasprache nicht reich genug ist, um die Grundlagen für eine unmittelbar einsichtige Verifizierung des Satzes bereitzustellen. Damit stellt sich die Frage, wie die Semantik einer Metaspra-

che formal zu behandeln ist, insbesondere in bezug auf ihre noch unanalysierten Teile.

Als Antwort konstruiert Tarski eine unendliche Hierarchie von Metasprachen.

2.3 Hierarchie der Metasprachen



Die Analyse der Wahrheit ist umso genauer, je mehr die Ausdrücke der Objektsprache auf verifizierbare Begriffe in der {Meta} + Sprache zurückgeführt werden. Daß mit dieser unendlichen Hierarchie der Metasprachen den Menschen ein umfassender Zugang zur Wahrheit letztlich verwehrt ist, spricht nicht gegen diese Theorie der Wahrheit – im Gegenteil, darin liegt der philosophische Reiz dieser Konstruktion.

Für die Semantik der Programmiersprachen und der natürlichen Sprachen ist eine Hierarchie von Metasprachen dagegen keine geeignete Grundlage. Betrachten wir z. B. die Gesetze der Grundrechenarten Addition, Multiplikation etc. Das Problem ist nicht, für diese einwandfreie metasprachliche Definitionen zu liefern. Der Punkt ist vielmehr, daß der Weg von einer solchen metasprachlichen Definition zu einer funktionierenden Rechenmaschine sehr weit ist und daß die Rechenmaschine am Ende mechanisch funktioniert, d. h. ohne irgendeinen Bezug auf die Metasprache und ohne irgendeine Notwendigkeit, die Metasprache zu verstehen.

Diese schlichte Tatsache ist Autonomie von der Metasprache genannt worden. Sie ist charakteristisch für alle Computerprogramme. Autonomie von der Metasprache bedeutet nicht, daß Computer auf uninterpretierte, rein syntaktische Deduktionssysteme beschränkt wären, sondern vielmehr, daß die Tarskische Methode der semantischen Interpretation nicht die einzige ist. Statt dem Tarskischen Zuordnungsprinzip über metasprachliche Definition verwenden Programmiersprachen ein *operationales* Zuordnungsprinzip: die Begriffe der Programmiersprache werden automatisch in elektronisch realisierte Operationen umgesetzt.²

3 Epimenides-Paradox

Da der praktische Einsatz von Programmiersprachen eine automatische (d. h. metasprachfreie) Umsetzung ihrer syntaktischen Befehle in entsprechende elektronische Prozesse erfordert, ist eine Tarskische Semantik für Programmiersprachen ungeeignet. Wie aber steht es mit einer Tarskischen Semantik für die natürlichen Sprachen?

Tarski selbst ließ keinen Zweifel daran, daß eine vollständige Analyse der natürlichen Sprachen im Rahmen der logischen Semantik prinzipiell unmöglich ist.

The attempt to set up a structural definition of the term 'true sentence' – applicable to colloquial language – is confronted with insuperable difficulties.

[Der Versuch, zu einer strukturellen Definition des Begriffs 'wahrer Satz' zu gelangen – anwendbar auf die Umgangssprache – steht vor unüberwindlichen Schwierigkeiten.]

TARSKI 1935, S. 164.

Tarski beweist diese Schlußfolgerung auf der Grundlage eines klassischen Paradoxons, das als Epimenides-, Eubulides- oder Lügner-Paradox bekannt ist.

Das Paradox beruht auf Selbstreferenz. In seiner ursprünglichen 'schwachen' Form lautet es wie folgt: Wenn ein Kreter sagt, Alle Kreter lügen (immer), gibt es zwei Möglichkeiten. Entweder der Kreter spricht die Wahrheit, in welchem Fall es falsch ist, daß *alle* Kreter lügen – da er ja selbst ein Kreter ist. Oder der Kreter lügt, d. h. es gibt mindestens einen anderen Kreter, der nicht lügt. In beiden Fällen ist der Satz falsch.³

Tarski 1935 verwendet das Paradox in der 'starken' Variante von Lesniewski und konstruiert daraus einen Beweis, wonach eine vollständige Analyse der natürlichen Sprachen im Rahmen der logischen Semantik prinzipiell unmöglich ist.

For the sake of greater perspicuity we shall use the symbol 'c' as a typological abbreviation of the expression 'the sentence printed in line 4 of the current quotation.' Consider now the following sentence:

c is not a true sentence (1935)

Having regard to the meaning of the symbol 'c', we can establish empirically:

(a) 'c is not a true sentence' is identical with c.

For the quotation-mark name of the sentence c we set up an explanation of type (2) [d. h. der T-Bedingung 1.1]:

(b) 'c is not a true sentence' is a true sentence if and only if c is not a true sentence.

The premise (a) and (b) together at once give a contradiction:

c is a true sentence if and only if c is not a true sentence.

[Um der besseren Klarheit willen werden wir das Symbol 'c' als typographische Abkürzung des Ausdrucks 'der Satz, der in der vierten Zeile dieses Zitats gedruckt steht' verwenden. Man betrachte nun den folgenden Satz:

c ist nicht ein wahrer Satz.

Bezüglich der Bedeutung des Symbols 'c' können wir empirisch feststellen:

(a) 'c ist nicht ein wahrer Satz' ist identisch mit c.

Für das Zitat des Satzes c geben wir eine Erklärung im Sinn der T-Bedingung 1.1:

(b) 'c ist nicht ein wahrer Satz' ist ein wahrer Satz dann – und nur dann – wenn c nicht ein wahrer Satz ist.

Die Prämissen (a) und (b) ergeben zusammen sofort einen Widerspruch:

c ist ein wahrer Satz dann und nur dann wenn c nicht ein wahrer Satz ist.]

TARSKI 1935

In dieser Konstruktion basiert die Selbstreferenz auf zwei Voraussetzungen. Erstens wird ein Satz, der in einer bestimmten Zeile an einer bestimmten Stelle steht, mit der Abkürzung 'c' gleichgesetzt.⁴

Zweitens kommt der Buchstabe 'c', mit dem der Satz in Zeile 4 des Zitats abgekürzt wurde, in der unabgekürzten Version eben dieses Satzes vor. Damit ergibt sich die Möglichkeit, das c in dem fraglichen Satz durch das zu ersetzen, was das 'andere' c abkürzt. Es gibt drei Möglichkeiten, diesen Widerspruch in der metasprachlichen T-Bedingung zu verhindern.

Die erste besteht darin, die Abkürzung und die daraus folgende Substitution zu verbieten. Dies lehnt Tarski ab, weil „es keinen rationalen Grund gibt, die Substitution im allgemeinen zu verbieten.“⁵

Die zweite besteht darin, die Wahrheitsprädikate ‘wahr^O’ der Objektsprache und ‘wahr^M’ der Metasprache zu unterscheiden. Damit wäre z. B.

c ist wahr^O dann und nur dann wenn c nicht wahr^M ist.

nicht widersprüchlich, weil wahr^O \neq wahr^M. Diese Möglichkeit zieht Tarski nicht in Betracht, vermutlich weil die Postulierung mehrerer Wahrheitsprädikate mit dem eigentlichen Anliegen der logischen Semantik, nämlich einer formalen Charakterisierung *der* Wahrheit, kollidiert.

Die dritte, von Tarski gewählte Möglichkeit, den durch das Epimenides-Paradox induzierten Widerspruch zu vermeiden, besteht darin, die Verwendung des Wahrheitsprädikats in der zu interpretierenden Objektsprache zu verbieten. Für die ursprünglichen Ziele der logischen Semantik stellt die Wahl dieser dritten Möglichkeit, also die Verbannung der Wörter *wahr* und *falsch* aus der Objektsprache, keinerlei Problem dar. Wenn es darum geht, naturwissenschaftliche Theorien wie die Physik als eine wahre Beziehung zwischen logischen Aussagen und Sachverhalten zu charakterisieren, benötigt man kein objektsprachliches Wahrheitsprädikat. Das gleiche gilt für formale Theorien wie die Mathematik.

Hinzu kommt, daß die Entwicklung semantisch interpretierter logischer Kalküle für viele mathematische Logiker vor allem den Zweck hat, die Vagheiten und Widersprüchlichkeiten der natürlichen Sprachen zu vermeiden. So schreibt z.B. Frege 1896 (1967, S. 221):

Der Grund, weshalb die Wortsprachen zu diesem Zweck [i. e. Schlüsse nur nach rein logischen Gesetzen zu ziehen] wenig geeignet sind, liegt nicht nur an der vorkommenden Vieldeutigkeit der Ausdrücke, sondern vor allem in dem Mangel fester Formen für das Schließen. Wörter wie >also<, >folglich<, >weil< deuten zwar darauf hin, daß geschlossen wird, sagen aber nichts über das Gesetz, nach dem geschlossen wird, und können ohne Sprachfehler auch gebraucht werden, wo gar kein logisch gerechtfertigter Schluß vorliegt.

Auch vor diesem Hintergrund ist es nur zu verständlich, wenn Tarski es entschieden von sich weist, sein System auf die natürlichen Sprachen zu übertragen.

Da die natürlichen Sprachen die Wörter *wahr* und *falsch* enthalten *müssen*,⁶ folgt, daß eine logisch-semantische Interpretation einer natürlichen (Objekt-)Sprache in ihrer Gesamtheit unvermeidlich zum Widerspruch führt. Tarskis Schüler RICHARD MONTAGUE (1930–1970) bestand jedoch trotzdem darauf, die logische Semantik auf die natürlichen Sprachen anzuwenden:

I reject the contention that an important theoretical difference exists between formal and natural languages. ... Like Donald Davidson I regard the construction of a theory of truth –or rather the more general notion of truth under an arbitrary interpretation– as the basic goal of serious syntax and semantics.

[Ich weise die Behauptung zurück, daß zwischen den logischen und den natürlichen Sprachen ein wichtiger theoretischer Unterschied bestünde. . . . Wie Donald Davidson betrachte ich die Konstruktion einer Theorie der Wahrheit – oder allgemeiner, der Wahrheit unter beliebigen Interpretationen – als das elementare Ziel ernsthafter Syntax und Semantik.]

MONTAGUE 1970, *English as a formal language*⁷

Natürlich kannte Montague das Epimenides-Paradox und Tarskis Folgerung. In seinen Arbeiten zur Semantik natürlicher Sprachen geht er jedoch mit keinem Wort auf diesen Problemkreis ein. Dafür ist der von Montague zitierte Davidson explizit:

Tarski's ... point is that we should have to reform natural language out of all recognition before we could apply formal semantic methods. If this is true, it is fatal to my project.

[Tarskis Punkt ist, daß wir die natürlichen Sprachen bis zur Unkenntlichkeit reformieren müßten, bevor wir die formal-semantische Methode auf sie anwenden könnten. Sollte dies wahr sein, so ist es tödlich für mein Projekt.]

DAVIDSON 1967

Das Fatale an einem logischen Paradox ist, daß es das ganze semantische System zerstört. Je nachdem, mit welchem Teil des Widerspruchs man eine Folgerungskette beginnt, kann man in einem solchen System zu jeder Aussage auch ihr Gegenteil beweisen. Und das ist gerade für eine Theorie der Wahrheit unerträglich.⁸

4 Propositionale Einstellung

Ein zweites Grundproblem einer logischen Semantik für die natürlichen Sprachen sind die propositionalen Einstellungen. Darunter versteht man Sätze, welche die Relation zwischen einem kognitiven Agenten und einem propositionalen Inhalt beschreiben. Beispielsweise beschreibt der Satz

Susanne glaubt, daß Cicero den Catilina angeklagt hat.

die propositionale Einstellung des Glaubens als Relation zwischen Susanne und der Proposition Cicero hat den Catilina angeklagt. Welche Wahrheitsbedingungen haben propositionale Einstellungen im Rahmen der logischen Semantik?

Nach den Intuitionen der modalen Logik denotiert ein Eigenname in allen möglichen Welten dasselbe Individuum (*rigid designator*, Kripke 1972). Weil Cicero und Tullius Namen für dieselbe Person sind, gilt notwendigerweise (also in allen möglichen Welten), daß Cicero = Tullius. Deshalb folgt aus Cicero klagte Catilina an notwendigerweise Tullius klagte Catilina an.

Wenn einer der beiden Sätze jedoch unter einem Prädikat der propositionalen Einstellung, z.B. glauben, eingebettet wird, gilt die Substitution *salva veritate* auch nicht bei Eigennamen. So folgt zum Beispiel aus Susanne glaubt, daß **Cicero** Catilina anklagte nicht, daß Susanne glaubt, daß **Tullius** Catilina anklagte. Denn obwohl Cicero notwendigerweise gleich Tullius ist, kann es ja sein, daß Susanne dies nicht weiß. Entsprechend würde eine Substitution *salva veritate* zusätzlich die Wahrheit von Susanne **glaubt**, daß Cicero = Tullius erfordern.

Weil verschiedene Menschen sehr unterschiedliche Vorstellungen von der externen Realität haben können, müßte eine Behandlung der propositionalen Einstellungen in der Carnap/Montagueschen Manier nicht nur die naturwissenschaftliche Realität, sondern auch die Glaubenswelten einzelner Individuen modellieren.⁹ Um festzustellen, was Individuen glauben, ist man jedoch auf deren Mitteilungen angewiesen. Dabei kann nicht objektiv überprüft werden, ob diese Mitteilungen wahr oder falsch sind. Deshalb stellen individuelle 'Glaubenswelten' einen Prototyp dessen dar, was außerhalb einer wissenschaftlichen Wahrheitsfindung liegt.¹⁰

Das Phänomen der propositionalen Einstellungen stellt eine modelltheoretische Semantik der natürlichen Sprachen vor die folgende Frage:

4.1 Ontologisches Grundproblem der Modelltheorie

Ist der Sprecher/Hörer Teil der Modellstruktur oder ist die Modellstruktur Teil des Sprecher/Hörers?

Wenn das Ziel eine Charakterisierung der Wahrheit ist, dann werden nur logische Bedeutungen verwendet, die als ewig und unmittelbar einsichtig vorausgesetzt werden. Wie alle anderen 'Gegenstände' auch, ist bei diesem Ansatz der Sprecher/Hörer Teil der Modellstruktur. Wahrheitsbeziehungen zwischen Aussagen und Sachverhalten bestehen unabhängig davon, ob sie von diesem oder jenem Sprecher/Hörer entdeckt werden oder nicht.¹¹

Steht dagegen die sprachliche Bedeutung im Vordergrund, wird das logische System, das anfänglich unter Verwendung logischer Bedeutungen zur Charakterisierung der Wahrheit entwickelt wurde, zu einer Charakterisierung sprachlicher Bedeutungen in der Form von Wahrheitsbedingungen verwendet. Das Ziel, die logischen Aspekte natürlicher Bedeutungen herauszuarbeiten, bringt aber nolens volens eine Veränderung der ursprünglichen ontologischen Annahmen mit sich.

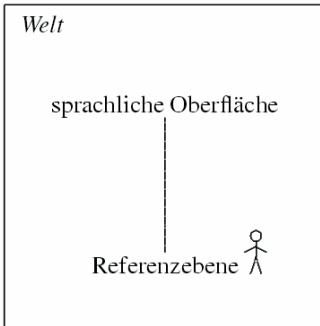
Sprachliche Bedeutungen können nämlich nur dann vom Sprecher/Hörer in der Kommunikation verwendet werden, wenn sie Teil seiner oder ihrer Kognition sind. Deshalb führt die Beschreibung sprachlicher Bedeutungen im Rahmen der logischen Semantik notwendigerweise dazu, daß das logische Modell als kognitive Struktur und damit als Teil des Sprecher/Hörers interpretiert wird.

Die kognitive Interpretation des Modells als Teil des Sprecher/Hörers ist mit den Zielen und Methoden einer traditionellen Theorie der Wahrheit unvereinbar. Umgekehrt ist die 'realistische' Interpretation des Modells im Rahmen einer Theorie der Wahrheit unvereinbar mit einer Bedeutungsanalyse natürlicher Sprache.¹²

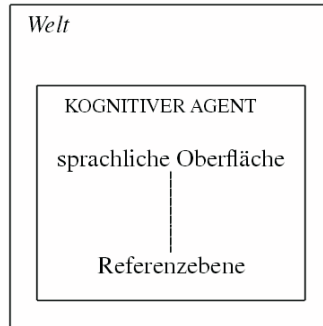
Die in 4.1 formulierte Alternative ist in 4.2 schematisch dargestellt, wobei der Unterschied mit dem Merkmal [±konstruktiv] bezeichnet wird.

4.2 Zwei Interpretationen der Modelltheorie

[-konstruktiv]



[+konstruktiv]



Bei der [-konstruktiv]-Interpretation findet die Beziehung zwischen der sprachlichen Oberfläche und der Referenzebene außerhalb des kognitiven Agenten in der Welt statt. Der Agent ist selbst ein Objekt der Referenzebene, der diese gewissermaßen gottgegebene, direkte Beziehung zwischen der Sprache und den Objekten der Welt betrachten und logisch nachbauen kann. Für die *Verwendung* von Sprache durch den Agenten ist in der [-konstruktiv]-Interpretation kein Raum.

Bei der [+konstruktiv]-Interpretation findet die Beziehung zwischen der sprachlichen Oberfläche und der Referenzebene dagegen ausschließlich innerhalb des kognitiven Agenten statt. Was der Agent in der Welt nicht wahrnehmen kann, spielt bei der sprachlichen Referenz keine Rolle, wohl aber das, was der Agent fühlt, wünscht, plant etc. Bei der [+konstruktiv]-Interpretation kommt Referenz *nur* über die Verwendung von Sprache durch den Agenten zustande.

Der fundamentalste Unterschied zwischen den beiden Ontologien besteht darin, daß [-konstruktiv]-Systeme eine metasprachliche und [+konstruktiv]-Systeme eine prozedurale Semantik haben *müssen*. In [-konstruktiv]-Systemen kann die Beziehung zwischen Aussage und Sachverhalt nur über die metasprachliche Definition etabliert werden, denn bei wissenschaftlichen Aussagen, die als ewig gültig und Sprecher/Hörer-unabhängig betrachtet werden, ist diese Beziehung nicht sinnvoll zu operationalisieren.¹³ Dagegen sind [+konstruktiv]-Systeme ohne

eine prozedurale Semantik nutzlos, denn mit einer metasprachlichen Semantik kann weder ein Computer noch ein kognitiver Agent praktisch funktionieren.

5 Ontologien semantischer Interpretation

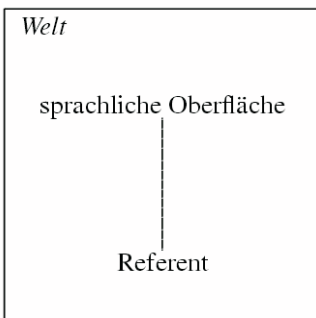
Ein zweiter Unterschied bei Systemen der semantischen Interpretation kann durch das Merkmal $[\pm \text{Sinn}]$ repräsentiert werden. Dieses bezieht sich darauf, ob die sprachliche Bedeutung mit den Referenzobjekten gleichgesetzt wird [-Sinn], oder ob die Bedeutung auf einer eigenen Stufe als Fregescher 'Sinn' behandelt wird, der sich von den Referenzobjekten unterscheidet [+Sinn].

Die binären Merkmale $[\pm \text{Sinn}]$ und $[\pm \text{konstruktiv}]$ sind unabhängig voneinander und können daher miteinander kombiniert werden. Daraus ergeben sich vier Typen der semantischen Interpretation, die auf vier verschiedenen Ontologien beruhen, nämlich [-Sinn, -konstruktiv], [+Sinn, -konstruktiv], [-Sinn, +konstruktiv] und [+Sinn, +konstruktiv].

5.1 Vier elementare Ontologien

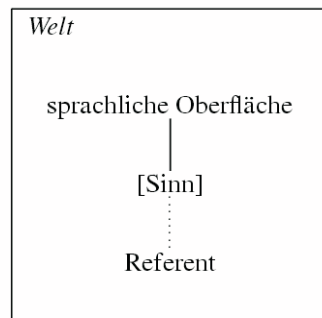
i [-Sinn, -konstruktiv]

Russell, Carnap, Quine, Montague



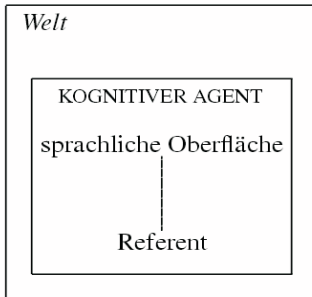
ii [+Sinn, -konstruktiv]

Frege

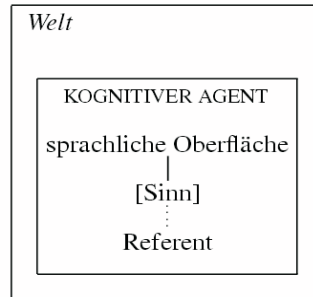


iii [-Sinn, +konstruktiv]

Newell & Simon, Winograd, Shank

**iv [+Sinn, +konstruktiv]**

Anderson, CURIOUS, SLIM-Maschine



Wie durch die Namen angedeutet, sind diese verschiedenen Ontologien von verschiedenen Schulen der semantischen Interpretation zugrunde gelegt worden.

Die [-Sinn, -konstruktiv]-Ontologie (i) entspricht der Semantik der Logik. In dem Bemühen um eine solide Fundierung der Wahrheit werden nur Dinge verwendet, die ontologisch unzweifelhaft 'real' sind. Im Nominalismus sind dies die sprachlichen Zeichen und die Sachverhalte, die aus konkreten Referenten aufgebaut sind. Im mathematischen Realismus wird die Ontologie um mathematische Objekte (z.B. Zahlen) und Strukturen (z.B. Mengen) erweitert. Beide Varianten der [-Sinn, -konstruktiv]-Ontologie haben gemeinsam, daß die Semantik als eine direkte, externe Beziehung zwischen 'Sprache und Welt' definiert ist. Dieser Semantiktyp entspricht der gesamten Hauptlinie der modernen philosophischen Logik, von Russell über den frühen Wittgenstein, Carnap, Montague¹⁴ bis Putnam.

Die [+Sinn, -konstruktiv]-Ontologie (ii) wurde von Frege verwendet und ist von dem Versuch motiviert, ungerade (opaque, intensionale) Lesarten in den natürlichen Sprachen zu analysieren. Für eine Modellierung der natürlichen Kommunikationsmechanik ist dies nur ein halber Schritt in der richtigen Richtung. Denn als Theorie der Wahrheit ist jede [-konstruktiv]-Semantik mit einer Darstellung kognitiver Zustände unvereinbar.¹⁵

Die [-Sinn, +konstruktiv]-Ontologie (iii) entspricht der Semantik der Programmiersprachen. Der Benutzer gibt in den Computer Befehle (programmiersprachliche Oberflächen) ein, die direkt in entsprechende elektronische Prozeduren umgesetzt werden. Wenn das Ergebnis errechnet worden ist, wird es dem Benutzer vom Computer mit Hilfe programmiersprachlicher Oberflächen mitgeteilt, z. B. auf dem Bildschirm. In dieser traditionellen Verwendung ist der Computer von einem kognitiven Agenten noch weit entfernt. Es liegt aber bereits eine Unterscheidung zwischen dem *task environment* in der Welt und dem computerinternen *problem space* vor, wobei die semantische Interpretation in letzterem stattfindet.

Aufgrund ihrer Entstehung als traditionelle Programme auf traditionellen Computern basieren die klassischen Systeme der künstlichen Intelligenz gewissermaßen unbewußt auf einer [-Sinn, +konstruktiv]-Ontologie. Dies gilt z. B. für SHRDLU (Winograd 1972), HEARSAY (Reddy et al. 1973) und SAM (Schank & Abelson 1977). Auch in der kognitiven Psychologie ist dieser ontologische Ansatz vertreten worden, z. B. bei den *mental models* von Johnson-Laird 1983.

In der künstlichen Intelligenz haben Newell & Simon 1972, S. 66, explizit gegen eine dem Sinn entsprechende Zwischenstufe argumentiert, und zwar rein ontologisch. Die Unterscheidung zwischen sprachlichen Bedeutungen ('Sinn') und den computer-internen Referenten würde „in einer unnötigen und unsparamen Vervielfältigung hypothetischer Entitäten, für die es keine Evidenz gibt“¹⁶, resultieren.

Eine direkte Verbindung zwischen sprachlichen Ausdrücken und ihren Referenten verhindert jedoch grundsätzlich die eigenständige Klassifikation neuer Gegenstände. Deshalb ist der [-Sinn, +konstruktiv] Semantiktyp auf geschlossene Spielzeugwelten (*toy worlds*) beschränkt, die vom Programmierer gebaut werden müssen.¹⁷ Es ist kein Zufall, daß diese Systeme keine Wahrnehmungskomponenten (z. B. künstliches Sehen) haben. Aufgrund der fehlenden Sinn- oder Konzeptebene wäre es ihnen strukturell unmöglich, neue Gegenstände zu erkennen und automatisch in ihr Weltbild zu integrieren.

Die [+Sinn, +konstruktiv]-Ontologie (iv) liegt der SLIM-Sprachtheorie¹⁸ zugrunde. SLIM ermöglicht die ontologische [+Sinn]-Eigenschaft durch die Abpassung zwischen Sprachbedeutung und kontextueller Referenten. Ihre ontologische [+konstruktiv]-Eigenschaft basiert darauf, daß diese Abpassung im Inneren der kognitiven Agenten stattfindet. In der kognitiven Psychologie findet sich der [+Sinn, +konstruktiv] Semantiktyp bei Anderson & Bower 1973 und 1980.

Das theoretische Verhältnis zwischen den vier verschiedenen Semantiktypen kann man analysieren, indem man entweder die ontologischen Unterschiede oder die formalen Gemeinsamkeiten betont. In letzterem Fall wird man seine Semantik als rein formales Gebilde präsentieren, dem die unterschiedlichsten Interpretationen zugeordnet werden können, ohne daß dabei die formale Essenz tangiert wird. Hierfür liegt es nahe, die verschiedenen ontologischen Grundlagen unter dem Gesichtspunkt der Spezialisierung bzw. Verallgemeinerung zueinander in Beziehung zu setzen.

Den Unterschied zwischen einer [+Sinn]- und einer [-Sinn]-Ontologie kann man minimieren, indem man letztere als eine Vereinfachung der ersteren interpretiert. Angenommen, (i) die Welt ist geschlossen, d. h. Objekte können weder verschwinden, noch neue hinzukommen¹⁹, (ii) die Beziehung zwischen sprachlichen Ausdrücken und ihren Referenten ist ein für allemal festgelegt²⁰ und (iii) eine spontane sprachliche Verwendung durch den Sprecher/Hörer ist nicht vorgesehen. Mit diesen vereinfachenden Annahmen verschwindet der Grund für die Verwendung einer Sinn-Ebene, womit aus einem [+Sinn]-System als Spezialfall ein [-Sinn]-System entsteht.

Letzteres mag aufgrund seiner Vereinfachung als gültiger oder essentieller angesehen werden als ein [+Sinn]-System. Man sollte dabei aber nicht vergessen, daß es empirische Phänomene gibt, die im Rahmen einer [-Sinn]-Ontologie grundsätzlich nicht modelliert werden können, insbesondere der referentielle Bezug auf neue Instanzen eines bestimmten Objekttyps.

Auch der Unterschied zwischen einer [+konstruktiv]- und einer [-konstruktiv]-Ontologie kann im Zuge einer Vereinfachung minimiert werden. Angenommen, die Wahrnehmung des kognitiven Agenten ist so perfekt, daß die Unterscheidung zwischen externen Zeichen bzw. Referenten und den entsprechenden kognitiven Repräsentationen vernachlässigt werden kann. Dann verschwindet der Grund für die Unterscheidung von kognitiver Repräsentation und externer Realität, womit der Verwendung einer [-konstruktiv]-Ontologie – als Spezialfall der [+konstruktiv]-Ontologie – nichts mehr im Wege steht.

Wiederum kann das System mit der [-konstruktiv]-Ontologie für das abstraktere, und daher gültigere und essentiellere, gehalten werden. Man sollte dabei aber nicht vergessen, daß es empirische Phänomene gibt, die im Rahmen einer [-konstruktiv]-Ontologie grundsätzlich nicht modelliert werden können, insbesondere das Phänomen der propositionalen Einstellungen.

Die Wahl zwischen den vier Semantiktypen hängt von der Anwendung ab. Deshalb sollte man (i) vor der *Erweiterung* einer Semantik auf neue Anwendungen und (ii) vor der *Übertragung* von Teilanalysen aus einer Anwendung in eine andere die strukturellen Unterschiede zwischen den vier Ontologien ebenso gut kennen wie ihre möglichen Äquivalenzen aufgrund von Vereinfachungen.

6 Sorites-Paradox und die Behandlung der Vagheit

Die Bedeutung der Ontologie für die empirische Analyse eines semantischen Phänomens zeigt das Beispiel der Vagheit. In der logischen Semantik nimmt die Behandlung der Vagheit ein klassisches Paradox aus der Antike als ihren Ausgangspunkt, nämlich das sogenannte Sorites- oder Haufen-Paradox.

