

## Zu den Grundlagen des IdS-Forschungsprojekts PLIDIS:

Formallogische Repräsentation für ein natürlichsprachliches Informationssystem, dargestellt am Beispiel der Quantifikation.

### 0. Ausgangspunkt und Argumentationsstruktur

Ausgangspunkt der folgenden Überlegungen ist die Feststellung, daß in dem Forschungsprojekt PLIDIS des IdS eine formale Sprache zur Sprachinterpretation, genauer zur Interpretation eines Fragments des Deutschen, benutzt wird.

Im folgenden soll versucht werden, die Wahl einer speziellen formalen Sprache zu begründen. Dies soll geschehen, indem die gewählte formale Sprache kontrastiert wird – wenn auch nicht durchgängig – mit anderen formalen Kandidaten-Sprachen, vorzugsweise vom Typ der Montague-Grammatiken. Die Kontrastierung gerade mit formalen Sprachen dieses Typs ist mit der unter formalen Sprachbeschreibern unbestritten hervorragenden Rolle der Montague-Grammatik zu erklären.

Der Vortrag wird sich daher große Strecken weit mit der Begründung bestimmter konzeptioneller Entscheidungen im Bereich der formal-sprachlichen Repräsentation befassen, und es werden nur gegen Ende Teile des Formalismus selbst vorgestellt, und zwar für den Bereich der Quantifikation im Deutschen oder genauer, der Repräsentation von mit *die, einige, alle* und *jeder* determinierten Nominalphrasen.

Dieser Vorgehensweise liegt folgende Überlegung zugrunde: Bei der gegenwärtigen Lage der Künstlichen-Intelligenz-Forschung in ihrem Verhältnis zur Linguistik und Logik erscheint es sinnvoller, argumentativ bestimmte Entscheidungen, die im Rahmen eines Projektes zur Künstlichen Intelligenz fallen, und die vom Standpunkt des Linguisten oder auch Logikers her zunächst unverständlich sind, abzuklopfen, um sie verständlicher zu machen, als einen perfekten Formalismus für einen Teilbereich zu präsentieren.<sup>1</sup>

#### 0.1. Das Informationssystem PLIDIS

Das System PLIDIS (*problemlösendes Informationssystem mit Deutsch als Interaktionssprache*) ist ein automatisches Informations- oder Frageantwortsystem (IS), das anwendungsorientiert ausgelegt ist und zur Zeit Daten zur Kontrolle von Industrieabwässern im Bereich des Landes Baden-

Württemberg enthält. Daten über diesen Anwendungsbereich, z.B. Daten über Betriebe, die der Überwachung durch eine Wasserwirtschaftsbehörde unterliegen, ihre Abwasseranlagen und die Protokolle der einzelnen Überwachungsvorgänge, der sogenannten Probeentnahmen, sind in der *D a t e n - b a s i s* des Systems abgespeichert. Anfragen, die sich auf die in der Datenbasis des Systems gespeicherten Daten beziehen, können in Deutsch gestellt werden. Sie werden vom System morphosyntaktisch analysiert, in eine interne Repräsentationssprache übersetzt und dann vermittels der Problemlösungskomponente gegenüber der Datenbasis ausgewertet, d.h. zu beantworten versucht. Dies geschieht durch einen Vergleich der in der internen Repräsentationssprache reformulierten Informationsanforderung mit den in der Datenbasis enthaltenen Informationen. Um zu einer Antwort zu gelangen, sind unter Umständen logische Verknüpfungen von abgespeicherten Einzelinformationen erforderlich.<sup>2</sup>

### 1. Formale Repräsentation im Informationssystem

Zurückkommend zum Ausgangspunkt stelle ich zunächst die vorgängige Frage: Warum wird hier, in einem System, das natürliche Sprache verarbeitet, überhaupt eine formale Sprache verwendet? Die Beantwortung dieser Frage führt diesmal nicht in die wohlbekannteren Rechtfertigungsstrategien dafür, warum wir formale Theorien machen. Bei solchen Strategien laufen die Argumentationen über die Begriffe Explizitmachen, Erklären, Disambiguieren. Die Antwort ist hier, zumindest auf einer ersten Ebene sehr viel nüchterner und praxisorientierter und führt anscheinend nicht an jenen gefährlichen Punkt, an dem den Formalisten die natürliche Sprache als letzte Metasprache vorgehalten wird. Sie lautet: Mit einer formalen Sprache kann in einem Computersystem besser umgegangen werden. Alle bekannten Verfahren zur Auswertung, Ableitung und Kombination von Informationen, spricht Aussagen über irgendwelche Fakten, sind für formale Repräsentationsmittel definiert. Dies schließt nicht aus, daß die interne Repräsentationssprache über diesen praktischen Begründungszusammenhang hinaus auch als formales Übersetzungskorrelat eines gewissen Ausschnittes einer natürlichen Sprache betrachtet werden kann. Auf das Verhältnis von Repräsentationssprache und natürlicher Sprache kann dann das von Montague eingeführte Interpretationsverfahren angewandt werden, bei dem die für die formalsprachlichen Ausdrücke gültige semantische Interpretation durch Induktion auf die entsprechenden natürlichsprachlichen Ausdrücke übertragen wird.<sup>3</sup>

In diesem Sinn kann dann die formale Sprache als *S e m a n t i k - s p r a c h e* des Deutschen betrachtet werden, die zum Zwecke der semantischen Beschreibung und Erklärung entworfen wurde.

### 1.1. Die Datenbasis des Systems als Modell

Wie verhalten sich die in der Datenbasis abgespeicherten Informationen und die in der formalen Sprache reformulierten deutschen Fragen zueinander? Um diese Frage zu beantworten, betrachten wir die stark vereinfachte Datenbasis eines fiktiven Informationssystems.<sup>4</sup>

Diese minimale fiktive Datenbasis bestehe aus folgenden Einträgen:

(Beispiel 1)

(BRUDER EVA HANS)  
(BRUDER EVA FRITZ)  
(BRUDER LINDA OTTO)  
(KOLLEGE LINDA KARL)  
(KOLLEGE LINDA EGON)

Diese Datenbasis ist die Kodierung eines Modells im modelltheoretischen Sinne, und zwar eines endlichen Modells. Abfrageformulierungen für dieses Modell sind theoretisch zu verstehen als Aufforderungen zur *Bewertung* der in ihnen enthaltenen Aussagen gegenüber dem Modell. D.h. wenn  $\alpha$  das in der Datenbasis repräsentierte Modell ist und  $\varphi$  eine Abfrageformulierung und  $d$  eine Belegungsfunktion für Variable, so ist

$\text{Val}(\varphi, \alpha, d)$

zu bestimmen.

Wenn  $\varphi$  eine ja/nein-Frage ist, so ist  $\text{Val}(\varphi, \alpha, d)$  ein Wahrheitswert. Er ist *W* oder *F* (wahr oder falsch), je nachdem, ob  $\varphi$  in  $\alpha$  gültige Relationen spezifiziert oder nicht.

(Beispiel 2)

*Ist Hans Bruder von Eva?*

$\text{Val}(\text{Hans ist Bruder von Eva}, \alpha, d) = W \rightarrow \text{ja}$

Wenn  $\varphi$  eine *W-Frage* (Ergänzungsfrage) ist, so ist  $\text{Val}(\varphi, \alpha, d)$  eine Menge von Entitäten, die  $\varphi$  gemäß  $\alpha$  erfüllen.

(Beispiel 3)

*Welche Brüder hat Eva?*

$\text{Val}(\text{die Brüder von Eva}, \alpha, d) = \{\text{Hans, Fritz}\}$

Dabei gehe ich von einer bestimmten Auffassung der *W-Frage* mit *welche* aus: Sie ist modelltheoretisch semantisch gleichzusetzen mit der bestimmten Deskription einer Menge von Elementen. Als pragmatischer Anteil der Bedeutung von *welche-Fragen* kommt noch die Aufforderungs-

handlung hinzu: Der Frager fordert den Adressaten auf, genau diese Menge von Elementen zu ermitteln bzw. zu benennen.

## 1.2. Das Verhältnis von deutschen Fragen und Aussagen zur internen Repräsentationssprache und zum Datenbasis-Modell

Als Abfragesprache wird in PLIDIS Deutsch verwendet. Die interne Repräsentationssprache muß somit als Semantiksprache für den zur Abfrage verwendeten Ausschnitt des Deutschen dienen. Sie sollte daher auch nicht durch den Anwendungsbereich und das Faktenwissen darüber bestimmt sein, sondern durch die semantische und kommunikative Struktur des Deutschen.

