

Lizentiatsarbeit
Universität Zürich
Philosophische Fakultät I

Ellipsenbehandlung im UIS

Verfasser: Simon Clematide
Referent: Prof. Dr. M. Hess

1.12.1998

Inhalt

1 Einleitung	1
1.1 UIS	1
1.2 Ellipsen: Ein Beispiel.....	3
1.3 „Ellipse“ und „elliptisch“ in der Sprachtheorie	4
2 Kommunikation und Ellipsen	9
2.1 Natürlich-sprachliche Mensch-Maschine-Kommunikation (MMK)	9
2.2 Arten von Ellipsen	11
2.2.1 Syntaktische Struktur von Frage- und Nachfrage-Ellipsen	11
2.2.2 Koordinationsellipsen.....	12
2.2.3 Koordinationsreduktion	13
2.2.4 Gapping und Stripping.....	15
2.2.5 VP-Ellipse	20
2.2.6 Sluicing	21
2.2.7 N-Ellipse.....	22
2.2.8 Satz-Ellipse.....	23
2.3 Kommunikationsfunktion.....	23
3 NLP und Ellipsenbehandlung	26
3.1 Ansätze zur Ellipsenbehandlung	26
3.2 Saarbrücker Dialogsystem	27
3.2.1 Benutzerinteraktion	27
3.2.2 Die Behandlung von Erweiterungsellipsen	28
3.2.3 Die Behandlung von Ersetzungsellipsen	30
3.2.4 Mischung von Erweiterungs- und Ersetzungsellipsen.....	43
3.2.5 Zur Architektur der Ellipsenbehandlung	43
3.2.6 Schlusskommentar	48
3.3 Core Language Engine (CLE).....	49
3.3.1 Überblick	49
3.3.2 Ellipsenbehandlung	51
3.4 HPSG-Ansatz.....	53
3.5 UIS	55
3.5.1 F-Struktur als Zielformat	56
3.5.2 Ellipsenauflösung als F-Struktur-Rekonstruktion.....	59
3.5.3 Umgang mit Mehrdeutigkeiten	61
3.5.4 Behandlung der Satznegation.....	64
3.5.5 Mögliche Erweiterungen der Ellipsenauflösung	65
3.5.6 Fazit	66
4 Schluss	67
5 Literatur	69
6 Anhänge	72

Abbildungsverzeichnis

<i>Abbildung 1: Passage retrieval im UIS</i>	2
<i>Abbildung 2: Vollformen-Lexikoneintrag zu „hält“</i>	32
<i>Abbildung 3: Vollformen-Lexikoneintrag zum Eigennamen „Barry“</i>	34
<i>Abbildung 4: Vollformen-Lexikoneintrag zu „Vorlesung“</i>	38
<i>Abbildung 5: Das Konzept concept</i>	40
<i>Abbildung 6: Schema der Anfrage-Beantwortung im Saarbrücker System</i>	44
<i>Abbildung 7: Schema der Satzanalyse in der CLE</i>	50
<i>Abbildung 9: Beschreibung der F-Struktur des Satzes im UIS</i>	57

1 Einleitung

Die vorliegende Arbeit „Ellipsenbehandlung im UIS“ beschäftigt sich mit dem syntaktischen und pragmatischen Verständnis von Ellipsen, stellt unterschiedliche Ansätze vor, mit denen elliptische Äußerungen in natürlich-sprachlichen Datenverarbeitungs-Systemen behandelt werden, und beschreibt die von mir implementierte Behandlung von Frage-Ellipsen im UIS.

1.1 UIS

Das UIS (*University Information System* = **Universitäts-Informationssystem**) ist ein Projekt der Computerlinguistik-Gruppe der Universität Zürich, das im Frühjahr 1996 gestartet wurde. Ziel des Projekts ist es, ein Abfragesystem zu schaffen, das natürlich-sprachliche Anfragen entgegennimmt und als Antwort eine oder mehrere relevante Textstellen aus einem gegebenen natürlich-sprachlichen Informationstext zurückliefert (*answer extraction*). Das Wiederauffinden der relevanten Textstellen (*passage retrieval*) wird allerdings nicht über statistische, indizierende oder andere klassische Formen des *information retrieval* geführt. Sondern es wird aufgrund einer möglichst exakten syntaktischen Analyse für den Informationstext und für die Anfrage eine logisch-semantische Struktur (*logical form*) konstruiert, welche die Bedeutung der natürlich-sprachlichen Texte repräsentiert. Per Beweisverfahren versucht das System danach, die logische Form der Anfrage aus der logischen Form des Informationstextes herzuleiten. Wenn die Herleitung glückt, werden diejenigen Textstellen als Antwort verwendet, welche die beim Beweis involvierten Axiome enthalten. Es gehört zur grundlegenden Idee des *passage retrieval*, dass keine Sprachgenerierung durchgeführt wird.

Das UIS ist ein deutschsprachiges System, das Studierenden einen einfachen Zugang zu administrativen Informationen der Universität Zürich bieten soll. Schriftliche Anfragen und die passenden Systemantworten erfolgen über eine Schnittstelle im *World Wide Web* (WWW) des Internets. Da der Informationstext viele Hyper-textverweise enthält, ergibt sich für die Benutzenden ein nahtloser Übergang zwischen UIS-interner Suche und Navigieren im WWW.

Unter der Bezeichnung *ExtrAns* verfolgt die Computerlinguistik-Gruppe der Universität Zürich ein ähnliches Projekt, bei dem nicht aufbereitete, englischsprachige Handbuchttexte des Unix-Betriebssystems (sogenannte *man pages*) als Informations-

texte dienen. Die wichtigsten Konzepte und Ansätze von *ExtrAns* werden in (Mollà Aliod Berri et al. forthcoming) beschrieben und treffen ebenfalls auf das UIS zu.

Abbildung 1 gibt einen Überblick über die Architektur des *passage retrieval*, wie es im UIS verwendet wird. Sowohl die Syntaxanalyse als auch das Beweisverfahren sind in PROLOG implementiert.

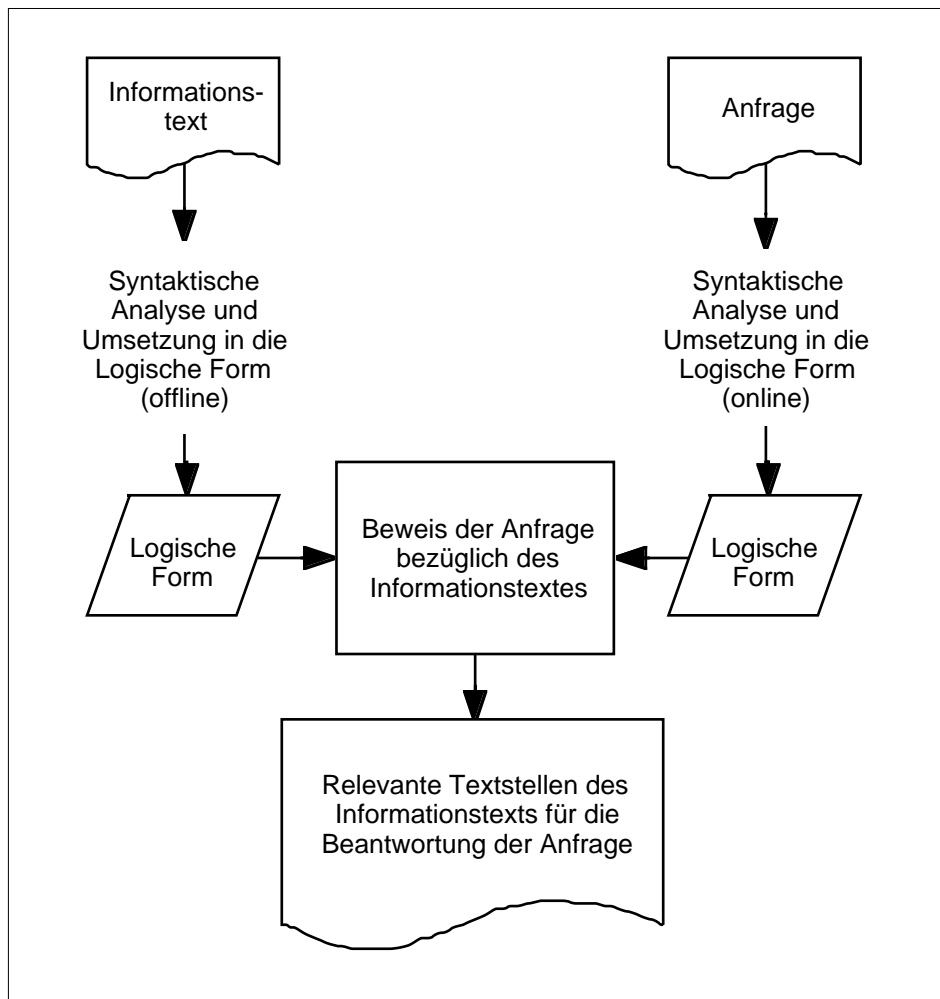


Abbildung 1: Passage retrieval im UIS

Die Passagensuche per Beweisverfahren erfordert bei der angestrebten feinkörnigen und linguistisch anspruchsvollen Analyse der Texte einen hohen Berechnungsaufwand; die Grösse der anvisierten Informationstexte beschränkt sich deshalb zum heutigen Zeitpunkt auf einige 10 bis 100 Kilobytes. Dies mag klein erscheinen im Vergleich zu *information retrieval* Systemen, wie sie standardmässig etwa als Suchmaschinen im WWW verwendet werden, und welche Anfragen in Dokumentenbeständen der Grössenordnung von vielen Gigabytes erlauben. Der Schwerpunkt von *passage retrieval* Systemen wie dem UIS liegt jedoch nicht bei einer möglichst

grossen Universalität und Skalierbarkeit. Es geht hier um Informationsbeschaffung, die in einem begrenzten Rahmen stattfindet, aber mit hoher Genauigkeit (*precision*) arbeitet, und den Verlust von Information vermeidet, d.h. einen hohen Wiedergewinnungsgrad (*recall*) aufweist. Textsorten wie Frage-Antwort-Zusammenstellungen (*FAQ*) oder Handbücher sind prototypisch für mögliche Anwendungsbeispiele, weil sie meist konzis und redundanzarm verfasst sind, d.h. einen gewissen Sachverhalt nur an einer Stelle erklären.