Ein Sandkorn bildet keinen Haufen. Wenn man ein zusätzliches Sandkorn daneben legt, besteht immer noch kein Haufen. Wenn n Körner keinen Haufen bilden, dann entsteht durch das Hinzufügen eines einzigen zusätzlichen Korns immer noch kein Haufen. Wenn dieser Prozeß des Hinzufügens jedoch lange genug fortgesetzt wird, dann befindet man sich irgendwann in einer Situation, wo tatsächlich ein Haufen entstanden ist.

Die Übertragung des Sorites-Paradox auf eine logische Semantik natürlicher Sprachen wird durch die Annahme bewerkstelligt, daß z. B. der Prozeß einer sich langsam schließenden Tür die Frage aufwerfe, bis zu welchem Punkt der Satz Die Tür ist offen noch wahr ist und an welchem Punkt der Satz falsch ist. Dann geht man einen Schritt weiter und fragt, zu welchem *Grad* der Satz bei den verschiedenen Stadien des Türschließens wahr oder falsch ist.

Sensitive students of language, especially psychologists and linguistic philosophers, have long been attuned to the fact that natural language concepts have vague boundaries and fuzzy edges and that, consequently, natural-language sentences will very often be neither true, nor false, nor nonsensical, but rather true to a certain extent and false to a certain extent, true in certain respects and false in other respects.

[Feinfühlig Sprachforscher, insbesondere Psychologen und Sprachphilosophen, haben sich seit langem auf die Tatsache eingestellt, daß die Konzepte der natürlichen Sprachen vage Grenzen und verschwommene Ränder haben, und daß deshalb Sätze der natürlichen

Sprachen sehr oft weder wahr noch falsch noch unsinnig sind, sondern vielmehr wahr zu einem gewissen Grad und falsch zu einem gewissen Grad, wahr in gewissen Hinsichten und falsch in anderen.]

George LAKOFF 1972, S. 183

Eine andere Situation, die als Beispiel wahrheitskonditionaler Vagheit präsentiert worden ist, ist die Klassifikation der Farben. Wenn ein Gegenstand in Kontext a als rot klassifiziert wird, aber als nicht-rot in Kontext b, folgt dann nicht, daß das natürlichsprachliche Konzept rot vage sein muß? Denn wenn das Prädikat 'x ist rot' auf den Übergang von rot zu orange in einem Farbspektrum angewendet wird, entsteht eine ähnliche Situation wie bei der sich langsam schließenden Tür.

Wenn diese Analysen akzeptiert werden, dann muß die traditionelle zweiwertige (bivalente) Logik in eine mehrwertige Logik erweitert werden. Die nicht-bivalenten Logiken können in zwei Gruppen aufgeteilt werden, nämlich die dreiwertigen Logiken, in denen eine Proposition wahr (1), falsch (0) oder unbestimmt (#) sein kann, und die vielwertigen Logiken, in denen eine unendliche Anzahl von Wahrheitswerten mit den reellen Zahlen zwischen 0 und 1, also z.B. 0,615, identifiziert wird. Sowohl bei den dreiwertigen als auch bei den vielwertigen Logiken stellt sich das folgende Grundproblem:

Wie soll der Wahrheitswert komplexer Propositionen aus Teilen bestimmt werden, die keine bivalenten Wahrheitswerte haben?

Welchen Wert soll z.B. 'A & B' erhalten, wenn A den Wert 1 und B den Wert # hat? Entsprechend in einem vielwertigen System: Wenn A beispielsweise den Wert 0,615 und B den Wert 0,423 hat, was soll dann der Wert von 'A & B' sein?

Auf diese Fragen gibt es eine unübersehbare Fülle verschiedener Antworten. Laut Rescher 1969 sind bis dato 51 verschiedene Systeme mehrwertiger Logiken in der Literatur belegt. Aus einem wissenschaftsgeschichtlichen Blickpunkt ist diese Vielfalt von alternativen Lösungen ein klarer Fall der Syndrome *deskriptiver Aporie* und *Qual der Wahl*. Sie sind ein untrügliches Zeichen, daß die elementaren Prämissen eines Ansatzes grundsätzlich verfehlt sind.

In nicht-bivalenten Logiken liegt der Irrtum in der von Lakoff formulierte Prämisse, daß Propositionen doch offensichtlich nicht-bivalente Wahrheitswerte haben können. Wer diese Prämisse akzeptiert, begibt sich zwangsläufig auf die aussichtslosen Suche nach einer adäquaten allgemeinen Wahrheitswertzuweisung

bei komplexen Propositionen, also z.B. der Frage, welchen Wert die Proposition 'A & B' haben soll, wenn A beispielsweise den Wert 0,615 und B den Wert 0,423 hat.

Anstatt die Prämisse zu akzeptieren, sollten wir uns lieber fragen, woher Propositionen so merkwürdige Wahrheitswerte wie 0,615 überhaupt bekommen. Und damit sind wir wieder bei der Frage nach der zugrundegelegten Ontologie. Genauer gesagt: Wie wirkt sich der strukturelle Unterschied zwischen der [-Sinn, -konstruktiv]-Ontologie der logischen Semantik und [+Sinn, +konstruktiv]-Ontologie der SLIM-Sprachtheorie auf die formale Analyse der Vagheit aus?

Wir beginnen mit der Analyse eines Beispiels im Rahmen der [-Sinn, -konstruktiv]-Ontologie der logischen Semantik. Angenommen 'A & B' ist eine Proposition, wobei A = [Die Tür ist offen] und B = [Die Tür ist rot] ist. Außerdem soll A den Wahrheitswert 0,615 und B den Wahrheitswert 0,423 haben. Dann hat diese Proposition in diesem Rahmen die folgende semantische Struktur:

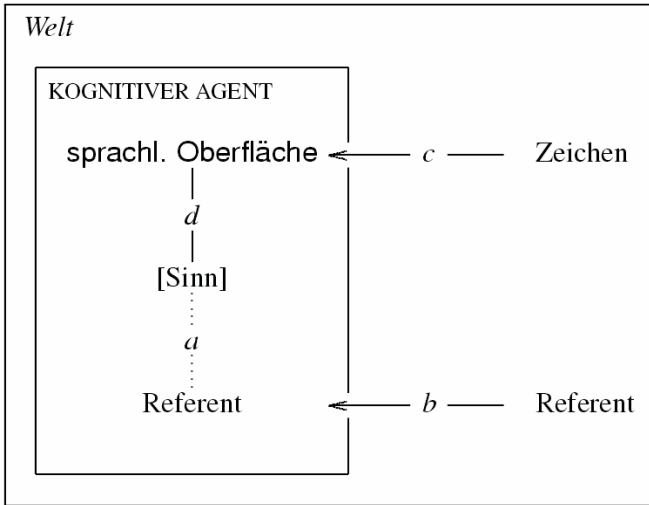
6.1 Vagheit in einer [-Sinn, -konstruktiv]-Semantik

<i>Welt</i>		
sprachliche Oberfläche:	[Die Tür ist offen]	und [die Tür ist rot]
Referenten:	0,615	0,423

In 6.1 werden den Propositionen A und B die Wahrheitswerte 0,615 und 0,423 als Referenzobjekte zugeordnet. Wie die Propositionen zu diesen Wahrheitswerten kommen, betrachtet man als etwas, das nicht im Zuständigkeitsbereich der logischen Theorie liegt. Stattdessen konzentriert sich das ganze Interesse auf die Frage, welchen Wahrheitswert die komplexe Proposition 'A & B' bekommen soll.

Ganz anders sieht die Analyse dieses Beispiels in einer [+Sinn, +konstruktiv]-Ontologie aus. Die Struktur dieser Ontologie weist vier verschiedene Stellen auf, aus denen das Phänomen der Vagheit hergeleitet werden kann.

6.2 Vagheit in einer [+Sinn, +konstr.]-Semantik



Die Stelle *d* entspricht der gepunkteten Linie in 6.1, weil hier die wahrheitskonditionalen Bedeutungseigenschaften der sprachlichen Oberflächen charakterisiert werden. Die Stellen *a*, *b* und *c* sind aufgrund der geänderten Ontologie neu hinzugekommen und haben im Unterschied zu *d* gemeinsam, daß sie Schnittstellen darstellen, die auf einer Abpassungsprozedur beruhen.

Im Rahmen der [+Sinn, +konstruktiv]-Ontologie ist die Stelle *a* am naheliegendsten für eine Behandlung der Vagheit. Denn dort wird eine sprachliche Bedeutung, z. B. das Konzept von rot, auf eine stark eingeschränkte Menge potentieller kontextueller Referenten abgepaßt. Dieser Abpassungsprozeß beruht auf dem pragmatischen Prinzip des *best match*.

So kann man beispielsweise mit dem Wort rot auf einen grau-rosa Stein referieren, wenn die anderen Steine im Verwendungskontext alle grau sind. Wird nun zusätzlich ein knallroter Stein in den Verwendungskontext gelegt, verändert sich der Kandidat für *best match* und der grau-rosa Stein ist im weiteren Verlauf des Dialogs zu den 'nicht-roten' Gegenständen zu rechnen. Dies liegt nicht an spezi-

ellen Eigenschaften des Farbkonzepts, sondern nur an einer Veränderung des Verwendungskontexts – und mit ihm des Kandidaten für *best match*.

Ein zweiter und dritter Aspekt der Vagheit kann an den Stellen *b* und *c* in 6.2 entstehen, und zwar aufgrund von Schwierigkeiten bei der Wahrnehmung. Bei der Stelle *b* ist es eine ungenaue Wahrnehmung des Kontexts. Bei Stelle *c* ist es dagegen eine ungenaues Hören oder Aussprechen. In beiden Fällen leiten sich Aspekte der Vagheit aus Problemen bei der Interaktion des kognitiven Agenten mit seiner Umwelt her, die aber in der sprachlichen Kommunikation auch auf den pragmatischen Prozeß *a* auswirken.

Somit folgt die vermeintliche Vagheit der Farbwörter nicht aus der Semantik ihrer Konzepte. Sie ist vielmehr ein Ergebnis der internen Abpassungsprozedur zwischen einem M-Konzept und kontextuellen Referenten (I-Konzepte_{loc}) im Rahmen der Pragmatik, insbesondere dem pragmatischen Prinzip des *best match* in der Situation eines eingeschränkten Teilkontexts. Diese Analyse der Semantik und Pragmatik der Farbwörter kann problemlos in der Konstruktion eines entsprechenden Roboters operational realisiert werden, wobei M-Konzepte wie rot als elektromagnetische Frequenzintervalle definiert werden.

7 Absolute und kontingente Wahrheit

In der Logik steht der Begriff der ‘Aussage’ für Sätze, bei deren semantischer Interpretation die konkreten Umstände der Äußerung möglichst vernachlässigt werden. Aus Sicht der natürlichen Sprachen ist dieser Begriff problematisch, weil er eine Zwitterstellung zwischen den Begriffen der *Äußerung* (pragmatisch interpretiertes oder interpretierbares Token) und des *Ausdrucks* (pragmatisch uninterpretierter Type) einnimmt. Diese Problematik tritt bei einer Betrachtung der traditionellen Unterscheidung zwischen absoluten und kontingenten Aussagen deutlich zutage.

Absolute Aussagen beziehen sich auf wissenschaftliche und mathematisch-logische Inhalte. Zum Beispiel haben bei der Aussage

Im rechtwinkligen Dreieck gilt für die Hypotenuse A und die Katheten B und C, daß $A^2 = B^2 + C^2$

die Umstände der Äußerung keinen Einfluß auf die Interpretation und den Wahrheitswert des in ihr verwendeten Satzes, weshalb sie vernachlässigt werden. Diesen speziellen Eigenschaften absoluter Aussagen entspricht der logische Wahrheitsbegriff. Er drückt sich darin aus, daß die metasprachlichen Wörter falsch

und wahr auf die mengentheoretischen Objekte \emptyset und $\{\emptyset\}$ der Modellstruktur referieren.

Kontingente Aussagen basieren dagegen auf alltäglicheren Sätzen, wie z. B.

Deinem Hund geht es gut.

Kontingente Aussagen können nur sinnvoll interpretiert – und dabei bzgl. ihres Wahrheitswerts geprüft – werden, wenn die relevanten Umstände der Äußerung bekannt sind und systematisch in den formalen Interpretationsprozeß eingebracht werden. Dies erfordert, daß die Parameter des Zeichenursprungs bekannt sind, also der Ort S, die Zeit T, die Person des Sprechers A und die Person des Adressaten R.

Den speziellen Eigenschaften der kontingenten Aussagen entspricht der natürliche Wahrheitsbegriff, im folgenden repräsentiert durch die Wahrheitswerte wahr^k und falsch^k. Intuitiv hat eine kontingente Aussage wie

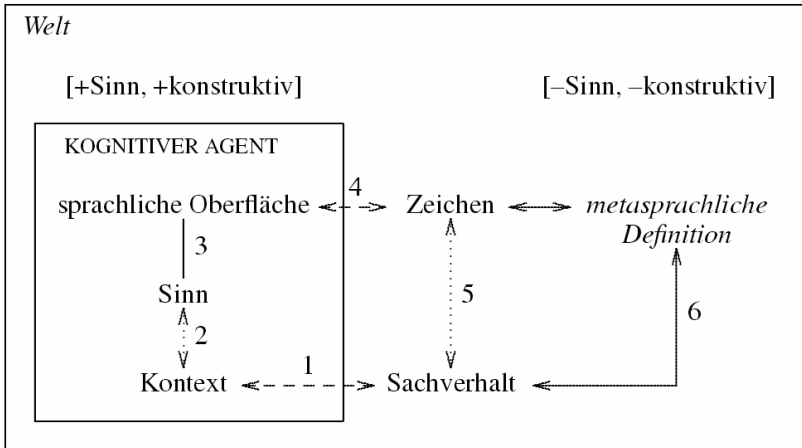
Die Perser haben die Schlacht verloren

z. B. den Wert wahr^k, wenn der Sprecher als Augenzeuge den Sachverhalt korrekt beurteilt und kommuniziert, oder wenn eine korrekt funktionierende Mitteilungskette zwischen dem Sprecher und einem zuverlässigen Augenzeugen existiert.

Die natürlichen Wahrheitswerte wahr^k und falsch^k haben eine prozedurale Definition: Die 'Aussage', oder – korrekt – die Äußerung, eines kognitiven Agenten erhält den Wert wahr^k, wenn alle an dem Kommunikationsvorgang beteiligten Prozesse korrekt funktionieren, und sonst den Wert falsch^k.

Die unterschiedlichen Wahrheitsbegriffe der natürlichen und der logischen Semantik leiten sich aus den strukturellen Unterschieden zwischen der [+Sinn, +konstruktiv]- und der [-Sinn, -konstruktiv]-Ontologie her, auf der sie jeweils beruhen.

7.1 Grundlagen von natürlicher und logischer Wahrheit



In beiden Systemen ist die Beziehung 5 zwischen dem externen sprachlichen Zeichen (Satz) und dem realen Sachverhalt entscheidend für die Wahrheit einer Aussage. Allerdings wird die Beziehung 5 von den beiden Ontologien sowohl methodisch als auch konzeptionell vollkommen unterschiedlich realisiert.

Ein [-Sinn, -konstruktiv]-System definiert die Beziehung 5 direkt mit Hilfe einer geeigneten Metasprache 6. Die Analyse liegt beim Logiker, der sich – im Einklang mit seinen ontologischen Voraussetzungen – allein auf die Wahrheitsbeziehung zwischen Aussage (bzw. Sprachzeichen) und Sachverhalt beschränkt, wobei er von allen Strukturaspekten der Kommunikation abstrahiert. Das logische Modell und die regelbasierte Interpretation des Zeichens vollzieht formal nach, was von vornherein als offensichtlich vorausgesetzt wird. Das System hat die explizite Ableitung von Wahrheitswerten als Ziel.

Bei einem [+Sinn, +konstruktiv]-System ist dagegen eine reale Handlungsumgebung vorgegeben. Sie wird vom kognitiven Agenten in bestimmten Teilaspekten automatisch analysiert, wobei eine entsprechende interne Kontextrepräsentation aufgebaut wird. Die Beziehung 5 zwischen sprachlichen Oberflächen und externen Sachverhalten wird hier also indirekt über kognitive Prozesse eta-

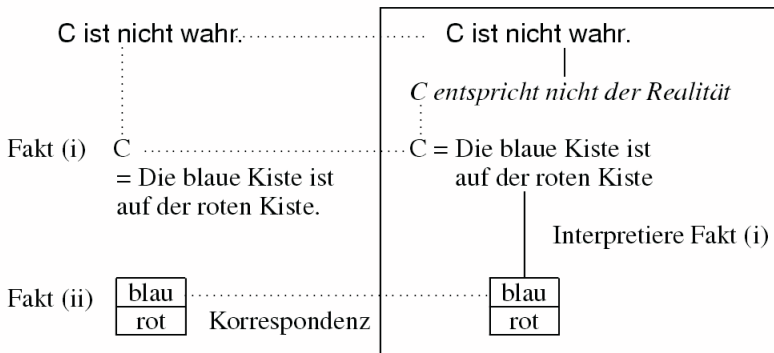
bliert, die sich aus den Komponenten 1 (nicht-verbale Erkennung/Handlung), 2 (pragmatische Interpretation), 3 (semantische Interpretation) und 4 (verbale Erkennung/Handlung) zusammensetzen. Das System hat die natürliche Kommunikation von kontextuellen Inhalten als Ziel.

8 Epimenides in einem [+Sinn,+konstruktiv]-System

Die natürlichen Wahrheitswerte wahr^k und falsch^k ermöglichen eine Reanalyse des Epimenides-Paradoxons. Sie erlaubt es, daß Objektsprachen die Wörter wahr und falsch enthalten können, ohne daß ihre Semantik deshalb inkonsistent wird.

Als Vorbereitung interpretieren wir zunächst eine gutartige Verwendung des Ausdrucks C ist nicht ein wahrer Satz. Dieser Ausdruck, den Tarski zur Ableitung des Epimenides-Paradoxons verwendet, besteht aus einer sprachlichen Abkürzung, C, und einer negativen Wahrheitsaussage. Seine legitime Verwendung im Rahmen eines [+Sinn, +konstruktiv]-Systems basiert auf der folgenden Struktur.

8.1 Gutartiger Fall einer Linguistischen Abkürzung



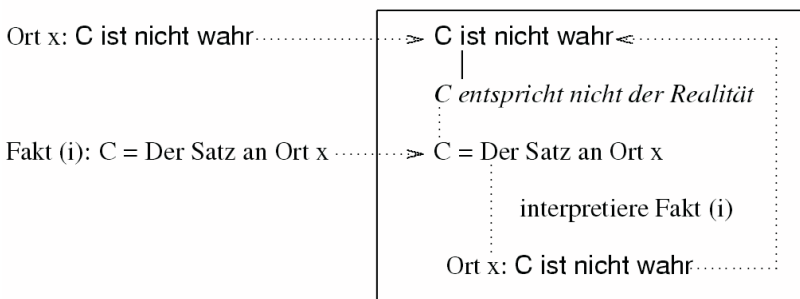
C dient als Abkürzung für Die blaue Kiste ist auf der roten Kiste. Die Abkürzung ist als Fakt (i) in der externen Handlungsumgebung dargestellt. Außerdem enthält sie als Fakt (ii) den Sachverhalt, den der mit C abgekürzte Satz beschreibt.

Wenn ein [+Sinn,+konstruktiv]-System, z.B. CURIOS, den Ausdruck C ist nicht (ein) wahr(er Satz) im Hörer-Modus verarbeitet, ordnet seine semantische Interpretationsprozedur der Oberfläche eine Meaning_1 (wörtliche Zeichenbedeutung) zu, die als *C entspricht nicht der Realität* paraphrasiert werden kann. Dieser Meaning_1 versucht die pragmatische Interpretationsprozedur eine entsprechende Kontextstruktur zuzuordnen.

Dabei stellt sich heraus, daß C über Fakt (i) als Abkürzung des Satzes Die blaue Kiste ist auf der roten Kiste definiert ist. Den verbleibenden Teil des Eingabesatzes, ist nicht wahr, verarbeitet die pragmatische Komponente, indem sie überprüft, ob der Inhalt der langen Version von C mit der Realität übereinstimmt. Der Meaning_1 von Die blaue Kiste ist auf der roten Kiste wird also die entsprechende kontextuelle Repräsentation der externen Handlungsumgebung, nämlich Fakt (ii), zugeordnet, wobei sich herausstellt, daß sie übereinstimmen. Somit hat die ursprüngliche Eingabe C ist nicht wahr den Wahrheitswert falsch^k. Es gibt beliebig viele Beispiele von linguistischen Abkürzungen in Kombination mit natürlichen Wahrheitsprädikaten, die ebenso gutartig wie normal sind.

Ein Spezialfall einer sprachbasierten Abkürzung ist das Epimenides-Paradox. Seine [+Sinn,+konstruktiv]-Reanalyse hat die folgende Struktur.

8.2 Rekonstruktion des Epimenides-Paradoxons



An einem deutlich gekennzeichneten Ort x liest der Roboter C ist nicht (ein) wahr(er Satz), was semantisch als C entspricht nicht der Realität repräsentiert wird. Wie in 8.1 versucht die Pragmatik, dieser semantischen Repräsentation eine entsprechende Kontextstruktur zuzuordnen.

Unter Berücksichtigung von Fakt (i) stellt sich heraus, daß C als Abkürzung von Der Satz an Ort x definiert ist. Den verbleibenden Teil des Eingabesatzes, ist nicht wahr, verarbeitet die pragmatische Komponente, indem sie überprüft, ob der Inhalt dessen, was C abkürzt, mit der Realität übereinstimmt. Der Meaning_1 von Der Satz an Ort x wird also die entsprechende kontextuelle Repräsentation der externen Handlungsumgebung zugeordnet. Im Unterschied zu 8.1, wo die Meaning_1 von Die blaue Kiste ist auf der roten Kiste auf den nicht-sprachlichen Fakt (ii) abgepaßt wird, verweist in 8.2 die Meaning_1 von Der Satz an Ort x auf das sprachliche Referenzobjekt (Zeichen) C ist nicht wahr.

An diesem Punkt kann das Referenzobjekt C ist nicht wahr von der Pragmatik als uninterpretiertes oder als interpretiertes Zeichen behandelt werden. Eine Behandlung als uninterpretiertes Zeichen wäre zum Beispiel pragmatisch sinnvoll in der Kombination mit *ist in Sans-Serif gedruckt*. In 8.2 würde eine Behandlung als uninterpretiertes Zeichen jedoch keinen Sinn machen. Vielmehr besteht die nun scheinbar naheliegenste Aktion darin, das Zeichen zu interpretieren – womit die semantisch-pragmatische Interpretationsprozedur wieder von vorne beginnt.

Somit kann ein [+Sinn,+konstruktiv]-System, das durch die äußeren Umstände in die spezielle Situation des Epimenides-Paradoxons gebracht wird, ohne zusätzliche Annahmen in einen blinden Kreislauf geraten und dort verharren. Wie in 8.2 schematisch dargestellt, kann das C in C ist nicht wahr immer wieder mit dem entsprechenden Satz an Ort x gleichgesetzt werden.

Unsere ontologisch basierte Reanalyse des Epimenides-Paradox resultiert nicht in seiner Auflösung, sondern in seiner Verwandlung. Was in Tarskis [-Sinn, -konstruktiv]-System ein logischer Widerspruch auf der semantischen Ebene ist, wird in der [+Sinn, +konstruktiv]-Reanalyse der SLIM-Sprachtheorie zu einer infiniten Rekursion der semantisch-pragmatischen Interpretation. Diese Reanalyse entschärft das Epimenides-Paradox sowohl semantisch als auch kommunikationstheoretisch:

- In einem [+Sinn,+konstruktiv]-System können die Wörter wahr^k und falsch^k in der zu interpretierenden Objektsprache enthalten sein, ohne daß deshalb in der Semantik der Objektsprache ein logischer Widerspruch auftritt.

- Die Rekursion, die das Epimenides-Paradox in seiner [+Sinn,+konstruktiv]-Analyse verursacht, kann in der Pragmatik erkannt und als scheiternder Interpretationsversuch²¹ entsorgt werden, ohne daß dadurch die kommunikative Funktionsfähigkeit des Systems beeinträchtigt würde.

Die Reanalyse vermeidet den Tarskischen Widerspruch in der Semantik, weil die Metasprache zwischen (i) den logischen Wahrheitswerten 1 und 0 aus der T-Bedingung, (ii) den natürlichen Wahrheitswerten wahr^k und falsch^k aus dem objektsprachlichen Satz C sowie (iii) deren prozeduralen metasprachlichen Korrelaten *entspricht (nicht) der Realität* unterscheidet. Wenn wir z. B. annehmen, daß CURIOUS eine logische Semantik wie die Montague-Grammatik verwenden würde, dann würde seine [+Sinn, +konstruktiv]-Interpretation des Epimenides-Paradoxons nicht in dem Tarskischen Widerspruch

a. C ist 1 dann und nur dann, wenn C nicht 1 ist.

resultieren, sondern stattdessen in der kontingenten Formulierung

b. C ist 1 dann und nur dann, wenn C nicht der Realität entspricht.

Im Gegensatz zu a weist b keinen logischen Widerspruch auf.

Für die Semantik der natürlichen Sprachen ist die Reanalyse des Epimenides-Paradox (kontingente Formulierung b) von großer Bedeutung: Indem sie den Tarskischen Widerspruch vermeidet, öffnet sie den Weg für eine *vollständige* formale Semantik der natürlichen Sprachen – also eine Semantik, die nicht von vornherein bestimmte Sätze ausschließen muß (weil sie die Wörter wahr oder falsch enthalten).

Für den Versuch einer logischen Semantik der natürlichen Sprachen ist diese Reanalyse dagegen kein Gewinn. Denn die natürlichen Wahrheitswerte wahr^k und falsch^k – die für die Vermeidung des Tarskischen Widerspruchs benötigt werden – können nur im Rahmen einer [+Sinn, +konstruktiv]-Ontologie konzeptionell motiviert und prozedural implementiert werden.

9 Wahrheit in den natürlichen Sprachen

Die Verwendung verschiedener Wahrheitsbegriffe für absolute und kontingente Äußerungen ('Aussagen') wäre aus logischer und linguistischer Sicht gleichermaßen suboptimal. Das Ziel ist vielmehr ein Gesamtsystem, dessen einheitliche Semantik in der Lage ist, beliebige Äußerungen ('Aussagen') der Form C ist

wahr korrekt zu interpretieren, unabhängig davon, ob es sich bei C um einen kontingenten oder einen absoluten Teilsatz handelt.

Eine naheliegende Methode, die Semantik absoluter und kontingenter Sätze zu vereinheitlichen, besteht darin, den einen Äußerungstyp (‘Aussagetyp’) als Spezialfall des anderen zu analysieren. Für den [-Sinn,-konstruktiv]-Ansatz wäre es daher wünschenswert, wenn die auf absolute Äußerungen (‘Aussagen’) ausgerichtete logische Semantik auf eine allgemeine Behandlung kontingenter Äußerungen ausgedehnt werden könnte. Umgekehrt wäre es für den [+Sinn,+konstruktiv]-Ansatz wünschenswert, wenn die auf kontingente Äußerungen ausgerichtete natürliche Semantik eine Behandlung absoluter Sätze als Spezialfall kontingenter Sätze erlaubt.

In der logischen Semantik kann die Behandlung der absoluten Äußerungen zwar in vielen Fällen auf eine Behandlung der kontingenten Äußerungen ausgedehnt werden – wie dies die modelltheoretische Analyse des Englischen bei Montague illustriert. Erstens zeigt jedoch das Beispiel der propositionalen Einstellungen (Sektion 4), daß eine befriedigende semantische Interpretation – also eine auch ontologisch vertretbare Zuordnung der logischen Wahrheitswerte 1 oder 0 – *nicht immer* möglich ist. Zweitens ist laut Tarski die Interpretation von objektsprachlichen Sätzen der Form C ist (nicht) wahr in der logischen Semantik ohnehin verboten. Aus diesen beiden Gründen ist eine generelle Analyse der kontingenten Äußerungen als Sonderfall der absoluten Äußerungen im Rahmen der logischen Semantik grundsätzlich ausgeschlossen.

In der natürlichen Semantik sind absolute Äußerungen dagegen nur insofern speziell, als sie (i) aufgrund ihres Inhalts unabhängig von der Kenntnis ihrer Ursprungsparameter interpretiert werden können und (ii) die kognitive Verantwortung für den Inhalt vom Sprecher auf die Gesellschaft und ihr historisch gewachsenes Weltbild übertragen wird. Somit hat z.B. eine absolute Äußerung wie Wasser hat die chemische Formel H_2O hat den Wert wahr^k, wenn eine korrekt funktionierende Mitteilungskette zwischen dem Sprecher und den in der Sprachgemeinschaft zuständigen Spezialisten existiert.²² Die wahren Sätze absoluter naturwissenschaftlicher und logisch-mathematischer Systeme werden also kontingent rekonstruiert, indem sie als kognitive Leistung der zugehörigen menschlichen – und somit fehlbaren – Gemeinschaft interpretiert werden.