Wenn man nun annimmt, daß auch abzuspeichernde Informationen als Aussagen des Deutschen formuliert werden können, so muß die zu wählende Repräsentationssprache auch für sie, für Aussagen, als Semantiksprache dienen. Eine in unser fiktives System einzugehende Information sei:

(Beispiel 4)

*Helga und Erna sind befreundet mit Eva.*

Diese Aussage, bzw. ihre interne Repräsentation, kann zerlegt werden in die atomaren Aussagen

(FREUND EVA HELGA)

(FREUND EVA ERNA)

Diese atomaren Formeln können nun als dynamische Erweiterung der Datenbasis aufgefaßt werden. Abgespeicherte Aussagen haben also in einem Informationssystem eine gewisse Ambiguität. Sie werden als elementare und als wahr markierte Aussagen der internen Repräsentationssprache verstanden; gleichzeitig dienen sie als Kodierung des Modells. An dieser Stelle wird also letztlich die natürliche Sprache auf dem Umweg über die Repräsentationssprache doch als Repräsentationsmittel für ihre eigene Semantik verwendet.

## 2. Bedingungen an eine interne Repräsentationssprache

Wir haben bisher festgestellt, daß eine formale Sprache als interne Repräsentationssprache für deutsche Ausdrücke verwendet wird, daß diese Sprache Semantiksprache für den entsprechenden Ausschnitt des Deutschen sein soll, daß sie aber für den Anwendungszweck des Informationssystems geeignet sein muß.

Es ist also nicht ausreichend, eine prospektive interne Repräsentationssprache so zu definieren, daß sie zur Erklärung der anstehenden seman-

tischen Phänomene des Fragments des Deutschen geeignet ist. Von den bisher genannten Gesichtspunkten her – modelltheoretische Interpretation, Übersetzungsrelation zur natürlichen Sprache, Faktenwissen-Unabhängigkeit – läge die Wahl einer an Montague orientierten Sprache nahe, denn Sprachen dieses Typs erfüllen die genannten Bedingungen.

Der Anwendungszweck des Systems schließt jedoch mit ein, das sagt schon der Name des Projekts, daß Problemlösungen mit Hilfe logischer Schlußfolgerungen möglich sein müssen. Automatische Theorembeweiser, die solche logischen Schlüsse syntaktisch umsetzen, stehen in vollem Umfang nur für die Prädikatenlogik erster Stufe ( $PL^1$ ) zur Verfügung.<sup>5</sup>

Dies ist das erste Eckdatum für die Bestimmung der Repräsentationssprache. Es bedeutet nicht, daß sie identisch sein muß mit  $PL^1$  – dies ist aus verschiedenen Gründen, zumal wegen der relativ starken Strukturähnlichkeit zwischen natürlichen Sprachen wie dem Deutschen und  $PL^1$ , die sich für den automatischen Übersetzer ungünstig auswirken würde, nicht sinnvoll. Es bedeutet wohl aber, daß für eine wohldefinierte Untermenge von formalsprachlichen Ausdrücken gilt, daß sie entweder direkt identifizierbar sind mit  $PL^1$ -Ausdrücken oder daß sie

- a) semantisch äquivalent sind mit  $PL^1$ -Ausdrücken und
- b) aufgrund von a) automatisch in  $PL^1$ -Ausdrücke überführt werden können bzw. daß auf sie prädikatenlogische Beweisverfahren induziert werden können.

Diese Untermenge ist genau bestimmt als die Untermenge über der man prädikatenlogische Schlußmöglichkeiten zulassen möchte. Wir bezeichnen sie als das  $PL^1$ -Analogon der internen Repräsentationssprache.

Diskutiert werden müssen also zwei Komplexe:

- (1) Welche Eigenschaften muß die Restmenge interne Repräsentationssprache minus  $PL^1$ -Analogon haben, um den Anforderungen des Informationssystems gerecht zu werden?
- (2) Rechtfertigen die zu (1) anzuführenden Eigenschaften eine Sprache, die bezüglich ihrer Konstruktionsprinzipien stark von  $PL^1$  abweicht, z.B. modallogische, intensionale oder typentheoretisch differenzierte Sprachen. Diese Gründe müssen schon recht gewichtig sein, denn sie müßten den wohl beträchtlichen Aufwand für die Überführung des  $PL^1$ -Analogons in  $PL^1$  oder eine gleichwertige Deduktionen erlaubende Behandlung dieses Ausschnittes aufwiegen.

### 3. Mögliche Eigenschaften, die für eine über PL<sup>1</sup> hinausgehende Repräsentationssprache sprechen

#### 3.1. Inhaltliche Aspekte

Ich möchte auf drei mögliche in der Repräsentationssprache zu berücksichtigende Eigenschaften natürlicher Sprache eingehen, die über die Repräsentation in Form von PL<sup>1</sup>-Propositionen hinausgehen würden:

- Sprechaktindikationen
- Deiktische Indizes
- Intensionale Kontexte

##### 3.1.1. Sprechaktindikationen

Der Handlungszusammenhang 'Informationssystem' ist, zumindest so, wie er anwendungsorientiert gebraucht wird und wie er in PLIDIS rekonstruiert vorliegt, ein Diskurs über Fakten, die als intersubjektiv gültig betrachtet werden. Einstellungen vom Sprecher – d.h. hier Informationssystem-Benutzerseite – zu diesen Fakten wie 'glauben', 'bezweifeln', 'hoffen' werden nicht thematisiert. Die einzigen zugelassenen Verhaltensweisen zu propositionalen Gehalten sind das Behaupten, das Negieren und das Fragen, eventuell die Rückfrage und ähnliche reaktive oder korrektive Handlungen, die dem Fortgang einer unbehinderten Interaktion dienen. *Behaupten*, das heißt im Kontext des IS, Eintragen des propositionalen Gehaltes in die Datenbasis. *Negieren* heißt Löschen des propositionalen Gehaltes aus der Datenbasis, *Fragen* heißt Bewerten des propositionalen Gehaltes gegenüber der Datenbasis. Diese Sprechakte haben also eine prozedurale Interpretation, d.h. sie schlagen sich nieder in konkreten Handlungen des Systems. Ihre Berücksichtigung in der Repräsentationssprache hat eine reine Markierungsfunktion für diese Systemhandlungen. Es kommt den sprechaktmarkierenden Ausdrücken also keine modelltheoretische Bedeutung zu, sie müssen nicht als Elemente alternativer Modellwelten, in denen es z.B. der Fall wäre, daß der Benutzer etwas behauptet, fragt usw., mitrepräsentiert werden – dies wäre sogar widersinnig.

##### 3.1.2. Deiktische Indizes

Mit "indexicals" im Sinne von Montague, d.h. also deiktischen Ausdrücken, deren Interpretation abhängig ist vom Gebrauchskontext ("context of use") wie *ich*, *du*, *jetzt*, *morgen*, *hier*, *dieses*, kann in einem Informationssystem ebenfalls prozedural verfahren werden. Dies geschieht dadurch, daß das System aus dem Sitzungsprotokoll oder mit Hilfe eines Perzeptionsorgans die Bezüglichkeiten deiktischer

Zeit-, Orts- und Sprecherbezeichnungen ermittelt und für die deiktischen Ausdrücke die entsprechenden Namen oder Deskriptionen einsetzt. Wenn z.B. in einer Benutzerfrage, die am 22.10.78 an das System gestellt wird, der Ausdruck *gestern* verwendet wird, z.B. in *Wurden gestern bei Firmen in Stuttgart Abwasserproben gezogen?* so ersetzt das System im Laufe seines automatischen Übersetzungsprozesses von der natürlichen in die formale Sprache den deiktischen deutschen Ausdruck *gestern* durch die absolute Zeitangabe 21.10.78.

### 3.1.3. Intensionale Kontexte

Intensionale Kontexte sind sprachliche Umgebungen, die es verhindern, die Bedeutung eines Ausdrucks, der in ihnen vorkommt, mit seiner Extension, d.h. dem Wahrheitswert bei Sätzen und der bezeichneten Entität bei Namen oder Kennzeichnungen, zu identifizieren. Die Intension eines Satzes nennen wir die ihm entsprechende Proposition, die Intension einer Kennzeichnung oder eines Namens das ihm entsprechende Individuenkonzept.

In einem Informationssystem der Art von PLIDIS ist nun weder mit Intensionalität aufgrund von "propositional attitudes" zu rechnen<sup>6</sup>, noch mit Intensionalität, die durch die Einführung unterschiedlicher Wissens- oder Vorstellungswelten entsteht. Aussagen nach dem Paradigma *seek a unicorn* sind in PLIDIS – oder in jedem anderen rein faktenbezogenen Informationssystem – sinnvollerweise nicht vorgesehen. Wenn jemand etwas sucht, so doch wohl etwas real, d.h. in der repräsentierten Modellwelt Existierendes. Die Verwendung von durch *wahrscheinlich, vielleicht* modifizierten Aussagen ist immerhin denkbar, sie kommt allerdings in PLIDIS nicht vor. Intensiv werden, zwar nicht in PLIDIS, aber in anderen Projekten zur Künstlichen-Intelligenz-Forschung, vage Ausdrücke behandelt<sup>7</sup>; sie führen aber wohl nicht zu intensionalen Kontexten. Intensionalität als ernstzunehmendes Problem tauchte in unserem Kontext bisher in zwei Fällen auf.