Die Beschränkung auf bestimmte Domänen erlaubt es wiederum, soviel domänenspezifisches Wissen ins System miteinzubeziehen, dass es für die Benutzenden genügend flexibel und robust wird. In natürlich-sprachlichen Systemen stellen Flexibilität und Robustheit entscheidende Kriterien dar für deren Nützlichkeit und Akzeptanz. Ein weiteres Kriterium, das ursprünglich für Dialogsysteme postuliert worden ist, aber trotzdem auch für das UIS zutrifft, lautet:

„Natürlichsprachliche Frage/Antwort-Systeme werden erst dann voll akzeptiert werden, wenn sie pragmatische Fähigkeiten beherrschen.“ (Fauser Rathke 1981: 93)

Mit Ellipsen umgehen zu können, wird gemeinhin als eine solche pragmatische Fähigkeit betrachtet, die die Mensch-Maschine-Kommunikation den Bedürfnissen der „menschlichen Schnittstelle“ näherbringt.¹

1.2 Ellipsen: Ein Beispiel

Wenn man eine deutschsprachige Person fragt, welche von den folgenden Äusserungen korrekte und zulässig Sätze sind,

Ich nicht.

Wer die Vorlesung?

Vielleicht Peter

dann würde die Antwort vermutlich lauten: „Das sind keine Sätze, sondern ziemlich unverständliche Aneinanderreihungen!“ Dieselbe Person hätte aber wahrscheinlich keinerlei Schwierigkeiten, die Äusserungen des folgenden Dialogs als akzeptabel zu bewerten:

¹ Vergleiche dazu etwa (Küpper Rösner et al. 1992: 31f.,54f.; Jönsson 1993: 2f.; Lewin Pulman 1995: 1).

A: „Wer besucht heute das Seminar?“

B: „Ich nicht.“

A: „Wer die Vorlesung?“

B: „Vielleicht Peter.“

Das Verständnis und die Wohlgeformtheit von Äusserungen hängt also nicht bloss von dem ab, was geäussert wird, sondern auch vom Kontext, in dem etwas geäussert wird. Wer eine elliptische Äusserung verstehen will, muss den Kontext präsent haben. Für Menschen bedeutet die Interpretation von kontextuell eingebetteten Ellipsen keine besondere Anstrengung. Es wäre im Gegenteil anstrengender, den obigen Dialog in vollständigen Sätzen zu führen:

A: „Wer besucht heute das Seminar?“

B: „Ich besuche es heute nicht.“

A: „Wer besucht heute die Vorlesung?“

B: „Vielleicht besucht sie Peter heute.“

Es ist der Zweck dieser Arbeit, den Komfort solcher elliptischer Interaktionen auch im UIS verfügbar zu machen. Da das UIS als Abfragesystem konzipiert ist, das die Antworten nicht selbst generiert, sondern bestehende Textausschnitte als Antworten verwendet, beschränkt sich der Anwendungsbereich des Ellipsenmoduls auf die Behandlung von Frageellipsen. Der Kontext, in dem diese Frageellipsen aufgelöst werden, umfasst dabei die vorgängigen Benutzerfragen, Systemantworten werden nicht berücksichtigt.

Inwieweit es sinnvoll und allenfalls machbar wäre, elliptische Fragen auch in bezug auf den Inhalt der vom System gelieferten Antwortpassagen zuzulassen, ist eine offene Frage.

1.3 „Ellipse“ und „elliptisch“ in der Sprachtheorie

In der Sprachtheorie und Grammatikschreibung beschreibt und erklärt der Begriff „Ellipse“ (lat. *ellipsis*: ‘Auslassung eines Wortes’ zu grch. *éllipsis*: ‘Mangel’) so viele unterschiedliche Phänomene, dass er für Dentler(1990: 4f.) eine schwammige Sammel- und Restkategorie darstellt. Manchmal bezeichnet das Wort „Ellipse“ einen Ausdruck, in dem etwas ausgelassen wurde, manchmal aber auch das Ausgelassene selbst. Die doppelte Verwendung von „Ellipse“ mag zusätzlich zur Begriffunschärfe beigetragen haben. Da sie gängig und praktisch ist, habe ich mich dieser Unsitte angeschlossen.

Wenn ich beispielsweise von Frage- oder Nachfrage-Ellipsen spreche, verwende ich den ersten Sinn, d.h. als Ausdruck in dem etwas ausgelassen wurde. Wenn ich

die meisten Nachfrage-Ellipsen zu den sogenannten Verbal-Ellipsen zähle, verwende ich es im zweiten Sinn.

Ortner (1987) hat die zum Teil verworrene Geschichte des ursprünglich stilistisch-rhetorischen Begriffs für die deutsche Sprachtheorie und Grammatikschreibung ausführlich aufgezeigt; Schuh (1974) macht dasselbe aus einer noch pragmatischeren Sichtweise der Textlinguistik für das Französische.

In der theoretischen Diskussion (Eikmeyer 1985) wird das Adjektiv „elliptisch“ oft in einen dualen Zusammenhang zu „vollständig“ gebracht. Die Vollständigkeit von sprachlichen Äußerungen muss dabei in mindestens zweifacher Hinsicht unterschieden werden (Strohner Rickheit 1985):

- **funktional kommunikativ.** Eine Äußerung ist dann kommunikativ vollständig, wenn sie von einem (meist kooperativen) Adressaten in der intendierten Form verstanden werden kann.
- **strukturell sprachlich.** Eine Äußerung ist dann strukturell vollständig, wenn sie einem (meist syntaktisch-semantisch spezifizierten) Satzmodell genügt.

Der Linguist Schreiber² fasst diese zwei Aspekte in seiner Definition von Satzellipse zusammen:

„Unter einer elliptischen Satzstruktur wird im allgemeinen verstanden, dass bestimmte, nach der Sprachnorm strukturell notwendige sprachliche Konstituenten vom Sender bei der Kodierung der Information weggelassen werden, ohne dass beim Empfänger eine Minderung des vom Sender gewünschten Kommunikationseffektes eintritt.“

Kommunikative Vervollständigung ist für Menschen oft auch bei der Verarbeitung von strukturell unvollständigen Äußerungen möglich; sogar sprachlich stark defektiv Äußerungen (Satzabbrüche, Satzfragmente, Einschübe, Wiederholungen, Fehlproduktionen usw.) sind aus der jeweiligen Kommunikationssituation gut interpretierbar.

Unter Ellipsen versteht man normalerweise aber keine sprachliche Unvollständigkeit durch fehlerhafte Auslassung, sondern usuelle, systematische und sprachökonomische Weglassung. Das können textsortenspezifische (Telegrammstil, Überschriften, Schlagzeilen usw.) oder situative und ritualisierte Kurzformen (*Das Skalpell!*) sein. In einem Sprachmodell, das für das Verb *essen* zwei Ergänzungen fordert, ist der Satz *Hans isst* elliptisch. In einem Sprachmodell, das für jeden Satz ein sichtbares Subjekt fordert, ist der Imperativ *Geh nach Hause!* elliptisch.³

² Zitiert nach (Ortner 1987: 132), der noch weitere Definitionen aufführt, die die Ellipsen als Produkt der Sprechhandlung und nicht der ‚normalen‘ Sprache betrachten.

³ Textsortenspezifische Ellipsen werden in (Ortner 1987: 140 ff.) beschrieben. Schwabe (1994) macht den (gewagten) Versuch, zu situativen Ellipsen eine generative Syntax und kompositionelle

Prototypisch für sogenannte ‘syntaktische Ellipsen’ sind Koordinationsellipsen: *Er trank Bier und sie Wein*. Bei Frage-Antwort-Paaren kann in der Antwort alles weggelassen werden, was nicht erfragt ist: *Wer hat was geschenkt bekommen? Hans ein Rasierwasser, Berta eine Flasche Wein*.⁴

Die obigen Beispiele verdeutlichen, dass die Unvollständigkeit von Ellipsen vom verwendeten Sprachmodell abhängt. Ortner (1987: 157f.) kritisiert die Sprachtheoretiker, welche vom „Masstab eines Normalsatzes“ ausgehen und alle abweichenden Strukturen in Paraphrasen von Normalsätzen charakterisieren; Ellipsen sind für ihn eine „autonome Versprachlichungsmöglichkeit“.

Die gegensätzliche Position vertreten die Reduktionisten:

„Für eine Ellipse soll gelten, dass ein Element eines Satzes weggelassen werden kann, ohne dass sich das Verständnis des Satzes und seine Bedeutung verändern.“⁵

Die weggelassenen Elemente werden dabei auf unterschiedlichen Ebenen angesiedelt. Syntaktische Reduktionisten gehen davon aus, dass auf der Ebene der syntaktischen Struktur weggelassen bzw. getilgt wurde. Phonologische Reduktionisten setzen zwischen elliptischem und dem entsprechenden vollständigen Satz identische Syntaxstrukturen voraus, wobei dann gewisse Elemente auf der phonologischen Ebene ausfallen.⁶

Der Weg von der vollständigen zur reduzierten Form entspricht der Sichtweise der Sprachproduktion. Aus der Perspektive des Sprachverstehens besteht das Problem nicht darin, wo man reduziert, sondern wie man rekonstruiert. Isoliert auftretende elliptische Äußerungen wie im Beispiel des vorhergehenden Abschnitts 1.2 benötigen zur Interpretation eine Vervollständigung. Wenn man diese Vervollständigung auf einer Repräsentationsebene ansiedeln möchte, bieten sich die Syntax, die Semantik oder allenfalls Kombinationsformen davon an.⁷ Um die Kontextinformation integrieren zu können, muss sie in einer geeigneten Form vorliegen. Im Verlauf

Semantik zu erstellen. Dentler (1990) untersucht das Fehlen von „notwendigen“ Verbergänzungen. Syntaktisch bedingte Nicht-Realisation von Subjekt oder Objekt (Stichwort *small clause*, *pro-drop* usw.) werden in den meisten moderneren Syntaxtheorien berücksichtigt; für Deutsch im Ansatz der Generativen Grammatik beispielsweise (Stechow Sternefeld 1988).