Aus dieser anthropologischen Sicht der Wissenschaftsgeschichte ist es normal, wenn absolute Äußerungen zu einer bestimmten Zeit aufgrund der Expertenmehrheit als wahr^k gelten, sich aber später als falsch^k herausstellen. Für derartige Irr-

tümer gibt es in der Wissenschaftsgeschichte zahlreiche Beispiele, z.B. Feuer beruht auf der materiellen Substanz des Phlogiston.

Anmerkungen

- 1 Tarski 1944 beklagt sich über diese Mißverständnisse und widmet die zweite Hälfte des Aufsatzes einer detaillierten Kritik seiner Kritiker.
- 2 Weil die Semantik der Programmiersprachen prozedural (d.h. metasprachunabhängig), die Semantik der Logiksprachen dagegen Tarskisch (d.h. metasprachabhängig) ist, erweist sich die Nachbildung logischer Kalküle auf dem Computer als bestenfalls schwierig – wie dies z.B. die programmiertechnische Realisierung des Prädikatenkalküls in der Form von Prolog illustriert.
- 3 Eine genauere Analyse der schwachen Variante(n) gibt C. Thiel, 1995, S. 325–7.
- 4 Die Seite und die Zeilennummer sind von Tarskis Orginaltext auf die hiesige Umgebung angepaßt worden. Diese Anpassung ist eine notwendige Voraussetzung dafür, daß die Selbstreferenz in der beabsichtigten Weise funktionieren kann.
- 5 "[N]o rational ground can be given why substitution should be forbidden in general."
- 6 Dies folgt aus der Rolle der natürlichen Sprachen als prätheoretische Metasprache der logischen Sprachen. Ohne die Wörter *wahr* und *falsch* in den natürlichen Sprachen könnte eine logische Semantik gar nicht definiert werden.
- 7 S. 188 in Montague 1974.
- 8 Als Vorschlag zur Güte meint Davidson, man könne sich doch darauf beschränken, nur möglichst große konsistente Teile in den natürlichen Sprachen zu interpretieren. Damit ist aber das Projekt einer logisch-semantischen Interpretation der natürlichen Sprachen als ganzes gescheitert.
 Versuche, das Epimenides-Paradox in der logischen Semantik zu vermeiden, sind Kripke 1975, Gupta 1982 und Herzberger 1982. Diese Systeme definieren jeweils eine künstliche Objektsprache (Prädikatenkalkül erster Stufe) mit Wahrheitsprädikaten. Daß diese Objektsprachen dennoch konsistent sind, beruht auf der Definition der Wahrheitsprädikate als *rekursive Valuationsschemata*.
 Rekursive Valuationsschemata basieren auf einer großen Anzahl von Valuationen (transfinit im Fall von Kripke 1975) – was an sich schon eine klare Verletzung der metasprachlichen Methode konstituiert. Hinzu kommt, daß rekursive Valuationsschemata den inhaltlichen Problempunkt des Epimenides-Paradoxes, verfehlen. Es handelt sich nämlich um ein Problem der Referenz: ein Symbol kann aufgrund seiner Bedeutung referieren und gleichzeitig aufgrund seiner Form (Oberfläche) als Referent dienen.

⁹ Rein formal könnte man zu diesem Zweck einen ‘Glaubensoperator’ G wie folgt definieren:

$G(x, p)^{M,i,j,g}$ ist 1 d.u.n.d., wenn $p^{M,b,j,g}$ 1 ist, wobei b eine Glaubenswelt von x am Index i,j ist.

Man sollte sich von der scheinbar anspruchsvollen Formalisierung, die sich notationell an Montagues PTQ hält, nicht täuschen lassen, denn als metasprachliche Definition ist diese T-Bedingung genauso inhaltsleer wie 2.1, solange die Verifikation relativ zu Glaubenswelten ungeklärt ist.

- ¹⁰ Wie Glaubensvorstellungen sind auch Empfindungen in der logischen Semantik ontologisch problematisch. In der sogenannten *double aspect theory* wird versucht, einen Begriff wie z.B. Schmerz, der im Gegensatz zu realen Dingen nicht in der Welt existiert, sondern auf die Empfindung einzelner Individuen beschränkt ist, durch das Messen von Gehirnströmen wahrheitskonditional zu objektivieren. Indem man dem Phänomen ‘Schmerz’ die beiden (*double*) Aspekte (i) Gefühl und (ii) korrespondierende Meßwerte zuordnet, soll dieses Phänomen ontologisch in einer Weise untermauert werden, die für die logische Semantik akzeptabel ist. Eine Übertragung dieses Ansatzes auf eine wahrheitskonditionale Analyse des *Glaubens* würde unfehlbare Lügendetektoren voraussetzen.
- ¹¹ Nach der noch immer gängigen – aber aus Sicht der heutigen Naturwissenschaften vollkommen überholten – Auffassung in der philosophischen Logik macht es die Darstellung wissenschaftlicher Wahrheit nicht erforderlich, den oder die Sprecher/Hörer in die Modellierung mit aufzunehmen. Der gelegentliche ‘Einbau’ von Sprecher/Hörern in die modelltheoretische Logik dient allein der Behandlung spezieller natürlich-sprachlicher Phänomene, insbesondere der Interpretation indexikalischer Pronomina wie *ich* und *du*. Dabei ist der Sprecher/Hörer grundsätzlich Teil der Modellstruktur – was eine wahrheitstheoretisch adäquate Behandlung der propositionalen Einstellungen aufgrund der dargestellten Schwierigkeiten ausschließt. Eine gut lesbare kritische Darstellung der veralteten *received view* in der Wissenschaftstheorie und ihrer Alternativen findet sich in F. Suppe 1977.
- ¹² Aus der Mathematik sind zwar Beispiele bekannt, wo eine formale Theorie verschiedene Interpretationen erlaubt, z.B. in der Geometrie. Dies bedeutet jedoch nicht, daß jede formale Theorie für jede gewünschte Interpretation verwendet werden kann. Speziell die logische Semantik ist ein Fall, wo der Formalismus nicht gleichzeitig für eine allgemeine Beschreibung der Wahrheit und der Bedeutung geeignet ist – wie das Phänomen der propositionalen Einstellungen zeigt.
- ¹³ In dem Sinn, daß sprachliche Aussagen automatisch in korrespondierende mechanische oder elektronische Operationen umgesetzt werden.
- ¹⁴ Die intensionale Logik von Carnap und Montague ist [-Sinn] weil sprachliche Ausdrücke dort auf die ‘Welt’ (Modellstruktur) referieren. Abgesehen von der Definition

von einigen zusätzlichen Operatoren, besteht der einzige Unterschied zwischen einer intensionalen und einer entsprechenden extensionalen Logik darin, daß in der intensionalen Logik die Welt nicht als ein Modell, sondern als eine Modellstruktur definiert wird. Die Modellstruktur repräsentiert die verschiedenen Zustände der Welt über eine Vielzahl von Modellen mit verschiedenen Indizes. Diese Indizes dienen als Domäne für Funktionen, die Carnap 'Intensionen' nennt.

- ¹⁵ In diesem Sinne verwehrt sich Frege ausdrücklich gegen eine Interpretation seines Systems als Darstellung kognitiver Zustände, die er als 'psychologistisch' bezeichnet. In neuerer Zeit wurde in der 'Situationssemantik' (Barwise & Perry 1983) und der 'Diskursesemantik' (Kamp & Reyle 1993) versucht, den Typ der [+Sinn, -konstruktiv]-Semantik wiederzubeleben. Seine inhärent antikognitive Sichtweise wird bei Barwise & Perry 1983 durch Diagramme (op.cit., S. 226) unmißverständlich verdeutlicht.
- ¹⁶ "[...] in an unnecessary and unparsimonious multiplication of hypothetical entities that has no evidential support."
- ¹⁷ Beispiele sind das Schachspiel (Newell & Simon, Reddy et al.) und die Blockswelt (Winograd).
- ¹⁸ Das Akronym SLIM steht für *Surface compositional Linear Internal Matching*.
- ¹⁹ Dies wird formal in der Definition der Mengen A, I und J einer Modelstruktur *MS* festgelegt.
- ²⁰ Dies wird formal in der Denotationsfunktion F einer Modelstruktur *MS* definiert.
- ²¹ Für die pragmatische Interpretation gilt ganz allgemein, daß eine ständige Wiederholung in der Analyse ein und desselben kontextuellen Objekts zu vermeiden ist – zum Beispiel mit Hilfe eines Zählers. Auf diese Weise wird auch die durch das Epimenides-Paradox verursachte Rekursion erkannt, was dazu führt, daß der Analyseversuch als uninterpretierbar abgebrochen wird. Abgebrochene Interpretationsversuche gehören in der Pragmatik zum Alltag und werden durch den systematischen Einsatz alternativer Interpretationsschemata sowie kommunikativer Strategien wie z.B. Rückfragen überwunden.
- ²² Das Konzept der 'kausalen Kette' von einem Sprecher zum nächsten, besonders bei der Referenz der Eigennamen und natürlichen Klassen, betont Kripke 1972. Die zentrale Rolle der 'Spezialisten' für die wissenschaftliche Spezifikation bestimmter Bedeutungen in der Gesellschaft – z.B. Wasser als H₂O – wird besonders von Putnam 1975a hervorgehoben, allerdings mit der absurden Folgerung, *that meanings just ain't in the head* (op.cit., S.227).
Bedeutung im Sinn von Referenzetablierung wird von diesen Autoren als Voraussetzung für die Fundierung der Wahrheit untersucht. Leider übersehen sie dabei aber

die notwendigen Unterscheidungen zwischen (i) logischer und natürlicher Semantik, zwischen (ii) [-Sinn,-konstruktiv]- und [+Sinn,+konstruktiv]-Ontologie, und zwischen (iii) absoluter und kontingenter Wahrheit.

Bibliographie

- Anderson, J.R. und G.H. Bower (1973) *Human Associative Memory*. V.H. Winston, Washington, D.C.
- Barwise, J. und J. Perry (1983) *Situations and Attitudes*. The MIT Press, Cambridge, Massachusetts.
- Davidson, D. (1967) "Truth and Meaning," *Synthese*, VII:304–323.
- Davidson, D. und G. Harman (Hg.) (1972) *Semantics of Natural Language*, D. Reidel Publishing Company, Dordrecht-Holland.
- Frege, G. (1967) *Kleine Schriften*, hrsg. von Ignacio Angelelli, Wiss. Buchgesellschaft, Darmstadt. Gupta, A. (1982) "Truth and Paradox," *Journal of Philosophical Logic*, 11:1–60.
- Hausser, R. (1989) *Computation of Language, An Essay on Syntax, Semantics and Pragmatics in Natural Man-Machine Communication*, Symbolic Computation: Artificial Intelligence, Springer-Verlag, Berlin–New York.
- Hausser, R. (1992) "Complexity in Left-Associative Grammar," *Theoretical Computer Science*, Vol. 103, Elsevier.
- Hausser, R. (1996) "A Database Interpretation of Natural Language," *Korean Journal of Linguistics*, Vol. 21, No. 1,2:29–55. Hausser, R. (1999) *Foundations of Computational Linguistics, Man-machine communication in natural language*, Springer-Verlag, Berlin–New York.
- Herzberger, H. (1982) "Notes on Naive Semantics," *Journal of Philosophical Logic*, 11:61–102.
- Johnson-Laird, P.N. (1983) *Mental Models*. Harvard University Press, Cambridge, Massachusetts.
- Kamp, J.A.W. und U. Reyle (1993) *From Discourse to Logic*, Part 1 und 2, Kluwer Academic Publishers, Dordrecht, Boston, London.
- Kripke, S. (1972) "Naming and Necessity," in Davidson und Harman (Hg.), p.253–355.
- Kripke, S. (1975) "Outline of a theory of truth," *The Journal of Philosophy*, 72:690–715.

- Lakoff, G. (1972) "Linguistics and natural logic," in Davidson, D. und G. Harman (Hg.), p. 545–665.
- Montague, R. (1974) *Formal Philosophy*, Yale University Press, New Haven, Connecticut.
- Pascal, B. (1951) *Pensées sur la Religion et sur quelque autre sujets*, éditions Du Luxembourg, Paris.
- Putnam, H. (1975a) "The meaning of 'meaning'," reprinted in Putnam 1975b, p. 215–271.
- Putnam, H. (1975b) *Mind, Language, and Reality 2*. Cambridge University Press.
- Quine, W.v.O. (1960) *Word and Object*. MIT Press, Cambridge, Massachusetts.
- Reddy, D.R., L.D. Erman, R.D. Fennell, und R.B. Neely (1973) "The Hearsay Speech Understanding System: An Example of the Recognition Process," *Proceedings of the Third International Joint Conference on Artificial Intelligence*, Stanford, California.
- Rescher, N. (1969) *Many-valued Logic*. McGraw-Hill, New York.
- Schank, R.C. und R. Abelson (1977) *Scripts, Plans, Goals, and Understanding*. Lawrence Earlbaum, Hillsdale, New Jersey.
- Scott, D. und C. Strachey (1971) "Toward a Mathematical Semantics of Computer Languages," Technical Monograph PRG–6, Oxford University Computing Laboratory, Programming Research Group, 45 Branbury Road, Oxford.
- Suppe, F. (Hg.) (1977) *The Structure of Scientific Theories*, University of Illinois Press.
- Tarski, A. (1935) „Der Wahrheitsbegriff in den Formalisierten Sprachen," *Studia Philosophica*, Vol. I, 262–405.
- Tarski, A. (1944) "The Semantic Concept of Truth," *Philosophy and Phenomenological Research* 4:341–375.
- Thiel, C. (1995) *Philosophie und Mathematik*, Wissenschaftliche Buchgesellschaft, Darmstadt.
- Winograd, T. (1972) *Understanding Natural Language*, Academic Press, Harcourt Brace Jovanovich, San Diego, New York.
- Weyhrauch, R. (1980) "Prolegomena to a Formal Theory of Mechanical Reasoning," *Artificial Intelligence*. Reprinted in Webber und Nilsson (eds.) 1981.

Zweimal zwei Ontologien für die Semantik?

Eine Replik auf Die vier Elementar-Ontologien semantischer Interpretation und ihre empirischen Folgen von Roland Hausser

Bernhard Schröder
IKP
Universität Bonn
B. Schroeder@uni-bonn.de

Machen semantische Theorien Annahmen über das, was es ist, was sprachliche Ausdrücke bedeuten? Und wenn ja, haben diese Annahmen Folgen für das, was die semantischen Theorien behaupten? Roland Hausser beantwortet in seinem Beitrag *Die vier Elementar-Ontologien semantischer Interpretation und ihre empirischen Folgen* in diesem Heft beide Fragen mit einem klaren Ja.

Hier soll dafür argumentiert werden, dass ontologische Unterscheidungen bei der Einordnung semantischer Theorien und der Diagnose und Lösung der von Hausser angesprochenen Probleme semantischer Ansätze eine geringe Rolle spielen. Insgesamt soll für drei Gegenthesen argumentiert werden:

I) Neben der [+/-Sinn]-Unterscheidung gibt es eine Unzahl weiterer ontologischer Unterscheidungsmöglichkeiten, die in ähnlich einschneidender Weise semantische Ansätze abgrenzen.

II) Erst die Vernetzung von Theorien der formalen Semantik mit kognitionswissenschaftlichen Theorien kann die Eigenschaften, die [+Sinn]-Theorien von Hausser zugeschrieben werden, erklären. Durch die Vernetzung ergeben sich für die formale Semantik jedoch keine neuen ontologischen Aspekte.

III) Theorien mit der Eigenschaft [+konstruktiv] sind in keiner wesentlich anderen Lage gegenüber dem Epimenides-Paradox als Theorien ohne diese Eigenschaft.

Nicht zuletzt mit Zeugnissen und Betrachtungen aus der sprachphilosophischen Diskussion stützt Hausser seine These. Darum soll zunächst skizzenhaft auf die sprachphilosophische Motivation ontologisch-semantischer Überlegungen eingegangen werden.

Ontologie und Sprachphilosophie

Frege's Sinne...

Ontologische Fragen spielten in der Entwicklung der „logischen“ Semantik stets eine bestimmende Rolle, zum einen als Ergebnis sprachphilosophischer Überlegungen – man denke nur an Gottlob Frege's Diktum „Und ein drittes Reich muß anerkannt werden“, mit dem er (nicht etwa in prophetischer Weise auf noch bevorstehende katastrophale politische Entwicklungen hinwies, sondern) das Ergebnis seiner sprachphilosophischen Untersuchung von nicht-extensionalen Kontexten, konkret Glaubenskontexten, zusammenfasst: Eine Reihe von sprachphilosophischen Problemen sind nur lösbar, wenn man solche Gegenstände wie „Gedanken“ oder allgemeiner „Sinne“ anerkennt, meint Frege.

Sprachphilosophie ist für den Mathematiker Frege aber nur Mittel zum Zweck der Durchführung seines mathematisch-logizistischen Programms. Er möchte zeigen, dass zumindest die Sätze der Algebra keine synthetischen Erkenntnisse a priori sind, wie Kant annahm, sondern analytisch sind, oder weniger kantisch ausgedrückt: Alle wahren Sätze der Algebra sind logisch wahr. Nur muss Frege einen möglichen Einwand gegen seine These entkräften: Wenn alle wahren Sätze der Algebra logisch wahr sind, wie kommt es dann, dass der Satz

$$(1) \quad 2^8=256$$

informativer erscheint als

$$(2) \quad \text{Die Venus ist die Venus.}$$

Auch in (1) bezeichnen ja schließlich 2^8 und 256 dieselbe Zahl, anderenfalls könnte die Identitätsbehauptung nicht wahr sein. Aber beide Ausdrücke bezeichnen diese Zahl in unterschiedlicher Weise. Darauf baut Frege die Lösung des Problems auf, indem er zur „Bedeutung“ eine zweite semantische Komponente ins Spiel bringt, den „Sinn“ eines Ausdrucks oder die „Gegebenheitsweise“ des durch diesen Ausdruck bezeichneten Gegenstands, und annimmt, dass diese bei den Ausdrücken zu beiden Seiten des Gleichheitszeichens in (1) verschieden sei. Die Tatsache, dass (1) eine informative Mitteilung darstellen kann, wird analog zur Tatsache erklärt, dass

$$(3) \quad \text{Der Morgenstern ist der Abendstern.}$$

informativ sein kann, wobei *Morgenstern* und *Abendstern* hier jeweils als abgekürzte Kennzeichnungen für einen in bestimmter Weise am Morgen- bzw. am Abendhimmel auszumachenden Himmelskörper zu verstehen sind. Um der Unterscheidung von Sinn und Bedeutung noch eine zusätzliche Rechtfertigung zu geben, argumentiert Frege in *Sinn und Bedeutung* dafür, dass in Glaubenskontexten wie

(4) Karl glaubt, dass der Morgenstern die Venus ist.

oder

(5) Karl glaubt, dass der Abendstern die Venus ist.

die Tatsache, dass (4) und (5) unterschiedliche Wahrheitswerte haben können, dadurch zu begründen sei, dass die Nebensätze jeweils unterschiedliche Sinne hätten und dass in bestimmten Kontexten der gewöhnliche Sinn als Bedeutung genommen würde.

... Carnaps Intensionen...

In *Meaning and Necessity* setzte der logische Empirist Rudolf Carnap genau an Freges Unterscheidung zweier Bedeutungskomponenten an und prägte seine einflussreiche Unterscheidung von Extension und Intension als semantischen Komponenten. Carnaps Anliegen ist hierbei nicht mathematische Grundlagenforschung, sondern eher die Verfolgung seines empiristischen Programms, demzufolge wissenschaftliche Aussagen, sofern sie überhaupt zu rechtfertigen sind, nur von potentiell Beobachtbarem handeln.¹ *Meaning and Necessity* stellt in diesem Programm einen wichtigen Baustein dar, indem Carnap hier zeigt, wie Aussagen z.B. über *mögliche* Ereignisse oder *notwendige* Zusammenhänge auf Aussagen über Ereignisse und Zusammenhänge in möglichen Situationen oder Weltzuständen oder – kurz in leibnizscher Sprechweise – in möglichen Welten zurückführbar seien. Intensionen sind mit Funktionen von möglichen Welten in Extensionen gleichzusetzen. *Morgenstern* und *Abendstern* haben verschiedene Intensionen, weil sie zwar in der tatsächlichen Welt dieselbe Extension, nämlich den Planeten Venus haben, weil es aber andererseits mögliche Welten gibt, in denen das nicht so ist. Es ist denkbar, dass *Morgenstern* und *Abendstern* unterschiedliche Himmelskörper bezeichnen.

Ist Carnap seinem reduktionistischen Ziel nähergekommen, wenn er nun über so ominöse Gegenstände wie mögliche Welten sprechen muss? Durchaus, denn ihrem metaphysisch anmutenden Namen zum Trotz lassen sich mögliche Welten als sehr profane Gegenstände auffassen, nämlich Mengen von Sätzen, und jede mögliche Welt als die Menge von Sätzen, die in dieser wahr ist. **(a)** Mögliche Welten im carnapschen Sinne sind für alle möglichen Sachverhalte dahingehend bestimmt, dass diese Sachverhalte in den möglichen Welten entweder zutreffen oder nicht. Und **(b)** mögliche Welten sind konsistent, d.h. in derselben möglichen Welt können keine zwei Sachverhalte bestehen, aus denen Widersprüchliches folgt. Konstruiert man mögliche Welten als maximal konsistente Satzmen- gen, sind (a) und (b) erfüllt.

In diesem Verständnis sind also mögliche Welten nichts als (maximal konsistente) Mengen möglicher – sprachlich ausgedrückter – Sachverhalte. Sieht man von den formalen Operationen der Mengen-, Relations- und Funktionsbildung ab, die zum „mathematischen“ Apparat formalsemantischer Theorien gehören, handelt die carnapsche Semantik nur von den sprachlichen Ausdrücken selbst und den Gegenständen, die sie ganz im fregeschen Verständnis bedeuten. Freges Sinne, soweit sie Carnaps Intensionen entsprechen, sind mit mengentheoretischen Mitteln aus diesen Gegenständen zu konstruieren.

Carnap selbst hat bereits bemerkt, dass seine Intensionen in einigen Hinsichten eine grobkörnigere semantische Komponente darstellen als Freges Sinne dies tun, da seine Intensionen kein Mittel bieten, die Informativität von (1) zu erklären, sofern (1) logisch wahr ist, oder zu erklären, wie beide Sätze

(6) Karl glaubt nicht, dass $2^8=256$.

(7) Karl glaubt, dass $16*16=256$.

gleichzeitig wahr sein können, da alle logisch wahren Sätze bei Carnap dieselbe Intension haben. Denn wäre die Wahrheit eines Satzes

(8) Karl glaubt, dass p .

nur von der Intension von p abhängig, so müssten (6) und (7) – da ihre Nebensätze als logisch wahre Sätze dieselbe Intension haben – entgegengesetzte Wahrheitswerte haben.

Durch eine Adaptation der Mögliche-Welten-Semantik kann man aber dieses Problem in carnapscher Weise behandeln, wenn man die möglichen Welten in Glaubenskontexten durch mögliche Glaubenszustände ersetzt. Auch diese kön-

nen wieder als Satzmenge aufgefasset werden, lässt man die Bedingungen (a) und (b) fallen und ersetzt sie ggf. durch schwächere Bedingungen (bei entsprechender Anpassung der Interpretationsfunktion). Carnap selbst skizziert einen anderen Ansatz, der sich eher an der syntaktischen Gestalt der Glaubenssätze orientiert, doch dies soll hier nicht Thema sein.

... und Ontologie...

Der Carnapsche Ansatz einer intensionalen Semantik dient hier nur als Beispiel dafür, dass die Postulierung bestimmter semantischer Entitäten, hier der Intensionen, als Bedeutungskomponenten sprachphilosophisch von einem ontologisch-erkenntnistheoretischen Programm motiviert sein kann: Die Absicht, in der Sprachphilosophie und Semantik hier betrieben wird, ist, zu zeigen, dass die postulierten Entitäten dem eigenen ontologischen Konzept nicht widersprechen und wie bestimmte Aussagetypen – bei Carnap alethische Modalaussagen – in das erkenntnistheoretische Konzept zu integrieren sind. Weder bei Frege noch bei Carnap also waren Fragen der Semantik natürlicher Sprachen der Ausgangspunkt. Der Einfluss, den beide Autoren auf die spätere Entwicklung der Semantik natürlicher Sprachen haben, zeigt, dass sie mit ihren sprachphilosophischen Überlegungen mächtige theoretische Werkzeuge bereitgestellt, die Nicht-Extensionalität einiger Kontexte richtig diagnostiziert und einen Ausweg aus den damit verbundenen Problemen gewiesen haben. Dass ihre *speziellen ontologischen* Überlegungen von großer Relevanz für die semantische Beschreibung weiter Teile der natürlichen Sprache sind, zeigt dies jedoch nicht.

Freges Ontologie ist von einem erkenntnistheoretischen Problem der mathematischen Grundlagenforschung motiviert; Carnaps intensionale Semantik dient dazu, mögliche Einwände gegen sein ontologisches Programm auszuräumen. Beide Ausgangspunkte liegen eher an der Peripherie des Interesses der natürlichsprachlichen Semantik. Selbst wenn Carnaps wissenschaftstheoretisches Programm durchführbar wäre und sich alle wissenschaftlichen Aussagen auf gleichbedeutende Aussagen über beobachtbare Gegenstände und beobachtbare Eigenschaften, Relationen usw. zurückführen ließen, so wäre damit keineswegs gezeigt, dass das für die vor- und unwissenschaftlichen Teile der Alltagssprache ebenso gilt. Wissenschaftstheoretische Reduktionsbemühungen stützen sich darauf, wie wissenschaftliche Aussagen empirisch gerechtfertigt werden. Alltagssprachliche Aussagen in unseren „Alltagstheorien“ werden i. d. R. nicht in derselben Weise empirisch gerechtfertigt. In der natürlichen Sprache wird nun

einmal von Gegenständen wie Röte, Demokratie, Kaufverträgen und Säugetierarten gesprochen, bei denen alles andere als klar ist, wie sie rückstandslos auf Beobachtbares oder Raum-Zeitliches zurückführbar sein sollen. Was bleibt der Semantik natürlicher Sprachen als deskriptiver und empirischer Wissenschaft anderes übrig, als bei der reichen und zuweilen eigenartigen alltagssprachlichen Ontologie anzusetzen, die Ontologie der natürlichen Sprache zunächst einmal ganz unreduktionistisch „ernst zu nehmen“.

Das Frege-Carnap-Beispiel illustriert allerdings auch, dass einzelne ontologische Kategorien, wie Freges Sinne, erfolgreich auf andere ontologische Kategorien rückführbar sind. Neben der bloßen Katalogisierung ontologischer Kategorien kann auch die Erforschung ihrer Relationen untereinander – und im Grenzfall die Reduzierung einzelner auf andere – eine Aufgabe der Semantik sein.

Ontologie der formalen Semantik

Zweimal Ontologie

Semantische Theorien nach ihren Ontologien zu unterscheiden, kann zweierlei heißen: Die Unterscheidung kann sich zum ersten darauf beziehen, welche Kategorien von nicht mehr weiter analysierbaren Entitäten man annimmt. Spricht die semantische Theorie von solchen Gegenständen wie fregeschen Sinnen oder Arten als nicht weiter analysierbaren Entitäten? Hier kann gelten, dass die nicht mehr weiter analysierbaren Entitäten schlicht noch nicht analysierte Entitäten sind, Analysierbarkeit also einfach einen semantischen Erkenntnisstand widerspiegelt, und die aktuell postulierte Ontologie nur einen Zustand des (Noch-) Nicht-weiter-Wissens darstellt.

Zum zweiten kann man Theorien danach ontologisch klassifizieren, welche mengentheoretischen oder sonstigen formalen Konstrukte aus diesen atomaren Entitäten aufgebaut werden.

Konkurrierende Theorien, z.B. die Diskursrepräsentationstheorie und die dynamische Prädikatenlogik im Bereich der dynamischen Semantik, können sich in der Wahl der formalen Mittel (im Beispiel: zur Beschreibung anaphorischer Phänomene) erheblich unterscheiden, auch da, wo sie bezüglich der Wahrheitsbedingungen von Sätzen zu gleichen Feststellungen gelangen. Das Beispiel der dynamischen Semantik zeigt auch: Theorien, die auf eine modelltheoretische Interpretation natürlichsprachlicher Ausdrücke zielen, können eine semantische Repräsentationsebene postulieren, wie im Falle der Diskursrepräsentationstheorie

die Ebene der Diskursrepräsentationsstrukturen, die die Oberflächen- oder Tiefenstrukturen natürlichsprachlicher Ausdrücke interpretiert und ihrerseits modelltheoretisch interpretierbar ist. Semantische Repräsentationen können selbst modelltheoretisch zu interpretierende Ausdrücke einer Formalsprache sein oder mententheoretische Strukturen, die auf sehr unterschiedliche Weise mit Modellen in Beziehung zu setzen sind. Beide Möglichkeiten sind beispielsweise in Theorien zur Repräsentation ambiger natürlichsprachlicher Ausdrücke realisiert. Und, wie das Verhältnis von unterspezifizierten Diskursrepräsentationsstrukturen ambiger Ausdrücke, die zu Diskursrepräsentationsstrukturen disambiguiert werden können, zeigt, müssen sich auch semantische Repräsentationen nicht auf eine Ebene beschränken.