#### 3.1.3.1. Ungültige Schlüsse

Wir haben einen quasi-deontischen Kontext, und zwar in Aussagen über die zulässigen Normen für Abwasserinhaltsstoffe. Eine solche Aussage ist z.B.:

(Beispiel 5)

*Der Grenzwert für den Anteil an Arsen in Abwasserproben ist 1 mg/l.*

*Der Anteil an Arsen in Abwasserproben darf 1 mg/l nicht überschreiten.*

Bei solchen Beispielen könnte Anlaß zu ungültigen Schlüssen gesehen werden. Gehen wir zunächst von folgenden Prämissen aus:

(1) *Für alle Abwasserproben gilt:*

*Der Grenzwert für den Anteil an Arsen in einer Probe ist 1 mg/l.*

(2) *Der Anteil an Arsen in der Probe bei Laumann ist 2,5 mg/l.*

Ersetzen wir in (1) den allquantifizierten Ausdruck durch die entsprechende Kennzeichnung aus (2), d.h. nehmen wir die Instanziierung durch einen Individuenterm vor, so gelangen wir zunächst zu

(3) *Der Grenzwert für den Anteil an Arsen in der Probe bei Laumann ist 1 mg/l.*

Zusammen mit (2) könnte folgender Schluß gezogen werden:

(4) *Der Grenzwert für 2,5 mg/l ist 1 mg/l.*

Dieser ungültige Schluß ähnelt sehr stark dem in Montague PTQ<sup>8</sup> erörtertem Beispiel des Schlusses:

(5) *The temperature is ninety.*

(6) *The temperature rises.*

---

(7) *Ninety rises.*

Montague erklärt die Ungültigkeit des Schlusses damit, daß das intransitive Prädikat *rise* nicht über Individuen prädiiziert, sondern über Individuenkonzepte, d.h. *rise* schaffe hier bezüglich seines Subjekts einen intensionalen Kontext.

Zwar ist im Vergleich zu unserer Beispielreihe (1) bis (4) in Montagues Fall kein deontischer Kontext im Spiel. Dieses Aspekt scheint mir für die Ungültigkeit des Schlusses (1) bis (4), d.h. für die Nicht-Substituierbarkeit der Kennzeichnung *der Anteil an Arsen in der Probe bei Laumann* und des Namens *2,5 mg/l*, auch nicht entscheidend zu sein. Ebenso wenig scheint mir in Montagues Beispiel der Versuch, einen versteckten Zeitindex als Ursache der Nicht-Substituierbarkeit und somit der Intensionalität der Kontextes anzunehmen, wie Montague es tut, ganz angemessen zu sein. Denn, auch wenn die "referential opacity"<sup>9</sup> beseitigt wird, die aus dem versteckten Zeitindex in *the temperature* und aus dem zeitindexverändernden Verb *rise* resultiert, bleiben entsprechende Schlüsse zumindest "seltsam":



(8) *Die Temperatur, die in diesem Raum um 9 Uhr herrschte, war angenehm.*

(9) *Die Temperatur, die in diesem Raum um 9 Uhr herrschte, war 20°.*

---

(10) *20° war angenehm*

Ich halte daher für die Ursache der Nicht-Substituierbarkeit in beiden Schlußreihen, (1) bis (4) und (5) bis (7), nicht irgendwelche Parameter wie Zeitindex, deontischer Kontext, die Intensionalität hervorrufen könnten, sondern die inhaltliche Inkommensurabilität zwischen einer beschreibenden Kennzeichnung wie *die Temperatur, der Anteil an Arsen in der Probe bei Laumann* und einem Zahlenwert wie 20° oder 2,5 mg/l.

Die Nicht-Substituierbarkeit von Kennzeichnung und Name scheint in dem vorliegenden Fall, wo Kennzeichnung und Zahl einander gegenüberstehen, besonders ausgeprägt zu sein. Sie ist aber auch in anderen Fällen beobachtbar, gerade auch in Kontexten, die nicht erkennbar intensional sind. So haben

*David Bronstein is David Bronstein.*

und

*David Bronstein is the author of 200 Open Games.*

nicht die gleiche Bedeutung, auch wenn *David Bronstein* und *the author of 200 Open Games* dasselbe Individuum bezeichnen.<sup>10</sup>

Eine Schlußfolgerung aus diesen Beobachtungen ist, daß es keine klar abgegrenzten Kontextparameter gibt, relativ zu denen bei der Interpretation von Termen nicht-extensional verfahren werden muß. Eine Konsequenz daraus wäre, generell einen nicht-extensionalen Ansatz zu wählen und nur mit Individuenkonzepten, nicht mit Individuen als Designaten von Termausdrücken der formalen Sprache zu arbeiten.<sup>11</sup>

Diese Vorgehensweise ist jedoch im Rahmen von PLIDIS unannehmbar, da sie unvereinbar ist mit dem Prinzip, am PL<sup>1</sup>-Analogon festzuhalten. Die Lösung, die für PLIDIS akzeptabel erschien, ist, Aussagen wie (2) nicht als Identitätsaussagen zu betrachten, was die Substitution von Kennzeichnung und Name erlauben würde, sondern als Wertzuweisung. Wertzuweisung als eine mögliche Interpretation des Verbs *sein* wie z.B. in

(2) Der Anteil an Arsen in der Probe bei Laumann, ist, 2,5 mg/l,  
Kennzeichnung Wertzuweisung Name

ist so definiert, daß sie keine Ersetzung der Kennzeichnung durch den Namen beim Schlußfolgern erlaubt.

### 3.1.3.2. Die Zeit als möglicher Grund für intensionale Kontexte

Die zweite und sehr viel wichtigere Quelle von Intensionalität in PLIDIS ist die Zeit. Sie spielt in einem Informationssystem, in dem es um exakte Fixierung von Ereignissen geht, eine große Rolle, allerdings nicht in der typischen Indexfunktion, weder bei Sätzen noch bei Kennzeichnungen. D.h. Sätze wie

*Es regnet.*

oder *Die Lage ist schlecht.*

deren Wahrheitswert relativ zu Zeitintervallen wechselt oder Kennzeichnungen wie

*die Probe bei Laumann*

*der Präsident der Vereinigten Staaten*

deren Designat zeitabhängig ist, sind im IS nicht zugelassen. Wir lassen nur Ausdrücke der internen Sprache zu, die entweder *zeitindexunabhängig* oder *zeitindexstabilisiert* sind.

Zeitindexunabhängig ist z.B. die Aussage

*Brecht ist Probenehmer.*

Diese Aussage gilt für alle Zeitintervalle des in der Datenbasis abgebildeten Modells.

Zeitindexunabhängig sind auch kennzeichnende Ausdrücke (Terme) wie

*der Vater von Hans*

*der 27. Präsident der Vereinigten Staaten.*

Zeitindexstabilisiert sind Ausdrücke, die absolute Zeitangaben enthalten. Beispiel für eine zeitindexstabilisierte Formel und einen zeitindexstabilisierten Term aus dem PLIDIS-Bereich sind jeweils:

*Am 25.7.77 wurde bei Laumann eine Abwasserprobe entnommen.*

und *die Probe bei Laumann vom 25.5.77.*

Damit haben wir bezüglich des Zeitindex extensionale Kontexte geschaffen: Die Interpretation von *die Probe bei Laumann vom 25.7.77* liefert diejenige Entität, die am 25.7.77 Probe bei Laumann ist. Da wir uns generell nur in einer Welt befinden, in der realen Welt des historischen Ablaufs, haben wir es aufgrund der zeitlichen Stufung mit einem Modell zu tun, das in sich strukturiert ist als Serie von Modellen, die eine besondere

Ähnlichkeitsstruktur haben. Bezüglich dieser zeitlichen Serie von Modellen sind Eigennamen selbstverständlich "rigid designators"<sup>12</sup>, dasselbe gilt für den referentiellen Gebrauch von zeitindexstabilen Kennzeichnungen, die wir ja als einzige zulassen. Der attributive<sup>13</sup> Gebrauch würde allerdings noch ein Problem darstellen – die Frage ist, ob wir ihn benötigen. De dicto und re re Interpretationen von in anderen Repräsentationsformen zeitindexabhängigen Kennzeichnungen sind in diesem Ansatz an der unterschiedlichen Quantifizierung von Zeitintervall-Variablen zu unterscheiden:

(Beispiel 6)

*The prime minister of New Zealand will always be a British subject*<sup>14</sup>:

de dicto: Für alle  $x^{\text{int}}$  gilt: Dasjenige

Individuum, das an  $x^{\text{int}}$  Premierminister von

Neuseeland ist, ist an  $x^{\text{int}}$  britischer Bürger.

de re: Für alle  $x^{\text{int}}$ , für alle  $y^{\text{int}}$  gilt:

Dasjenige Individuum, das an  $x^{\text{int}}$  Premierminister

von Neuseeland ist, ist an  $y^{\text{int}}$  britischer Bürger.