⁴ Issatschenko (1978) enthält gutes Material zu den Bedingungen von Frage-Antwort-Ellipsen im Deutschen und in weiteren Sprachen.

⁵ Aus R. Steinnitz (1969): „Adverbial-Syntax“, zitiert nach (Ortner 1987: 99).

⁶ Klein (1991) vertritt eine Position der phonologischen Reduktion, bespricht aber die wichtigen syntaktischen Ansätze.

⁷ Wegen der Unterschiedlichkeit der Phänomene, welche mit dem Prädikat „elliptisch“ bezeichnet werden, muss man damit rechnen, dass unterschiedliche Komponenten beteiligt sind oder zusammenspielen.

dieser Arbeit sollen Beispiele für syntaktische, semantische und mischende Ansätze vorgestellt werden.

Zum Schluss möchte ich noch Kindts pragmatisch ausgerichtete Explikation des intuitiven Ellipsenbegriffs präsentieren und auf Frage-Ellipsen anwenden:

„Eine angemessene Explikation des intuitiven Ellipsenbegriffs ist durch folgende Definition gegeben: Eine nicht-leere Sequenz S ist in der Umgebung USV elliptisch relativ zu der Expansionserwartung E genau dann, wenn erstens S in USV relativ zu E syntaktisch unvollständig ist, wenn zweitens USV grammatisch akzeptabel ist und wenn drittens S in USV dieselbe Bedeutung hat wie mindestens eine der E entsprechenden Vervollständigungen von S.“
(Kindt 1985: 181)

Die Explikation mag etwas hölzern klingen. Sie ist aber genügend allgemein und trotzdem explizit, so dass sie die Anforderungen an ein ellipsenverarbeitendes Sprachsystem spezifiziert. Unter Umgebung USV wird sowohl sprachlicher wie sprachbegleitender Kontext verstanden. Eine pragmatische und sprachliche Kompetenz muss semantisch äquivalente Vervollständigungen E zu S erzeugen können, wobei offengelassen ist, wie und auf welcher Ebene eines Sprachsystems E repräsentiert ist. Das Kriterium der grammatischen Akzeptabilität von USV stellt einige Ansprüche an die Sprachkompetenz bzw. an syntaktisches Sprachwissen; es muss beurteilt – oder besser noch erklärt – werden können, warum von den folgenden Satzpaaren nur (a) und (b) akzeptabel sind:

- a) Zuerst hat er Maria das Buch nur geliehen, dann geschenkt.
- a') *Zuerst leiht er Maria das Buch nur, dann schenkt.
- b) Dann kam der Jäger und schoss den Hasen.
- b') *Dann kam der Jäger und den Hasen schoss.

Die Frage, warum gewisse Auslassungen möglich sind und andere nicht, ist eine spannende Frage, die die gesamte Grammatik tangiert.

Das Explikationsschema von Kindt soll noch auf Frage-Ellipsen angewendet werden: In einem Ausdruck wie *Vielleicht Peter?* ist syntaktisch Unvollständiges gegeben. Grammatische Akzeptabilität kann durch einen Kontext wie *Kommt Hans heute? Nein.* gesichert sein. Die pragmatische Kompetenz wird eine Expansionserwartung sinnvollerweise so steuern, dass eine Parallelkonstruktion zur Vorfrage im Sinn von *Kommt Peter vielleicht heute?* erzeugt wird.

In einem anderen Kontext könnte sich eine Ellipse wie *Vielleicht Peter?* nicht auf die Vorfrage, sondern auf die Antwort beziehen: *Kommt Hans heute? Das weiss ich doch nicht.* Die Expansionserwartung im Sinn von *Vielleicht weiss das Peter?* wäre in dieser Situation wahrscheinlicher.

Die Interpretation von Ellipsen ist also in einem stärkeren und grundsätzlicheren Sinn vom Kontext und der Situation abhängig als die Interpretation von sogenannt „vollständigen“ Äusserungen. Je nach Abhängigkeit teilt man die Ellipsen in zwei grobe Klassen ein⁸:

- In **syntaktische Ellipsen**, zu deren Interpretation das sprachliche Umfeld ausreicht. Von ‚syntaktisch kontrollierten‘ Ellipsen spricht man, wenn das grammatische Umfeld die Interpretation determiniert.
- In **situative Ellipsen**, zu deren Interpretation Wissen über den nicht-sprachlichen Kontext – spricht die Situation und die Welt – benützt wird.

⁸ Vergleiche (Klein 1991: 765) sowie (Schwabe 1994:1ff.).

2 Kommunikation und Ellipsen

2.1 Natürlich-sprachliche Mensch-Maschine-Kommunikation (MMK)

Erst wenn man – etwa wie Rath(1979) – zwischenmenschliche Dialoge in Alltagsgesprächen auf Verkürzungen hin anschaut, fällt auf, wie die Dialogbeiträge verkürzt, verstümmelt und ineinander verwoben verwendet sind.

Auch bei informationsabfragenden Dialogen gibt es viele Abweichungen von Sprachnormen. Für die Modellierung einer deutschsprachigen Dialogkomponente im EVAR wurde ein Korpus aus solchen zwischenmenschlichen Dialogen erstellt: Als auffallendste Abweichungen von der Standardsprache wertet Mast(1993: 48f.) die Änderungen von syntaktischen Konstruktionen mitten im Satz, Unterbrechungen und Wiederholungen.

Beim Entwurf von Frage-Antwort-Systemen hat man verschiedentlich Erhebungen zur natürlich-sprachlichen Schnittstellengestaltung gemacht:

„Natural language interfaces can be built only when the functional and linguistic demands of the user can be determined.“ (Jönsson 1993: 29)

Eine beliebte Form, um zu den nötigen Daten zu kommen, besteht in sogenannten *Wizard-of-Oz* oder *hidden-operator* Experimenten, bei denen Menschen Leistungen künftiger natürlich-sprachlicher Zugangssysteme simulieren. Jönsson (1993) kommt aufgrund seiner so erhobenen, empirischen Untersuchungen zum Schluss, dass sich die Benutzenden bei der Mensch-Maschine-Kommunikation in ihrer Sprache bzw. Sprachverhalten einschränken und somit viele komplizierte Merkmale zwischenmenschlicher Dialoge nicht auftreten. Die Benutzenden adaptieren ihr sprachliches Verhalten an die vermuteten oder erlebten Beschränkungen des Systems und passen sich kooperativ den möglichen Kommunikationsmustern an.

In (Krause 1982) sind *Wizard-of-Oz* Experimente für deutschsprachige Dialoge mit mündlicher und schriftlicher Interaktion beschrieben, die im Zusammenhang vom DICOS System zur Bahn- und Bibliotheksauskunft geführt worden sind. Ein Schwerpunkt bestand dabei im Vergleich von zwischenmenschlicher Kommunikation mit Mensch-Maschinen-Kommunikation bei unterschiedlich restringierten Sprachmöglichkeiten. Die wichtigsten Ergebnisse: Einfache, direkte und modellhaft

verwendete Fragemuster sind typisch für ein Sprachregister, das *computer talk*⁹ genannt wird. Bezüglich elliptischen Fragen beobachtet Krause (1982: 145ff.), dass ihr Gebrauch unabhängig von mündlicher oder schriftlicher Interaktionsformen ist.¹⁰ Sie nehmen bei der Mensch-Maschinen-Kommunikation stark zu, sobald nicht-restringierte Sprache zugelassen ist.

Küpper (1992: 32) schliesst aus einer vergleichenden Untersuchung deutschsprachiger Frage-Antwort-Systeme:

„Die Kompaktheit natürlicher Dialoge beruht z.T. auf der Möglichkeit, Sätze elliptisch zu verkürzen, wenn der Sinn aus dem Kontext zu erschliessen ist. Für natürlich-sprachliche Zugangssysteme ist diese Möglichkeit interessant, z.B. um knappe spezifizierende Nachfragen stellen zu können. Auch bei Systemantworten ist die Möglichkeit, Ellipsen generieren und damit überflüssige Textteile auslassen zu können, ein wichtiges Mittel, um die Aufmerksamkeit des Benutzers auf die relevanten Information zu lenken.“

Lewin (1995) erwähnt für den *Wizard-of-Oz*-Korpus, welcher für ein Dialogsystem zur Routenplanung erstellt wurde, bei bestimmten Systemfragen elliptische Benutzerantworten bis über 90%. Es hat sich gezeigt, dass auch bei vollständigen Entscheidungsfragen wie *Do you want the quickest journey or the shortest journey?* die meisten Benutzer in Form einer Nominal-Ellipse geantwortet haben: *quickest, the quickest, the quickest route*. Um eine solche Antwort zu verstehen muss zuerst die Nominal-Ellipse aufgelöst werden, dann noch die Verbal-Ellipse; bei blosser *quickest* muss zusätzlich der Artikel rekonstruiert werden. Was für Menschen konzise und einfache Kommunikation ist, stellt einige Ansprüche an pragmatisch ansprechende Sprachsysteme.

Dem *passage-retrieval*, wie es im UIS konzipiert ist, entspricht im engeren Sinn keine gängige zwischenmenschliche Dialogart. Es verhält sich ähnlich wie ein Abfragesystem:

„Abfragesysteme sind dadurch charakterisiert, dass sie eine natürlichsprachliche Anfrage in eine formalsprachliche Datenbankabfrage übersetzen und dass der Benutzer die Antwort direkt vom Datenbanksystem erhält.“ (Küpper, Rösner et al. 1992: 54)

Die formalsprachliche Datenbankabfrage entspricht der *PROLOG-query*, welche aus der Logischen Form der natürlich-sprachlichen Frage berechnet wird. Die Datenbank besteht aus den Logischen Formen der Sätze des Informationstextes bzw.