Neben diesen für den Theorieaufbau sehr grundlegenden Unterschieden gibt es aber auch solche, die auf mehr oder weniger arbiträren formalen Entscheidungen beruhen. Wenn Carnap Prädikate extensional als Mengen versteht, Montague aber zu charakteristischen Funktionen, also Funktionen, die den Elementen der carnapschen Mengen den Wahrheitswert wahr, den Nichtelementen den Wahrheitswert falsch zuordnen, übergeht, so ist das ein formaler Schachzug, der ihm eine vereinfachte Formulierung des Kompositionalitätsprinzips gestattet und der im Zuge eines verallgemeinerten typenlogischen Ansatzes vorgenommen wird. Diese Entscheidung darf man jedoch eher als eine dem Streben nach formaler Eleganz als auf sachlichen Differenzen beruhend einordnen.

Aus der Vielfalt möglicher ontologischer Klassifikationen semantischer Theorien werden in Haussers Beitrag zwei Unterscheidungsdimensionen hervorgehoben: Die Unterscheidung, ob die Semantik eine Sinnkomponente aufweist [+Sinn] oder nicht [-Sinn], und die Unterscheidung, ob die Semantik „konstruktiv“ ist [+konstruktiv] oder nicht [-konstruktiv]. Auf letztere Unterscheidung wird unten noch einzugehen sein.

Sinn in der Semantik

Die [+/-Sinn]-Unterscheidung verläuft bemerkenswerterweise zwischen Carnap/Montague ([-Sinn]) und Frege ([+Sinn]), obgleich alles, was im Abschnitt 5 über Freges Semantik gesagt wird, unverändert auch für die Carnapsche intensionale Semantik gilt: Auch sie dient der Interpretation „ungerader“ Kontexte; dass dies zunächst gerade keine Glaubenskontexte sind wie bei Frege, liegt an den unterschiedlichen Ausgangspunkten beider Autoren; Carnaps Vorgehensweise ist jedoch wie oben angedeutet auch auf den Fall der Glaubenskontexte übertragbar.

Tatsächlich scheint die Angemessenheit einer Bedeutungskomponente beim Umgang mit „ungeraden“ Kontexten für die [+/-Sinn]-Unterscheidung bei Hausser nur eine untergeordnete Rolle zu spielen und wird hier wohl nur als Motiv für die Theorie-Entwicklung erwähnt.

Im Zentrum des weiteren Interesses steht eher etwas anderes: Ein scheinbar neuer Aspekt der [+/-Sinn]-Unterscheidung wird bei der Diskussion der konstruktiven Semantiken angesprochen: Dort heißt es, dass „die [+Sinn]-Eigenschaft durch die Abpassung zwischen Sprachbedeutung und kontextuellen Referenten“ ermöglicht werde. Und weiter unten werden drei Bedingungen genannt, unter denen der Unterschied zwischen [+/-Sinn]-Ontologien verschwinde, nämlich

(i) die Welt ist geschlossen, d.h. Objekte können weder verschwinden, noch neue hinzukommen, (ii) die Beziehung zwischen sprachlichen Ausdrücken und ihren Referenten ist ein für allemal festgelegt und (iii) eine spontane sprachliche Verwendung durch den Sprecher/Hörer ist nicht vorgesehen. (Hausser 1998, Abschn. 5)

Unter Bedingung (i) und (ii) werden Problembereiche angesprochen, die in der intensionalen Semantik behandelt werden, und zwar durchaus von der intensionalen Semantik in der Carnap/Montague-Tradition: In einer temporalen Semantik können Zeitpunkte wie mögliche Welten behandelt werden. Das zeitliche Verschwinden und Entstehen von Gegenständen wird durch unterschiedliche Mengen „existierender“ Gegenstände in den jeweiligen Welten modelliert. Auch die Veränderlichkeit der Beziehung zwischen sprachlichen Ausdrücken und ihren Referenten gehört zur ureigensten Erklärungsdomäne der intensionalen Semantik: Weil Intensionen Funktionen oder Abbildungen von möglichen Welten (oder auch Zeitpunkten in einer temporalen Semantik) auf Extensionen, also u. a. auf die Referenten „referierender“ Ausdrücke, sind, stellen sie genau das semantische Instrumentarium zur Modellierung dieser Veränderlichkeit zur Verfügung. Dass der Ausdruck „der deutsche Bundeskanzler“ zum Jahresbeginn 1998 auf eine andere Person referierte als zum Jahresbeginn 1999, liegt eben daran, dass die Intension dieses Ausdrucks so beschaffen ist, dass unterschiedlichen Zeitpunkten unterschiedliche Personen als Referenten des Ausdrucks zugewiesen werden.

Schwerer zu fassen ist Bedingung (iii). Was meint eine „spontane sprachliche Verwendung“ durch den Sprecher/Hörer? Und was steht im Gegensatz dazu? Bezieht man Hauszers Bemerkung ein, dass „SLIM [...] die ontologische [+Sinn]-

Eigenschaft durch die Abpassung zwischen Sprachbedeutung und kontextuellen Referenten“ ermögli­che, so wird hier offensichtlich Bezug auf einen Bedeutungs­begriff genommen, der sich weder in einer extensionalen noch einer intensionalen noch einer Zusammenfassung beider Komponenten erschöpft. Der Zusammen­hang zwischen dieser dritten Bedeutungskomponente und den Referenten wird dieser Charakterisierung zufolge durch eine „Abpassungs“-Relation herge­stellt. Wie passt dies mit dem fregeschen Sinn zusammen?

Frege bezeichnet Sinne, wie oben erwähnt, auch als „Gegebenheitsweisen“ von Gegenständen. Dieselbe Person kann im Oktober 1999 einmal als „der Bundeskanzler der Bundesrepublik Deutschland“ und als „der Vorsitzende der SPD“ (an-)gegeben werden. Es ist klar, dass beide Ausdrücke nicht notwendigerweise und zu allen Zeiten genau diese Person oder auch nur dieselbe Person bezeichnen. Mit Frege könnte man sagen, dass dies daran liegt, dass beide Ausdrücke dieselbe Person auf unterschiedliche Weise (an-)geben, also unterschiedliche Eigenschaften oder Erkennungsmerkmale heranziehen, um die gemeinte Person zu identifizieren. Mit Hausser könnte man davon sprechen, dass beide Ausdrücke unterschiedliche *Bedeutungen* haben, die jede für sich mit den kontextuell gegebenen Referenten „abgepasst“ werden.

Im Licht dieser Interpretation betrachtet, ist der Mehrwert der [+Sinn]-Komponente darin zu suchen, dass sie *erklärt*, warum zwei Ausdrücke mal denselben Referenten, mal verschiedene bezeichnen und wie ein Ausdruck zu seinem Referenten gelangt, während die intensionale Semantik nur *konstatiert*, dass es für jeden Ausdruck eine bestimmte Abbildung von möglichen Welten auf Referenten gibt. Wie diese im Vermögen des Sprecher/Hörers etabliert wird, bleibt offen. Die intensionale Semantik im Sinne Carnaps und Montagues scheint keine Erklärung für spontane Subsumtionen von Gegenständen unter Begriffe zu bieten; denn entweder kennt der Sprecher/Hörer die einem Begriff *P* entsprechende Abbildung zwischen möglichen Welten und Begriffsextensionen bereits vollständig – dann ist es keine spontan zu entwickelnde neue Erkenntnis, dass $P(a)$ für einen irgendeinen Gegenstand *a* gilt. Oder der Sprecher/Hörer kennt die einem Begriff *P* entsprechende Abbildung zwischen möglichen Welten und Begriffsextensionen noch nicht vollständig, dann bleibt es offen, wie die Kenntnis dieser Abbildung bei Konfrontation mit einem „neuen“ Gegenstand entsprechend erweitert werden soll. Die Erklärung der spontanen Subsumtion von Gegenständen unter Begriffe erfordert also die Einbeziehung einer semantischen „inhaltlichen“ Komponente.

Übersieht die intensionale Semantik diesen Bedeutungsaspekt also, wenn sie Intensionen so versteht wie oben angedeutet? Nein; denn von der Kenntnis von Funktionen zu sprechen, kann fundamental anderes bedeuten als von den Funktionen selbst zu sprechen, ohne dass ein Widerspruch vorläge.

Eine Abbildung oder eine Funktion zu *kennen*, kann sehr Unterschiedliches meinen. Im mengentheoretischen Sinne *ist* die Quadrierungsfunktion beispielsweise schlicht eine Menge von Zahlenpaaren, die Paare $\langle 1,1 \rangle$, $\langle 2,4 \rangle$, $\langle 3,9 \rangle$ gehören z.B. dazu. Die Quadrierungsfunktion zu *kennen*, bedeutet für uns aber nicht, alle Paare „gespeichert“ zu haben (was wegen der unendlichen Anzahl ja auch unmöglich wäre), sondern eine Rechenvorschrift zu kennen, die es uns erlaubt, zu einer gegebenen Zahl ihr Quadrat zu finden, möglicherweise auch eine Vorschrift, die es uns erlaubt, zu einer Quadratzahl eine Wurzel zu finden. Verschiedene Personen können über unterschiedliche Verfahren zu Quadratzahlen gelangen. In einem weniger mental besetzten Sinne von *kennen* kann vielleicht auch eine Rechenmaschine die Funktion „kennen“, sofern auf ihr eine Rechenvorschrift für diese Funktion implementiert ist. Und vielleicht wären wir sogar bereit, einer Person Kenntnis der Quadratfunktion zuzubilligen, die zwar weiß, dass das Quadrat einer Zahl das Ergebnis der Multiplikation dieser Zahl mit sich selbst ist, bei der sich dann aber herausstellt, dass sie die relevanten Multiplikationsschritte nicht fehlerfrei durchführen kann.

Das Beispiel soll illustrieren: Kenntnis einer Funktion kann Unterschiedliches bedeuten, ist aber i. d. R. von der bloßen Speicherung der Instanzen verschieden. Insofern muss die Behauptung, diese oder jene Bedeutungskomponente *sei* diese oder jene Funktion und damit diese oder jene Menge von Paaren, nicht der Behauptung widersprechen, dass unsere *Kenntnis* dieser Funktion oder unser Wissen über die Funktion auf bestimmten bewussten oder unbewussten Verfahren beruhe, die Funktionsargumente und Werte zueinander in Beziehung setzten. Im engeren Sinne ontologisch, also auf das, was *ist*, bezogen, ist nur die Frage, was die Bedeutung eines Ausdrucks *ist*; die Frage, wie wir die Bedeutung *kennen* oder wie wir über sie *wissen* können, ist nach der üblichen philosophischen Nomenklatur eine *epistemologische* Frage. Oder, insofern sie sich empirischen Untersuchungen erschließt, eine psychologische, neurologische oder allgemein kognitionswissenschaftliche.

So besehen bietet es sich eher an, Unterschiede zwischen [+/-Sinn]-Semantiken epistemologisch oder durch ihre *Vernetzung* mit psychologischen, neurologischen oder kognitionswissenschaftlichen Theorien zu bestimmen als eine ontologische Differenz zu postulieren. Dann ist allerdings auch Frege kein besonders

typischer Vertreter einer [+Sinn]-Semantik, da er zwar in seiner Erkenntnistheorie annehmen musste, dass verschiedene Gegebenheitsweisen sich auch epistemisch unterscheiden, ihn aber die Frage, worin der epistemische Unterschied bestehe, kaum interessierte.

Kognition und Verifikation

Die formale Semantik hat zur Frage der kognitiven Implementierung unseres semantischen Vermögens zunächst wenig zu sagen: Sie bietet den formalen Rahmen, in dem sich die Komposition der Bedeutung komplexerer Ausdrücke aus den Bedeutungen einfacherer Ausdrücke beschreiben lässt. Sie analysiert die Typverhältnisse, die zwischen Bedeutungen bestehen. Und sie bietet einen formalen Rahmen für die Darstellung von Bedeutungsrelationen; hierin besteht eine Schnittstelle zur lexikalischen Semantik. In einigen Fällen liefert sie allerdings komputationell umsetzbare Modelle für bestimmte Interpretationsprozesse, wie z.B. die Diskursrepräsentationstheorie es für das inkrementelle Verstehen eines Diskurses leistet.

Aber erst durch die Verknüpfung mit kognitionswissenschaftlichen Theorien, die von anderen Dingen als sprachlicher Bedeutung handeln, lässt sich unsere *Kenntnis* der sprachlichen Bedeutung erklären. Die Frage nach der Bedeutung von Ausdrücken und die nach unserer Kenntnis darüber, nach den Lernprozessen, mit denen wir uns diese Kenntnis verschaffen, nach den Vagheiten und Ungewissheiten unserer Kenntnis, nach den Reaktionszeiten, mit denen wir Bedeutungswissen abrufen, nach den Hirnregionen, die an bedeutungsverarbeitenden Prozessen beteiligt sind und nach der physiologischen Struktur ihrer Neuronen sind jeweils ganz unterschiedliche Fragen, und es ist nicht der Fehler einer Theorie, nicht auf alle Fragen zugleich zu antworten. Wenn sich allerdings herausstellte, dass eine bestimmte semantische Theorie sich der Vernetzung mit einer empirisch fundierten kognitionswissenschaftlichen Theorie der Bedeutungsverarbeitung widersetze oder ihr sogar widerspräche, dürfte das linguistische und kognitionswissenschaftliche Interesse an dieser semantischen Theorie abnehmen. Dass dies beim heutigen kognitionswissenschaftlichen Kenntnisstand bereits für irgendeinen einflussreichen theoretischen Ansatz in der formalen Semantik feststellbar sei, ist sehr zu bezweifeln.

Man könnte der These von der weitgehenden Neutralität der formalen Semantik gegenüber Theorien der kognitiven Implementierung entgegenhalten, dass das Interesse des logischen Empirismus an der Semantik ja gerade darin bestand,

über die Bedeutungsanalyse zu Verfahren der empirischen Verifikation oder Falsifikation von Sätzen zu gelangen. Die logischen Empiristen wollten also durchaus zu Aussagen darüber gelangen, wie wir über die Interpretation von Sätzen zu deren Wahrheitswert gelangen (können). Im ersten Abschnitt seines Beitrags weist Hausser sehr nachdrücklich auf dieses Interesse an der Semantik hin. Allerdings gehört auch dies zum Bereich der epistemologischen und ontologischen Motivation zur Entwicklung einer formalen Semantik. Als theoretisches Paradigma zur Bedeutungsbeschreibung natürlicher Sprachen hat sich die formale Semantik jedoch von ihrer Motivationsbasis emanzipiert.

Formale Semantik kann ganz ohne Bezug auf verifikationistische oder falsifikationistische Auffassungen betrieben werden. In der Theorie der generalisierten Quantoren beispielsweise wird die Bedeutung des Determinierers „die meisten“ mit mengentheoretischen Mitteln beschrieben, der Determinierer kann hinsichtlich seiner Monotonie-Eigenschaften untersucht werden usw., ohne dass Spekulationen darüber angestellt werden müssten, wie ein Satz wie

(9) In den meisten Sonnensystemen gibt es kein Leben

zu verifizieren oder zu falsifizieren oder auch nur zu stützen oder zu erschüttern sein sollte. Durch eine modelltheoretische Interpretation des Satzes erfährt man zwar, welche Modelle den Satz selbst und welche seine Negation wahr machen, oder anders: man erfährt, unter welchen Bedingungen der Satz wahr ist; die Bedingungen benennen zu können, die einen Satz wahr machen, heißt aber nicht, diese Bedingungen auch empirisch überprüfen zu können. Empirische Verifikation und Falsifikation kann an vielen Klippen scheitern: daran, dass nicht alle Instanzen von Gegenständen, deren Untersuchung nötig wäre, aus faktisch (Unerreichbarkeit der Gegenstände) oder prinzipiell (unendliche Anzahl) bestehenden Gründen untersucht werden können; oder daran, dass die fragliche Behauptung über Dinge ohne empirischen Bezug spricht.

(10) Zeus lebt ewig

ist gegenüber einem Proponenten dieser Behauptung, der konsequent leugnet, dass es irgendwelche beobachtbaren Konsequenzen dieser Behauptung gibt, nicht zu widerlegen, solange er sich jedenfalls nicht selbst widerspricht. Dennoch lässt sich über die unterschiedlichen Modelle eines Zeusanhängers und eines Zeusleugners durchaus Sinnvolles sagen.

Konstruktive Semantik

Wenn oben im Zusammenhang mit der [+/-Sinn]-Unterscheidung bezug auf „unsere“ Kognition genommen wurde, wurde bewusst offengelassen, ob hier jeweils individuelle kognitive Eigenschaften oder ein überindividuelles oder interindividuelles kognitives Abstraktum gemeint war. Freges Sinne oder Gegebenheitsweisen jedenfalls sind keine individuell oder subjektiv anzusiedelnden Entitäten. Wären sie dies, so argumentiert Frege, und würde jeder über seine eigenen Sinne oder Gegebenheitsweisen sprechen, weil sie ihm allein zugänglich sind, so gäbe es gar keine Verständigungsgrundlage dafür, mit anderen darüber zu sprechen, ob ein Gedanke (der Sinn eines Satzes) wahr oder falsch wäre, weil jeder nur über seinen eigenen und damit einen jeweils anderen Gedanken spräche (vgl. Freges Aufsatz *Der Gedanke*, S. 42 ff.). Diese Feststellung Freges stellt für jeden, der semantische Interpretation auf innerindividuelle Vorgänge und Repräsentationen reduzieren will, eine Herausforderung dar: Es muss geklärt werden, wie sich individuelle Interpretation auf öffentliche Gegenstände bezieht, auf die sich auch andere mit ihren je individuellen Interpretationen beziehen können. Jemand, der von dem Exemplar des LDV-Forums, das er gerade in der Hand hält, spricht, meint ja das Heft selbst, das auch für andere sichtbar etc. ist, und spricht nicht von seiner Vorstellung, Repräsentation o.ä. des Heftes; denn es mag ja sein, dass der Umschlag des Heftes in der individuellen Vorstellung oder Repräsentation blau ist, es wäre aber extrem kontraintuitiv, den Satz

- (11) Der Umschlag des vorliegenden Heftes des LDV-Forums
ist weiß

für falsch zu halten, von wem er auch immer geäußert werden mag.

Wie individuelle Interpretation und Repräsentation mit dem Realismus der natürlichen Sprache in Einklang gebracht werden sollen, ist nicht Thema des Beitrags Haussers. Dieser Aspekt wird nur an einigen Stellen, wie der Beziehung *b* in Abbildung 6.2, angedeutet. Klammern wir also im Folgenden diesen Aspekt des Verhältnisses konstruktiver und „realistischer“ Semantik aus.

Eine konstruktive Semantik im Sinne Haussers betrachtet die Interpretationsvorgänge in einem individuell zu verstehenden Agenten, wie seine Argumentation bezüglich des Epimenides-Paradoxes zeigt. Für die Überlegenheit von [+konstruktiv]-Ansätzen führt Hausser insbesondere zwei Problembereiche ins Feld: den Umgang mit Vagheit von Bedeutung und das Epimenides-Paradox.

Soll man Vagheit eher der Seite der Bedeutung selbst oder der Seite der Bedeutungsverarbeitung des kognitiven Systems zuschlagen? Was ist Vagheit überhaupt? Eine Eigenschaft von Zuordnungsinstanzen zwischen sprachlichen Ausdrücken und Bedeutungen? Eine Eigenschaft von Bedeutungen selbst? Und wie verhält sich Vagheit zu (epistemischer) Unsicherheit. Bezieht man die semantischen Verarbeitungsvorgänge in die Betrachtung ein, gewinnt man neue mögliche Ansatzpunkte für Vagheit, wie Hausser zeigt. Zu fragen ist allerdings, ob die vagheitsanfälligen Beziehungen *b* und *c* zwischen Welt und Agent nicht am ehesten Ansatzpunkte für *Unsicherheit* sind. Unzweifelhaft gehört die Behandlung der Vagheit und der Kontextbedingungen, die geeignet sind, Vagheit einzuschränken (neben der kontextabhängigen Auflösung von Anaphern, Ellipsen und Ambiguitäten), zu den Bereichen der Semantik, die von kognitionswissenschaftlichen Theorien der Begriffsverarbeitung am meisten profitieren werden.

„Dekonstruktion“ des Epimenides-Paradox?

Hilft jedoch ein [+Sinn,+konstruktiv]-System beim Entschärfen des Epimenides-Paradox? Setzt man voraus, dass (A) in einem Modell, das den Satz

(12) *P* ist falsch

wahr macht, der Satz *P* selbst immer falsch ist, und dies für beliebige Sätze *P*, so gelangt man zum Ergebnis, dass auch in einem Modell, in dem

(13) (13) ist falsch

wahr ist, (13) falsch ist. Ersetzt man (13) jedoch durch den Satz, für den die Abkürzung steht, so erhält man

(14) Dass (13) falsch ist, ist falsch

Unter der Voraussetzung der Gültigkeit des Substitutionsprinzips (B), das die wahrheitswerterhaltende Ersetzung gleichbedeutender Ausdrücke erlaubt, muss dieser Satz in denselben Modellen wahr sein, in denen auch (13) wahr ist. Das bedeutet aber nach (A), dass (13) in diesen Modellen falsch ist, was der Voraussetzung widerspricht.

Ist (13) also falsch? Dies angenommen und nimmt man weiter an, dass (C) jedes Modell, in dem

(15) *P* ist falsch

falsch ist, P selbst wahr macht, so folgt daraus die Wahrheit von (13) in jedem Modell, in dem (13) falsch ist. Beide Annahmen, die der Wahrheit und die der Falschheit von (13), führen unter den Voraussetzungen (A–C) also zum Widerspruch. Soweit die Antinomie.

Ein Deduktionsverfahren, dass die Folgerungen gemäß (A–C) vollständig realisiert, erlaubt den Beweis sowohl der Wahrheit als auch der Falschheit von (13). Anderenfalls liefert das Deduktionsverfahren nur ein oder gar kein Ergebnis und würde damit zu einem unvollständigen Verfahren.

Aus Logiksprachen kann dieser Widerspruch verbannt werden, indem man direkt oder indirekt selbstbezügliche Sätze des Typs (13) syntaktisch ausschließt oder die Interpretation des Systems unvollständig gestaltet, indem man Sätzen dieses Typs keinen Wahrheitswert in Modellen zuordnet, aber auch indem man eines der Prinzipien (A–C) in dieser Allgemeinheit fallen lässt.

Die syntaktische Ausschlussoption ist bezüglich der natürlichen Sprache sicherlich die unattraktivste und kontraintuitivste. Sätze wie (13) sind offensichtlich syntaktisch möglich. Die Möglichkeit über die mit diesen Sätzen verbundene Antinomie zu diskutieren, kann man auch als Hinweis darauf sehen, dass wir jedenfalls für gewöhnlich und vortheoretisch nicht von einer unvollständigen Interpretation ausgehen. Nur spielen in der Alltagskommunikation Sätze, die die Gefahr des Epimenides-Paradox in sich bergen, keine Rolle, da kaum Situationen vorstellbar sind, in denen ein solcher Satz eine pragmatisch sinnvolle Mitteilungsfunktion hätte. Aus rein praktischen Gründen beschränken wir uns in der Alltagskommunikation also in einem widersprüchlichen System auf ein nicht-widersprüchliches Fragment. In diesem Sinne kann man sagen, dass das Epimenides-Paradox für die praktische Semantik ohne Belang ist, nur für ihre formale Systematisierung ein Problem darstellen kann, allerdings ein auf einem der skizzierten Lösungswege pragmatisch lösbares.

Allerdings kann das Problem nicht schon durch die Unterscheidung eines kontingenten und eines logischen Wahrheitsbegriffs oder für das vorliegende Problem wichtiger noch: eines kontingenten und eines logischen Falschheitsbegriffs zum Verschwinden gebracht werden. Wenn die oben genannten Bedingungen (A–C) auch für Haussers wahr^k und falsch^k gelten, lässt sich aus der Annahme, dass

(13) ist falsch^k

in einem Modell wahr^k ist, ebensogut folgern, dass (13) falsch^k ist. *Prima facie* jedenfalls ist kein Grund zu sehen, warum (A–C) für wahr^k und falsch^k weniger plausibel als für logisches Wahr und Falsch sein sollten. „Modell“ kann bei diesen Überlegungen übrigens nach Belieben durch „Wirklichkeit“ oder „Glaubensinhalte eines Agenten“ ersetzt werden.

Das Verhalten des von Hausser skizzierten interpretierenden Agenten ist, wenn der Agent durch einen dem Epimenides-Paradox unterliegenden Satz in eine infinite Rekursion gerät, ein Spezialfall eines unvollständigen Deduktionsverfahrens. Für dem Epimenides-Paradox unterliegende Aussagen liefert das Interpretationsverfahren kein Ergebnis. Wird das Scheitern der Interpretation durch ein Kontrollprogramm erkannt und die infinite Rekursion beendet, so kann das Kontrollprogramm alle für Deduktionsverfahren aufgezählten Möglichkeiten realisieren: Es kann sowohl den Satz selbst als auch seine Negation als zutreffend interpretieren, was einem vollständigen aber inkonsistenten Deduktionsverfahren entspräche, oder nur eins oder keins von beiden als zutreffend interpretieren, was einem unvollständigen Deduktionsverfahren gleichkäme. Die Implementation nicht ganz vollständiger oder nicht ganz konsistenter Deduktionsverfahren ist jedoch nicht an ein Agentenmodell gebunden.

Semantisch ist die Antinomie also weder durch die Einführung von kontingenten Wahrheitswerten noch durch eine agentenorientierte Interpretation zu entschärfen, solange die Bedingungen (A–C) gelten, und damit ist es ebenso aussichtslos wie ohne die Einführung kontingenter Wahrheitswerte und Agenten ein korrektes und vollständiges Deduktionsverfahren zu finden.

Das Epimenides-Paradox lässt sich demnach in [+/-konstruktiv]-Semantiken auf dieselbe Weise entschärfen: Durch Aufgabe einer der Bedingungen (A–C) oder durch den Ausschluss einiger Sätze von der Interpretation.

Zusammenfassung

Die [+/-Sinn]-Unterscheidung Haussers scheint ihren Grund eher in der Einbeziehung epistemologischer oder kognitionswissenschaftlicher Aspekte von Bedeutung zu haben als in unterschiedlichen Ontologien. In [+Sinn]-Semantiken werden gewisse semantische Entitäten zu kognitiven Entitäten in Beziehung gesetzt; dadurch rücken Aspekte unserer Kenntnis und unseres kognitiven Umgangs mit semantischen Entitäten ins Blickfeld, dies ändert aber nichts daran, was die semantischen Entitäten *sind*.

Auch ein Agentenmodell in einer [+konstruktiv]-Semantik kann als eine zu formalsemantischen „realistischen“ Ansätzen hinzutretende Betrachtung gesehen werden, die beispielsweise Strategien zur Auflösung von Vagheiten und Ambiguitäten beisteuert. Die durch das Epimenides-Paradox aufgeworfenen (vermeidbaren) logischen Probleme stellen sich jedoch auch in einem Agentenmodell nicht anders dar als in einer nicht agentenbasierten Interpretation.

Anmerkungen

¹Die Beziehung seiner semantischen und modallogischen Studie zum logischen Empirismus wird u. a. durch den ersten Artikel des Supplements der 2. Auflage von *Meaning and Necessity* deutlich.

Literatur

- Jens Allwood; Peter Gärdenfors: *Cognitive Semantics. Meaning and Cognition*. Amsterdam; Philadelphia: 1998.
- Wolfgang Balzer; C. Ulises Moulines; Joseph D. Sneed: *An Architectonic for Science*. Dordrecht etc.: 1987.
- Rudolf Carnap: *Meaning and Necessity. A Study in Semantics and Modal Logic*. 2nd edition. Chicago; London. 1946.
- Gottlob Frege: *Über Sinn und Bedeutung*. In: *Funktion, Begriff, Bedeutung*. Hrsg. v. Günther Patzig. 6. Aufl. Göttingen: 1986. S. 40–65.
- Gottlob Frege: *Der Gedanke. Eine logische Untersuchung*. In: *Logische Untersuchungen*. Hrsg. v. Günther Patzig. 3. Aufl. Göttingen: 1986. S. 30–53.
- L.T.F. Gamut: *Logic, Language, and Meaning*. 2 vols. Chicago; London: 1991.
- J. Groenendijk; M. Stockhof: *Dynamic Predicate Logic*. *Linguistics and Philosophy* 14. 1991. S. 39–100.
- Hans Kamp; Uwe Reyle: *From Discourse to Logic*. 2 vols. Dordrecht; Boston; London: 1993.
- Richard Montague: *Formal Philosophy*. New Haven: 1974.
- Franz von Kutschera: *Einführung in die intensionale Semantik*. Berlin; New York: 1976.