Bezüglich der realen Welt und den Zeitindizes bezeichnen alle Entitätsterme somit entweder eine zu dem Zeitindex existierende Entität oder die Nullentität.<sup>15</sup> Es ist nicht notwendig, mit Individuenkonzepten zu arbeiten. Bei der Interpretation von Prädikaten muß in der Weise vorgegangen werden, daß sie jeweils die passenden Indizes "treffen", d.h. diejenigen Indizes, an denen die beteiligten Individuen existieren. Dabei wird vorausgesetzt, daß Prädikate nur "wohlgeformt" gebraucht werden. Für diesen wohlgeformten Gebrauch von Prädikaten ist jeweils ein spezielles Indexverhältnis gültig, an dem die beteiligten Entitäten zu existieren haben. Erfüllen die beteiligten Entitäten dieses Indexverhältnis, ist der Satz semantisch wohlgeformt, d.h., er kann wahr oder falsch sein.

Dieser Sachverhalt soll am Beispiel einer Verwendungsweise von *erzählen* erläutert werden. Wir nehmen an, daß *erzählen* ein 3-stelliges Prädikat mit folgendem Inhalt ist: Eine Person erzählt zu einem Zeitpunkt bzw. in einem Zeitintervall  $j_j$ , der innerhalb der Lebensdauer der Person liegt, von einer Person oder einem Geschehnis, wobei dasjenige worüber erzählt wird, bereits existiert hat bzw. zu existieren begonnen hat, d.h. die betreffende Person lebt an  $j_j$  bereits, bzw. hat gelebt oder das entsprechende Geschehnis ist an  $j_j$  bereits vorbei bzw. es hat begonnen sich zu ereignen.

Wir analysieren mit diesen Vorgaben den Satz *Eva erzählt am 23.7.79 von Hans*, ordnen den Argumenten  $arg1$ ,  $arg2$  und  $arg3$  die Zeitintervalle, die die Dauer ihrer "Existenz" ausmachen, zu und überprüfen, ob dieser Satz das für *erzählen* festgelegte Indexverhältnis erfüllt:

(Beispiel 7)

*Eva erzählt am 23.7.79 von Hans.*

$arg1$                        $arg2$                        $arg3$

$j_1 = 20.1.1960$        $j_2 = 20.7.79$        $j_3 = 1780 - 1830$

Das für *erzählen* gültige Indexverhältnis ist:

$j(arg1) \subseteq j(arg2) \wedge j(arg3) <+ j(arg2)$

Dabei bezeichnet ' $\subseteq$ ' die Teilmengenbeziehung zwischen Zeitintervallen und '<+' die Relation 'vorher sein oder vorher beginnen'.

Es zeigt sich, daß unser Beispiel das Indexverhältnis erfüllt, wobei  $j(arg1)$  den Wert  $j_1$  annimmt,  $j(arg2)$  den Wert  $j_2$  und  $j(arg3)$  den Wert  $j_3$ .

Dagegen würde das Beispiel

(Beispiel 8)

*Eva erzählt am 23.7.79 von Hans.*

$arg1$                        $arg2$                        $arg3$

$j_1 = 1730-1799$        $j_2 = 23.7.79$        $j_3 = 1780-1830$

das Indexverhältnis von *erzählen* nicht erfüllen. Solche Sätze wollen wir im Rahmen des Informationssystems als nicht wohlgeformt betrachten.

Diese Überlegungen zur Intervallsemantik für ein IS sind noch sehr vorläufig. Wir können jedoch zusammenfassend festhalten: Wir brauchen eine Sprache, deren Ausdrücke relativ zu einem konstanten Weltindex  $w_0$  und einer Menge von Zeitindizes bzw. Tupeln von Zeitindizes interpretiert werden. Da der Zeitindex, bzw. das Tupel von Zeitindizes, bei der Interpretation des zu interpretierenden Ausdrucks jeweils bereits einen festen Wert angenommen haben, z.B. das Zeitindextupel  $\langle j_1, \dots, j_n \rangle$  sind der Interpretationsfunktion mit  $(w_0, \langle j_1, \dots, j_n \rangle)$  die Argumente vorgegeben und sie kann berechnet werden. Die Funktion (Intension) kann mit ihrem Wert (Extension) identifiziert werden.

### 3.2. Typenlogischer Aspekt

Nachdem diese Vorklärung bezüglich der Intensionalität erreicht ist, ist zu überprüfen, ob eine extensionale Typenlogik das geeignete Instrument

zur internen Repräsentation in einem Informationssystem ist. Eine typenlogische Sprache hätte den Vorteil, daß die Kategoriendifferenzierungen der Syntax der natürlichen Sprache auf die interne Repräsentationssprache als Typen übertragen werden könnten. Darüberhinaus kann man argumentieren, daß natürliche Sprachen höherstufig im Sinne der Typentheorie sind, sie enthalten Prädikate über Prädikate (*quickly*), Quantifikationen über Prädikate, Mengen und ähnliches – zumindest kann man bestimmte Ausdrücke so interpretieren.

Für die Verwendung in einem Informationssystem stellen sich zwei Fragen:

- Liefert die automatische Syntaxanalyse bereits die Analysestrukturen einer Kategorialgrammatik, aufgrund deren Typenzuordnungen erfolgen können, oder auch liefert sie die der Montagueschen Kategorialsyntax äquivalente transformationsgrammatische Tiefenstruktur, die eine ebenso eindeutige Zuordnung von Teilbäumen und typenspezifischen Ausdrücken der formalen Sprache zuläßt? <sup>16</sup>
- Werden im Kontext des IS Entitäten höheren Typs gebraucht?

Zu der ersten Frage, Syntax, muß ich im Falle PLIDIS sagen, daß wir weder über ein kategorialgrammatisches Analyzesystem verfügen noch über eine genügend differenzierte phrasenstrukturelle Syntaxanalyse, um Typenzuordnungen zu machen. Kategorialgrammatische Analyse-systeme sind noch kaum im Gebrauch. <sup>17</sup>

Auf die zweite Frage wird wieder die für den Kontext 'Informationssystem' typische pragmatische Antwort gegeben: Zwar kann nicht ausgeschlossen werden, daß Ausdrücke im System verwendet werden, die als höherstufig einzustufen wären, wie z.B. *Die Farbe der Probe ist die Farbe gelblich. Die Anzahl der Proben ist 5.*; diese Gebräuche spielen jedoch eine marginale Rolle.

Im Kernbereich des IS, den abzuspeichernden und abzufragenden Informationen, allerdings, in dem gewissen Entitäten gewisse Eigenschaften zugesprochen werden, d.h. im Bereich der Prädikation, Quantifikation über Entitäten und der Referenz auf sie, da sollte die interne Repräsentationssprache möglichst nicht mit Entitäten höheren Typs arbeiten, da ja gerade dieser Bereich die Domäne von PL<sup>1</sup> ist, an der der Deduktionsapparat von PLIDIS ansetzt. Aus sicher guten Gründen arbeitet die Montague-Grammatik gerade in diesem Bereich nicht einfach mit Prädikaten und Individuen, sondern mit Entitäten höheren Typs.

Terme (Eigennamen und Kennzeichnungen) werden in UG und PTQ nicht als Individuennamen oder als Individuenkonzeptnamen aufgefaßt, also

vom Typ  $e$  bzw.  $\langle s, e \rangle$ , sondern als "properties of properties" (Eigenschaften von Mengen von Individuenkonzepten), also vom Typ  $\langle s, \langle \langle s, \langle \langle s, e \rangle, t \rangle \rangle, t \rangle \rangle$ . Sie werden funktional angewandt auf Prädikatsausdrücke; das Verhältnis von Argument und Prädikat wird umgedreht.

Aufgrund der bei Montague aufgeführten Bedeutungspostulate läßt sich jeder rein extensionale Fall der Kombination von Prädikat und Termen aus der komplexen Struktur auf  $PL^1$  zurückführen. Dies läßt sich an der schrittweisen Extensionalisierung des folgenden Beispiels zeigen:

(Beispiel 9)

*Alle Männer lieben Eva.*

$\lambda P \wedge x [\text{mann}'(x) \rightarrow P \{x\}] (\text{lieb}'(eva^*))$

$\langle s, \langle \langle s, e \rangle, t \rangle \rangle$

$\langle s, \langle \langle s, e \rangle, t \rangle \rangle, t \rangle \rangle$

$\wedge x (\text{mann}'(x) \rightarrow (\text{lieb}'(eva^*)) \{x\})$

$\wedge x (\text{mann}'(x) \rightarrow (\text{lieb}'_*(x, eva^*))$

Die explizite Rückführung solcher Fälle innerhalb des Informationssystems mittels Regeln, die den Montagueschen Bedeutungspostulaten entsprechen, wäre nur sinnvoll, wenn sie nicht generell stattfände, d.h., wenn intensionale Kontexte und "echte" Höhertypigkeit in größerem Umfang vorlägen.<sup>18</sup>

Andernfalls, d.h., wenn keine intensionalen Kontexte und Höhertypigkeit in höherem Umfang vorliegen, ist eher die umgekehrte Strategie empfehlenswert. Sie wird in PLIDIS verfolgt: Anstelle einer umfangreichen  $PL^1$ -Reduzierung aller "unechten" Höhertypisierungen werden in der internen Repräsentationssprache intensionale Kontexte, nicht-extensionale Adjektive, u.a. gesondert gekennzeichnet, sie fallen nicht in das  $PL^1$ -Korrelat der Sprache.