⁹ Der Ausdruck *computer talk* wurde bewusst in Analogie zu Sprachregistern wie *baby talk* gebildet. Krause (1982: 157) geht davon aus, dass sich mit der Weiterentwicklung der Informationssysteme und Schnittstellen auch der *computer talk* weiterentwickelt.

¹⁰ Leider differenziert Krause nicht zwischen verkürzten, fragmentarischen Äusserungen und ‚regulären‘ Ellipsen. Insofern sind seine Ergebnisse bezüglich Frage-Ellipsen mit Vorsicht zu geniessen.

der Verknüpfungen der Logischen Formen mit den Sätzen. Die Systemantwort ist ein Textausschnitt, der die zur Beantwortung der Frage relevanten Informationen enthält.

Die Interaktionsform des UIS lässt sich allenfalls mit einem Informationsschalter vergleichen, wo man schriftliche Fragen stellen darf, und dann eine Broschüre zurückbekommt, in der die entscheidenden Stellen mit Leuchtstift markiert sind.

Inwiefern im UIS Nachfrage-Ellipsen von den Benutzenden überhaupt verwendet werden, ist eine empirische Frage. Sie hängt von mehreren Faktoren ab: Die Qualität der Beantwortung spielt eine wichtige Rolle. Dies hängt u.a. damit zusammen, wie gut das UIS auf die jeweilige Domäne eingestellt ist. Die Quantität der Beantwortung ist ebenfalls ein gewichtiger Faktor. Bei mono-modalen Schnittstellen, wie etwa bei 'mündlicher' Telefonkommunikation, sollten die Antworten knapp sein. Elliptische Vergewisserungsfragen sind dann für eine einigermaßen natürliche Interaktionsform unerlässlich. Multi-modalen Schnittstellen, wie sie etwa auf den Seiten des WWWs gestaltet werden können, erlauben es, viel Antwortinformation via Hyperlinks in kompakter Form zurückzugeben. Graphische Schnittstellen bringen deshalb eine andere Ergonomie als die älteren befehlszeilen-orientierten Systeme. So kann das Überschreiben bestimmter Wörter in einem Eingabefeld weniger aufwendig sein als das neu Schreiben einer elliptischen Nachfrage.

2.2 Arten von Ellipsen

2.2.1 Syntaktische Struktur von Frage- und Nachfrage-Ellipsen

Der Ausdruck ‚Frage-Ellipse‘ besagt eigentlich nur, dass eine elliptische Frage vorliegt. ‚Nachfrage-Ellipse‘ (*follow-on question*) ist der spezifischere Begriff, der den Vervollständigungsbezug auf eine Vorfrage bedingt. Über die syntaktische Struktur der Frage und die elliptische Konstruktionsklassifizierung ist damit noch nichts gesagt. In den folgenden Abschnitten soll eine Zusammenstellung zu den wichtigsten sogenannten „syntaktischen Ellipsen“ gegeben werden, die in einem möglichen Zusammenhang mit Frage-Ellipse-Phänomenen stehen. Ich stütze mich dabei vor allem auf den Übersichtsartikel von Klein(1991).

Wie schon angesprochen, gibt es im wesentlichen zwei Ansätze, um Ellipsen zu analysieren: Die reduktionistische Sicht, die in jeder Ellipse einen strukturell vollständigen Satz sucht und meist auch findet. Neben diesem *main-stream* gibt es noch die anti-reduktionistische Position, die das, was die reduktionistische Fraktion als „elliptisch“ apostrophiert, einigermaßen gleichgestellt neben dem „vollständigen“ Satz akzeptiert. Aus der Sicht der Syntax stellen sich die Reduktionisten die an-

spruchsvollere Aufgabe, da sie für Ellipsen eine differenziertere und reichere Struktur bestimmen müssen. Der Aufwand macht sich allerdings bezahlt, indem dieselben Regeln sowohl für Ellipsen wie auch für die entsprechenden vollständigen Formen verwendet werden können. Die Nicht-Reduktionisten dagegen verlagern den Aufwand, den die Behandlung und Interpretation von Ellipsen erfordert, aus der Satz-Syntax in andere Verarbeitungs- oder Repräsentationsstufen.

Wer sich für die syntaktischen Struktur von Ellipsen interessiert, dem bietet die reduktionistische Sichtweise das interessantere Paradigma. Deshalb erstaunt es nicht, dass die syntaktischen Analyseansätze stark reduktionistisch geprägt sind. Die reduktionistische Perspektive erleichtert zudem die Klassifikation, da immer gefragt werden kann, was wurde ausgelassen bzw. warum kann etwas nicht ausgelassen werden.

Erschwerend für eine neutrale Zusammenstellung wirkt, dass die Analysen – wie sich Klein (1991: 770) ausdrückt – „stark theoriegebunden sind“. Das bedeutet, dass die verschiedenen Ansätze oft nur im engen Rahmen einer bestimmten Grammatiktheorie oder eher eines Stadiums derselben zu verstehen und auch funktional sind. Eine detaillierte Aufsicht hätte den Rahmen dieser Arbeit gesprengt, die wichtigsten Positionen sollten dennoch deutlich werden.

2.2.2 Koordinationsellipsen

Eine Koordinationsellipse liegt nach Klein (1991: 768) dann vor, wenn „der elliptische Ausdruck mit dem kontrollierenden [Ausdruck] innerhalb eines Satzes durch Koordination (im weitesten Sinne) verbunden“ ist. Prototypisch ist dabei ein Satz wie (a), aber auch ein Vergleichsatz wie (b) wird manchmal als Koordinationsellipse aufgefasst.¹¹

- a) In Stuttgart ~~regnet~~ es und in München regnet es.
- b) Fritz schwimmt schneller als Otto ~~schwimmt~~.

Im Fall (a) wäre „In Stuttgart“ der elliptische Ausdruck und „in München regnet es“ der kontrollierende, verbunden durch die Koordination „und“. Bei (b) wäre „Fritz schwimmt schneller“ der kontrollierende und „Otto“ der elliptische Ausdruck, wobei „als“ als koordinierendes Element aufgefasst wird. Aufgrund dieser Analyse, deren Plausibilität hier noch nicht zur Diskussion steht, kann man zwei Arten von Ellipsen unterscheiden:

¹¹ Im folgenden wird das als elliptisch betrachtete Material typographisch mit Durchstreichen ausgezeichnet.

- Eine **Rückwärtsellipse** liegt vor, wenn der kontrollierende Ausdruck *hinter* dem elliptischen Ausdruck steht.
- Eine **Vorwärtsellipse** liegt vor, wenn der kontrollierende Ausdruck *vor* dem elliptischen Ausdruck steht.

Im Satz (a) liegt demnach eine Rückwärtsellipse vor, in (b) dagegen eine Vorwärtsellipse.

Es ist entscheidend, dass man Vorwärts- und Rückwärtsellipsen auf einzelne Ausdrücke und nicht auf Sätze bezieht. In einem Satz wie (c) erscheint gleichzeitig eine Vorwärts- und eine Rückwärtsellipse.

c) Otto hat vier Bier ~~getrunken~~ und ~~Otto~~ hat drei Glas Wein getrunken.

„getrunken“ kontrolliert nach hinten und „Otto hat“ nach vorne.

2.2.3 Koordinationsreduktion

Das Konzept der Koordinationsreduktion ist eng verbunden mit der Generativen Grammatik. Ein erstes Beispiel findet man bei Chomsky in einer Regel zur Einführung der Konjunktion *and*:

„If S1 and S2 are grammatical sentences, and S1 differs from S2 only in that X appears in S1 where Y appears in S2 (i.e., $S1 \rightarrow \dots X \dots$ and $S2 \rightarrow \dots Y \dots$), and X and Y are constituents of the same type in S1 and S2 respectively, then S3 is a sentence, where S3 is the result of replacing X by X *and* Y in S1 (i.e., $S3 \rightarrow \dots X + \textit{and} + Y \dots$).“ (Chomsky 1957: 36)

Der entscheidende Punkt dieser Regel liegt darin, dass die Koordination beliebiger Konstituenten auf der Ebene des Satzes definiert wird, d.h. durch Kombination zweier Sätze. In der Logik des Theorieansatzes von 1957 ist die Regel als induktive Definition gehalten, welche die Menge der zulässigen Sätze einer Sprache charakterisiert. Die Sätze S1 und S2 brauchen effektiv gar nicht syntaktisch koordiniert vorzuliegen.

Die Weiterentwicklung der Generativen Grammatik zur Transformationsgrammatik in den 60-er Jahren und insbesondere die Arbeit von Ross (1970) führten zum Begriff der Koordinationsreduktion (*coordination reduction*). Er besagt grob, dass bei zwei koordinierten Sätzen, die bis auf eine Konstituente identisch sind, das identische Material *getilgt* werden kann. Hankamer(1973) enthält den Versuch, diese Tilgung explizit und detailliert im Rahmen der zeitgenössischen Transformationsgrammatik fürs Englische zu regeln. Wie Klein (1991: 784f.) aufzeigt, lassen sich diese Regeln äusserst schlecht aufs Deutsche übertragen. Der Ansatz ist zu sprachspezifisch und letztlich zu theoriespezifisch, um brauchbar zu sein.

Schliesslich kann man sich wie Klein (1991: 771) fragen, ob es wirklich sinnvoll ist, wie von Chomsky initiiert, alle Koordinationen auf der Ebene des Satzes einzuführen. Ein Satz wie

Otto hat vier Bier und drei Glas Wein getrunken.

lässt sich ganz einfach analysieren, wenn man davon ausgeht, dass es koordinierte Nominalphrasen wie „vier Bier und drei Glas Wein“ gibt. Versucht man denselben Satz dagegen mittels iterativer Koordinationsreduktion aus

Otto hat vier Bier ~~getrunken~~ und ~~Otto~~ hat drei Glas Wein getrunken.

herzuleiten, verkomplizieren sich die Dinge wesentlich.