Empirische Grundlagen für Computer- linguisten unter den Informatikern

Lehr-Erfahrungen mit der Grammatik „Vierineins“ (das ist: 4 in 1)

M. Lutz-Hensel

Universität Koblenz-Landau, Abt. Koblenz

Institut für Computerlinguistik

hensel@informatik.uni-koblenz.de

1. Kontext

Hintergrund und Anlaß

In den Jahren 1995 und 1996 haben im Institut für Computerlinguistik der Universität Koblenz-Landau, Abt. Koblenz, Diskussionsrunden zur Studienreform stattgefunden. Daran war außer Dozentinnen und Dozenten eine größere beständige Gruppe von Studierenden des Hauptstudiums beteiligt. Die Initiative ging von Studierenden aus, die angeregt durch Veranstaltungsrückmeldungen und aufgrund ihrer Studienerfahrungen im Ausland ihre Wünsche für Veränderungen anmeldeten.

Dabei wurde vieles andiskutiert, was erhebliche Modifikationen der Studien- und Prüfungsordnung zur Folge gehabt hätte und deshalb nicht ohne weiteres machbar ist. Als positive Bilanz ist zu verzeichnen, daß erstmals in der Gemeinschaft von Lernenden und Lehrenden Verständigung über Lern- und Lehrerfahrungen und Lehrinhalte gesucht und in Teilen auch erreicht wurde. Daß das Gewünschte durch das Machbare ersetzt werden mußte, wenn spürbare Folgen für das Curriculum entstehen sollten, war eine wichtige, aber nicht lähmende Erkenntnis.

Es kam zu einem Antrag für ein Lehrprojekt, der von einer Vierergruppe erarbeitet und auf Bundesebene eingereicht wurde. Gegenüber Konkurrenten, die „integrierte Auslandsaufenthalte“ oder „Tele-Teaching“ auf ihre Fahnen geschrieben hatten und mit „neuen Berufsmöglichkeiten“ winken konnten, nahm sich unser Antrag bescheiden aus, und unser Vorhaben wurde nicht gefördert. Ebenso muß-

ten wir auf inneruniversitäre Sonderzuwendungen zur Unterstützung der Lehre verzichten.

Die gewonnenen Einsichten wollten wir¹ jedoch nutzen: Wir machten uns von Drittmittelfinanzierung dadurch unabhängig, daß wir Modifizierungen vorerst auf Teile einer wiederkehrenden Lehrveranstaltung des Grundstudiums einschränkten, ohne deren Grundkonzeption zu verlassen.

1.1 Curriculum-Konzept eines CL-Schwerpunkts im Grundstudium eines Informatikstudiengangs

Unser Auftrag besteht darin, Studierende der Informatik mit einem Anwendungsfach bekanntzumachen, in dessen Zentrum die Mensch-Computer-Interaktion (MCI) mithilfe natürlicher Sprache steht. Wir legen diese Aufgabe so aus, daß sie den folgenden Vorgaben genügt:

- i) Die natürlichsprachliche Interaktion zwischen Mensch und Computer ist mit interpersonaler Kommunikation darin vergleichbar, daß sprachlich agierende Menschen in beiden Fällen Handelnde im vollen Sinne des Wortes sind.

Das heißt unter anderem, daß Menschen im Rahmen der MCI Einstellungen sowohl zum Interaktionsprozeß wie auch zum interagierenden Partner entwickeln, auf ihnen bestehen, sie revidieren oder sie aufgeben. Zu diesen Einstellungen gehört, daß Menschen – ob zu Recht oder zu Unrecht – dem Partner vergleichbare Einstellungen zuschreiben.

Es heißt weiterhin, daß Menschen absichtsvoll und zielgerichtet sprachlich handeln, so wie sie ihrerseits intentionales Vorgehen ihrer Partner schließend oder konstruierend voraussetzen.

- ii) Da unsere Studierenden künftig nicht nur kritische Nutzer von sprachverarbeitenden Systemen sein werden, sondern vielmehr Konstrukteure und Konstrukteurinnen, werden sie die Möglichkeit haben, Interaktionsprozesse nicht nur im vorgegebenen Rahmen eines Systems sprachlich zu steuern, sondern sie werden als Systemarchitekten grundlegenden Einfluß auf die Interaktion von Nutzern mit Computern nehmen.

Unangemessen gegenüber i) wäre es, wollten wir menschliche Individuen auf ein Reagieren als Dekoder reduzieren. Zwar ist das Modell des Informationstransports wegen seiner bestechenden Simplität immer noch en vogue², nicht abseh-

bar sind aber seine deformierenden Auswirkungen, wenn die natürlichsprachliche Mensch-Computer-Interaktion in Beruf und außerberuflichem Alltag weiter Platz greift.

Unangemessen gegenüber ii) wäre es, wollten wir fortschreitenden Erkenntnissen im Bereich kommunikativen sprachlichen Handelns mit Rechnern dadurch den Weg verbauen, daß wir uns auf das in absehbarer Zeit Machbare beschränken.

Deshalb haben wir uns für ein Konzept für das Grundstudium entschieden, das durch zwei Grundlinien gekennzeichnet ist:

1.1.1 Die thematische Grundlinie

Trotz des Titels einer zweisemestrigen Übung, die 'Linguistische Grundlagen' heißt, ist es nicht ausreichend, diese Grundlagen in einer globalen Übersicht zu vermitteln. Unser Augenmerk ist in einem fächerintegrierenden Studiengang vielmehr auf alle Wissenschaften gerichtet, die sich – und sei es am Rande – mit Sprache, Sprechen und Denken befassen. Auf der anderen Seite können nicht alle Teildisziplinen der Linguistik in unserem Anwendungsfach gleiche Aufmerksamkeit beanspruchen.

Die empirischen Grundlagen werden auch nicht dargestellt als linguistische Einheiten, die sich innerhalb von Regelsystemen zu immer komplexeren Einheiten eines Anwendungssystems fügen: von Phonemen zu Morphemen, zu Wortformen, zu Sätzen, zu semantischen Einheiten, zu pragmatischen Einheiten. Auch wenn diese Sichtweise für unsere Adressaten vielleicht naheliegend wäre, empfiehlt sie sich nicht, weil sie ein vereinfachendes, einsträngiges Denken unangebracht befördert.

Statt dessen werden die empirischen Grundlagen in vielfach zusammenhängenden Themenkomplexen dargestellt, beginnend bei der kommunikativen Nutzung natürlicher Sprache durch individuelle Sprecherinnen und Hörer, Schreiber oder Leserinnen. Die Anordnung der Themenkomplexe folgt einem Prinzip der fortschreitenden Abstraktion bis hin zur Syntax. Danach werden kommunikative Aspekte wieder einbezogen, so daß die Gedankenbewegung am Ende mit „Präsuppositionen in der Argumentation“ geschlossen werden kann.

Da die Themenkomplexe problembezogen behandelt werden, mögen sie durch die folgenden Leitfragen charakterisiert sein:

- Inwiefern unterscheidet sich die automatische Verarbeitung gesprochener Sprache von der geschriebener Sprache?
- In welchen Fällen wird man davon sprechen wollen, daß ein natürlichsprachlicher Text verstanden wurde?
- Wir lassen Computer mathematische Handlungen vollziehen wie Differenzieren oder Wurzelziehen, wir lassen sie zeichnen; können wir sie auch sprachliche Handlungen vollziehen lassen?
- Was ändert sich in einer Sachverhaltsbeschreibung in natürlicher Sprache oder in formaler Sprache gegenüber dem ursprünglichen Sachverhalt?
- Unter welchen Bedingungen ist eine Ansammlung von Wortformen als Text anzusehen? Und gibt es Oberflächenindikatoren, die sich für den Computierzugriff eignen?
- Welche natürlichsprachlichen Kennzeichen gibt es für die Einordnung eines Sachverhalts in ein Zeitschema; wie ließe sich die sprecherbezogene Einordnung formalisieren?
- Wie kann eine lauffähige formale Grammatik für einen Ausschnitt des Deutschen aussehen, und durch welche Arbeitsschritte wird sie erreicht?
- Welche Vorbereitungen sind notwendig, wenn man alle syntaktisch korrekten Wortstellungsvarianten in Deklarativsätzen automatisch erzeugen möchte?
- Welche verborgenen Informationen dürfen Hörerinnen bzw. Adressaten von Äußerungen in Rechnung stellen, oder: Was geben Sprecher außer dem Gesagten sonst noch zu verstehen? In welcher Weise beschränken verborgene Informationen die Strategien von Dialogpartnern in argumentativen Dialogen derart, daß sie in Dialogmodellen berücksichtigt werden müßten?

1.1.2 Die didaktische Grundlinie

Die Lernsituation³ ist bestimmt als

eine gemeinschaftliche Unternehmung
von
über linguistische Grundlagen
im Bereich sprachlicher Kommunikation
mithilfe sprachlicher Kommunikation
agierenden Partnern.

Dabei kann ein Rechner einen Part übernehmen, jedoch nicht so, daß er einen menschlichen Kommunikator ersetzt, sondern so, daß der Dialogkreis in Richtung auf ein sprachverarbeitendes System hin geöffnet wird.

Dieser Ansatz hat mehrerlei Folgen:

Die Studierenden treten mit dem für die Lehrveranstaltung Verantwortlichen gemeinsam in eine Arbeitsgruppe ein, in der vorrangig im Gespräch Verständigung über natürlichsprachliche Erscheinungen und ihre wissenschaftliche Erfassung gesucht wird.

Das beinhaltet einerseits, daß Vorverständnisse bei einzelnen nicht ignoriert, sondern bewußt gemacht und in einer sprachlichen Ausdrucksweise für alle Gesprächsteilnehmer (nicht zu übersehen: auch für den Redenden selbst) offengelegt werden. Auf dieser Basis soll es sich als nützlich erweisen, mithilfe von Fachtermini Differenzierungen vorzunehmen und sich auf den Gebrauch dieser Termini zum Zweck präzisen Redens zu einigen. Andererseits verlangt dieses Vorgehen von allen Teilnehmern die ständige Beobachtung des kommunikativen Prozesses, der immer wieder selbst zum Untersuchungsgegenstand werden kann. Auf diese Weise wird für alle spürbar, welchen kognitiven Aufwand es macht, die natürliche Sprache als Mitteilungssprache von der natürlichen Sprache als Objektsprache zu unterscheiden.

Diejenigen, die unbefragt gewohnt sind, formale Sprachen als Beschreibungssprachen zu verwenden, sehen die Trennung von Sprachentypen nicht länger als etwas Unabdingbares, zu Lernendes, sondern sie sehen, wie mühsam und zugleich nutzbringend die Unterscheidung ist. In der Konsequenz des Vorgehens liegt, daß Gewußtes mehrfach und von verschiedenen Zugängen her zum Erschließen neuen Wissens genutzt wird.

Für explorative Arbeitsverfahren wird jeweils ein Fundus von Material bereitgestellt, der zu Hypothesenformulierung und Hypothesenprüfung eingesetzt wird. So wird einsichtig, inwiefern ein Einzelbeispiel als erinnerungs- und demonstrierungstechnisches Mittel für eine Beispielmeng e eintreten darf.

Selbstbeobachtung beim Aufgabenlösen soll den bewußten Einsatz von Lösungsstrategien unterstützen, dagegen ist die Präsentationsform der Lösungen freigestellt und kann sich als mündlicher Vortrag, Diskussionsbeitrag, handschriftliche Ausarbeitung oder Computeroutput darbieten. Hausaufgaben zu den Themengebieten sind untereinander abgestimmt, jede einzelne wiederum setzt sich aus Strategieschritten zusammen; so liegt es nicht in unserem Interesse, daß die Bearbeitung eines gewissen Prozentsatzes oder von Aufgabenteilen für ein er-

folgreiches Ergebnis ausreichend ist – wie die Studierenden das von Aufgabenstellungen in ihrem Hauptfach vielfach gewohnt sind.

Die Übung in der Dialogführung ist ein indirektes Studienziel und hat gegenüber den mitteilenden Lehrstilen einigen Neuigkeitswert für die Studierenden. Damit werden unter der Hand Erinnerungsleistung und Wiedererkennen durch Vergleichen erforderlich, der Aufbau von Begründungsketten sowie das Eingehen auf nicht-eigene Argumentationsbeiträge im Koreferat oder in der freien Diskussion. Viele Formen von Verstehenssicherung, u. a. durch Exemplifizieren, Analogiebildung, Paraphrasierungen, Abstraktion können eingesetzt werden und erweisen sich ad hoc als mehr oder wenig erfolgreich für den Fortgang von Verständigungsprozessen.

Die Studierenden werden angeleitet, sich Wissensgebiete eigenständig anzureichern, später: sich eigenständig zu öffnen, sie werden 'Fachreferentin für...'. Damit zusätzlich soll der Eindruck vermieden werden, daß (auch nur partiell) vollständiges, vorgefertigtes und abrufbares Wissen zu erwerben ist, und der Eindruck gefördert werden, daß vorläufiges Wissen neue Fragen eröffnet, eine geeignete Basis für Erkundungen darstellt und den Umgang mit Wissen zu einer lebendigen Auseinandersetzung macht.

Wir sehen unser Vorgehen als wirklichkeitsnah an und lassen es mit neueren Modellen vom Text- und Sprachverstehen korrespondieren. Unsere Studierenden gaben uns in Feedback-Runden Grund zur Annahme, daß es von ihnen als – wenn auch forderndes – Kontrastprogramm geschätzt wird.

Zur Position des Rechners in der skizzierten Lernsituation möchte ich im Zusammenhang mit dem detailliert dargestellten Lernkomplex in der Folge Stellung nehmen.

1.1.3 Die Aufgabe

In den Runden zur Studienreform hatten Studierende den Wunsch geäußert, im Grundstudium auch im Anwendungsfach mehr Arbeiten am Rechner durchführen zu können. Im Zusammenhang mit den 'Linguistischen Grundlagen' bot sich dafür zunächst der Themenkomplex 'Syntax'⁴ an.

Inhalte, die bis dahin auf syntaktische Relationen, Konstituenten- und Phrasenstrukturgrammatik und zentral auf Verbvalenz ausgerichtet waren, waren also zu verändern. Die zugehörige Hausaufgabe, in der anhand eines vorgegebenen Textkorpus in bestimmten Arbeitsschritten die Valenzrahmen von drei nicht in

Valenzlexika verzeichneten Verben ermittelt und nach Art von Schumachers 'Verben in Feldern' notiert werden sollten, mußte aufgegeben werden. Stattdessen sollte nun eine Aufgabenlösung mithilfe des Rechners gesucht werden. Dafür wurde die Fragestellung zum Themenkomplex geändert in:

Wie kann eine lauffähige formale Grammatik für einen Ausschnitt des Deutschen aussehen,

und wie wird sie erreicht?

und das brachte die folgenden Weiterungen mit sich:

Reduzierung von Inhalten

- Verzicht auf Einordnung der Verben in die Verbfelder und auf die semantische Explikation samt ihrer Beschreibungssprache,
- Kürzung der historischen Entwicklung zur Phrasenstrukturgrammatik (Immediate Constituent Analysis und Konstituentenstrukturgrammatik),
- Verzicht auf die Typen der Phrasenstrukturgrammatik (verschoben auf das Hauptfachstudium),
- Kürzung von Anteilen der traditionellen Schulgrammatik (Wortklassen, Gliedsätze u. a.);

Ergänzung von Inhalten

- Konstruktion einer formalen, leicht zu implementierenden Grammatik,
- Auswahl von relevanten sprachlichen Erscheinungen, die in dieser Grammatik darstellbar sind,
- Verknüpfung von bisherigen Inhaltsteilen mit der implementierten Grammatik;
- Neufassung der Hausaufgabe mit einer Lösung, die die Nutzung der implementierten Grammatik und des Rechners notwendig macht.

Eine wichtige Rahmenbedingung war, daß bei den Studierenden des vierten Semesters (und des 2. Anwendungsfachsemesters) noch keine Prolog-Kenntnisse⁵ voraussetzen waren, daß wir uns andererseits aber nicht von bereits erlernten anderen Programmiersprachen abhängig machen konnten, da die Programmiersprachenauswahl im Grundstudium des Hauptfachs (teils dozentenabhängig) wechselt. Für die Hausaufgabe schied eine Programmieraufgabe demnach aus.

Die formale Grammatik stand unter folgenden Zielvorgaben:

1. Die Grammatik ist linguistisch und programmiertechnisch überschaubar,
2. sie ist linguistisch nicht trivial.
3. Die Grammatik ist veränderbar, ohne daß es sich um bloße Zufügung neuen Vokabulars handelt.
4. Die Grammatik ist transparent, d. h. sie erzeugt sichtbare syntaktische Strukturen.
5. Die Grammatik ist beurteilungsgeeignet, d. h. sowohl erzeugte Sätze als auch deren Strukturen können gemäß den erworbenen Vorkenntnissen der Studierenden als syntaktisch korrekt und linguistisch angemessen eingeschätzt werden.

2. Die Grammatik 'Vierineins'

Die Grammatik 'Vierineins' wurde in Nebenarbeit während eines halben Jahres von zwei wissenschaftlichen und einem studentischen Mitarbeiter konzipiert, von dem studentischen Mitarbeiter implementiert, schließlich wurde sie getestet, bevor sie im SS 1996 zum ersten Mal zum Einsatz kam.

Die studentische Mitarbeit erstreckte sich auch auf weitere Begleitung des Vorhabens und auf die Modifizierungen, die notwendig wurden.

2.1 Bedeutung des Namens

Die Grammatik 'Vierineins' ist eine Definite Clause Grammar, die in Prolog Eclipse implementiert ist. Sie ist das Endprodukt, in das linguistische und linguistisch-methodische Kenntnisse aus zwei nonformalen und zwei formalen Grammatiken eingegangen sind. Daß vier Grammatiktypen integriert wurden, ist weniger aus historischem als aus entwicklungstechnischem und kognitivem Interesse geschehen.

Für die Erkennung syntaktischer Strukturen des Deutschen wird der Begriff der Konstituente benötigt. Eisenbergs Festlegung „Eine Konstituente ist jeder Teil einer syntaktischen Einheit, der bei einer gegebenen Konstituentenstruktur einer Konstituentenkategorie zugeordnet ist.“ (Eisenberg, 48; s. Anm.11) kann leicht akzeptiert werden, wenn die Konstituentenstruktur vorgegeben ist. Für die Studierenden stellt sich aber zunächst die Frage, wie Konstituenten eines individuellen Satzes überhaupt gefunden werden können.

Dafür kann man auf die **traditionelle Schulgrammatik** zurückgreifen, die den Teilnehmenden aus dem Fremdsprachen- oder dem Deutschunterricht bekannt war bzw. ist. Syntaktische Relationen, als Relationen zwischen speziellen Syntagmen zu einem speziellen Satz⁶, können dadurch ermittelt werden, daß mithilfe des Satzmaterials nach einem speziellen Syntagma gefragt wird. Das Problem der Konstituentengrenzen wird damit zwar umgangen, aber das Ergebnis ist erst einmal weiterführend: Gelingt es, ein akzeptables Frage-Antwort-Paar zu finden, ist mit diesem Paar die Art der syntaktischen Relation bestimmt. Wichtig ist dabei zunächst die erfüllte Voraussetzung: Das Syntagma, das in einer syntaktischen Relation zu seinem Satz steht, ist auf jeden Fall eine Konstituente des Satzes.⁷ (Die Auskunft über die Art der syntaktischen Relation wird für den Entwurf des Strukturbaums aufbewahrt, wo sie latent eingeht). Die Zuordnung von Wortformen eines Satzes zu Wortklassen der traditionellen Grammatik liefert lexikalische Konstituenten und ihre Kategorienamen. Die Erinnerung an syntagmatische Relationen wie Rektion und Kongruenz verhilft zur Fixierung von Konstituentengrenzen bei vielen phrasalen Konstituenten. Flexionsmorphematische Analyse und die Nutzung von Rektions- und Kongruenzregeln erfassen Strukturindikatoren der Oberfläche und geben erste Anhaltspunkte für automatisierbare Analyseschritte.⁸

Die **Phrasenstrukturgrammatik (PSG)** kann mit ihren Phrasenkategorien die fehlenden Ebenen der unmittelbaren Dominanz ergänzen und liefert die restlichen Kategorienamen, auch für die zuerst gefundenen Konstituenten.

Weitere linguistische Operationen wie Substitutionstests (incl. Deletion und Permutation) und Koordinationstests werden zur Ergebniskontrolle und weiteren Bestimmung von Konstituenten hinzugezogen.

Entscheidend sind natürlich die formalen Eigenschaften der PSG, die den Studierenden von Lehrveranstaltungen ihres Hauptfaches bekannt sind, nun aber auf breiterer empirischer Basis zum Tragen kommen. Hauptproblem in dieser Phase der Bearbeitung ist: Wie kann der formale Apparat der PSG mit linguistischen Strukturkenntnissen zusammengeführt werden? Da es darüber keine Vorschriften gibt und es laut Chomsky⁹ gleichgültig ist, auf welchem Wege (mithilfe von Intuition, oder Raten, aufgrund praktischer oder methodischer Tips, durch Erfahrungen) die Syntaxregeln erreicht werden, sind die angegebenen heuristischen Methoden für die Studierenden hilfreich – zumindest dann, wenn man linguistisch sinnvolle Ergebnisse bevorzugt. Beim Grammatikschreiben bleibt ausreichend Raum für weitere Maßgaben, etwa Maßgaben der Ökonomie, der Eleganz, der Simplität, der Erweiterbarkeit usf.

Der Beitrag der **Valenzgrammatik** (eingeschränkt auf Verbvalenz) schafft die Grundlage für weitere linguistische Differenzierungen. Daß es sich hierbei um eine andere Sichtweise handelt, die gleichwohl mit den beiden anderen Ansätzen verträglich ist, ist entscheidend für die Gesamtkonstruktion gewesen. Zwar wird mit der Valenztheorie noch eine Anzahl konkurrierender Termini eingeführt, die Verknüpfung mit der PSG ist aber bereits darin angelegt, weil (valenznotwendige) Ergänzungen als Phrasen verschiedener Kategorien realisiert sein können.¹⁰ Außerdem wird die Verbvalenz (mit Eisenberg¹¹) als eine Art der Rektion aufgefaßt.

Die Übertragung der PSG in eine Form der **Definite Clause Grammar (DCG)** bringt einige günstige Modifikationsmöglichkeiten mit sich:

- Das DCG-Format in Prolog erlaubt, durch Hinzufügung eines Hilfsprädikats, die Rekursionstiefe einer rekursiven Regel (hier: auf 3) festzulegen.
- Rektions- und Kongruenzbedingungen können auf einfache Weise in die DCG-Regeln inkorporiert werden. (Das ist besonders deswegen wünschenswert, weil Valenzrahmen Realisierungen in bestimmten Kasus vorschreiben.)

Die mnemotechnisch günstigen Kategorienamen der Phrasenstrukturgrammatik können beibehalten werden.

2.2. Leistung der Grammatik

Die DCG 'Vierineins' umfaßt insgesamt zehn Valenzrahmen, wobei neun verschieden voneinander sind. Es handelt sich dabei um zwei einstellige, vier zweistellige und vier dreistellige Valenzrahmen zu deutschen Verben. Die Ergänzungen erfassen ein Spektrum von Akkusativ-, Dativ-, Nominativ-Ergänzung, Adverbial-Ergänzung, Präpositional-Ergänzung und Satz-Ergänzung. Satz-Ergänzungen, die von zwei Valenzrahmen gefordert werden, sind durch die Regeln auf Sätze mit einstelligen Verben, ein zweistelliges und ein dreistelliges, eingeschränkt. Durch letzteres werden Einbettungen in geringem Maße gestattet. (Zielvorgabe 2)

In der DCG finden die folgenden Wortklassen Berücksichtigung: Verben, Substantive, Präpositionen, Determinantien, die Subjunktion DASS und das Reflexivpronomen SICH (im Dativ und Akkusativ). Ein Verb wird mit einem abtrennbaren Verbpräfix gebildet.

Für die Menge der Substantiva haben wir Feminina im Singular in den Kasus Akkusativ, Dativ und Nominativ ausgewählt, um die Regelmenge zu reduzieren. Wir gebrauchen jedoch den definiten und den indefiniten Artikel, um Kasusunterschiede sichtbar zu machen und um die erzeugten Sätze natürlicher erscheinen zu lassen.

Die Grammatik arbeitet nur mit einem synthetisch gebildeten Tempus, um das Auftreten diskontinuierlicher Elemente zu vermeiden. Das ausgewählte Präsens führt bei manchen der Verben zu Sätzen, die nicht als Beschreibungen gegenwärtiger Sachverhalte benutzt werden können. Gegen den Einsatz des Präteritums sprach, daß die Sätze gesucht klangen. Natürlichkeit und syntaktische Korrektheit ließen sich nur schwerlich vereinen.

Aber das ausgewählte terminale Vokabular bringt es ohnehin zwangsläufig mit sich, daß neben semantisch sinnvollen auch syntaktisch korrekte, aber semantisch wenig sinnvolle Sätze erzeugt werden.

Mit 15 nonterminalen und 41 terminalen Regeln bleibt die formale Grammatik übersichtlich. (Zielvorgabe 1)

Die DCG ist sowohl für die Analyse einzugebender Sätze geeignet als für die Generierung, wobei Sätze direkt als ganze oder unter der Vorgabe bestimmter Variablenbelegungen erzeugt werden können. Ein Hilfsprädikat sorgt dafür, daß zum ersten Kennenlernen nummerierte Testsätze angezeigt werden, die die Grammatik akzeptiert, und solche, die sie nicht akzeptiert. Für die ersten werden außerdem die Strukturbäume ausgegeben, wie das auch für alle erzeugbaren Sätze gilt. Die Baumstrukturen werden vertikal, in Spalten eingerückt notiert, so daß Schwesterknoten in derselben Spalte zu finden sind. (Zielvorgabe 4).

2.3 Angeschlossene Übungsaufgaben

Die erste der mit der DCG verbundenen Übungsaufgaben verlangt von den Studierenden die Ergänzung des terminalen Vokabulars um zwei weitere Verben und die entsprechende Ergänzung der Syntaxregeln (per analogiam). Dazu muß auf der Basis eines kleinen Satzkorpus die Anzahl und Art der Valenzrahmen beider Verben festgestellt werden. Durch die Angabe eines solchen Korpus kann die Aufgabe gezielt auf bestimmte Valenzphänomene ausgerichtet werden, zumal hier auch solche Sätze eingefügt werden können, die gerade nicht abgeleitet werden sollen.

Arbeitsschritte, die die Studierenden finden und anwenden müssen, sind die folgenden:

- o) ggf. Übertragung der Passiv-Sätze in die Aktiv-Gegenstücke;
- i) ggf. Feststellung der nicht-valenznotwendigen Konstituenten (mithilfe von Substitutions- und Deletionsproben, evtl. Paraphrasenbildung) und Streichung aller freien Angaben sowie weiteren nicht-valenz-notwendigen Materials;
- ii) Valenz-Analyse der Sätze des Korpus (mithilfe der Deletions-, Permutations-, Substitutionsprobe);
- iii) Zusammenstellung der Sätze mit gleichartiger Ergänzung, d. h. u. a. Entscheidung darüber, ob eine Präpositional-Ergänzung oder eine Adverbial-Ergänzung vorliegt, Gegenkontrolle etwa durch Substitution einer Adverbialphrase;
- iv) ggf. Revision der Valenz-Analyse;
- v) Festlegung auf je einen oder mehrere Valenzrahmen beider Verben aufgrund der Art der Ergänzungen,
ggf. abhängig von der Verbsemantik (Zuhilfenahme von Paraphrasenbildungen);
- vi) als Hilfsschritt: Notierung der entsprechenden PSG-Regeln für VP, ggf. mit Zusatzkonditionen über Kasuswahl, Präpositionenauswahl und bedingte oder unabhängige Fakultativität der Ergänzungen;
- vii) Übertragung in je eine oder mehrere nonterminale DCG-Regeln;
- viii) Ergänzung der DCG um die notwendigen terminalen Regeln.

Zweifellos können die Arbeitsschritte – korpusabhängig – weiter unterteilt werden, z.B. auch zur Trennung polysemer von homonymen Verben.

Wird das Bewußtsein der Studierenden an einer solchen Aufgabe dafür geschult, wie schrittweise vorzugehen ist, so wird auch einer Algorithmisierung der Arbeiten dort Vorschub geleistet, wo es möglich ist.