#### 4. Einführung der internen Repräsentationssprache KS

Auf die in PLIDIS verwendete interne Repräsentationssprache, die "Konstruktsprache" (KS), treffen einige der angeführten Argumente zu. Allerdings ist die Entwicklung eines Darstellungsmittels auch eine Sache der Praxis, und manche der zuvor vorgetragenen Argumente erscheinen auch mir selbst als Argumente ex post facto.

KS ist ein Lambda-Kalkül mit einer auf die Semantik von  $PL^1$  reduzierten Semantik.

Die Bedingung des  $PL^1$ -Analogons ist hier der effektivsten Weise erfüllt, insofern als  $PL^1$ , bzw. eine sortenlogische Variante von  $PL^1$ , eine echte Untermenge von KS darstellt.

Ich kann hier keine volle Regelgrammatik für KS einführen.<sup>19</sup> Ich vernachlässige auch den Aspekt der Sortiertheit vollkommen. Es werden nur diejenigen Aspekte der Sprache eingeführt, die für die Behandlung der Quantifizierung, die als Beispielbereich dienen soll, notwendig sind.

Die zentrale Überlegung dabei ist folgende: Entsprechend den Gegebenheiten des Informationssystems, dessen Datenbasis ein endliches Modell darstellt, können quantifizierte Ausdrücke innerhalb von Frageformulierungen gegenüber der Datenbasis zu Verknüpfungen, sogenannten Listen, von Individuen ausgewertet werden. Eine Theorie der Interpretation quantifizierter Ausdrücke für ein Informationssystem hat daher eine der Bedeutung unterschiedlich determinierter und quantifizierter Nominalphrasen (mit und ohne Plural-Artikel, quantifiziert durch *alle*, *jeder*, *einige*, *ein* usw.) entsprechende Auswahl und Verknüpfung von Entitäten aus dem Modell zu ermöglichen.

Die Ausdrucksseite dieses Quantifikationsteils der Sprache sollte diesem Ansatz der Reduktion der Quantifikation durch Aufzählung entgegenkommen, indem sie bezeichnende Ausdrücke (Terme) für Verknüpfungen von Entitäten zuläßt. Die Konstrukte der Sprache, die dies ermöglichen, werden im folgenden beschrieben.

#### 4.1. Zur Konstruktion elementarer KS-Ausdrücke

KS unterscheidet zwischen Prädikatsausdrücken und Termausdrücken. Diese Unterscheidung ist jedoch keine absolute, sondern eine kontextbezogene. Auf der Ebene der Basiskategorien ist noch eine Trennung da:

Basisprädikatsausdrücke: BRUDER, KOLLEGE, FREUND, ... (m-stellig)  
Metavariablen: P, Q

Basistermausdrücke: HANS, EVA, LINDA, ..... (Basiskonstante)  
x.per, x.ort, ..... (sortierte Variable)  
Metavariablen:  $t_i$ , ( $i \in \mathbb{N}$ )

Basistermausdrücke sind Termausdrücke, Basisprädikatsausdrücke sind Prädikatsausdrücke.

Aus einem m-stelligen Prädikatsausdruck gefolgt von m Termausdrücken in der Form  
( P  $t_1$  ....  $t_m$  )

bilden wir eine atomare Formel. Atomare Formeln sind Formeln.<sup>20</sup>  
 Eine atomare Formel, in der P ein Basisprädikatsausdruck ist und  $t_1, \dots, t_m$  Basistermausdrücke sind, ist eine atomare Basisformel.  
 Die Elemente der Datenbasis sind atomare Basisformeln, die keine Variablen enthalten.

Wenn  $\varphi$  eine Formel ist oder ein komplexer Ausdruck, so ist

(LAMBDA x.a  $\varphi$ )

ein komplexer Ausdruck. Wird ein komplexer Ausdruck als Argument verwendet, d.h. steht er nicht unmittelbar rechts von einer formeleinleitenden Klammer, so ist er ein komplexer Termausdruck, sonst ein komplexer Prädikatsausdruck. Wir lassen derzeit allerdings nur 1-stellige LAMBDA-Abstrakte als komplexe Termausdrücke zu.

In

- (1) (KOLLEGE (LAMBDA x.per (BRUDER x.per OTTO)) KARL)  
 'Karl ist Kollege derjenigen, deren Bruder Otto ist'

ist der LAMBDA-Ausdruck Termausdruck, in

- (2) ((LAMBDA x.per (KOLLEGE LINDA x.per)) KARL)  $\leftrightarrow$   
 (KOLLEGE LINDA KARL)  
 'Karl ist Kollege von Linda'

Prädikatsausdruck.

Es wird hier systematisch von der Möglichkeit Gebrauch gemacht, LAMBDA-Abstrakte als Prädikate bzw. je nach Kalkül als Funktionen zu betrachten oder auch als deren Extension, d.h. als Mengen bzw. Tupelmengen. Semantisch heißt das, wir betrachten LAMBDA-Ausdrücke einerseits als Bezeichnungen für "properties"

'die Eigenschaft, Kollege von Linda zu sein'.

Als solche können sie auf Entitäten zutreffen oder nicht zutreffen. Andererseits gelten sie als Bezeichnungen für Mengen von Entitäten, die bestimmte Formeln erfüllen:

'die Menge der Kollegen von Linda'

Ähnliche Doppelfunktion hat LAMBDA-Abstraktion auch bei Montague.

Komplexe Termausdrücke wie in (1) sind nun jedoch nicht als höher-typig im Sinne von Mengenbezeichnungen gegenüber Individuenbezeichnungen (Basistermausdrücken) aufzufassen. Vielmehr sind alle Termausdrücke gleichen Typs, wenn wir auch nicht von Typen sprechen.



Das entspricht dem semantiksprachlichen Gebrauch, den wir von komplexen Termausdrücken machen wollen: Sie sollen kennzeichnende Nominalphrasen, u.zw. singularische und pluralische, bezeichnen. Da im "normalen" Gebrauch pluralischer NP, im sogenannten distributiven Gebrauch, nicht über die höherstufige Entität 'Menge', sondern über die Elemente einer Menge geredet wird, besteht kein Typenunterschied zu Basiskonstantenausdrücken:

In

*Karl ist Kollege von Linda.*

und

*Die Brüder von Eva sind keine Kollegen von Linda.*

wird gleichermaßen über Individuen gesprochen.

Um die Einheitlichkeit des Formalismus zu gewährleisten, sagen wir, daß alle Termausdrücke ein- oder mehrelementige Listen (i.e. nicht-höherstufige Mengen) bezeichnen:

Hans  $\vec{u}$  → HANS kurz für (LISTE HANS)

*die Brüder von Eva* ("distributiv")  $\vec{u}$  →

(LAMBDA x.per (BRUDER EVA x.per))<sup>21</sup>

Jeder "beschreibende" komplexe Termausdruck kann gegenüber der Datenbasis ausgewertet werden, u.zw. zu einer "Liste":

[Hans, Fritz]

Für diese Liste wiederum gibt es eine aufzählende Bezeichnung, indem jeder Entität ihr Name zugeordnet wird und ein listenbildender Operator zur Verknüpfung benutzt wird.

(LISTE HANS FRITZ)

Wir gehen im folgenden davon aus, daß jeder beschreibende komplexe Termausdruck gegenüber der Datenbasis in dieser Weise aufgelöst werden kann. Wir nennen dieses Vorgehen *Reduktion durch Aufzählung* (RA). Dies ist möglich in einem endlichen Modell.

Wir untersuchen jetzt die Interpretation natürlichsprachlicher pluralischer NP, die als komplexe beschreibende Termausdrücke repräsentiert werden.

Wir nehmen an, die Aussage

(3) *Die Brüder von Eva sind Beamte.*

sei abzuspeichern. Sie wird übersetzt:

$\ddot{u} \rightarrow$  (BEAMTER (LAMBDA x.per (BRUDER EVA x.per)))

wobei BEAMTER ein 1-stelliger Basisprädikatsausdruck ist.