Ein weiteres Argument gegen Koordinationsreduktion liefern sogenannte „distributive“ Prädikate. Wenn man davon ausgeht, dass syntaktische Umformungen bedeutungserhaltend sein müssen, darf ein Satz wie

Männer und Frauen füllten die Kirche zur Hälfte.

auf keinen Fall aus folgender Satzkoordination hergeleitet werden:

Männer füllten die Kirche zur Hälfte und Frauen füllten die Kirche zur Hälfte.

Beide Sätze habe unterschiedliche Wahrheitsbedingungen. Koordination muss demnach auf Phrasenebene direkt einführbar sein. Für „und“ kann das beispielsweise mit folgender Phrasenstrukturregel geschehen:

$XP \rightarrow XP$ „und“ XP

Mit Regeln dieser Art lösen sich dann einige Klassen ‘vermeintlicher’ Ellipsenkonstruktionen in Luft auf.

Es gibt allerdings noch genügend Fälle, wo die phrasale Koordination nicht ausreicht. Dazu gehört prominent das *Gapping*, welches im nächsten Abschnitt separat diskutiert wird. Komparativkonstruktionen, die im vorhergehenden Abschnitt ange-tippt wurden, lassen sich damit ebenfalls nicht erschlagen. Auf jeden Fall elliptisch ist folgender Satz:

Peter stand VOR ~~dem Sofa~~ und Maria lag unter dem Sofa.

Die Grossbuchstaben markieren, dass das letzte nicht ausgelassene Element eine besondere Betonung braucht. Eine phrasale Lösung ist ausgeschlossen, da „Peter stand vor“ im obigen Sinn nie als Phrase klassifiziert werden kann. Solche Rückwärtsellipsen brauchen nicht einmal vor Wortgrenzen haltzumachen:

Ihm wurde vom Hinauf-~~sehen~~ und vom Hinunterschauen schlecht.¹²

¹² Zumindest nach der alten Rechtschreibung. „Er wurde sowohl vom Ein- wie vom Ausgang her bedrängt“ ist eine passendes Beispiel nach neuer Rechtschreibung. – Solche morphologischen Ellipsen bieten bei mehrfachen Kompositagrenzen Disambiguierungsprobleme wie etwa in „Regierungs- und Gemeinderatswahlen“.

Interessanterweise funktionieren diese Beispiele nicht als Vorwärtsellipsen:

*Peter stand vor dem Sofa und Maria lag unter ~~dem~~ Sofa.

*Ihm wurde vom Hinaufschauen und vom Hinunter-~~sehen~~ schlecht.

Ein weiterer systematischer Unterschied von Vorwärts- und Rückwärtsellipsen liegt im anaphorischen Verhalten. In Satz (a) mit der Rückwärtsellipse kann „jemand“ unterschiedliche Personen betreffen, während es in (b) dieselbe Person sein muss. Das sichtbare „Jemand“ in (b) kontrolliert die Referenzidentität des elliptischen „jemand“.

a) Um vier Uhr kam ~~jemand~~_i und um fünf Uhr ging jemand_i.

b) Jemand_i kam um vier Uhr und ~~jemand~~_i ging um fünf Uhr.

Diese Kontrolle wirkt sogar in satzübergreifenden Frage-Paaren, wie (b') zeigt. Der parallele Effekt (a') lässt sich bei Rückwärtsellipsen nicht bilden, da die elliptische Frage ungrammatisch ist.

a') F: *Um vier Uhr kam ~~jemand~~? A: Ja. F: Und um fünf Uhr ging jemand.

b') F: Jemand_i kam um vier Uhr? A: Ja. F: Und ~~jemand~~_i ging um fünf Uhr?

Im Gegensatz zu Vorwärtsellipsen sind Rückwärtsellipsen in satzübergreifenden Kontexten generell ausgeschlossen. Satzübergreifende elliptische Frage-Paare mit koordinativem Anschluss wie in (b') sollten aufgrund der aufgezeigten Regularitäten als koordinationsreduzierte Vorwärtsellipsen analysiert werden.

Als Fazit zum Thema „Koordinationsellipsen und Koordinationsreduktion“ möchte ich festhalten: Koordination von gleichwertigen Phrasen geschieht besser ohne Rückgriff auf Ellipsenkonstruktionen.¹³ Wo Elemente ungleicher grammatischer Kategorien koordiniert werden, braucht es elliptische Mechanismen. Allerdings müssen Phänomene wie Vorwärts- und Rückwärtsellipsen differenziert betrachtet werden.

2.2.4 Gapping und Stripping

Mit *Gapping* („Lochen“ oder „Lücken schaffen“) bezeichnet man seit Ross(1970) elliptische Koordinationskonstruktionen, bei denen das Verb ausgelassen ist. Ein typisches Beispiel ist (a), wo das „ausgestanzte“ Loch zwischen Subjekt und Objekt gut zur Geltung kommt. *Gapping* wird auch bei rechtsperipheren Rückwärtsellipsen wie in (b) angesetzt.

a) Karl besorgte das Bier und Otto ~~besorgte~~ den Wein.

b) weil Karl das Bier ~~besorgte~~ und Otto den Wein besorgte

¹³ Die Grammatik, die im UIS verwendet wird, erlaubt die direkte Koordination von gleichwertigen Phrasen wie NPs und PPs.

Wie schon bei den oben untersuchten Koordinationsellipsen, unterscheiden sich Vorwärts- und Rückwärtsellipsen auch beim *Gapping*. So verlangen Rückwärtsellipsen bezüglich morphologischer Ausprägung Identität, während Vorwärtsellipsen problemlos mit morphosyntaktischen Varianten funktionieren. In (c) ist „trinke“ mit „trinkst“ unverträglich, während Satz (d) völlig grammatisch ist.

c) *weil ich BIER ~~trinke~~ und du Wein trinkst

d) weil ich Bier trinke und du Wein ~~trinkst~~

Dieser und die im letzten Abschnitt erwähnten Unterschiede erhärten den Verdacht, dass bei Vorwärts- und Rückwärtsellipsen trotz Gleichheit der ausgelassenen Phrasen ganz verschiedene Mechanismen am Werk sind.¹⁴ Da Rückwärtsellipsen nie satzübergreifend auftreten, möchte ich sie nun an dieser Stelle ausblenden. Es soll hier auch keine genaue Analyse der englischen *Gapping*-Phänomene gemacht werden, die doch eingeschränkter und geregelter scheinen als im Deutschen.¹⁵ Ich möchte den Blick auf den Zusammenhang zwischen vorwärtselliptischem *Gapping* und satzübergreifenden Ellipsen richten, ganz im Sinn von Kindt(1985: 167):

„Allgemein kann man nachweisen, daß das im Fall von Frage-Antwort-Sequenzen unter der Rubrik ‘Ellipsen’ eingeordnete Phänomen identisch mit dem Phänomen des Gapping ist, mit dem sich die Satzgrammatiker in jüngster Vergangenheit sogar sehr intensiv, wenn auch nicht mit letztem Erfolg auseinandergesetzt haben.“

Die folgenden Beispielsatz-Paare variieren das Zitat von Kindt. Es spielt keine Rolle, ob das *Gapping* satzintern, in Antwortellipsen (‘) oder in Nachfrage-Ellipsen (‘’) auftaucht. Alle Konstruktionen erlauben auch Objekt-Topikalisierung (b) oder ein Fragewort im elliptischen Teil (c).

a) Karl besorgt das Bier und Otto ~~besorgt~~ den Wein.

a’) F: Besorgt Karl das Bier? A: Ja, und Otto ~~besorgt~~ den Wein.¹⁶

a’’) F: Besorgt Karl das Bier? A: Ja. F: Und Otto ~~besorgt~~ den Wein?

b) Karl besorgt das Bier, aber den Wein ~~besorgt~~ Otto.

b’) F: Besorgt Karl das Bier? A: Ja, aber den Wein ~~besorgt~~ Otto.

¹⁴ Ross (1970) wollte das *Gapping* mit universalgrammatischen Beschränkungen regeln, wonach Rückwärts- und Vorwärtsellipsen über die Links- bzw. Rechtsperipherität des Verbs in der Verbalphrase getriggert werden. Gemäss Klein (1991: 784) liess sich weder die Originalthese von Ross noch spätere Verfeinerungen derselben halten.

¹⁵ In Chao(1988) wird versucht – im Rahmen des späteren *Government and Binding* Ansatzes der Generativen Grammatik – *Gapping* und andere syntaktisch kontrollierte Ellipsenformen in einer einheitlichen Theorie zu fassen. Die Idee von Chao, mit „defektiven“ Phrasenstrukturregeln, leeren Kategorien und unterschiedlichen Lizenzierungsbedingungen zu operieren, ist zwar attraktiv, nach Klein (1991: 787) letztlich zu ungenau und für mich leider unbrauchbar, da zu theoriegebunden.

¹⁶ Ich habe das elliptische Prädikat „besorgt“ in Verb-Zweitstellung gestellt, obwohl es genauso gut parallel zur Vorfrage in Verb-Erststellung ‘erscheinen’ könnte.

- b'') *F*: Besorgt Karl das Bier? *A*: Ja. *F*: Aber den Wein ~~besorgt~~ Otto?
 c) Karl besorgt das Bier und wer ~~besorgt~~ den Wein.
 c') *F*: Besorgt Karl das Bier? *A*: Ja, und wer ~~besorgt~~ den Wein?
 c'') *F*: Besorgt Karl das Bier? *A*: Ja. *F*: Und wer ~~besorgt~~ den Wein?

Im engen Sinn verwendet man den Begriff *Gapping*, wenn das Verb ausfällt. Von *Stripping* spricht man, wenn nebst Verb noch weitere Konstituenten ausfallen.¹⁷ Die Aussage-, Frage-Antwort- und Frage-Nachfrage-Varianten sind dann nicht mehr unbedingt gleichwertig. So sind die Sätze (e) und (e') bei normaler Betonung nicht mehr akzeptabel. Satz (e'') ist möglich, vermutlich wegen des Frageakzents auf „Otto“:

- e) *Karl hat Bier besorgt und Otto ~~hat Bier besorgt~~.
 e') **F*: Hat Karl Bier besorgt? *A*: Ja, und Otto ~~hat Bier besorgt~~.
 e'') *F*: Karl hat Bier besorgt? *A*: Ja. *F*: Und Otto ~~hat Bier besorgt~~?