Die Grammatik erzeugt fehlerhaft einen Hauptsatz, in dem die Abtrennung eines abtrennbaren Verbpräfixes unterbleibt. Die Studierenden stellen fest, daß die DCG diese – der Beobachtung von Sprachdaten zuwiderlaufende – Erscheinung als syntaktisch korrekt ableitet und anbietet. Einige bemerken, daß es eine Konkurrenzform des Verbs gibt, bei der dieselbe Zeichenfolge nicht als Präfix, sondern als Präposition zum Anschluß eines Präpositionalobjekts steht; sie weisen darauf hin, daß es sich um eine semantische Verbvariante handelt. (Zielvorgabe 5)

Wenn für den ersten Teil der Hausaufgabe die Unterscheidung von Adverbial-Ergänzung, Präpositional-Ergänzung und freier Angabe, alle als Präpositionalphrasen realisiert, zu leisten war, so sind im letzten Teil Anschlußüberlegungen für den Fall anzustellen, daß die Präpositionalphrase attributiv steht. Während die DCG 'Vierineins' Adverbial-Ergänzungen von Präpositional-Ergänzungen über Regelinhalte unterscheidet, während sie freie Angaben von valenznotwendigen Ergänzungen strukturell im Strukturbaum disambiguiert, können Attribute mit der gegebenen DCG gar nicht abgeleitet werden. So kann den strukturellen Homonymien etwa in

Die Kinder besuchen die Großmutter im Haus an der Brücke.

nur zu einem Teil Rechnung getragen werden. Weiteres Experimentieren mit dem gegebenen formalen Apparat wirft außerdem die Frage auf, ob beide Präpositionalphrasen gleichgeordnete freie Angaben oder ob sie Attribut (zur Akkusativ-Ergänzung) + freie Angabe sein können. Hier zeigt sich den Studierenden, daß formale Grammatiken auch dazu geeignet sind, strukturelle Ambiguitäten aufzudecken, während sie dem normalen Benutzer der natürlichen Sprache unbeachtlich erscheinen. (Zielvorgabe 5)

In demselben Zusammenhang wird auch festgestellt, daß die an sich wünschenswerte Disambiguierung von (valenznotwendigen) Ergänzungen und freien Angaben im Strukturbaum bei zwei- und dreistelligen Verben mit einer zwar syntaktisch akzeptablen, aber unüblichen Konstituentenreihenfolge erkauft wird.¹² (Zielvorgabe 5)

3. Erfahrungen aufgrund des Einsatzes von 'Vierineins'

Lösungsangebote

Die Aufgabenlösungen der drei Durchgänge in den Sommersemestern 1996, 1997 und 1998 werden miteinander verglichen. Ich beschränke mich dabei auf den wichtigsten, den ersten Teil der Übungsaufgaben, der sich auf die Zufügung von weiteren Valenzrahmen zweier Verben zur DCG bezieht.

3.1.1 Sachlage

Die Aufgabenlösung für Verb A machte eine Trennung in V1 und V2 (mit Präpositionalergänzung) notwendig; beide Verben unterscheiden sich semantisch. Zudem mußte V1 in zwei Regeln gefaßt werden, da die beiden fakultativen Ergänzungen nicht beide zugleich unrealisiert bleiben können. Insgesamt waren also drei neue nonterminale Regeln zu schreiben und zwei terminale.

Die Aufgabenlösung für Verb B machte eine Trennung in V1 und V2 (mit unechtem Reflexivum) notwendig; beide Verben unterscheiden sich semantisch erheblich. Es waren zwei neue nonterminale Regeln zu schreiben und zwei terminale, wenn beide Lesarten grundsätzlich voneinander getrennt werden sollten.

3.1.2 Lösungstendenzen

Im Sommersemester 1996 hat nur eine Arbeitsgruppe (von insgesamt 5) von dem Angebot Gebrauch gemacht, die Lösung in PSG-Regeln zu notieren. Daraufhin konnte dieses Angebot bei den Folgejahrgängen vernachlässigt werden, zumal dann mehr Zeit für die Einführung in die implementierte Grammatik war.

Während eine Arbeitsgruppe die DCG für Verb A um drei nonterminale und eine terminale Regel (für V1 und V2 zugleich) ergänzte, haben die übrigen Arbeitsgruppen ein Splitting vorgenommen und zwei Lesarten von V1 kurzerhand in zwei bereits vorhandenen nonterminalen Regeln untergebracht. Die nicht ausdrücklich verlangten Valenzrahmen wurden von lediglich einer Gruppe sichtbar bei der Lösung der Aufgabe eingesetzt. Ein Splitting wurde auch für Verb B trotz seiner semantisch stark unterschiedlichen Lesarten benutzt; die Kommentare erschöpften sich in „siehe Klasse...“ und die Ergänzungen erschienen nur in einzelnen terminalen Regeln. Die Studierenden favorisierten diese Strategie so sehr, daß ihnen sogar Analysefehler entgingen. Auf die Frage des Feedback-Fragebogens „Wäre es Ihnen lieber gewesen, wenn auf die Valenz-Grammatik verzichtet worden wäre?“, antworteten 4 der Teilnehmer mit „Nein“, einer mit „Ja“, die restlichen blieben indifferent.

Die Lösungstendenz liegt darin, die Bearbeitung des einzelnen Satzes irgendwo im bestehenden Regelsystem durchführen zu lassen. Um diesen Weg zu versperren, wurden für den 2. Durchgang zwei nonterminale Regeln in der DCG geändert.

Im Fragebogen war mehrmals der Wunsch aufgetaucht, mehr Gelegenheit zum Kennenlernen der DCG zu haben.

Nun hatten den Studierenden als Lösungszeit elf Arbeitstage (Pfingstferien) zur Verfügung gestanden, was einige erklärtermaßen als wichtig anmerkten. Im entsprechenden Handout waren verstreut Hinweise zur Charakteristik der Grammatik zu finden, allerdings keine gefällige Übersicht darüber, was die Grammatik bewirkt. Die Unterlagen wurden von allen als ausreichend empfunden, also kamen wir dem Wunsch auf andere Weise entgegen (s.u. Spielgrammatik).

Im Sommersemester 1997 zeigten die Aufgabenlösungen zum gleichen Aufgabenteil weitgehende Absprachen zwischen den Arbeitsgruppen, was bis zu Inkonsistenzen ging, wenn eigene und fremde Gedankengänge vermischt wurden. Deswegen lassen sich Lösungstendenzen dabei kaum festhalten.

Im Grunde lagen zwei Lösungstypen vor:

- eine nonterminale Einzelregel für eine Sonderbehandlung der Satz-Ergänzungen anstelle einer Erfassung als Variante der Akkusativ-Ergänzung bei Verb A (V1) und eine fehlerhafte Lösung für die bedingt fakultativen Ergänzungen bei Verb A;
- vier nonterminale Regeln für V1, wobei für die bedingt fakultativen Ergänzungen eine Lösung geschrieben wurde, die als ‘sophisticated’ anzusehen ist: Beide Ergänzungen wurden in einer Regel als ungebunden fakultative notiert, allerdings im Kontext der Satz-Ergänzung als obligatorischer Ergänzung. So wurde verhindert, daß das verbotene Syntagma (nur mit Nominativ-Ergänzung) abgeleitet werden konnte. Diese Lösung zeigt weitgehende Einsicht in den Regel-Mechanismus und seine inneren Abhängigkeiten – leider wurde Verb A dadurch aber zu einem vierstelligen Verb erklärt. Diese Tatsache war den Erfindern entgangen, weil sie mit ihrer Grammatikregel einen passablen deutschen Satz ableiten konnten, bei dem der Daß-Satz strukturell als Attribut zur Akkusativ-Ergänzung auftrat. Diese Struktur wurde allerdings durch zufällig gut passende lexikalische Elemente verschleiert. Der Generierungstest konnte die Erfinder deshalb nicht warnen, und auf die Ausgabe des Strukturbaums hatten sie daraufhin verzichtet.

Die starken semantischen Unterschiede in den Lesarten von Verb B wurden nicht wesentlich anders behandelt als im Vorjahr und hatten daher keine merklichen Rückwirkungen auf die Behandlung von Verb A.

Für die Aufgabenlösung war den Studierenden eine Wochenfrist gegeben. Die Kürze der Zeit mag ein Grund für das Einschwenken auf zwei Lösungstypen gewesen sein. Einzelsitzungen mit den Arbeitsgruppen am Terminal, wo mit den

angebotenen Lösungen experimentiert wurde, konnten im nachhinein das Verständnis für die Aufgabenstellung fördern. Die Sitzungen mußten außerhalb der regulären Übungsstunden erfolgen.

Auch bei diesem Durchgang wurde die Valenztheorie explizit kaum in die Lösungskommentare übernommen, weswegen wir im Folgejahr die Aufgabe in dieser Richtung pointierten.

Das Sommersemester 1998 unterschied sich von den beiden Vorgängern besonders dadurch, daß das wahlfreie Tutorium zur Lehrveranstaltung dieses Mal von den meisten Teilnehmern lückenlos besucht wurde. In dieser zusätzlichen Zeit konnten wir uns Fragen ausführlich widmen, Diskussionen zuende führen und Übungen an wechselndem Material durchführen. Insofern sind die Rahmenbedingungen andere als vorher.

Die Aufgabenlösungen sind erstmals in Einzelbearbeitung entstanden, obwohl es den Studierenden freigestellt war, Arbeitsgruppen zu bilden, was für andere Übungsgelegenheiten auch geschah. Die Lösungen zeichneten sich bei insgesamt breiterem Spektrum dadurch aus, daß weniger Generalisierungen gemacht wurden, mit einer Tendenz „eine neue Regel für jeden Satz“. Anstelle einer non-terminalen Regel mit zwei fakultativen Ergänzungen wurden drei Regeln mit obligatorischen Ergänzungen geschrieben. Das zeigt ein Ausweichen gegenüber dem Zugang, die bedingt fakultativen Ergänzungen bei Verb A (V1) direkt anzugehen. Stattdessen wurde von vielen eine umständliche, aber dafür korrekte Lösung bevorzugt. Vermutlich fehlte bei der Bearbeitung das Gruppenmitglied, das gesagt hätte: „Aber, diese Regeln können wir doch noch zusammenfassen.“ Die wenigen, die jetzt noch das Splitting einsetzten, mußten grobe Übergenerierung in Kauf nehmen, was sie auch kommentarlos taten. Eine ausgezeichnete Lösung entstand dadurch, daß die Ergebnisse der Valenzanalyse für Einzelsätze auf wenige Valenzrahmen unter Angabe der Begründungen zusammengeführt wurden. Dort war der Entstehungsprozeß der Regeln mühelos nachzuvollziehen.

Daraus wird die Lehre gezogen, beim nächsten Mal getrennt Valenzanalysen zu üben und/oder die Valenzanalysen von Verb A und Verb B als einen Teil der Aufgabe ausdrücklich zu fordern und/oder die benutzten Arbeitsschritte (etwa wie in 2.3 o)–viii)) aufgrund von Selbstbeobachtung aufschreiben zu lassen.

Eine Spielgrammatik zum freien Experimentieren

Die Rückmeldungen aus dem Sommersemester 1996 veranlaßten uns, optionales Übungsmaterial für den experimentellen Zugang zu implementierten formalen Grammatiken bereitzustellen. Wir sahen es als ungünstig an, wenn wir die DCG ‘Vierineins’ erweitert hätten, denn das lief der Zielvorgabe ‘Überschaubarkeit’ zuwider. Auch konnten wir die Grammatik, wie sie war, nicht schon vorher zum Einüben benutzen; damit hätten wir die Aufgabenlösung nahegelegt und ihr den Entdecker-Reiz genommen.

So schalteten wir der DCG eine vereinfachte Spielgrammatik vor, die ebenfalls in Prolog (Eclipse) implementiert ist, die aber den DCG-Formalismus nicht benötigt, weil sie sehr vereinfacht ist und beispielsweise keine Kasus kennt. Die Grammatik erzeugt und analysiert Sätze des Deutschen und gibt die entsprechenden Baumstrukturen aus.

Die Existenz dieser Grammatik und die Zugangsmöglichkeit wurde den Studierenden des Sommersemesters 1997 zu Beginn des Themenkomplexes Syntax über eine zweiseitige schriftliche Vorstellung bekannt gemacht. Auf der zweiten Seite gibt es einen Punkt 4, der mit „Experimentieren mit der Grammatik“ überschrieben ist und erklärt, daß eine Kopie der Grammatik editiert und in eine neue Grammatik umgeschrieben werden kann, die ohne weiteres zu testen ist. Darin wurden die Studierenden auch auf den Kommentar (in der Datei) zur Grammatik verwiesen, der ihnen Einblick in Regeln und Vokabular gibt. Im Feedback am Semesterende, das neun der dreizehn Teilnehmer abgaben, zeigte sich: Drei hatten die Spielgrammatik zum Generieren genutzt; zu den Einsichten befragt, die sie dabei gewonnen hätten, schrieben zwei „keine“, einer aber: „,Ich habe erst zu spät gesehen, daß ich hätte damit experimentieren können.“ Ein solches Ergebnis fanden wir unbefriedigend und für die Zukunft änderungsbedürftig – auch ein Grund, sich im Tutorium mit dem Experimentieren zu befassen.

Die Spielgrammatik enthält Regeln für Sätze, die auf zwei fächerartige Strukturen zurückgehen, einen Drei-Wort-Typ und einen Vier-Wort-Typ, jeweils mit Verbzweitstellung. Gebraucht wurden ein Ausgangssymbol und Einzelbuchstaben¹³ als Klassenbezeichnungen für die terminalen Elemente. Es ist möglich und gewünscht, die vorgegebene Kategorisierung der Grammatik zu ändern.

Die Menge der terminalen Elemente besteht aus Eigennamen und einem singularen Substantiv (Gattungsname) in der Wortform des Nominativ/Akkusativ Plural ohne Artikel, das auf ersten Anschein einem Eigennamen ähnlich sehen könnte. Die Menge dieser Wortformen ist mit der Nutzung nichtleerer Schnitt-

mengen auf drei Mengen verteilt. Zu den terminalen Elementen gehört eine vierte Menge, die aus 12 Verbformen besteht, alle in der dritten Person Singular Präteritum Indikativ Aktiv, stark bzw. schwach konjugiert, sowie dazu eine Verbform abweichend im Plural. Es wurden – wiederum bis auf eines – solche Verben ausgewählt, die mit zwei, alternativ mit drei, Eigennamen zusammen einen sinnvollen Satz bilden.

Kritischen Betrachtern kann auffallen, daß die Klassen der Eigennamen auf eine hätten reduziert werden können, daß das Substantiv im Plural eine eigene Klasse verdient hätte und daß im Gegenzug die Klasse der Verbformen am besten in mehrere Klassen geteilt wäre. Bei entsprechender Regeländerung wäre so zu erreichen, daß ausschließlich korrekte und – bis auf Exemplare mit Wortwiederholungen – auch sinnvolle Sätze durch die Grammatik erzeugt werden. Auf dem Wege zu diesem Ziel kommt man an vier Einsichten nicht vorbei:

- a) Es gibt Verbformen, die zwei und solche die drei Ergänzungen verlangen. Beide Mengen sind disjunkt.
- b) Das Substantiv an dritter Stelle im Satz vom Drei-Wort-Typ bildet nur mit einer Teilmenge der Verben korrekte Sätze, – bei anderen müßte eine andere Wortform aus demselben Wortparadigma stehen: die anderslautende Dativ-Form.
- c) Nur mit der Pluralverbform ist das Substantiv an erster Stelle eines Satzes auch das Subjekt. Diese Verbform sollte daher in einer eigenen Klasse stehen.
- d) In syntaktisch korrekten Sätzen, in denen Substantiv (Gattungsname) und Eigennamen je einen Platz einnehmen, können beide die Plätze tauschen, ohne daß sich die Bedeutung der Sätze ändert. Wenn ein vergleichbarer Stellentausch nur unter Eigennamen vorgenommen wird, **kann** sich die Bedeutung der Sätze ändern, muß es aber nicht; d.h. diese Sätze sind strukturell homonym.

Aus diesen Einsichten können wiederum andere abgeleitet werden:

- aus a): Verben fordern unterschiedlich viele Ergänzungen;
- aus b): Die geforderten Ergänzungen sind nicht kasus-indifferent, und weiter:
- Daß Konflikte nur bei Ergänzungen durch den Gattungsnamen auftreten, kann zwei Gründe haben:

- Entweder Verben verhalten sich gegenüber Ergänzungen in Eigennamenform anders als gegenüber Ergänzungen in Gattungsnamenform;
- oder Verben verhalten sich gegenüber Ergänzungen unabhängig von ihrer Repräsentationsform gleich, dann sind die Wortformen der Eigennamen homograph zueinander. (Das Homographenproblem ist den Studierenden aus dem Vorsemester bekannt.)

Mit anderen Worten: Das Problem ist vorhanden, es wird an Eigennamen nur nicht sichtbar.

aus c): Konjugationsmorpheme helfen bei der Disambiguierung von Satzstrukturen.

aus d): Es gibt Wortstellungsvarianten von syntaktisch korrekten Sätzen. (Und demnach: An der Oberfläche unterscheidbare Kasus sind zum Experimentieren den homographen Wortformen eines Paradigmas vorzuziehen.)

Punkt a) und b) bereiten die Valenztheorie vor, d) ist ein erster Beitrag für den folgenden Themenkomplex 'Topologie'.

Last but not least: Selbst bei einer simplen formalen Grammatik kann man durch geschickte Kategoriebildung Übergenerierungen teilweise unterbinden.

Der Rechner in der experimentierenden Arbeitsgruppe

Im Sommersemester 1998 wurde das Experimentieren mit der Spielgrammatik Gegenstand des Tutoriums. Nachdem die Studierenden ihre ersten Erfahrungen mit dieser Grammatik allein gesammelt hatten, ebenfalls ziemlich hilflos fragend, wozu diese Grammatik nütze sein soll, folgte eine Gruppensitzung am Terminal mit Rollenverteilung.

Der Sitzung wurden zwei relativ allgemeine Fragen richtungweisend vorangestellt.

- 1) Die Spielgrammatik erzeugt Sätze, die syntaktisch inkorrekt sind; welche Fehler treten auf?
- 2) Die Spielgrammatik erzeugt oder analysiert nicht alle Deklarativsätze, die aufgrund der terminalen Elemente zu erwarten wären; woran liegt das?

Während die erste Frage eine Bestandsaufnahme auslösen sollte, mußten für die Beantwortung der zweiten Materialien zur Begründung gefunden werden.¹⁴

Beide Aufgaben hätten bei entsprechender betreuender Anleitung an zwei vier- bis fünfköpfige Arbeitsgruppen delegiert werden können, wenn zwei Workstations in ausreichend großen benachbarten Räumen zur Verfügung gestanden hätten und wenn zusätzlich eine Gelegenheit hätte organisiert werden können, so daß beide Arbeitsgruppen am Ende die Ergebnisse ihrer Arbeit hätten mündlich und mit Demonstrationen am Bildschirm austauschen können. Dazu wird nicht nur von seiten der Studierenden ein gewisses Zeitkontingent mitgebracht werden müssen, die vorbereitende Organisation von Betreuerseite, auch die des reibungslosen technischen Ablaufs, ist nicht zu unterschätzen. Dieses didaktische Experiment steht noch aus.

Stattdessen starteten wir einen anderen Versuch, bei dem einiges dem Zufall überlassen blieb:

Die Arbeitsteilung, bei der ein Gruppenmitglied die Tastatur bediente, ein anderes Kontrolle und Hilfestellung für das erforderliche Input-Format übernahm, erwies sich als sinnvoll. Beide zusammen waren die Mittler für das Einbeziehen des Rechners in die gemeinsame Aufgabe, weil sie nicht nur für den Input, sondern auch – wie es sich ergab – für die Weitergabe des Rechner-Outputs an die Gruppe zuständig waren. Bei den übrigen Mitgliedern wechselten nach anfänglichem Suchen die Rollen: einige notierten Teilbefunde auf der Wandtafel bzw. gut lesbar auf Endlospapier, einige achteten auf strategisches Vorgehen, einige sammelten Teilfragen, andere setzten sie in Testaufgaben an das System um. Einige wollten von anderen etwas erklärt haben, andere boten verschiedene oder verschieden formulierte Erklärungen an. Einer achtete auf den Zeitrahmen. Die Rolle des Regisseurs war vakant, weswegen es Phasen der Unsicherheit gab. Gefehlt haben uns Protokollanten und jemand, der die aktuellen Ergebnisse zusammengefaßt hätte, zum Mitnehmen.

Über die Beantwortung der beiden Leitfragen sind wir in 90 Minuten nicht hinausgekommen, jedenfalls nicht so, daß wir die Spielgrammatik-Regeln geändert hätten. Die in a) – d) genannten Einsichten wurden berührt, ohne daß man eine Aussage darüber machen dürfte, ob und wie weit sie von allen Mitgliedern geteilt wurden.

Eine Block-Sitzung des Tutoriums diente der individuellen Nachbesprechung der Syntax-Aufgaben, die mithilfe der DCG 'Vierineins' gelöst worden waren. Dabei wurden zu einem Teil die in der Aufgabenlösung vorgeschlagenen Syntax-

Regeln in ihren Auswirkungen mithilfe des Rechners getestet, – oft aber mußten zunächst die linguistischen Folgen aus den Vorschlägen gemeinsam durchdacht werden. Das geschah teils in Einzelgesprächen, teils in Kleingruppen, wenn sich dieselbe Angelegenheit bei mehreren als problematisch erwiesen hatte.

Insgesamt: Der Rechner war auf drei Weisen bei der Bearbeitung der DCG ‘Vierineins’ und der Spielgrammatik eingesetzt:

- in von außen ungesteuerter Interaktion zwischen einem Menschen und dem Rechner,
- als Testinstrument in der sich selbst überlassenen experimentierenden Arbeitsgruppe, also eingefügt in interpersonale Kommunikation zum Zweck des gemeinsamen Lösens von Übungsaufgaben,
- als Test- und Kontrollinstrument in angeleiteten problemzentrierten, experimentellen Kommunikationssituationen der Arbeitsgruppe.

Während für die ersten beiden Konstellationen Beobachtungen fehlen und allenfalls aus Aufgabenlösungen und Feedback-Reaktionen vorsichtige Rückschlüsse möglich sind, kann man von der dritten Konstellation sagen, daß sie in das übrige didaktische Vorgehen gut eingepaßt und daß sie für die Art des angestrebten Lernens fruchtbar war.

Ohne Gründe dafür erforscht zu haben, ist zu konstatieren, daß beide zum Experimentieren implementierten Grammatiken keine „Selbstläufer“ sind, vielmehr erst in Erörterungssituationen von Kleingruppen zu ihrem endgültigen Ziel gelangten. (Da es sich um kein didaktisches Forschungsprojekt handelt, kann Stichhaltigeres dazu nicht gesagt werden.)

4. Rückblick

Soll ein Fazit aus den drei Durchgängen mit der Grammatik ‘Vierineins’ gezogen werden, so bietet sich das indirekte Urteil der Studierenden an: Auf Nachfrage erklärten sie den Themenkomplex ‘Syntax’ für den für sie persönlich am wichtigsten aus der Grundlagenveranstaltung. Außerdem sei die Lehrveranstaltung interessant gewesen und man habe unerwartetes Neues gelernt.

Von seiten derer, die das Lehrprojekt verantwortlich vorbereitet haben, bleibt einiges zu bedenken:

Das thematisch gebundene Lehrziel, zu vermitteln, wie man natürlichsprachliche syntaktische Strukturen erkennt, beschreibt und die Beschreibungen schließlich in eine formale, implementierte Syntax umformt, ist von einem Informatik-Standpunkt der Studierenden her anscheinend erreicht. Von einem computerlinguistischen Standpunkt gesehen, fällt die mangelnde Generalisierung aufgrund linguistischer Prinzipien auf (Lösungstendenzen: Splitting, Ein-Satz-Regel). Es wäre aber wohl unpassend zu behaupten, daß Vorhaben sei nicht geglückt, wenn die Lösungen den ursprünglichen Absichten der Lehrenden nicht völlig entsprechen. Bringt es doch diese Weise des Lernens mit sich, daß das Gewicht von einer Ergebnisorientierung auf die individuellen Lernwege, die die Studierenden wählen, verlagert ist. So muß den Studierenden zugestanden werden, daß sie berechtigt anmahnten, ihre Leistung dürfe eigentlich nicht nur an Lösungen auf den Übungsblättern gemessen werden, da sie in den Übungen mehr gelernt hätten, als abfragbar sei, – wobei ‘mehr’ vermutlich ‘anderes’ bedeutet.

Außerdem müssen sich auch Lehrende vergegenwärtigen, daß sie im Ansatz auf die Vorstellung von einem abgeschlossenen Wissensfundus verzichtet und die Vorstellung von einem nach vorn offenen Wissen propagiert haben. Auch für die Leistungsbewertung müssen wahrscheinlich neue Wege gesucht werden.

Wie wir bei dem Themenkomplex ‘Syntax’ und anderen computerlinguistischen Themenkomplexen gesehen haben, kann es die einzig richtige Lösung für eine Aufgabe oft nicht geben, vielmehr gilt es, abzuwägen, unter welchen Kriterien eine Lösung ‘erfolgreich’, ‘akzeptabel’ oder nur ‘neutral gegenüber den Anforderungen’ genannt zu werden verdient. Das ist für die zukünftigen Praktiker und Praktikerinnen ein entscheidender Lernschritt.

Gelernt wurde weiter, daß Wissen, das nicht nur abgespeichert werden soll, sondern das dem Erschließen neuen Wissens dient und das handlungssteuernd (z.B. beim Experimentieren) wirkt, nicht ohne ein gutes Maß an Zeit und Energie, oft auch an Beständigkeit, zu erwerben ist, – eine Einsicht, die besonders Teilzeit-Studierenden neu sein kann. Die Studierenden fügten aus ihrer Beobachtung hinzu, daß es zum Verstehen (im Sinne von: Beherrschen) meist nicht ausreicht, wenn die Lösung einer Aufgabe exemplarisch vorgeführt wird.

Wenn gemeinsames Entdecken an die Stelle von Wissenspräsentation rückt, so müssen Stundenplanungen an mehreren Stellen reversibel sein. Studierende schlugen vor, Unbearbeitetes und offene Fragen, die sich während der Sitzungen ergaben, gegen Ende der Stunde zu sammeln, explizit zu benennen und dann zu

entscheiden, ob und gegebenenfalls wann dieses gemeinsam oder von einem Teil der Gruppe bearbeitet wird.

Die Abhängigkeit vom Rechner zeigt sich noch auf unerwartete Weise: Die Studierenden hatten das DCG-Programm auf ihre PC's daheim geladen, um sich die Interna der Grammatik anzusehen. Einigen fiel später aus Vergleichen auf, daß Prolog Eclipse nur teil-akzeptiert worden war: die Regel, die Rekursionsbeschränkung durchführt, wurde ohne besondere Meldung nicht übernommen. Glücklicherweise war diese Regel nicht aufgabenrelevant, jedoch ändert das nichts an dem Faktum, daß nicht vorhersehbare, für einige Adressaten veränderte Lernsituationen aufgrund technischer Bedingungen auftreten können.

Defizite sind auf drei Ebenen zu verzeichnen:

- 1) curricular: Das Lehrkonzept ist nicht ohne ein Tutorium durchzuführen, das bedeutet für die Studierenden Bereitschaft zu längerer als der curricular vorgesehenen Arbeitszeit, auch wenn sich diese Rechnung bezogen auf die Gesamt-Studienzeit wahrscheinlich wieder ausgleicht.
- 2) inhaltlich: Das DCG-Format als solches ist nicht in die Vorbereitungen einbegriffen, Herkunft und logische Fundierung der DCG wurden nur gestreift.
- 3) kritisch: Inkonsistenzen (linguistischer Art) in der Formulierung von DCG-Regeln wurden aus praktischen Gründen mit Rücksicht auf den Formalismus in Kauf genommen, aber nicht thematisiert.

Insgesamt konnte in mehreren Zusammenhängen nachgewiesen werden, daß sich Lerninhalte und didaktisches Vorgehen wechselseitig bedingen, angefangen bei der Basis-Entscheidung, daß man schlecht moderne Erklärungen für Sprach- und Textverstehen lehren kann, ohne daß diese das praktische Vorgehen beeinflussen, – es sei denn, man wagt es, die Adressatinnen einem schwerwiegenden Widerspruch auszusetzen, der noch dazu unausgesprochen bleibt. Das aber widerspricht unserem wissenschaftlichen Streben nach Klarheit.

Nachbemerkung

Leserinnen und Leser mögen den Eindruck gewonnen haben, hier sei etwas Selbstverständliches zum Gegenstand eines Lehrprojektes gemacht worden, zumal wenn sie selbst erfahrungsgelitetes und erörterungsorientiertes Lernen bevorzugen oder praktizieren. Veränderte Lernumgebungen in der Gesellschaft wie die Nutzung von sprachlichen Informationssystemen und Intellimedia-Präsentationssystemen

können etwas Selbstverständliches zu etwas Bemerkenswertem machen. Damit das weniger geschieht und Lernen ohne Amputation menschlicher Fähigkeiten und Bedürfnisse auskommt, mögen diese Leser und Leserinnen den Bericht als Bestätigung oder Ermutigung nehmen.