(RA) (BEAMTER (LISTE HANS FRITZ))

Zerlegung in atomare Basisformeln:

(BEAMTER HANS)

$\wedge$ (BEAMTER FRITZ)

Interessant wird die Kombination von mehreren pluralischen definiten NP:

(4) *Die Brüder von Eva sind mit den Kollegen von Linda befreundet.*

$\ddot{u} \rightarrow$  (FREUND (LAMBDA x.per (BRUDER EVA x.per))  
(LAMBDA y. per (KOLLEGE LINDA y. per)))

(RA) (FREUND) (LISTE HANS FRITZ)  
(LISTE KARL EGON))

#### 4.2. "Tiefenstrukturelle" KS-Repräsentation

An dieser Stelle müssen wir uns zunächst noch mit einem mehr formalen Problem befassen. Wir sind bisher davon ausgegangen, daß ein komplexer Termausdruck wie

(LAMBDA x.per (BRUDER x.per))

die Interpretation einer definiten pluralischen Nominalphrase ist. Der "Quantifikator" *die* wird nicht mitrepräsentiert. Wir sagen daher, der komplexe Termausdruck sei 'unquantifiziert'. Um die Parallelität des unquantifizierten Falles mit den anderen Fällen, in denen z.B. ALL (*alle*) oder EIN (*einige*) als Quantifikator des komplexen Termausdrucks auftritt, zu gewährleisten, schreiben wir als Marker des quantifizierten Falles 'U'. Wie semantische Überlegungen, auf denen unten z.T. eingegangen wird, zeigen, sind die Marker U, ALL und EIN nicht als Termpräfixe zu betrachten, sondern als Operatoren, bzw. in ihrer Kombination als Operatorenwörter, die aus dem Paar, das gebildet wird aus dem Prädikatsausdruck und dem Tupel der Argumentsausdrücke, eine Formel machen.

D.h. eine Formel

(P  $t_1 \dots t_m$ )

wo  $t_1, \dots, t_m$  komplexe Termausdrücke sind, betrachten wir als KS-Oberflächenstruktur der tiefenstrukturellen Formel

$$\underbrace{((U \dots U)}_{m\text{-mal}} P \langle t_1, \dots, t_m \rangle)$$

Dabei bezieht sich das Präfix U an der Stelle  $i$  ( $i \in \{1, \dots, m\}$ ) auf den Argumentterm  $t_i$ .

Im Beispielsfall ist die Tiefenstruktur von (4)

(4-TSTR)

$$((U U) \text{FREUND} \langle (\text{LAMBDA } x.\text{per} (\text{BRUDER EVA } x.\text{per}), (\text{LAMBDA } y.\text{per} (\text{KOLLEGE LINDA } y.\text{per})) \rangle).$$

Generell gilt für die Bildung von quantifizierten KS-Formeln die Regel:

Wenn  $QU_1, \dots, QU_m$  Quantifikatoren von KS sind, P ein Prädikatsausdruck ist und  $t_1, \dots, t_m$  komplexe Termausdrücke sind, so ist

$$((QU_1 \dots QU_m) P \langle t_1, \dots, t_m \rangle)$$

eine tiefenstrukturelle KS-Formel.

Kommen wir nach diesem Exkurs zur KS-Syntax auf die inhaltliche Problematik von (4) zurück.

#### 4.3. Die Interpretation definiter pluralischer Nominalphrasen

Wir müssen uns hier für eine bestimmte Interpretation kookkurierender definiter pluralischer NP entscheiden. Sie werden im Deutschen sicherlich in mehreren verschiedenen Weisen gebraucht, wobei diese Gebrauchsweisen kontextabhängig variieren und zudem eine gewisse Vagheit aufweisen. Ich stütze mich bei der Festlegung, die ich hier vornehme, auf Langendoen 1978.

Er interpretiert Sätze mit 2 pluralischen NP – sein Beispiel ist

*The women released the prisoners.*

wie folgt: Um den Satz wahr zu machen, muß jedes Element der einen Liste (*die Frauen*) zu mindestens einem Element der zweiten Liste (*die Soldaten*) in der genannten Relation (*released*) stehen und jedes Element der zweiten Liste muß zu mindestens einem Element der ersten Liste in der Umkehrrelation stehen.<sup>22</sup> Langendoen beschäftigt sich übrigens selbst mit dem Problem der Vagheit des Gebrauchs: Ist nicht auch die Wahrheitsbedingung erfüllt, wenn fast alle Frauen die Soldaten befreit haben – z.B. alle bis auf eine?

Nach diesem Interpretationsschema kommen wir zu folgender Zerlegung für (4):

$$\begin{aligned} & ( ((\text{FREUND HANS KARL}) \vee (\text{FREUND HANS EGON})) \\ & \wedge ((\text{FREUND FRITZ KARL}) \vee (\text{FREUND FRITZ EGON})) ) \\ & \wedge ( ((\text{FREUND HANS KARL}) \vee (\text{FREUND FRITZ KARL})) \\ & \wedge ((\text{FREUND HANS EGON}) \vee (\text{FREUND FRITZ EGON})) ) \end{aligned}$$

Machen wir nun die Reduktion durch Aufzählung rückgängig, so ergibt sich als quantorenlogische Umschreibung für (4):

$$\begin{aligned} (4\text{-PL}^1) & (\forall x.\text{per} (\text{BRUDER EVA } x.\text{per}) \rightarrow \exists y.\text{per} (\text{KOLLEGE LINDA } y.\text{per}) \\ & \wedge (\text{FREUND } x.\text{per } y.\text{per})) \\ & \wedge (\forall y.\text{per} (\text{KOLLEGE LINDA } y.\text{per}) \rightarrow \exists x.\text{per} (\text{BRUDER EVA } x.\text{per}) \\ & \wedge (\text{FREUND } x.\text{per } y.\text{per})) \end{aligned}$$

Wir wollen nun diese prädikatenlogische Umformulierung jedoch nicht syntaktisch vollziehen, wie eben demonstriert, sondern semantisch, d.h. wir führen die Interpretation von (4) als Reduktion auf die Semantik von  $\text{PL}^1$  durch. Genau das geschieht durch den Auswertungsalgorithmus von KS-Ausdrücken gegenüber der Datenbasis.

$$\begin{aligned} \text{Val}(((U \ U) \text{FREUND} <(\text{LAMBDA } x.\text{per} (\text{BRUDER EVA } x.\text{per})), \\ & (\text{LAMBDA } y.\text{per} (\text{KOLLEGE LINDA } y.\text{per}))>)) \\ & \alpha, d) = T \end{aligned}$$

gdw

$$\begin{aligned} \in_U U(\text{Val}(\text{FREUND}, \alpha, d), <\text{Val}((\text{LAMBDA } x.\text{per} (\text{BRUDER EVA } x.\text{per})), \\ & \alpha, d), \\ & \text{Val}((\text{LAMBDA } y.\text{per} (\text{KOLLEGE LINDA } y.\text{per})), \\ & \alpha, d)>)) \end{aligned}$$

Hier zeigt sich nun die durch die tiefenstrukturelle Umformung von (4) zu (4-TSTR) erreichte Analogie von Syntax und Semantik:  $\in_U U$  ist die semantische Interpretation des Quantifikatorenwortes '(U U)',  $\in_U U$  ist eine zweistellige Relation zwischen einer Tupelmenge, der Interpretation des Prädikats, und einem Tupel von Entitäten, der Interpretation des Argumenttermtupels. Die Relation  $\in_U U$  ist rückführbar auf eine bestimmte wahrheitsfunktionale Verknüpfung von  $\in$ -Aussagen. Dabei wird eine  $\in$ -Aussage der Form

$$\in(a_1, a_2)$$

wie üblich interpretiert als 'a<sub>2</sub> ist Element von a<sub>1</sub>'.

$\epsilon_U U$  wird auf eine  $\wedge$ -Verknüpfung von  $\vee$ -Verknüpfungen von  $\epsilon$ -Formeln zurückgeführt, die im Falle der Interpretation von (4-TSTR) zu folgender Zerlegung führt. Dabei nehmen wir an, daß die beiden eingebetteten Val-Ausdrücke bereits ausgewertet sind zu jeweils [Hans, Fritz] und [Karl, Egon]. Dies geschieht rekursiv ebenfalls mit Hilfe des  $\epsilon$ -Prädikats.

$\epsilon_U U (\text{Val} (\text{FREUND}, \alpha, d), \langle [\text{Hans, Fritz}], [\text{Karl, Egon}] \rangle) =$

$$\begin{aligned} & ( (\epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Hans, Karl} \rangle) \\ & \quad \vee \epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Hans, Egon} \rangle)) \\ & \wedge (\epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Fritz, Karl} \rangle) \\ & \quad \vee \epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Fritz, Egon} \rangle))) \\ \wedge & ( (\epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Hans, Karl} \rangle) \\ & \quad \vee \epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Fritz, Karl} \rangle)) \\ & \wedge (\epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Hans, Egon} \rangle) \\ & \quad \vee \epsilon (\text{Val} (\text{FREUND}, \alpha, d), \langle \text{Fritz, Egon} \rangle))) \end{aligned}$$

Mit dieser Zerlegung haben wir das modelltheoretische Analogon von (4-PL<sup>1</sup>) erreicht.

#### 4.4. Die Interpretation von *alle / jeder* und die *alle-ein*-Ambiguität

In Abhängigkeit von der Stelligkeit des Prädikats, d.h. der Anzahl der eingebetteten komplexen Terme und dem Präfix der eingebetteten Terme sind die Indizes der  $\epsilon$ -Relation definierbar. Der Index 'U U' steht, wie wir gesehen haben, für zwei unquantifizierte komplexe Terme, die uns zwei pluralische *die*-Artikel repräsentierten.