Damit das *Stripping* wieder in allen Spielarten funktioniert, braucht es eine kontrastierende Aussage wie in (f) oder ein beschwerendes Element wie das nachgestellte „auch“ in (g).

- f) Karl hat kein Bier besorgt, aber Otto ~~hat Bier besorgt~~.
 f') *F*: Hat Karl kein Bier besorgt? *A*: Nein, aber Otto ~~hat Bier besorgt~~.
 f'') *F*: Karl hat kein Bier besorgt? *A*: Nein. *F*: Aber Otto ~~hat Bier besorgt~~?
 g) Karl hat Bier besorgt und Otto ~~hat auch Bier besorgt~~.
 g') *F*: Hat Karl Bier besorgt? *A*: Ja, und Otto ~~hat auch Bier besorgt~~.
 g'') *F*: Karl hat Bier besorgt? *A*: Ja. *F*: Und Otto auch ~~hat Bier besorgt~~?

Stripping oder *Gapping* ist im Deutschen fast uneingeschränkt möglich. Wenn erst einmal das finite Prädikatswort ausgelassen ist, können beliebige Konstituenten eines Satzes verschwinden. Seien es Subjekt (a), direktes Objekt (b), indirektes Objekt (c) oder Satzmodifikatoren (d).

- a) Hans gab seiner Frau Wein und ~~Hans gab~~ seinem Sohn Bier.
 b) Hans gab seiner Frau Wein und Otto ~~gab~~ seinem Sohn ~~Wein~~.
 c) Hans gab seiner Frau Wein und Otto ~~gab seiner Frau~~ Bier.
 d) Hans gab seiner Frau letztes Jahr Wein und Otto ~~gab~~ seinem Sohn ~~letztes Jahr~~ Bier.

Eine andere Frage ist, welche Elemente sichtbar zurückbleiben dürfen in einer elliptischen Teiläußerung. Eine erste Antwort fällt leicht, wenn man folgendes Beispiel betrachtet:

- a') *Hans gab seiner Frau Wein und Hans ~~gab~~ seinem Sohn Bier.

¹⁷ In der neueren angelsächsischen Literatur wie in (Lappin Shih 1996) spricht man anstelle von *stripping* auch von *bare ellipsis*.

Die explizite Erwähnung von „Hans“ in (a') macht den Satz inakzeptabel.¹⁸ In der elliptischen Teiläusserung dürfen nur noch die Konstituenten erwähnt werden, die dem Teilsatz zu einer neuen Bedeutung verhelfen. Personalpronomen erfüllen diese Bedingung nicht:

*Hans_i hat Wein getrunken und er_i hat auch Bier ~~getrunken~~.

Obwohl „Hans“ und „er“ unterschiedliche Formen sind, bedeuten sie in diesem Zusammenhang dasselbe. Gerade umgekehrt liegt der Fall bei deiktischen Elementen: Wenn sich dieselbe deiktische Wortform in einer Situation auf zwei unterschiedliche Referenten bezieht, darf sie im elliptischen Teil nicht ausgelassen werden. Im folgendem Beispiel wird „aus diesem Glas“ zwar wiederholt, dies ist jedoch korrekt und notwendig, wenn damit auf zwei unterschiedliche Gläser referiert wird:

Hans hat aus diesem Glas Wein getrunken und Hans hat aus diesem Glas Bier ~~getrunken~~.

Einen Sonderfall scheinen Reflexivpronomen darzustellen. Im folgenden Beispiel bezeichnet „sich“ zwei unterschiedliche Personen, trotzdem kann das zweite „sich“ ausgelassen werden¹⁹:

Hans_i hat sich_i im Fernsehen gesehen, aber Otto_k hat sich_k nicht ~~im Fernsehen~~ gesehen.

Fragewörter dürfen ebenfalls wiederholt werden, da sie ja nicht etwas Fixiertes bezeichnen:

F: Wer hat den Wein getrunken? A: ... F: Und wer das Bier?

Was genau unter „unterschiedlicher Bedeutung“ verstanden wird, haben verschiedene Autoren unterschiedlich ausbuchstabiert. Für Klein(1991) hängt das zusammen mit der Topik-Fokus-Struktur, was ich in Abschnitt 2.3 etwas explizieren werde. Sag (1976) hat in bezug auf Verbalphrasen-Ellipsen einen Ansatz formuliert, der Bedingungen möglicher Ellipsen auf der Ebene der „Logischen Form“ spezifiziert. Ich werde die Idee im Abschnitt 2.2.5 ansprechen.

Zunächst möchte ich noch Eigenheiten der Satznegation diskutieren. Beim *Gapping* und *Stripping* müssen normalerweise Satzadjektive, Satzpartikel oder Adverbien aus dem kontrollierenden Satz vom elliptischen Satz übernommen werden. Satz (a) bedeutet darum dasselbe wie (a') und nicht wie (a'').

a) Otto trinkt gerne Wein und Hans Bier.

a') Otto trinkt gerne Wein und Hans trinkt gerne Bier.

a'') Otto trinkt gerne Wein und Hans trinkt Bier.

¹⁸ Ob in (a') die Inakzeptabilität syntaktisch oder eher semantisch zu begründen ist, bleibt offen.

¹⁹ Allerdings kann das zweite „sich“ auch sichtbar bleiben, was den vorgängig formulierten Regeln entsprechen würde. – Der Grammatikduden (Drosdowski 1984: §1143) verbietet im Standarddeutsch die Auslassung des Reflexivpronomens.

Bei Satznegation und gekoppelt mit „aber“ verhält es sich anders. Der Satz (b) bedeutet dasselbe wie (b') und nicht wie (b'').

b) Hans trinkt nicht gerne Wein, aber Otto.

b') Hans trinkt nicht gerne Wein, aber Otto trinkt gerne Wein.

b'') Hans trinkt nicht gerne Wein, aber Otto trinkt nicht gerne Wein.

Der gleiche Effekt wirkt auch, wenn die Satznegation verschmolzen ist.²⁰ Im Fall (c) mit dem Artikel „ein“ und in (d) mit so etwas wie „manchmal“²¹.

c) Hans trinkt keinen Wein, aber Otto ~~trinkt Wein~~.

d) Hans trinkt nie Wein, aber Otto ~~trinkt manchmal Wein~~.

Ein Erklärungsansatz für diese Effekte wäre, dass der koordinierende Anschluss mit „aber“ hier aus zwei Teilen besteht, nämlich „...nicht..., aber...“. Ähnlich wie das zweigliedrige „nicht nur..., sondern auch...“. Allerdings kann das „nicht“ auch im mit „aber“ koordinierten Teil auftauchen (e):

e) Otto trinkt gerne Wein, aber Hans ~~trinkt nicht gerne Wein~~.

Ein anderer Ansatz wäre, dass die Satznegation sowieso ein spezielles Phänomen ist, nicht mit anderen Satzmodifikatoren vergleichbar. Dafür spricht, dass sich „nicht“ auch bei „und“-Koordination zwiespältig verhält. Satz (f) ist nach meinem Empfinden nicht besonders gut. Im elliptischen Teilsatz (f') ist das „nicht“ nicht zwingend mitverstanden. Interessanterweise stellt ein parallel dazu konstruiertes *Stripping* zusammen mit einem expliziten „ebenfalls“ wie in (g) keine derartigen Probleme.

f) ?Karl trinkt das Bier nicht und Otto den Wein.

f') Karl trinkt das Bier nicht und Otto ~~trinkt~~ den Wein ?~~nicht~~.

g) Karl trinkt das Bier nicht und Otto ~~trinkt das Bier~~ ebenfalls nicht.

Um die Sache noch etwas zu komplizieren, verhält sich ein parallel zu (f) konstruiertes *Gapping* mit einem Hilfsverb – siehe (h) – nochmals anders. Als Begründung für diesen Effekt könnte man allenfalls die Wortstellung ins Spiel bringen.

h) Karl hat das Bier nicht getrunken und Otto ~~hat~~ den Wein ~~nicht getrunken~~.

Bei satzübergreifenden Frage-Ellipsen ist die Lage nicht viel besser. Je nach Antwort scheint mir für (i) die Lesart (i') und für (j) die Lesart (j') wahrscheinlicher – sofern man die Antwort in (j) als „Nein, Karl besorgt das Bier nicht“ auffasst. Eine mögliche Erklärung dafür wäre, dass sich die Nachfrage weniger auf die Vorfrage bezieht, sondern auf die in der Antwort zum Ausdruck gebrachten Aussage.

i) F: Besorgt Karl das Bier nicht? A: Doch. F: Und den Wein?

²⁰ Vergleiche zur Satznegation etwa den Grammatikduden (Drosdowski 1984: §1148ff.).

²¹ Gedacht als Lexikalisierung für das Gegenteil von „nie“.

i') *F*: Besorgt Karl das Bier nicht? *A*: Doch. *F*: Und ~~besorgt Karl~~ den Wein?

j) *F*: Besorgt Karl das Bier nicht? *A*: Nein. *F*: Und den Wein?

j') *F*: Besorgt Karl das Bier nicht? *A*: Nein. *F*: Und ~~besorgt Karl~~ den Wein ~~nicht~~?