Anmerkungen

1. „Wir“, das sind Dozent(inn)en des Grundstudiums. An der Entwicklung und Erprobung der Grammatik ‘Vierineins’ waren H.-D.Lutz und U.Koch maßgeblich beteiligt. Ihnen, lieber Herr Lutz und lieber Herr Koch, sei hier für Ihre Arbeit gedankt. Ebenso richte ich meinen Dank an Sie, die Studierenden der letzten drei Jahrgänge, und an die Institutsleitung unter der Geschäftsführung von Herrn Prof.Dr. Bători, die uns eine studentische Mitarbeiterstelle zur Verfügung stellte.
2. Auch wenn das Transportmodell der Kommunikation nicht in reiner Form auftritt, scheint es zählbar zu sein: Es bedeutet nämlich keine grundsätzliche Veränderung, wenn z.B. in Forschungsprojekten nicht mehr von Information verlangt wird, eingängig zu sein, sondern neuerdings nach der „effektiven“ Informationspräsentation gesucht wird, – was heißen soll, daß nun einer Präsentationsform die Eigenschaften zugeschrieben werden, „leicht verständlich“, „schnell wahrnehmbar“, „einprägsam“ und „ansprechend“ zu sein. (So: André, Elisabeth; Rist, Thomas: Von Textgeneratoren zu den Intellimedia-Präsentationssystemen. [DFKI Research Report. 94–02], Kaiserslautern, Saarbrücken: DFKI, 1994., S. 3.)
Damit ist man einer sprachlichen Täuschung erlegen: Verdeckt wird, daß die natürlichsprachliche Formulierung eine Übertragung erlaubt: Das Ergebnis eines Kommunikationsprozesses wird als ein Attribut einer der Prozeßkonstituenten angehängt. Es ist nicht länger davon die Rede, daß es einer Person gelungen ist, eine andere Person auf eine bestimmte Weise in einer Sache anzusprechen, sondern sogar Sender und Empfänger – um im Vokabular des Transportmodells zu bleiben – sind in der Vorstellung etwa von einer „leicht verständlichen Instruktion“ aufgehoben, – „aufgehoben“ im schillernden Doppelsinn von „aufbewahrt“ und „suspendiert“.
3. Zweifellos könnte hier von ‘Lehr-/Lernsituation’ geredet werden, angemessen erscheint trotzdem der verkürzte Begriff, weil eine Akzentverschiebung auf das gegenseitige Voneinander-Lernen durchaus intendiert ist.
4. Zu ähnlichem Zweck haben wir in einem zweiten Vorhaben versucht, den Themenkomplex ‘Topologie’ zu ändern. Über Ergebnisse davon wird demnächst zu berichten sein.
5. Auch das Erlernen von Prolog in einem Vorkurs verbot sich, da die Studierenden in diesem Semester im Anwendungsfach mit dem Erlernen und Anwenden von LISP

beschäftigt waren.

6. Vergleichbares gilt auch für die syntaktischen Relationen 2. Ordnung.
7. Auch Konstituenten sind nur zu einem Teil von individuellen Sätzen unabhängig: Während die Konstituenten *der* und *Mensch* in vermutlich allen Vorkommen **eine** Konstituente bilden, ist das bei den Konstituenten *natürlich* und *weiß* nicht immer der Fall, z.B. nicht in *Natürlich weiß der Mensch, was er wirklich braucht*.
8. Die Dekomposition nach Art der Immediate Constituents hat sich für unsere Studierenden als schwieriger erwiesen, wenngleich wir die Schreibweisen nutzen. Auch distributionsanalytische Verfahren sind versucht worden; sie wurden gut akzeptiert, ihre Einführung ist aber zeitintensiv, und sie lassen sich deshalb nicht in den vorgegebenen Zeitrahmen bringen. Der Einsatz von Substitutionstests zur Kategorienbildung allein führt nur dann weiter, wenn er stark reglementiert wird.
9. Chomsky, Noam: *Syntactic Structures*. (Janua Linguarum. Series Minor. 4) The Hague: Mouton, 1956., S.56.
10. Schumacher, Helmut (Hg.): *Verben in Feldern. Valenzwörterbuch zur Syntax und Semantik deutscher Verben*. (Schriften des IDS. 1) Berlin, New York: de Gruyter, 1986. Das Abkürzungsverzeichnis nennt 'Nominalgruppe', 'Präpositionalgruppe', 'Adjektivgruppe', 'Adverbgruppe' als Realisierungen von Präpositional-Ergänzungen, Prädikativ-Ergänzungen, Adverbial-Ergänzungen u.a. Bei der Unterscheidung von Adverbial (früher: 'Umstandsbestimmung') – Adverbialphrase bzw. Adverbialgruppe (PSG) – Adverbialergänzung (Valenztheorie) ist allerdings Aufmerksamkeit geboten.
11. Eisenberg, Peter: *Grundriß der deutschen Grammatik*. 3. überarb. Aufl., Stuttgart, Weimar: Metzler, 1994., S. 54.
12. Nicht ausgenutzt wurde bisher die Erscheinung, daß freie Angaben in Form von lokalen Angaben nicht so frei jedem Satz zuzufügen sind, wie man auf den ersten Blick vielleicht annehmen könnte:
 - a) steht z.B. bereits eine lokale Adverbialergänzung in Form einer Präpositionalphrase, so ist eine weitere Präpositionalphrase (mit lokal interpretierbarer Präposition) schwerlich als freie Angabe zu identifizieren:
cf. *Die Spieler trafen sich in einem Stadion in Gelsenkirchen*. Paraphrase I: *Die Spieler trafen sich in einem Stadion. Das Stadion liegt in Gelsenkirchen*. Paraphrase II: *Die Spieler trafen sich in einem Stadion. Und das geschah in Gelsenkirchen*. – wobei die zweite Lesart semantisch nur gewaltsam von der ersten unterschieden werden kann.
 - b) Verben, die semantisch in die Klasse der kognitiven Verben, Verben der Gemütsbewegung oder der inneren Einstellung gehören, werden nicht

ohne weiteres mit lokalen Angaben verknüpft:

cf. *?Ich weiß die Lösung in der Straßenbahn.* (wohl dagegen: *In der Straßenbahn wußte ich die Lösung noch nicht.*) und: *?Er bezweifelt in Köln, daß die Angaben richtig sind.*

Allerdings würden sich solche Bedenken zu einer Kritik an der Valenztheorie ausweiten, zumindest an der Valenztheorie in der von uns zugrundegelegten Form, die keine semantischen Rücksichten kennt.

13. Die Bezeichnungen sind Buchstaben des Alphabets und keine üblichen Konstituentenkategorienamen, um die Studierenden einerseits auf die Beliebigkeit der Namen hinzuweisen, zugleich aber, um bei ihnen die Erinnerung an Darstellungen der formalen Chomsky-Grammatiken (im Hauptfach) zu evozieren.
14. Zur Beantwortung beider Fragen muß eine Analyse dessen gemacht werden, was die Grammatik generiert.

Dazu gehört

- a) die Übersicht über Regeln und Vokabular,
- b) die Betrachtung der Baumstrukturen von Sätzen, und vor allem:
- c) die gezielte Erzeugung bestimmter Sätze der Grammatik.

Für a) genügt das Aufrufen der Datei und das Lesen des dortigen Kommentars, für b) müssen Sätze eingegeben werden, die das System parsen kann. Für alle analysierbaren Sätze erscheint der entsprechende Strukturbaum vertikal eingerückt in Spalten ohne weiteren Befehl auf dem Bildschirm.

Für c) muß entweder das Ausgangssymbol eingegeben werden – so werden sukzessiv auf Anforderung alle ableitbaren Sätze erzeugt und angezeigt – oder die Benutzerin gibt eine oder mehrere Variablenbelegungen vor und läßt die freie(n) Variable(n) vom Programm nacheinander mit allen möglichen Belegungen versehen. Dadurch werden systematisch variierte Sätze erzeugt und angezeigt; das geschieht ebenfalls auf jeweilige Anforderung. (Durch den Anforderungsschritt kann die Benutzerin die Geschwindigkeit der Interaktion mit bestimmen oder die Interaktion abbrechen.) Auf diese Weise kann man z.B. erreichen, daß alle Sätze gezeigt werden, die mit einer ausgewählten Verbform gebildet werden. Die Auswahl beim Generierungsvorgang setzt voraus, daß die Benutzerin ausschließlich Sätze mit bestimmten syntaktischen Eigenschaften sehen will, was allerdings wiederum zur Voraussetzung hat, daß die menschliche Interaktionspartnerin diese gezielte Auswahl auch bewußt treffen kann. Darin liegt wahrscheinlich die Ursache dafür, daß man ohne Fragestellung nicht sofort einen Zugang zum zielgerichteten Experimentieren findet.

Informationszentrum Sozialwissenschaften als europäische Trägerinstitution für TREC Erfolgreiche erste Teilnahme an TREC-7

Das IZ Sozialwissenschaften (Bonn) hat erfolgreich an TREC-7, der Text Retrieval Conference als Partner des National Institute of Standards and Technology (dem US-Normungsinstitut) teilgenommen. TREC ist eine seit sieben Jahren durchgeführte Konferenz- und Testserie für moderne Indexierungs- und Retrievalsysteme, an der sich die führenden Forschungsinstitute und Softwareentwickler der Welt beteiligen. Die Testergebnisse werden vielfach als Benchmarking für die beteiligten Institute und Firmen, aber auch darüber hinaus angesehen. Auch für Hersteller, die sich nicht offiziell an dem Wettbewerb beteiligen, stellen die TREC-Ergebnisse den Maßstab für den eigenen Entwicklungsstand dar.

Für TREC-7 hat das IZ im Rahmen des Aufgabenfeldes „Cross-lingual information retrieval“ deutschsprachige sozialwissenschaftliche Datenbestände (d. h. Auszüge aus FORIS und SOLIS) zu Testzwecken zur Verfügung gestellt. Diese Daten wurden parallel mit Datenbeständen in englischer, französischer und italienischer Sprache bei den Retrievaltests durchsucht, wobei die Frageformulierung selbst in nur einer der Sprachen erfolgte. Die Ergebnisse sollten dann aus allen Sprachen stammen.

Ferner hat das IZ deutschsprachige Fragestellungen (Topics) für die Tests entwickelt und Fragen, die in Englisch, Französisch oder Italienisch vorlagen, ins Deutsche übersetzt. Nach Durchführung der Tests durch die beteiligten Gruppen hat das IZ die Relevanzbewertung der gefundenen deutschsprachigen Dokumente vorgenommen, wodurch die Basis für die Leistungsbewertung der einzelnen Softwarepakete für sich und zwischen den verschiedenen Systemen gelegt wurde. Die Beteiligung des IZ wurde auf der Abschlußkonferenz in Gaithersburg (9.–11. November 1998) gewürdigt und dadurch positiv bewertet, daß es auch für TREC-8 im Jahr 1999 in gleicher Weise wieder aktiv werden wird. Außerdem soll die Koordination und die Verstärkung des europäischen Engagements im multilingualen Bereich vom IZ forciert werden.

(Quelle: IZ-Telegramm, Dez. 1998, S. 2)

Lernout & Hauspie übernimmt TikSoft

Erweiterung des Angebotes an PC-basierten Sprachanwendungen

Lernout & Hauspie (L&H) ist einer der führenden Anbieter von Sprach- und linguistischen Technologien, Produkten und Dienstleistungen, mit Schwerpunkten in den Bereichen Automatische Spracherkennung, Text-zu-Sprache, digitale Sprach- und Musikkompression und Text-zu-Text (Übersetzung). Das Unternehmen hat nun bekanntgegeben, daß es im Zuge des weiteren Ausbaus seiner Dictation Division am 30. September die Firma TikSoft übernommen hat. TikSoft mit Hauptsitz in Burlingame, Kalifornien (USA), ist Entwickler von Spracherkennungsprodukten für Unternehmen und Consumer-Märkte. Durch die Übernahme erwirbt Lernout & Hauspie zahlreiche Desktopsoftware-Produkte, die auf der Spracherkennungstechnologie von L&H basieren, einen Mitarbeiterstamm mit umfassenden Erfahrungen im Bereich Handel und Direktmarketing sowie ein engagiertes Entwicklungs-Team.

In engem Zusammenhang mit dieser Übernahme kündigte Lernout & Hauspie zudem an, daß das Unternehmen die Entwicklung und Vermarktung von PC-basierten Produkten für private und professionelle Anwender weiter ausbauen will. Diese Produkte werden über den Handel und via Direkt-Marketing verkauft. Im Zuge dieser Entwicklung will L&H die für private Anwender und Unternehmen konzipierten Produkte und Technologien aus den Unternehmensbereichen Language Division, Core Division und Dictation Division kombinieren und zu einem neu formierten Bereich unter der Bezeichnung PC Desktop Applications Unit zusammenfassen, die zu L&H's Dictation Division gehören wird.

Schwerpunkt: Marktwachstum im Bereich PC-Applikationen Die PC Desktop Applications Unit soll zur Verbreitung von neuen und fortschrittlichen Desktop-Applikationen auf der Basis von L&H's führenden Sprach- und linguistischen Technologien beitragen. Die neue Unit richtet sich damit vor allem an die schnell wachsende Nachfrage der Verbraucher nach sprachfähigen Applikationen für PCs und Desktops. Nach Angabe von L&H ist diese Nachfrage vor allem auf die sinkenden Kosten für PCs und Rechnerleistung zurückzuführen, aber auch auf immer weiter entwickelte und benutzerfreundlichere Sprachtechnologien wie z.B. die Natural Language Technology sowie eine zunehmende Würdigung von Sprachbenutzeroberflächen bei Verbrauchern und kleineren Unternehmen.

„Als Pionier und eines der führenden Unternehmen auf dem Markt für Spracherkennung waren wir schon immer davon überzeugt, daß eine Sprachbenutzeroberfläche die Benutzerfreundlichkeit und Leistungsfähigkeit jeder Art von Software verbessert. Jetzt zeichnet sich bei den Verbrauchern, und hier vor allem bei Windows-Nutzern, eine verstärkte Nachfrage nach diesen Vorteilen ab“, sagt Gaston Bastiaens, President und CEO von Lernout & Hauspie. „Unsere neue Unit wird unsere umfassenden, auf Verbraucher und kleinere Unternehmen ausgerichteten Produkte mit den Produkten sowie den profunden Marktkenntnissen und -erfahrungen von TikSoft verbinden. Wir werden die vereinten Ressourcen dafür einsetzen, die Vorteile von Sprachbenutzeroberflächen möglichst schnell einer größeren Anzahl potentieller Anwender nahezubringen.“

Produkte und Marketing von TikSoft Durch die Übernahme von TikSoft gewinnt Lernout & Hauspie neue Produkte dazu, die mit Hilfe von L&H's Sprachtechnologie entwickelt wurden. Die Produkte erweitern L&H's bestehende Produktpalette und beinhalten Windows-basierte Produktivsoftware sowie ein Produkt für den Unterhaltungsmarkt. Beide Produkte sollen Ende dieses Jahres auf den Markt kommen. Die Akquisition wird zudem das Entwicklungs-Team von L&H vergrößern und die Bemühungen des Unternehmens bei Direkt-Mails und im Handelsmarketing stärken. In beiden Bereichen hat TikSoft in der Vergangenheit erhebliche Anstrengungen unternommen, um den Absatz seiner Produkte zu unterstützen.

Zu den Spracherkennungsprodukten von TikSoft gehört eine proprietäre Software, mit der alle Anwendungen in Microsoft® Office 95 und 97 via Sprache gesteuert werden können d. h. eigentlich alle Windows-Applikationen und eine ganze Reihe von sprachfähigen Hilfsprogrammen für Produktivsoftware. TikSoft verfügt darüber hinaus über ausgefeilte Technologien, die es den Anwendern ermöglichen, im Gesprächsstil mit Software zu arbeiten und Anwendern zu antworten - unabhängig von der verwendeten Sprache.

„Wir freuen uns sehr darüber, jetzt zu Lernout & Hauspie zu gehören. Das Unternehmen besitzt unserer Meinung nach die Leistungsfähigkeit und die Vision für die Entwicklung sprachfähiger Lösungen, die die Art und Weise verändern werden, wie Menschen mit ihren PCs umgehen“, sagt Anatoly Tikhman, President und CEO von TikSoft. „Lernout & Hauspies führende Position und die Ressourcen des Unternehmens bieten eine hervorragende Möglichkeit, die Produkte und Technologien von TikSoft zu unterstützen und sie zum einem Bestandteil dieser Vision zu machen.“

Nach dem Zusammenschluß wird TikSoft in L&H's neu formierte PC Desktop Applications Unit integriert, zu der auch die VoiceXpress-Familie von L&H gehört. Die neue Unit ist Bestandteil von L&H's Dictation Division und wird von Anatoly Tikhman geleitet. Koen Bouwers, der derzeitige President und General Manager der Dictation Division wird auch weiterhin der Professional Solution Group vorstehen und seinen neuen Kollegen mit der Division vertraut machen. Tikhman wird President und General Manager der PC Desktop Applications Unit und berichtet in dieser Funktion direkt an Gaston Bastiaens. Koen Bouwers hat äußerst erfolgreich die Geschäfte in der Dictation Division und die Einführung von L&H's ersten Produktangeboten im Bereich kontinuierliche Sprache angekurbelt.

(Quelle: Lernout und Hauspie, Pressemeldung)

Computerlinguistische Analyse mehrsprachiger Fachtexte

Ingeborg Blank

*Centrum für Informations- und Sprachverarbeitung
Universität München, Dissertation*

Durch die zunehmende internationale Zusammenarbeit und wirtschaftliche Verflechtung ist der Bedarf an Übersetzungen von Fachtexten in den letzten Jahren ständig gestiegen. So ist es verständlich, daß gegenwärtig ein starkes Interesse an effizienten Werkzeugen zur Unterstützung des Übersetzungsprozesses bzw. an Maschinellem Übersetzung besteht. Ein multilinguales Textkorpus (eine Sammlung von Texten und deren Übersetzungen in mehrere Sprachen) stellt eine wichtige Informationsquelle für einen Übersetzer, insbesondere bei Terminologiefragen, dar. Ein solches Textkorpus bedarf aber einer entsprechenden Aufbereitung, um seinen vollen Nutzen entfalten zu können. Diese Idee war der Ausgangspunkt für die vorliegende Arbeit. Das Ziel besteht darin, ein Verfahren zu entwickeln, mit dem ein bilinguales Fachtextkorpus maschinell so aufbereitet werden kann, daß die darin enthaltenen lexikalischen bzw. terminologischen Ressourcen explizit verfügbar werden. Die Aufbereitung sollte es ermöglichen,

- Vorkommen möglicher Fachtermini aus den Daten zu extrahieren und
- nach Auffinden eines relevanten Beispiels in einem Quelltext die entsprechende Stelle im Zieltext zu lokalisieren.

Für die Untersuchung standen dreisprachige (dt.-engl.-frz.) Texte aus der Patentdokumentation (insgesamt 12 Mio. Wörter) zur Verfügung. Eine maschinelle Aufbereitung unter terminologischen Gesichtspunkten erfordert

1. eine Parallelisierung des bilingualen Korpus, d. h. eine Alignierung der Texte mit ihren Übersetzungen auf Satzebene,
2. eine Definition von Fachtermini mit formalisierbaren Kriterien, d. h. eine Definition, die in einem automatischen Verfahren umsetzbar ist.

Eine solche Definition liegt beispielsweise dann vor, wenn ein Fachterminus als Nominalphrase mit sprachspezifischen Bildungsmustern definiert wird, z. B. im Deutschen als Nominalkompositum (Beschwerdeverfahren), im Englischen als eine Nomen-Nomen-Verbindung (appeals procedure) und im Französischen als Verbindung aus einem Nomen und einer Präpositionalgruppe (proddure de recours).

3. die Extraktion der Fachtermini, d. h. die automatische Lokalisierung der Fachtermini in den Texten.

Zwei Implementierungen von Satzalignierungsverfahren wurden für alle Sprachpaare des Korpus evaluiert, die höchste Korrektheit lag bei ca. 97 %. Da die Extraktion von Termini eng an die einzelsprachlichen Besonderheiten der Wortbildung gebunden ist, wurde im Rahmen einer Untersuchung, die den Schwerpunkt mehr auf den qualitativen als auf den quantitativen Aspekt legt, die Extraktion von Termini auf das Sprachpaar Deutsch-Französisch begrenzt.

Die Extraktion der Fachtermini erfordert eine Vorverarbeitung der Texte, deren wichtigste Bestandteile die Lemmatisierung (Verfahren, das die Wortformen eines Textes auf eine kanonische Grundform zurückführt, z. B. Verben auf den Infinitiv, Substantive auf die Form des Nominativ Singulars) und das POS-Tagging (Verfahren, das die Wortformen eines Textes mit Wortklassen [engl. *part of speech* bzw. POS] annotiert) sind. Nach der eigentlichen Extraktion muß geprüft werden, ob die für jede Einzelsprache extrahierten Sequenzen linguistisch korrekte Nominalphrasen sind, um in einem zweiten Schritt der Frage nachzugehen, ob es sich inhaltlich und funktional um Fachterminologie handelt. Diese zweite Frage kann letztlich nur von einem Terminologen beantwortet werden. Eine Auswertung der extrahierten Sequenzen nach statistischen Kriterien kann allenfalls Hinweise auf mögliche Fachtermini geben.

Auf der Grundlage von Literatur aus der Terminologielehre, der Übersetzungswissenschaft und der Computerlinguistik wurden potentielle Fachtermini im Deutschen und im Französischen als Nominalphrasen bestimmter Bildungsmuster definiert. Die Vorverarbeitung und die Extraktion wurden für das Französische mit dem INTEX-System (Silberztein 1993) und eigenen Programmen durchgeführt; bei den deutschen Texten erfolgte die Vorverarbeitung mit dem CISLEX-System (Maier-Meyer 1995) und die Extraktion wurde mit eigenen Programmen durchgeführt. Im Vorfeld wurde die Eignung der beiden Systeme für die jeweilige Aufgabe untersucht und, soweit erforderlich, eine Anpassung durchgeführt.

Im Französischen waren ca. 80 % der extrahierten Bildungsmuster linguistisch korrekte Nominalphrasen. Fehler in der Extraktion sind zum größten Teil auf inkorrektes POS-Tagging und auf Ambiguitäten bei der Bestimmung der Bildungsmuster zurückzuführen. Letztere sind nur durch eine intellektuelle Bearbeitung zu beheben. Im Deutschen waren ca. 95 % der extrahierten einfachen Nomina, Nominalkomposita und Adjektiv-Nomen-Verbindungen linguistisch korrekte Nominalphrasen. Weitere Bildungsmuster konnten nur mit einer Korrektheit von 65 % extrahiert werden; dieses Ergebnis kann maschinell nur durch eine zumindest partielle syntaktische Analyse der Texte behoben werden.

Als mögliche Anwendung wird ein Konkordanzprogramm vorgestellt, das drei Aufgaben erfüllt:

- Lokalisierung von Vorkommen einzelner Fachtermini im Quell- und im Zieltext für die Ermittlung von Übersetzungsäquivalenten
- Analyse spezieller Verwendungskontexte einzelner Fachtermini (Präpositionen, Verben, Adjektive usw.)
- Zusammenstellung von Termini mit gemeinsamen Konstituenten in einer netzwerkartigen Struktur.

Die Ergebnisse dieser Arbeit können für die Erstellung mehrsprachiger Glossare und Terminologiedatenbanken sowie für die Entwicklung eines Übersetzungsspeichers (translation memory) genutzt werden. Mehrsprachige Terminologie wird auch bei der Entwicklung und Anpassung von Systemen zur Maschinellen Übersetzung für neue Fachgebiete benötigt.

Auswahlbibliographie

- Arntz, Reiner; Pieht, Heribert: Einführung in die übersetzungsbezogene Terminologiearbeit, Hildesheim: Olms, 1982.
- Bourigault, Didier: LEXTER, un Logiciel d'EXtraction de TERminologie: Application à l'acquisition des connaissances à partir de textes. Diss. Ecole des Hautes Etudes en Sciences Sociales, Paris, 1994.
- Chureh, Kenneth W.; Dagan, Ido; Termight: Identifying and translating Technical Terminology. In: Proc. of the 4th Conference on Applied Natural Language Processing, Stuttgart, 1994, 5. 34–40. Daille, Béatrice: Approche mixte pour l'extraction de terminologie: statistique lexicale et filtres lin-

- guistiques. Thèse de doctorat en informatique fondamentale, Université Paris VII, 1994.
- Eijk, Pim van der: Automating the acquisition of Bilingual Terminology. In: Proc. of the Meeting of the European Chapter of the Association for Computational Linguistics, Utrecht, 1993, S. 113–119.
- Felber, Helmut; Budin, Gerhard: Terminologie in Theorie und Praxis, Tübingen: Narr, 1989.
- Fluek, Hans-Rüdiger: Fachsprachen, Tübingen: Francke, 3. Auflage, 1985.
- Gale, William A.; Chureh, Kenneth W.: A program for aligning sentences in bilingual corpora. In: Proc. of the 29th Annual Meeting of the Association for Computational Linguistics, Berkeley, 1991.
- Gaussier, Eric: Extraction automatique de lexiques bilingues par des méthodes statistiques. Thèse de doctorat en informatique fondamentale, Université Paris VII, 1995.
- Isabelle, Pierre: Bi-textual aids for translators. In: Proc. of the Annual Conference of the UW Center for the New OED and Text Research, 1992.
- Kay, Martin; Röscheisen, Martin: Text-Translation Alignment. In: Computational Linguistics, Volume 19, Number 1, 1993, 5. 121–142.
- Krause, Jürgen (Hrsg.): Inhaltserschließung von Massentexten, Hildesheim: Olms, 1988.
- Maier-Meyer, Petra: Lexikon und automatische Lemmatisierung. Centrum für Informations- und Sprachverarbeitung (CIS-Bericht 95-84), München, 1995.
- Reinart, Sylvia: Terminologie und Einzelsprache, Frankfurt: Lang, 1993.
- Silberztein, Max: Dictionnaires électroniques et reconnaissance lexicale automatique, Paris: Masson, 1993.
- Sta, Jean-David: Comportement statistique des termes et acquisition terminologique à partir de corpus. In: Traitement automatique des langues, Vol. 36, 1995, 5. 119–132.
- Waliner, Margot: Untersuchung von Text-Alignment im Rahmen computergestützter Textübersetzung auf der Basis des Kay-Röscheisen-Algorithmus. Diplomarbeit Fachhochschule München, Fachbereich Informatik, 1994.

Machine Translation: Theory, Applications, and Evaluation. An Assessment of the State-of-the-art

Herausgegeben von Nico WEBER, Gardez! Verlag, 1. Auflage 1998. 196 S. DIN A5. Broschur. ISBN 3-928624-71-7. (gardez@pobox.com)

Machine Translation (MT) has made a comeback during the 1990s. Though systems, as to their linguistic performance, are much what they were 10 years earlier, there have been significant changes on the technical level. Systems that were once costly and running on mainframes and expensive workstations only are now licensed for rather modest sums to run on personal computers. The comeback of MT has been furthered by substantial progress in the development of auxiliary linguistic tools for text authoring and editing, as well as for translation (machine-aided translation, MAT). In order to achieve similar progress in MT we have to gain more knowledge and a better understanding of how (human and machine) translation works. We may learn something about this from cognitive science and translation theory. A theory of MT is only just developing. Another primary information source is system evaluation and translation quality assessment. MT output quality greatly varies with different formal and content types of input.

The state-of-the-art in the domains mentioned is documented in this volume by individual contributions about (machine) translation research, system development and application, performance evaluation, and translation quality assessment. All papers are in English with significant summaries in German. The authors are confirmed experts in the domain, with individual accents on research and development, teaching and training, industrial and business applications, or evaluation and consulting.

Contents

1. Preface
2. Mental Information Processing in Human Translation *Isabelle Schrade*
3. Machine Translation, Evaluation, and Translation Quality Assessment
Nico Weber
4. MT Evaluation in Research and Industry: Two Case Studies *Rita Nübel*
5. Transfer in Machine Translation with OO-LPL *Jürgen Rolshoven*

6. Linguistic Features of Instructional Texts and Their Treatment by Machine Translation Systems *Uta M. Seewald-Heeg*
7. The Automatic Translation of Idioms: Machine Translation vs. Translation Memory Systems *Martin Volk*

Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering

Herausgegeben von Henning Lobin, Westdeutscher Verlag 1999

1997 ist in der GLDV der Arbeitskreis Texttechnologie gegründet worden (vgl. <http://www.uni-giessen.de/~g91062/GLDV/>). Der Arbeitskreis hat unter anderem das Ziel, den Teilnehmerinnen und Teilnehmern einen Überblick über Aktivitäten zu verschaffen, die zum Thema SGML/XML in der linguistischen Datenverarbeitung von GLDV-Mitgliedern durchgeführt werden (dazu gab es im März ein erstes Arbeitstreffen). Mitglieder des Arbeitskreises haben wesentlich zu einer aktuellen Buchpublikation beigetragen, auf die hier hingewiesen werden soll:

Henning Lobin (Hrsg.):

Text im digitalen Medium. Linguistische Aspekte von Textdesign, Texttechnologie und Hypertext Engineering.

Wiesbaden: Westdeutscher Verlag, 1999

IX, 229 Seiten, Broschur DM 68,-

Nähere Informationen zu dem Band enthält die Web-Site <http://coli.lili.uni-bielefeld.de/Text-Band/>. Der Herausgeber ist per e-Mail unter lobin@lili.uni-bielefeld.de zu erreichen.