Wollen wir hingegen den Satz

- (5) *Die Brüder von Eva sind mit jedem der Kollegen von Linda befreundet.*

interpretieren, so gelangen wir zunächst zu folgender Repräsentation:

(5-TSTR)

$$((U \text{ ALL}) \text{ FREUND} \langle (\text{LAMBDA } x.\text{per} (\text{BRUDER } \text{EVA } x.\text{per})), (\text{LAMBDA } y.\text{per} (\text{KOLLEGE } \text{LINDA } y.\text{per})) \rangle)$$

Die Interpretation von (5-TSTR) lautet:

$$\begin{aligned} \text{Val } & (((U \text{ ALL}) \text{ FREUND} \langle (\text{LAMBDA } x.\text{per} (\text{BRUDER } \text{EVA } x.\text{per})), \\ & \quad (\text{LAMBDA } y.\text{per} (\text{KOLLEGE } \text{LINDA } y.\text{per})) \rangle), \\ & \alpha, d) = T \text{ gdw} \end{aligned}$$

$$\in U_{ALL}(\text{Val}(\text{FREUND}, \alpha, d), \langle \text{Val}((\text{LAMBDA } x.\text{per}(\text{BRUDER EVA } x.\text{per})), \alpha, d), \text{Val}((\text{LAMBDA } y.\text{per}(\text{KOLLEGE LINDA } y.\text{per})), \alpha, d) \rangle)$$

Nach Auswertung der eingebetteten Val-Ausdrücke gegenüber der Datenbasis ergibt sich:

$$\begin{aligned} \in U_{ALL}(\text{Val}(\text{FREUND}, \alpha, d), \langle [\text{Hans}, \text{Fritz}], [\text{Karl}, \text{Egon}] \rangle) = \\ \in (\text{Val}(\text{FREUND}, \alpha, d), \langle \text{Hans}, \text{Karl} \rangle) \\ \wedge \in (\text{Val}(\text{FREUND}, \alpha, d), \langle \text{Hans}, \text{Egon} \rangle) \\ \wedge \in (\text{Val}(\text{FREUND}, \alpha, d), \langle \text{Fritz}, \text{Karl} \rangle) \\ \wedge \in (\text{Val}(\text{FREUND}, \alpha, d), \langle \text{Fritz}, \text{Egon} \rangle) \end{aligned}$$

Diese Zerlegung entspricht dem intuitiven Verständnis von (5), gemäß dem beide Brüder von Eva mit beiden Kollegen von Linda befreundet sein müssen.

Ich kann hier nicht mehr ausführlich darauf eingehen, wie mithilfe von LAMBDA-Abstrakten, die als Prädikatsausdrücke gebraucht werden, Skopusambiguitäten bewältigt werden können.

Nur ein Beispiel sei angeführt:

(6) *Alle Brüder von Eva lieben eine Freundin von Eva.*

Die Art der Ambiguität solcher Beispiele ist vielfach erörtert worden.

Wir bedienen uns zur Disambiguierung der beiden komplexen Prädikatsausdrücke

(6a) (LAMBDA x.per (LAMBDA y.per (LIEB x.per y.per)))

und

(6b) (LAMBDA y.per (LAMBDA x.per (LIEB x.per y.per))).

(6a) ist ein Prädikat, dessen Semantik informal so verstanden werden kann, daß es von einer Person (x.per) prädiziert, daß sie jemanden (y.per) liebt, während (6b) so zu verstehen ist, daß es von einer Person (y.per) prädiziert, daß sie von jemandem (x.per) geliebt wird.

Es handelt sich also um zwei semantisch verschiedene Prädikate, deren Verwendung in Formeln wir wie folgt interpretieren: In Analogie zu der Verwendungsweise der LAMBDA-Abstraktion und der Applikation von LAMBDA-Abstrakten bei ähnlichen Fällen in Cresswell 1973 sagen wir, daß bei einer Verwendung von (6a) in einer tiefenstrukturellen Formel das Argument, das als Wert von y.per erscheint, so interpretiert wird, als stünde es im Skopus des Argumentes, das als Wert von x.per erscheint, in (6b) dagegen umgekehrt.

In der ersten Lesart von (6):

1. Lesart von (6) = *Die Brüder von Eva lieben eine, jeweils verschiedene Freundin von Eva.*

ist der Ausdruck *eine Freundin von Eva*, der als Wert des y.per-Argumentes aufzufassen ist, im Skopus des Ausdrucks *alle Brüder von Eva*, der als Wert des x.per-Argumentes aufzufassen ist, denn wir umschreiben hier so: *Für alle Brüder gibt es eine Freundin, sodaß ...*

Daher ergibt sich für die erste Lesart die folgende tiefenstrukturelle KS-Repräsentation:

(6-1 TSTR)

$$\begin{aligned} & (( \text{EIN ALL} ) (\text{LAMBDA } x.\text{per} (\text{LAMBDA } y.\text{per} (\text{LIEB } x.\text{per } y.\text{per}))) \\ & \quad < (\text{LAMBDA } u.\text{per} (\text{FREUND } \text{EVA } u.\text{per})), \\ & \quad (\text{LAMBDA } z.\text{per} (\text{BRUDER } \text{EVA } z.\text{per})) > ) \end{aligned}$$

Ich übergehe jetzt die Schritte der Anwendung der Val-Funktion und der Auswertung der eingebetteten Terme gegenüber der Datenbasis.

Danach wird ausgewertet zu:

$$\begin{aligned} \in \text{EIN ALL} (\text{Val} ((\text{LAMBDA } x.\text{per} (\text{LAMBDA } y.\text{per} (\text{LIEB } x.\text{per } y.\text{per}))) \alpha, d) \\ < [\text{Helga}, \text{Erna}], [\text{Hans}, \text{Fritz}] > ) = \text{T} \text{ gdw} \end{aligned}$$

$$\begin{aligned} & ( \quad \in (\text{Val} (\text{LIEB}, \alpha, d), < \text{Hans}, \text{Helga} > ) \\ & \quad \vee \in (\text{Val} (\text{LIEB}, \alpha, d), < \text{Hans}, \text{Erna} > ) ) \\ \wedge & ( \quad \in (\text{Val} (\text{LIEB}, \alpha, d), < \text{Fritz}, \text{Helga} > ) \\ & \quad \vee \in (\text{Val} (\text{LIEB}, \alpha, d), < \text{Fritz}, \text{Erna} > ) ) \end{aligned}$$

2. Lesart von (6) = *Die Brüder von Eva lieben alle ein und dieselbe Freundin von Eva.*

Hier ist dagegen *alle Brüder von Eva* (Wert von x.per) im Skopus von *eine Freundin*:

$$\begin{aligned} & (( \text{ALL EIN} ) (\text{LAMBDA } y.\text{per} (\text{LAMBDA } x.\text{per} (\text{LIEB } x.\text{per } y.\text{per}))) \\ & \quad < (\text{LAMBDA } z.\text{per} (\text{BRUDER } \text{EVA } z.\text{per})), \\ & \quad (\text{LAMBDA } u.\text{per} (\text{FREUND } \text{EVA } u.\text{per})) > ) \end{aligned}$$

Dieser Ausdruck wird interpretiert als:

$$\begin{aligned} \in \text{ALL EIN} (\text{Val} ((\text{LAMBDA } y.\text{per} (\text{LAMBDA } x.\text{per} (\text{LIEB } x.\text{per } y.\text{per}))), \\ \alpha, d), \\ < [\text{Hans}, \text{Fritz}], [\text{Helga}, \text{Erna}] > ) = \text{T} \text{ gdw} \end{aligned}$$

- (  $\in$  (Val (LIEB,  $\alpha$ , d), <Hans, Helga>))
- $\wedge$   $\in$  (Val (LIEB,  $\alpha$ , d), <Fritz, Helga>))
- $\vee$  (  $\in$  (Val (LIEB,  $\alpha$ , d), <Hans, Erna>))
- $\wedge$   $\in$  (Val (LIEB,  $\alpha$ , d), <Fritz, Erna>))

Die beiden Auflösungen in Verknüpfungen von  $\in$ -Aussagen spiegeln genau den Bedeutungsunterschied zwischen den beiden Lesarten in der konkreten Anwendung auf ein endliches Modell wieder: In der Interpretation von Lesart 1 steht Hans in der Relation 'Lieben' zu Helga oder Erna, dasselbe gilt für Fritz; in der Interpretation von Lesart 2 stehen sowohl Hans als auch Fritz in der Relation 'Lieben' zu Helga oder in der Relation 'Lieben' zu Erna.

Der Übergang von indizierten  $\in$ -Aussagen zu den jeweiligen Verknüpfungen von einfachen  $\in$ -Aussagen ist regelhaft beschreibbar und somit auch algorithmisierbar.