Die Unklarheiten, welche die Negation im Zusammenhang mit koordinierenden Ellipsen stellt, müssen offenbleiben. Die Sprechenden haben es jedoch immer in der Hand, ihre Intentionen durch ein „auch nicht“, „schon“ oder „doch“ Partikel explizit zu machen.²²

2.2.5 VP-Ellipse

Die Verbalphrasen-Ellipse (*verbal phrase ellipsis*) ist besonders im Englischen verbreitet und wird deshalb in der angelsächsischen Literatur häufig behandelt. „*do-support*“ in Frage und Antwort (a) und in koordinierenden Anschlüssen (b) sind im Standarddeutsch jedoch kaum machbar. Die Konstruktion funktioniert am ehesten, wenn sie kontrastierend mit einem Modalverb (c) oder Hilfsverb (d) verwendet wird:

a) *Q*: Did John buy a new car? *A*: Yes, he did ~~buy a new car~~.

b) John bought a new car, and Mary did ~~buy a new car~~, too.²³

c) Peter möchte seine Frau nicht anlügen, aber er muss ~~seine Frau anlügen~~.

d) *F*: Hat Peter seine Frau angelogen? *A*: Nein, aber Hans hat ~~seine Frau angelogen~~.

Nachfrage-Ellipsen sind bei VP-Ellipsen wie (e-f)) möglich, aber deutlich markierter als bei *Gapping* oder *Stripping*.

e) *F*: Hat Peter seine Frau angelogen? *A*: Ja. *F*: Und hat Hans auch ~~seine Frau angelogen~~?

f) *F*: Hat Peter seine Frau angelogen? *A*: Ja. *F*: Und hat Iris ihren Chef auch ~~angelogen~~?

Neben der koordinierenden VP-Ellipse gibt es im Englischen noch eine subordinierende Form, die oft mit ACE (*Antecedent Contained Ellipsis*) bezeichnet wird.²⁴ Eine Konstruktion wie in (g) ist allerdings im Deutschen nicht möglich.

g) John liked the girls who Mary did.

Ich möchte noch, wie versprochen, die Analyse der elliptischen Reflexivpronomen von Sag (1976) ansprechen.²⁵ Aus reduktionistischer Sichtweise ist bei der VP-El-

²² In Abschnitt 3.5.4 auf Seite 64 ist die Methode beschrieben, die ich im UIS für die Behandlung von Satznegation gewählt habe.

²³ Falls die Betonung nicht auf „bought“ liegt, kann das „did“ wegfallen. Dann liegt aber *stripping* vor und nicht mehr VP-Ellipse.

²⁴ Vergleiche (Lappin 1998). – ACE ist vor allem durch May (1985) in den Blickwinkel der Linguistik gekommen. So entsteht nach May im Satz

„John [_{VP} liked the girls who Mary did [_{VP}]]“

ein unendlicher Regress, falls die elliptische VP im Relativsatz mit der VP des Hauptsatzes identifiziert wird. Die Ellipse ist selbst Bestandteil der Struktur, auf die sie sich bezieht.

lipse wie beim *Gapping* keine wortwörtliche Identität zwischen kontrollierendem und elliptischen Element notwendig. Im Englischen kombiniert sich das bei Reflexivpronomen damit, dass keine referentielle Identität nötig ist:

John accused himself, and Mary did ~~accuse herself~~.

„himself“ und „herself“ referieren auf je verschiedene Personen. Für Sag ist nun die entscheidende Ebene, auf der kontrollierende und elliptische Elemente übereinstimmen müssen, die ‚Logische Form‘. Darin stellt er Prädikate als lambda-abstrahierte Terme dar. „Sich anklagen“ formalisiert er beispielsweise mit

$$\lambda x (x \text{ accuse } x)$$

Bei der Applikation der Lambda-Terme benutzt er ebenfalls eine ‚natürlich-sprachlichere‘ Schreibweise, die das Subjekt-Argument vor den Funktionsterm stellt. Obiger Satz wird deshalb als ‚Logische Form‘ etwa folgendermassen aussehen:

$$(\text{John}) \lambda x (x \text{ accuse } x) \text{ and } (\text{Mary}) \lambda y (y \text{ accuse } y)$$

Die VP-Ellipse ist nun genau darum zulässig, weil die Prädikatstruktur der Teilsätze auf der Ebene der logischen Form identisch sind – modulo Umbenennung der gebundenen Variablen x und y .

Diese Art der Ellipsenanalyse kann auch auf weitere Ellipsenkonstruktionen angewendet werden. Sie ist ein Versuch, die Mittel der typtheoretischen Semantik in der Syntaxtheorie nutzbar zu machen.²⁶ Im Abschnitt 3.3 über die CLE wird ein Verfahren erklärt, das für die Rekonstruktion elliptischer Sätze ähnliche Mittel verwendet.

2.2.6 Sluicing

Sluicing wurde als Begriff von Ross (1969) in einem Aufsatz mit dem bezeichnenden Titel „Guess who?“ eingeführt. Gemeint ist damit das Kappen eines untergeordneten Fragesatzes unmittelbar nach dem einleitenden Fragewort. Als Beispiele für *Sluicing* im Deutschen werden folgende Sätze betrachtet:

- a) Es findet eine Feier statt, aber ich weiss nicht, wann ~~eine Feier stattfindet~~.
- b) Jemand hat angerufen, aber ich weiss nicht, wer ~~angerufen hat~~.

²⁵ Vergleiche (Klein 1991: 786). – Sag(1976) hat mit seinem Ansatz via (Dalrymple Shieber et al. 1991) auf die Konzeption der *Core Language Engine* (Alshawi Carter et al. 1992b) gewirkt.

²⁶ Vergleiche Pinkal (1995: 456), der ein einfaches Beispiel für die typtheoretische Semantikkonstruktion von Koordinationsellipsen gibt. Die Bedeutung eines Satz wie „Peter arbeitet und sieht Maria“ wird durch folgende Formel ausgedrückt:

$$\lambda x (\text{arbeitet}(x) \wedge \text{sieht}(\text{Maria})(x)) (\text{Peter})$$

Dank der Lambda-Abstraktion lässt sich ein komplexes Prädikat „jmd. arbeitet und sieht Maria“ bilden, das mit „Peter“ als Subjekt zu einem Satz wird.

Im Satz (a) kommt mit „wann“ eine zusätzliche Konstituente zur Sprache. In (b) wird das Subjekt ersetzt. Klein(1991: 779) lockert die Bedingung, dass es ein Fragesatz sein muss, der gekappt wird, und betrachtet auch Fälle wie (c) als *Sluicing*.

- c) Er kann kein Englisch, obwohl er in die Schule gegangen ist, sondern weil er ~~in die Schule~~ gegangen ist.

Im Zusammenhang mit Nachfrage-Ellipsen sind Fälle wie (a-b) interessanter, weil diese Sätze ja eine Frage enthalten, wenn auch eine indirekt gestellte:

- a') F: Findet eine Feier statt? A: Ja. F: Weisst du, wann ~~eine Feier stattfindet~~?

Oberflächlich betrachtet mag *Sluicing* Ähnlichkeit mit *Stripping* aufweisen. Doch ein Versuch *stripping*-artige Sätze wie (d) zu bauen, scheitert:

- d) *Gerda heiratet ihren Mann im Mai, aber niemand kann mir sagen, wann Beatrice ~~ihren Mann~~ heiratet.

Beim *Sluicing* darf Material aus dem Hauptsatz im Nebensatz elliptisch verwendet werden. So etwas ist beim *Gapping* – sogar ohne Fragewort – schlecht möglich:

- e) *Peter heiratet Maria, aber niemand weiss, dass Werner Ulrike ~~heiratet~~.

2.2.7 N-Ellipse

N-Ellipsen sind ein verbreitetes Phänomen. Sie unterscheiden sich deutlich von den bisher behandelten Ellipsen, bei denen immer Phrasen ausgelassen wurde.²⁷ Bei N-Ellipsen verschwindet keine NP, PP oder VP, sondern der lexikalische Kopf einer Nominalphrase. Deutsch ist dabei freizügiger als beispielsweise Englisch, das in ähnlichen Fällen eine Pronominalisierung mit „one“ verlangt:

Hans hat ein blaues Auto und Priska hat ein rotes ~~Auto~~.

John has a blue car and Priscilla has a red one.

Weiter gibt es keine syntaktischen Restriktionen, wie sie etwa beim *Sluicing* bzw. *Gapping* auszumachen sind. Der kontrollierende Ausdruck kann ziemlich weit weg und verschachtelt sein:

Hans kaufte sich ein blaues Auto. Dazu hat ihn Priska überredet. Ich glaube, dass ihm ein rotes ~~Auto~~ besser gefallen hätte.

Klein (1991: 782) fragt deshalb, ob N-Ellipsen tatsächlich „syntaktisch kontrollierte“ Konstruktionen sind, da sie sich wie pronominale Anaphern verhalten. Aus einer NP darf zudem nicht in jedem Fall einfach der lexikalische Kern getilgt werden. Wie die zwei folgenden Beispiele zeigen, wird durch die N-Ellipse bei „ein“ starke Flektion notwendig.

- *Hans kaufte ein Auto, und Priska wollte auch ein ~~Auto~~ haben.

²⁷ Auch wenn letztlich die Vorstellungen darüber, was Phrasen sind, ganz unterschiedlich sind.

Hans kaufte ein Auto, und Priska wollte auch eines ??? haben.

Da N-Ellipsen frei sind, können sie selbstverständlich auch in Nachfragen auftauchen. Interessant sind dabei kombinierte Ellipsen, beispielsweise mit *Stripping* und N-Ellipse:

F: Kauft Hans ein Auto? *A:* Ja. *F:* ~~Kauft Hans~~ Ein rotes ~~Auto~~?

Die Auflösung solcher Ellipsen muss iterativ vollzogen werden. Beim Suchen nach einem passenden Antezedensausdruck steht nebst Kasus noch Genus als Auswahlkriterium zur Verfügung. Wenn anstelle von einem Auto von einem Wagen die Rede ist, verändert sich auch das grammatische Geschlecht der N-Ellipse, was an der Verwendung von „Einen“ deutlich wird:

F: Kauft Hans einen Wagen? *A:* Ja. *F:* ~~Kauft Hans~~ Einen roten ~~Wagen~~?

2.2.8 Satz-Ellipse

Als letzte syntaktische Ellipse möchte ich noch die Satz-Ellipse erwähnen. Klein(1991: 783) gibt folgendes Beispiel:

F: Kommst du? *A:* Ja, ~~ich komme~~, wenn ich Zeit habe.