Sinnentstellender Fehler in letztem LDV-Forum: Rezension zu „Lexikon und Text“ (Feldweg/Hinrichs) sowie „Lexikonimport, Lexikonexport“ (Hoetker/Ludewig)

In der letzten Ausgabe ist auf schwer nachvollziehbare Weise in der Sammelrezension der beiden Niemeyer-Bände über Lexika (Feldweg/Hinrichs: Lexikon und Text sowie Hoetker/Ludewig: Lexikonimport, Lexikonexport) ein schwerer, systematischer Fehler unterlaufen, für den Autor und Redaktion alle Beteiligten und alle Leser um Entschuldigung bitten. Die Kürzel, mit denen auf die beiden Werke verwiesen wurde, waren nämlich grundsätzlich vertauscht verwendet worden. Um jegliche Unklarheit auszuschließen, folgt nun direkt die korrekte Fassung der Rezension als Nachdruck. Außerdem liegt dem Heft ein Korrekturblatt bei, der in die Ausgabe LDV-Forum 1998/1 einsortiert werden sollte.

Feldweg, H. und Erhard W. Hinrichs (Hrsg.):

Lexikon und Text. Wiederverwendbare Methoden und Ressourcen zur linguistischen Erschliessung des Deutschen. [LSM]

Lexicographica Series Maior 73. Tübingen 1996, Niemeyer Verlag

Hoetker, W. und Petra Ludewig (Hrsg.):

Lexikonimport, Lexikonexport. Studien zur Wiederverwendbarkeit lexikalischer Informationen. [SI]

Sprache und Information, Bd. 31. Tübingen 1996, Niemeyer Verlag

Rezensiert von Johann Haller, Saarbrücken (e-mail: hans@iai.uni-sb.de).

Gleich zwei Bände mit fast gleichem (Unter-)titel zur gleichen Zeit im gleichen Verlag: Das muß doch ein wichtiges Thema sein – oder hat da einer bei Niemeyer nicht aufgepaßt? Aber bei näherem Hinschauen ergibt sich, daß nur ein Artikel gleichzeitig in beiden Bänden zu finden ist, und immerhin einmal in Englisch und einmal in Deutsch.

Aus welchen theoretischen und praktischen Gründen das Thema der Wiederverwendbarkeit so wichtig ist, erfährt man in sehr übersichtlicher und systematischer Weise in der Einleitung von SI, das neun längere und in drei nach Forschergruppen getrennte Aufsätze enthält. Zu einem theoretischen Umschwung (mehr lexikalistisch begründete Sprachtheorien) kommt die Zunahme der praktischen Relevanz lexikalischer Ressourcen, die sich auch in der steigenden Anzahl von EU-Projekten zu diesem Thema zeigt. Einige davon (EAGLES, DELIS, MULTEXT, ET10-51) sind zwar dem Spezialisten bekannt, es ist aber sehr verdienstvoll, daß gerade die Ergebnisse dieser Projekte ausführlich und im Zusammenhang in deutscher Sprache zugänglich gemacht werden. Gerade die mangelnde Sichtbarkeit dieser und ähnlicher Projekte in der Öffentlichkeit war mehrmals kritisiert worden. Das aktuelle mehrsprachige Korpus- und Lexikonprojekt der EU (PAROLE) sollte eigentlich die Fortsetzung sein; daß die Fortwirkungen dieser Projekte sich besonders im deutschen Teilprojekt (PAROLE-D) in Grenzen halten, mag zunächst darauf zurückzuführen sein, daß sich die genannten EU-Projekte hauptsächlich mit der englischen Sprache beschäftigten. Hier und da taucht mal ein deutsches Beispiel zwischen den „be“-Formen und den syntaktischen Verwendungsmöglichkeiten von „taste“ auf. Außerdem sind zwei Jahre in der Computerlinguistik ein langer Zeitraum (die entsprechenden Workshops fanden 1994 statt), in dem Ergebnisse schnell veralten. Beide Bände (SI und LSM) machen eigentlich neugierig auf die weiteren Fortschritte; das studentische Beispielprojekt in einem Osnabrücker SI-Artikel, das sich mit der Anwendung des mehrfach verwendbaren Lexikons in Sprachlernsystemen befaßt, hofft darauf, daß die (eigentlich altbekannten) Schwierigkeiten bald überwunden werden: Disjunktions- und Negations-Operatoren für den Formalismus, Anbindung an eine große lexikalische Datenbank für partielle Lexikoneinträge etc. Ein wichtiger Schritt in diese Richtung ist sicher mit der in den drei Bochumer Beiträgen beschriebenen Bearbeitung des Cobuild-Lexikons geschehen; dieser wird von Schnelle auf der Hypothese der Bedeutung als Implikation theoretisch fundiert, und von den beiden anderen Autoren am Beispiel des HPSG-Formalismus ausgearbeitet und technisch umgesetzt. Ein Teil des Ergebnisses ist (noch kosten-

frei) im WWW zu besichtigen (<http://ww.linguistics.ruhr-uni-bochum.de/ccsd/>).

Einen diffuseren Eindruck macht zunächst LSM, das eine Reihe kürzerer Artikel enthält; die in der kurzen Einleitung unternommene Gliederung in sechs Bereiche hilft da nur wenig weiter: Lexikon-Text, Wörterbücher, Wissensrepräsentation, Korpora, Linguistische Annotation, Statistische Disambiguierung. Zur lakonischen Kürze kommt auch noch ein Druck- bzw. Grammatikfehler in der ersten Zeile. Um die formale Kritik gleich auf einmal loszuwerden: Druckfehler, leider auch sinnentstellende, gibt es eine ganze Menge. Daß mit „Beispielgruppenabgabe“ eigentlich die entsprechende „Angabe“ gemeint ist, bedarf schon einiger Überlegung, bei „müßen“ mit scharfem S war wohl die neue Rechtschreibung irgendwie schuld. Und schließlich meinen noch ein paar Autoren, sie müßten ihre Lateinkenntnisse an den Mann bringen. Auch hier kann man „computativ“ noch erschließen, aber bei „Fehlende Lexika sind ein Desiderat“ hätte man es auch einfacher haben können. Vom Inhalt her findet man in manchen Artikeln Interessantes: immer noch die Bonner Wortdatenbank (eine Pionierleistung, wenn auch inzwischen ein bißchen betagt), endlich ein paar Lichtblicke aus dem Institut für deutsche Sprache in Mannheim (jetzt mit dem Projekt COSMAS auch im Internet zu erproben: <http://www.ids-mannheim.de/ldv/cosmas/>), von dem man schon seit vielen Jahren computerlinguistische Aktivitäten erhofft hätte, Information über das eindrucksvolle CISLEX in München (die deutsche Entsprechung zum LADL-Lexikon von Maurice Gross in Frankreich), Information zu einem großen Korpus-Projekt der EU (MULTEXT) und für Programmierfans eine Reihe technischer Beschreibungen, wie man zu einsatzfähigen Programmen kommt, die auch mit großen Text- und Wortmengen umgehen können. Schwedisch-Fans werden auch bedient; für diese Sprache – und dann auch für Deutsch – werden Vorschläge für Tagsets gemacht, über die aber die Zeit vermutlich schon wieder hinweggegangen ist. Daß doch gelegentlich auch auf alte Materialien zurückgegriffen wird, zeigen die vereinzelt Verweise auf alte Saarbrücker Wörterbücher, Analyse- und Lemmatisierungsprogramme, die bei der Konfektion umfangreicher Vorlagen zur Anwendung kommen. Solche werden bei den Versuchen des stochastischen Tagging benötigt, deren Beschreibung den Abschluß von LSM bildet. LSM hat am Schluß einen lieblos gemachten (und damit ein bißchen nutzlosen) Index, in dem ‚Wiederverwertbarkeit‘ gerade einmal auftaucht, dafür „leere“ Wörter wie ‚Zeichen‘, ‚ZEIT‘ und einige Varianten linguistischer Begriffe wie ‚Wortartklassifikation‘, ‚Wortartenklassifikation‘ etc. Hier wäre es sicher angebracht, sich einmal auf eine Schreibweise festzulegen und diese dann beizu-

behalten. Da hätte man es lieber wie SI machen sollen und den Index ganz weglassen.

Insgesamt betrachtet, kann der fortgeschrittene Student der Computerlinguistik oder der Forscher, der etwas über den Stand der maschinellen Lexika in Deutschland wissen will, durch die Lektüre der beiden Bände einen recht guten und vollständigen Überblick bekommen; es fragt sich natürlich, ob in Zeiten der elektronischen Publikation wirklich fast zwei Jahre vergehen müssen, bis die Ergebnisse eines schnell fortschreitenden Gebietes auf Papier vorliegen. Der Rezensent schwankt noch, was ihm mehr Spaß macht: die schnelle und aktuelle Info via Internet oder die bequeme Lektüre der Bücher. Nach der Lektüre kann man noch ein bißchen träumen: Wie wäre es, wenn COSMAS eine (syntaktische) Kontextsuche à la DELIS hätte, wenn es ein standardisiertes Tagset für das Deutsche gäbe, das auch in der von Bochum ins Internet gestellten DUDEN-Version enthalten wäre und so weiter ...

34th Colloquium of Linguistics

September 1999, University of Mainz, Germany

The 34th Colloquium of Linguistics took place at the Johannes-Gutenberg-Universität Mainz, Faculty of Applied Linguistics and Cultural Studies, Germersheim, from September 7 to September 10, 1999. The motto of this year's conference is "Linguistics on the Way into the New Millennium". Continuing the tradition of the colloquium, there were no restrictions regarding the choice of topics. The conference languages were English, German, and French. A volume of abstracts was available at the conference. The proceedings with the full papers are published after the conference with Peter-Lang-Verlag. In a break with tradition, this year's conference program is supplemented by a number of tutorials. Each tutorial comprises three hours and gives a concise introduction to a specific field for audiences with a different focus of research. More information can be found on our website at <http://www.fask.uni-mainz.de/lk/> (Prof. Dr. Dieter Huber, Dr. Reinhard Rapp)

TUTORIALS:

- * Prof. Peter Hellwig: Natural Language Parsing
- * Christian Otto: Sprachtechnologie für das Internet
- * Prof. Uta Seewald-Heeg: Maschinelle Übersetzung

The Tutorials took place during the first full day of the conference. Therefore, the official opening was on the second day. For each tutorial, a description can be found at <http://www.fask.uni-mainz.de/lk/>.

CONFERENCE SITE The Faculty of Applied Linguistics and Cultural Studies (FASK) of the University of Mainz is located in Germersheim in the southwest of Germany on the Rhine between the cities of Mannheim and Karlsruhe. It can easily be reached by car, train, and airplane. With its 2500 students, more than 1000 of whom coming from 70 different foreign countries, the faculty is known as one of the world's largest institutions dedicated to the training and education of translators and interpreters. Within walking distance from the faculty, six hotels and the student's residence are available with single rooms priced

between 30 and 90 DM per night. Meals are provided by the university canteen, the cafeteria, and by a number of restaurants (with lunch specials). The main attractions in and around Germersheim are its fortress, the German wine road, the Palatinate Forest with Hambach Castle, Speyer, Heidelberg, the Castle Park in Schwetzingen and the Old Abbey in Maulbronn.

INTERNATIONAL ORGANIZING COMMITTEE Dr. Susanne Beckmann, University of Münster; Dr. Abraham P. ten Cate, Rijksuniversiteit Groningen; Dr. Tadeusz Danilewicz, Gdansk University; Prof. Dr. K. Dorfmueller-Karpusa, Univ. of Thessaloniki; Dieter W. Halwachs, University of Graz; Prof. Dr. Mikhail Kotin, State University Moscow; Prof. Dr. Wim Klooster, University of Amsterdam; Prof. Dr. Wilfried Kürschner, University of Osnabrück; Dr. Reinhard Rapp, University of Mainz (Conference Chair); Prof. Dr. Hans Otto Spillmann, University of Kassel; Prof. Dr. Kazimierz A. Sroka, Gdansk University; Dr. Jürg Straessler, University of Bern; Prof. Dr. Zygmunt Vetulani, University of Poznan; Prof. Dr. Richard J. Watts, University of Bern; Prof. Dr. Heinrich Weber, University of Tübingen; Dr. Ingo Warnke, University of Kassel; Dr. habil. Lew Zybatow, University of Bielefeld.

CONFERENCE ADDRESS Please send all correspondence to the following address: 34th Colloquium of Linguistics <http://www.fask.uni-mainz.de/1k/>, c/o Dr. Reinhard Rapp, Universität Mainz, FASK, D-76711 Germersheim. rapp@usun2.fask.uni-mainz.de Phone: (+49) 7274 / 508-457, Fax: (+49) 7274 / 508-429

Workshop

Natürlichsprachliche Schnittstellen – Dialog- und Partnermodellierung

Im Rahmen der Fachtagung für Künstliche Intelligenz,
Bonn 13.9.–14.9.1999

Über natürlichsprachliche Schnittstellen können Benutzer verbal mit Dialogsystemen interagieren: Sie stellen eine Anfrage über Mikrofon oder Tastatur und erhalten die gewünschte Auskunft durch Ausgabe eines Textes in geschriebener

oder gesprochener Form. Schnittstellen dieser Art existieren schon seit einiger Zeit; die Herausforderung liegt jetzt darin, sie so robust und effizient wie möglich zu gestalten, ohne Syntax und Vokabular unnötig zu beschränken.

Inwiefern können neue Ansätze in der Dialog- und Partnermodellierung dazu beitragen, dieses Ziel zu erreichen? Inwieweit können dabei linguistische Forschungsergebnisse aus Unterspezifizierungstheorie, Rhetorical Structure Theory, dynamischer Semantik oder Centering Theory genutzt werden? Insbesondere interessierten Ansätze, die bereits in Dialogsystemen integriert und im Kontext dieser Systeme evaluiert worden sind.

Der Workshop sollte die Möglichkeit bieten zu einem regen Dialog zwischen Linguisten und Informatikern, zwischen Theorie und Praxis. Deshalb waren Vorträge auf 15 Minuten Länge beschränkt, gefolgt von 15 Minuten Diskussion; abschließend eine anderthalbstündige Diskussion.

Organisation: Organisatoren waren Bernhard Schröder (b.schroeder@uni-bonn.de) und Maria Wolters(wolters@ikp.uni-bonn.de). Das Programmkomitee bildeten: Elisabeth Andre (Saarbrücken), Harry Bunt (Tilburg), Paul Dekker (Amsterdam), Roland Hausser (Erlangen), Janet Hitzeman (Edinburgh), Winfried Lenders (Bonn), Paul McKeivitt (Aalborg).

Workshop SGML/XML – Einsatz in der Lexikographie

Auf Einladung der Arbeitskreise Hypermedia, Lexikographie und Sprachtechnologie der Gesellschaft für Linguistische Datenverarbeitung (GLDV) fand am 21. September 1999 in den Räumen der Heidelberger Akademie der Wissenschaften ein Workshop zum „Einsatz von SGML und XML in der Lexikographie“ statt. Wie das letztjährige Symposium bot dieser Workshop wieder einen geeigneten Rahmen für den Gedankenaustausch zwischen Theoretikern und Praktikern.

Dies wurde schon in der Vormittagssitzung deutlich, in der Kurzvorträge von Ingrid Schmid / Carolin Müller, Angelika Storrer, Ingrid Lemberg und Claudia Kunze diskutiert wurden.

Ingrid Schmidt und Carolin Müller stellten ein Publikationsmodell vor, das, ausgehend von der zentralen Rolle der Inhaltsstrukturmodellierung, die Ebenen des Informationspools, der redaktionellen Arbeit und des jeweils produktbezogenen Ausschnittes aus dem Informationspool trennt. Darüber hinaus zeigt dieses Modell, wo die Erstellung und Bearbeitung dieses und eine sich am „Multiple-Media-Publishing-Konzept“ orientierende Produktgenerierung angesiedelt sind.

Angelika Storrer zeigte anhand verschiedener Verfahren der digitalen Aufbereitung von Wörterbuchdaten Desiderate für die Architektur lexikalischer Informationssysteme auf: der Weg von den Quellen, der Wörterbuchbasis, zur lexikographischen Beschreibung muss nachvollziehbar sein, die Datenbasis sollte die lexikologischen Objekte und lexikalisch-semantischen Relationen widerspiegeln, der lexikographische Prozess sollte sich am Gegenstand und nicht an den Gegebenheiten des Präsentationsmediums orientieren, der Zugriff auf lexikalische Daten sollte durch möglichst intelligente Navigationswerkzeuge unterstützt werden. Schließlich sollte ein lexikalisches Informationssystem offen sein für die Anbindung zu anderen digitalen Ressourcen. Ingrid Lemberg referierte anhand des Beispiels des Deutschen Rechtswörterbuchs über mögliche Informationstypen im elektronischen Wörterbuch, die sich aus dem Beschreibungsgegenstand eines Wörterbuchs und seinen Artikelstrukturen ergeben und den Benutzern eines elektronischen Wörterbuchs zur Verfügung stehen, im Gegensatz zum gedruckten Wörterbuch und seiner meist einfachen, lemmatischen Zugriffsstruktur, in dem viele Informationen „versteckt“ bleiben (vgl. Ulrich Goebel, Ingrid Lemberg, Oskar Reichmann: Versteckte lexikographische Informationen. Niemeyer: Tübingen 1995). Auch hier ist eine vernünftige, benutzerorientierte lexikographische Konzeption die Voraussetzung jeder semantischen Datenmodellierung.

Claudia Kunze stellte ein in Tübingen nach lexikalisch-semantischen Prinzipien organisiertes Wortnetz als lexikalische Datenbasis vor. Dieses „Wortnetz“ (GermaNet) wird spätestens mit Ende dieses Jahren verfügbar sein.

In der anschließenden Diskussion wurde deutlich, dass die von Frau Storrer benannten Desiderate an eine Lexikographie der Zukunft erst einmal als solche in das Bewusstsein der Verlage und der dort tätigen Lexikographen gehoben werden muss. Dies scheint ein aus der Sicht der Benutzer von Wörterbüchern sicher

lohnenswertes Unterfangen. Zugleich wurde die Liste der Desiderate in Redebeiträgen aus dem Publikum weiter ergänzt. Genannt wurden hier Aspekte der Qualitätssicherung einer lexikalischen Datenbasis. Diese wird um so schwieriger und zugleich umso wichtiger, je größer und komplexer diese lexikalische Datenbasis ist.

In einem den Vormittag abschließenden Vortrag stellte Martin Volk den Terminologiekodierungsstandard MARTIF vor und stellte diesen in den Kontext anderer Normierungsvorhaben für den Austausch terminologischer Daten. Im Anschluss an diesen Vortrag wurden zwei Fragen diskutiert, die auch späterhin noch eine Rolle spielten: wie verhalten sich die verschiedenen unter dem Dach von SGML entworfenen Standards, die zum Teil die gleichen Gegenstandsbereiche berühren, zu einander (Fragen der Kompatibilität)? Wie verhält sich die Datenkodierung in einem SGML konformen Dokument zu anderen Formen der Datenmodellierung, z. B. dem Entity-Relationship-Modell für die Planung der konzeptuellen Struktur von Datenbanken. Die letzte Frage ergab sich aus einer durchaus praxisrelevanten Designentscheidung bei der Erstellung einer DTD: wann sollen Elemente und wann Elementattribute bei der Darstellung von Eigenschaften von Textobjekten verwendet werden?

Alexander Geyken eröffnete die Nachmittagssitzung. Er berichtete von den Bemühungen an der Berliner Akademie, ein in Bezug auf den Gegenstand „deutsche Sprache des 20. Jahrhunderts“ ausgewogenes Korpus als Textbasis für ein „digitales Wörterbuch des Deutschen“ zu erstellen. Ein Testkorpus von 10 Millionen Token wurde in Anlehnung an den EAGLES Corpus Encoding Standard annotiert, so dass zunächst seitens der Lexikographen präzisere Suchfragen möglich sind, als es dies ohne eine linguistische Aufbereitung der Daten, mit reiner Volltextsuche, möglich wäre. Die Akademie strebt als Wörterbuchbasis ein Korpus von 600 Millionen Token an.

Bernhard Schröder war einer der wenigen, die XML als Format für die Kodierung von Daten tatsächlich verwenden. Er stellte die „Voice Extensible Markup Language“ (VXML) und ihre Beziehungen zum Kodierungsstandard der TEI und den Darstellungsmitteln, die SAMPA und UNICODE für die Darstellung phonetischer Zeichen bieten, vor.

Angelika Storrer skizzierte einen Versuch, in Anlehnung an Blumenthal / Lemnitzer / Storrer 1988 (Blumenthal, Andreas/ Lemnitzer, Lothar/ Storrer, Angelika: Was ist eigentlich ein Verweis? – Konzeptuelle Datenmodellierung als Voraussetzung computergestützter Verweisbehandlung. In: Harras, G. (ed.): Das

Wörterbuch: Artikel und Verweisstrukturen, Düsseldorf: Schwann, pp. 351-373) und diesen Ansatz weiter entwickelnd, die Konzepte eines Verweises im traditionellen Printwörterbuch und eines Links in einem lexikalischen Informationssystem grundsätzlich zu trennen. Für Verweise und für Hyperlinks müssen unabhängig voneinander Klassifikationen entwickelt zu werden. Auf der Basis dieser Klassifikationen lässt sich erst die Frage beantworten, ob und wie ein Verweis in einem gedruckten Wörterbuch als Link in einem lexikalischen Informationssystem repräsentiert werden kann und soll und welche zusätzlichen Links in und zwischen primären und sekundären lexikalischen Daten den Nutzwert eines solchen Informationssystem erhöhen können. Ein wichtiges Mittel hierbei ist die Typisierung von Links. Diese ermöglichen es, die Daten für den Anwender transparenter zu machen. Als gutes Beispiel eines „Hypertextes auf Papier“ nannte Frau Storrer den „Longman Language Activator“. Henning Lobin aus Bielefeld stellte Ideen für eine effektive Präsentation von enzyklopädischen Daten vor. Die Überlegungen entwickelten sich aus Projekten in Zusammenarbeit mit der Lexikonredaktion von Bertelsmann. Was Henning Lobin entwickelte waren allerdings eher Visionen für die Zukunft als Ansätze, die bereits in der Arbeit des Lexikonverlages umgesetzt worden wären. Sie orientieren sich an dem Ziel, einen hohen Nutzen bzw. Mehrwert für Benutzer dieser Daten zu erzielen. Als wichtige Desiderate nannte er eine adaptive Benutzeroberfläche, die den Benutzer mit genau den Informationen bedient, die für sie / ihn relevant sind, die Verwendung von semantischen Distanzen zwischen Termen als Mittel der Navigation und die Erstellung von Hypertrails als thematische Routen durch die Datenlandschaft. Eine wichtige Frage, die sich an eine solche Art der Aufbereitung lexikalischer Daten anknüpft ist die, an welcher Stelle diese zum Teil benutzerspezifischen Metadaten kodiert werden sollen. In der Datenbasis selber, als zugeordnete Metadaten oder direkt in der Anwendung, die dem Benutzer die entsprechenden Daten präsentiert? Sonja Müller Landmann berichtete über das Projekt „LEXXIS“ am Institut für deutsche Sprache. Alle lexikalischen Daten dieses Instituts sollen unter die gemeinsame Oberfläche eines Hypertextsystems mit darunter liegender Datenbank gebracht werden. Ein wichtiger Aspekt dieses Projektes ist, angesichts der heterogenen Datenquellen, der Entwurf eines konzeptuellen Modells für die Artikel-Mikrostruktur. Außerdem soll, mit Blick auf maschinelle Anwendungen, die in den Daten implizit vorhandene Information explizit gemacht werden.

Abschließend berichtete Frau Schmidt von einem in nächster Zeit zu verabschiedenden Standard für „Topic Maps“ (ISO/IEC 13250). Dieser Standard legt eine Syntax für die Repräsentation von Topics und der Beziehungen zwischen ihnen, zwischen ihren Instanzen und ihren Benennungen fest. Ein Ziel, der mit diesem Standard

verfolgt wird, ist die Definition von Benutzerschnittstellen als Datensichten. Topic Maps sind sicher auch ein geeignetes Mittel, die von Henning Lobin dargestellten „Hypertrails“ über lexikalischen und enzyklopädischen Daten zu organisieren. Ganz im Sinne dieses Workshops, in dessen Zentrum Wege des „cross media publishing“ standen, wird sich Bernhard Schröder darum bemühen, die „Topics“ dieses Workshops zusammen mit relevanten Links in das World-Wide Web auf einer WWW-Seite aufzubereiten. Der URL lautet: http://www.gldv.org/Veranstaltungen/WS_Sept99.html.

Dr. Lothar Lemnitzer
Seminar für Sprachwissenschaft / Abt. Computerlinguistik
Universität Tübingen
<http://www.sfs.nphil.uni-tuebingen.de/~lothar/>

Konvens 2000

Die Konvens 2000 wird vom 9. bis zum 12. Oktober 2000 an der TU Ilmenau zusammen mit der 6. ITG-Fachtagung „Sprachkommunikation“ stattfinden.

Ueber den Call for Papers können Sie sich auf der WWW-Seite <http://ikmcip1.e-technik.tu-ilmenau.de/dsv/callfp2000.html> informieren. Das Schwerpunktthema der Tagung wird „Robustheit und Fehlertoleranz“ sein. Wie üblich sind jedoch Beiträge aus „alle[n] Bereiche[n] der maschinellen Verarbeitung von Sprache in gesprochener als auch in geschriebener Form“ gewünscht. –Im Programmkomitee der Konvens 2000 ist die GLDV vertreten durch:

Prof. Dr. Roland Hausser, Erlangen (GLDV)

Prof. Dr. Winfried Lenders, Bonn (GLDV)

Bis zum 1. Februar 2000 sollen Kurzfassungen von Vortrags-, Poster- und Workshopvorschlägen eingereicht werden. Über eine rege Beteiligung von GLDV-Mitgliedern, insbesondere über Workshopvorschläge von Seiten der GLDV-Arbeitskreise, würde ich mich sehr freuen.

Bernhard Schröder
B.Schroeder@uni-bonn.de

Communicating Agents

**Workshop of the GLDV special interest group on
generation and parsing in morphology, syntax and
semantics**

**IKP, University of Bonn
Feb 15, 2000**

http://www.gldv.org/Veranstaltungen/WS_Feb00/

The focus of this workshop will be approaches to the formal description and eventually implementation of communicating agents. We invite papers on the following aspects:

1. incremental analysis / interpretation
 - datastructure of the context (e.g. dynamic semantics)
 - computational reconstruction of reference
 - correlation of semantic and pragmatic interpretation
2. partner oriented generation
 - intentionally directed generation
 - partner knowledge and attribution of belief
3. evolution of systems and strategies of communication
 - concept evolution / formation / construction
 - evolution of linguistic norms
 - repair strategies in disturbed communication

Lectures should not exceed 20 minutes. They will be followed by a 10-minute discussion each. Workshop languages are English and German. We plan to publish workshop proceedings with the full papers. System demonstrations will be possible on demand. Abstracts of no more than 600 words should be sent – preferably by email in ASCII, HTML, PS or PDF format – to Bernhard Schröder before Jan 10, 2000.

Timetable

- Jan 10, 2000 Deadline for abstracts
Jan 24, 2000 Notification on acceptance
Feb 15, 2000 Workshop

Please register before Feb. 8, 2000, if possible.

Organizing Committee

Roland Hausser, University of Erlangen; Hans-Christian Schmitz and Bernhard Schröder, University of Bonn

Workshop Fees

Members of the GLDV, non-students:	DM 20
Student members of the GLDV:	DM 10
Others, non-students:	DM 40
Other students:	DM 20

Contact

Bernhard Schröder
Institut für Kommunikationsforschung und Phonetik
Poppelsdorfer Allee 47
D-53115 Bonn
email B.Schroeder@uni-bonn.de
phone +49 228 735621
fax +49 228 735639

Leistungsfähigkeit und Einsatzmöglichkeiten von Translation-Memory-Systemen

Der Arbeitskreis „Maschinelle Übersetzung“ veranstaltete am 9. und 10. Juli 1999 im Rahmen der diesjährigen Jahrestagung der GLDV in Frankfurt am Main ein Arbeitskreistreffen zum Thema „Leistungsfähigkeit und Einsatzmög-

lichkeiten von Translation-Memory-Systemen“, zu dem wir Sie herzlich einladen (vgl. auch <http://www.heeg.de/~uta/Einladung.html>).

Das Treffen des Arbeitskreises soll nach einer ersten dem Themenkreis Translation Memories (TM) gewidmeten Arbeitskreissitzung im Februar 1999 ein weiterer Schritt auf dem Weg zur Evaluierung von TMs sein. Wir möchten Sie daher einladen, sich mit einem Beitrag, der sowohl aus dem Bereich der Forschung als auch der Entwicklung, insbesondere aber auch aus dem Bereich des praktischen Einsatzes von Translation Memories kommen kann, an dem Arbeitskreistreffen zu beteiligen.

Kontakt: Rita Nübel (rita@iai.uni-sb.de) und Uta Seewald-Heeg (seewald@heeg.de).