## 5. Schlußbemerkung

Das Fragment einer Analyse und Interpretation quantifizierter deutscher Nominalphrasen für ein Informationssystem, das hier gegeben werden konnte, sollte einige Eigenschaften von KS verdeutlichen: KS bietet gegenüber PL<sup>1</sup> eine strukturell angemessenere Repräsentation von Nominalphrasen, indem sie sie als Argumente – somit als Besetzer von "Valenzstellen" – der Prädikate einbettet. Die Repräsentation von Nominalphrasen als komplexe Termausdrücke, die gegenüber dem Modell zu Listen auswertbar sind, bietet den Ansatz zu einer Operationalisierung modelltheoretischer Interpretation. Die Berücksichtigung der Quantifikatorenwörter ermöglicht den semantisch angemessenen Übergang von listenenthaltenden Relationen zu elementaren, nur Individuen enthaltenden Relationen.

Theoretische Reflexion der ersten Kapitel und beispielhafte Analyse der Quantifikation zeigen, daß der Anwendungsfall 'semantische Analyse für ein Informationssystem' besondere Bedingungen setzt. Diese Bedingungen, Kommunikation in einem faktischen Diskurs, Möglichkeiten der Extensionalisierung, Rückbezug auf ein endliches Modell, und die sich daraus ergebenden Konsequenzen für die Konstruktion semantischer Beschreibungssprachen sind nicht ohne Interesse für die Bedeutungsanalyse auch außerhalb dieses Kontextes.



## Anmerkungen

- 1 Zur Forschungsrichtung 'Künstliche Intelligenz', speziell ihrem sprachorientierten Zweig, vgl. z.B. Wulz/Zifonun 1979. Siehe dort auch weitere Literaturangaben.
- 2 Zum Informationssystem PLIDIS vgl. den Beitrag von Berry-Rogghe/Lutz 1979.
- 3 Zu diesem Interpretationsverfahren vgl. Montague UG und Kamp 1978.
- 4 Dieses fiktive Informationssystem teilt mit einem anwendungsorientierten System wie PLIDIS diejenigen Eigenschaften, die unter den eher theoretischen Gesichtspunkten, die hier erörtert werden sollen, relevant sind. Auch der in Kapitel 4. dargestellte sprachliche Anwendungsteil, die Repräsentation quantifizierter Ausdrücke, bezieht sich auf den allgemeinsprachlichen Bereich jedes beliebigen Informationssystems mit Deutsch als Interaktionssprache, nicht auf die spezielle Fachsprache der Abwasserüberwachung.
- 5 Es gibt allerdings Versuche, modallogische Theorembeweiser und Beweisverfahren für typentheoretische Sprachen zu entwickeln, vgl. dazu z.B. Stephan 1975.
- 6 Zur Nicht-Berücksichtigung von propositionalen Einstellungen wie 'glauben', 'wissen', 'erwarten', 'hoffen' vgl. oben im Zusammenhang mit Sprechaktindikationen.
- 7 Vage Ausdrücke werden z.B. in dem Künstlichen-Intelligenz-Projekt HAM-RPM ('Hamburger Redepartner-Modell') behandelt, vgl. dazu Wahlster 1977.
- 8 Montague PTQ, S. 263 f., vgl. dazu auch die Bemerkungen Thomasons in der "Introduction" zu dem Band, S. 62 ff.
- 9 Montague 1974, Introduction, S. 62.
- 10 Das Beispiel stammt aus Mondadori 1978, S. 17.
- 11 Diese Konsequenz zieht Montague, sie ist für eine rein spracherkklärende Theorie ohne Anwendungsbezug wohl angemessen.
- 12 "Rigid designators" sind Ausdrücke, die bezüglich aller in Frage stehenden Referenzpunkte oder möglichen Welten dasselbe Individuum bezeichnen, vgl. dazu Mondadori 1978, S. 21 ff.
- 13 Attributiver Gebrauch liegt z.B. vor in *The man who murdered Smith is insane.* (Partee 1972, S. 418).
- 14 Das Beispiel stammt aus Cresswell 1973, S. 68 ff.
- 15 Die Nullentität wird z.B. in Bressan 1978 in diesem Sinne benutzt.
- 16 Die Zuordnung von Teilausdrücken einer intensionalen Sprache zu einer transformationsgrammatischen Tiefenstruktur beschreibt Cooper 1978.
- 17 In dem Projekt von J. Friedman und ihren Mitarbeitern wird ein automatisches Syntaxanalyse-system für kategorialgrammatische Strukturen entwickelt, vgl. Friedman 1978, 2.

- 18 Solche Rückführungen ("extensional reductions") geschehen in Computersystemen auf der Basis von Montague-Grammatiken, so in den Ansätzen von Friedman 1978, 1, Sondheimer 1978 u. Janssen 1978. Möller-Pantleon 1978 beschreibt ebenfalls einen theoretischen Ansatz zur Extensionalisierung intentionaler Ausdrücke für ein Frage-Antwort-System. Sie gibt auch einen Überblick über die Probleme logischer Repräsentation für Informationssysteme.
- 19 Zu einer detaillierteren Beschreibung vgl. Zifonun 1979 und Dilger/Zifonun 1978.
- 20 Zur Bildung nicht-atomarer Formeln, die nach prädikatenlogischem Vorbild erfolgt, vgl. Zifonun 1979.
- 21 Das Zeichen '↳' steht für 'wird in KS übersetzt als'.
- 22 Langendoen 1978, S. 185

## Literatur

- Berry-Rogge, G.L. / Lutz, H. (1979): Das Informationssystem PLIDIS. In: Kolvenbach/Lötscher/Lutz (1979).
- Bressan, A. (1978): The Interpreted Modal Calculus MC and Some of Its Applications. In: Guentner/Rohrer (1978), S. 75 - 118.
- Cooper, R. (1978): Montague's Theory of Translation and Transformational Syntax. In: Guentner/Guentner-Reutter (1978), S. 307-325.
- Cresswell, M.J. (1973): Logics and Languages. London.
- Dilger, W./Zifonun, G. (1978): Eine Regelgrammatik für KS. Unveröffentlichtes Arbeitspapier IdS, Abt. LDV.
- Friedman, J./Moran, D.B./Warren, D.S. (1978): Two Papers on Semantic Interpretation in Montague Grammar, in: American Journal of Computational Linguistics, Microfiche 74 (=Friedman 1978, 1).
- Friedman, J./Warren, D.S. (1978): A Parsing System for Montague Grammar, in: Linguistics and Philosophy 2, S. 347-373 (= Friedman 1978, 2).
- Guentner, F./Guentner-Reutter, M. (Hg.) (1978): Meaning and Translation. Philosophical and Linguistic Approaches. London.
- Guentner, F./Rohrer, Ch. (Hg.) (1978): Studies in Formal Semantics. Intensionality, Temporality, Negation. Amsterdam.
- Janssen, T.M.V. (1978): Logical Investigations on PTQ Arising from Programming Requirements. Department of Mathematics, University of Amsterdam (preprint).
- Kamp, H. (1978): The Adequacy of Translation between Formal and Natural Language. In: Guentner/Guentner-Reutter (1978), S. 275 - 306.
- Kolvenbach, M./Lötscher, A./Lutz, H.(Hg.) (1979): Künstliche Intelligenz und natürliche Sprache. Sprachverstehen und Problemlösen mit dem Computer = Forschungsberichte des IdS 42, Tübingen. In Druck.

- Langendoen, D.T. (1978): The Logic of Reciprocity, in: *Linguistic Inquiry* 9, 2, S. 177 - 197.
- Möller-Pantleon, G. (1978): *Montague-Grammatiken und Frage-Antwort-Systeme*. Unveröffentlichte Magisterarbeit, Fachbereich Linguistik und Literaturwissenschaften, Universität Stuttgart.
- Mondadori, F. (1978): Interpreting Modal Semantics. In: Guenther/Rohrer (1978), S. 13 - 40.
- Montague, R. (1974): *Formal Philosophy. Selected Papers*. Ed. by R. Thomason. New Haven and London. Darin: *Universal Grammar*, S. 222 - 246 (=UG), *The Proper Treatment of Quantification in Ordinary English*, S. 247 - 270 (=PTQ).
- Partee, B. (1973): Opacity, Coreference and Pronouns. In: Davidson, D./ Harman, G. (Hg.) (1973), *Semantics of Natural Languages*, Dordrecht, S. 415 - 441.
- Sondheimer, N.K./Gunji, T. (1978): Applying Model-Theoretic Semantics to Natural Language Understanding: Representation and Question Answering. In: *Proceedings of 7th COLING*, Bergen, Norway (to appear).
- Stephan, W. (1975):  $\lambda$ -Kalkül und Logik höherer Stufe. Universität Karlsruhe, Fakultät für Informatik, Interner Bericht Nr. 17.
- Wahlster, W. (1977): Die Repräsentation von vagem Wissen in natürlich-sprachlichen Systemen der künstlichen Intelligenz. Institut für Informatik, Universität Hamburg, Bericht Nr. 38.
- Wulz, H./Zifonun, G. (1979): Automatische Problemlösung und Sprachverarbeitung als Forschungsgegenstände. In: Kolvenbach/Lötscher/Lutz (1979).
- Zifonun, G. (1979): Formale Repräsentation natürlichsprachlicher Äußerungen. In: Kolvenbach/Lötscher/Lutz (1979).