Wer „ja“ als satzwertiges anaphorisches Element betrachtet, könnte behaupten, dass hier gar keine Ellipse vorkommt. Klein hält dagegen, dass „ich komme“ durchaus mit „ja“ zusammen wiederholt werden könnte. Das bessere Argument scheint mir, dass das „ja“ gar nicht aufzutauchen braucht. Eine satzübergreifende Variante, die das belegt:

F: Kommst du? *A:* Vielleicht. *F:* ~~Kommst du~~, Wenn du Zeit hast?

Eine passende Antwort auf die Nachfrage ist nur möglich, wenn letztere als „Kommst du, wenn du Zeit hast?“ verstanden wird. Insofern stellen Satz-Ellipsen für die Interpretation von Dialogen ein wichtiges Phänomen dar.

2.3 Kommunikationsfunktion

Ellipsen sind ein Phänomen, das in verschiedenen Formen in unterschiedlichsten Sprachen vorkommt. Chao(1988: 125ff.) enthält eine vergleichende Übersicht bezüglich möglicher Ellipsen. Vom Standpunkt der rationalen kooperativen Kommunikationsprinzipien, wie sie von Grice (1975) formuliert wurden, fungieren Ellipsen bestens, um die Maxime der Kürze und der minimalen Informationsquantität zu erfüllen. Allerdings bieten sie auch Probleme und verstossen gegen Konversationsmaximen, wenn durch den elliptischen Sprachgebrauch Mehrdeutigkeiten oder Unklarheiten entstehen. Ich möchte hier aber nicht weiter diese allgemeinen Funktionen untersuchen, sondern auf die Informationsverteilung innerhalb elliptischer

Konstruktionen zu sprechen kommen. Genauer: Inwiefern mithilfe der Topik-Fokus-Unterscheidung bezüglich Ellipsen etwas gewonnen werden kann.

Für Klein (1991: 790f.) sind Ellipsen stark gekoppelt mit Topik- und Fokus-Gliederung. Um das auszuführen, müssen zunächst die beiden Begriffe „Fokus“ und „Topik“ genauer bestimmt werden:

„Ich bezeichne die Menge von Alternativen, die zur Entscheidung ansteht, als Topik und jene, die davon spezifiziert wird, als Fokus einer Äusserung.“ (Klein 1991: 791)

Am Beispiel von Frage-Antwort-Paaren kann diese Definition am einfachsten nachvollzogen werden. Eine Frage wie (a) stellt eine Topik in den Raum, nämlich die Menge der Alternativen all jener Personen, welche möglicherweise einen Steinadler gekauft haben. Eine Antwort wie (b) spezifiziert dann eine Möglichkeit aus dieser Alternativenmenge.

- a) Wer hat gestern einen Steinadler gekauft?
- b) Mein Bruder ~~hat gestern einen Steinadler gekauft~~.

Topik-Fokus-Gliederungen von Aussagesätzen kann man erheben, indem man sich die möglichen Fragen konstruiert, auf die sie Antwort geben. Sätze können darum nach Äusserungsintention unterschiedliche Topik-Fokus-Gliederung aufweisen.

Welcher Bezug besteht nun zwischen Topik-Fokus-Gliederung und Ellipsen? Klein (1991: 791) formuliert ihn in der ‚Regel E‘ für Vorwärtsellipsen folgendermassen:

„Genau jene lexikalische Einheiten, die eine beibehaltene Topik ausdrücken, können p-reduziert werden.“

Mit „p-reduziert“ meint Klein „phonologisch reduziert“, da Ellipsen für ihn syntaktisch immer vollständig bleiben und Material nur auf der phonologischen Ebene ausgelassen wird. Die Regel tönt bestechend einfach und klassifiziert ein bis anhin vor allem syntaktisch betrachtetes Problem neu als ein Epiphänomen der Topik-Fokus-Gliederung. Direkten Nutzen wirft sie jedoch kaum ab, da Klein keine genaue Form der Repräsentation der Topik-Fokus-Gliederung gibt. So bleibt auch völlig unklar, was „beibehaltene Topik“ in bezug auf Nominalphrasen bedeutet. Die Frage, warum im folgenden Beispiel „sie“ und „ihn“ in der Nachfrage nicht ausgelassen werden dürfen, kann so nicht beantwortet werden.

F: Liebt er sie? *A:* Vielleicht. *F:* Und sie ihn?

Klein sieht den Konnex zwischen Ellipsen und Topik-Fokus-Gliederung zurecht nur bezüglich Topik. Zwischen Ellipsen und Fokus lässt sich keine Abhängigkeit feststellen. In (a) wechselt der *Fokus* auf ganz andere Weise als in (b).

F: Trinkt Otto Wein? *A:* Ja. *F:* Wann?

F: Trinkt Otto Wein? A: Ja. F: Und Bier?

Solange kein expliziter und systematischer Zusammenhang zwischen syntaktischen Strukturen und Topik-Fokus-Gliederung ausgemacht ist, gewinnt man durch die Umformulierung des Problems wenig. Für die Auflösung von Ellipsen bietet der Ansatz darum keine grosse Hilfe.

3 NLP und Ellipsenbehandlung

3.1 Ansätze zur Ellipsenbehandlung

Die Aufmerksamkeit der Computerlinguistik verdanken Ellipsen vor allem dem Trend zu Frage-Antwort-Systemen, der in den späten 70-er Jahren aufblühte. Küpper (1992: 54-57) hält die Behandlung von Dialogphänomenen für einen „Forschungsschwerpunkt der Verarbeitung natürlicher Sprache in den 80-Jahren“. Von den 7 deutschsprachigen Dialogsystemen, die er untersucht, können 5 mit Ellipsen umgehen.

Dieses Kapitel stellt unterschiedliche Ansätze vor, wie Ellipsen in NLP-Systemen (*natural language processing*) behandelt werden. Als Einstieg habe ich das Saarbrücker Dialogsystem ausgewählt. Damit lassen sich verschiedene Aspekte, die bei der Auflösung von Frage-Ellipsen entstehen, anschaulich darstellen und diskutieren. Es zeigt modellhaft, was durch eine gezielt Einschränkung des Anwendungsbereichs und der Sprache in einem Frage-Antwort-System möglich ist.

Danach folgt eine Beschreibung der Ellipsenbehandlung in der *Core Language Engine (CLE)*. Sie findet wie beim Saarbrücker Dialogsystem auf einer semantischen Ebene statt. Die Ziele und Ansprüche der *CLE* als ein langjähriges professionelles Projekt sind ehrgeiziger. Die *CLE* hat ihre allgemeine linguistische Fundiertheit gemäss Rayner(1996) auch durch die recht einfache Übertragung in andere Sprachen wie Schwedisch, Französisch und Spanisch erwiesen.

Lappin hat zusammen mit anderen Forschenden Algorithmen zur Ellipsenbehandlung auf der syntaktischen Ebene vorgeschlagen. Dieser Ansatz, der in einer HPSG-Implementierung für das Englische verwendet wird, geht ähnlich vor wie die Ellipsenbehandlung im UIS.

Schliesslich präsentiere ich noch Vorgehen und Eigenheiten der Auflösung von Frageellipsen im UIS. Ein syntaktischer Ansatz hat sich allein schon empfohlen, da zum Zeitpunkt der Implementation der Ellipsenauflösung kein Modul zur semantischen Verarbeitung der Satzstrukturen existierte. Strukturelle Mehrdeutigkeiten durch Ellipsen, was ja im Deutschen weit häufiger ist als im Englischen, erfordern für ihre Auflösung in die intendierte Form trotzdem gewisse semantisch-ontologische Information. So musste doch eine gewisse semantische Komponente vorgesehen werden.

3.2 Saarbrücker Dialogsystem

Das Saarbrücker Dialogsystem ist ein natürlich-sprachliches Frage-Antwort-System, das auf einer Datenbank über das Lehrangebot der Computerlinguistik und Phonetik aufsetzt.²⁸ Die Eingabe erfolgt schriftlich über die Tastatur. Die Systemausgaben werden in gesprochener Form via Lautsprecher gemacht, wobei Ganzwort-Sprachsynthese verwendet wird. Sowohl die Datenbank als auch die natürlich-sprachliche Schnittstelle sind in PROLOG auf Apple Macintosh implementiert.

Über das System wurde nichts veröffentlicht, der Zweck des Systems war in erster Linie ein funktionierendes Ganzes zu implementieren mit Ansätzen, wie sie in Lehrbüchern mit gutem Standard abgehandelt sind. So kann man vieles, was den semantischen Aufbau betrifft, in Pinkal(1995) finden.

Das Saarbrücker Dialogsystem eignet sich trotz einiger Einschränkungen sehr gut, gewisse Problemstellungen und Lösungsansätze bezüglich Ellipsenbehandlung aufzuzeigen. Als typischer Vertreter einer natürlich-sprachlichen Datenbankschnittstelle teilt es einige Eigenschaften mit seinem berühmten Vorfahren CHAT-80²⁹ von Warren (1982), da die Beschränkung auf eine kleine Domäne mit einfacher, klar strukturierter Ontologie eine darauf zugeschnittene und vereinfachte Sprache zulässt.

3.2.1 Benutzerinteraktion

Im Normalfall wechseln sich Fragen der Benutzenden mit den Antworten des Systems ab. Bei unbekanntem Wörtern gibt das System eine Fehlermeldung aus; wenn ein Wort als Tippfehlervariante eines bekannten Worts erkannt wird, macht das System einen Korrekturvorschlag. Wenn die syntaktische Analyse scheitert, wird zur Neuformulierung der Frage aufgefordert.

Ich kann Sie leider nicht verstehen, bitte formulieren Sie Ihre Frage neu.³⁰

Wenn die Syntaxanalyse und die Semantikkonstruktion gelingt, bei der Datenbankabfrage aber nichts herauskommt, antwortet das System mit:

Ich kann dazu keine Angaben machen.

²⁸ Studierende der Computerlinguistik der Universität Saarbrücken haben es 1993/94 in einem Projektseminar mit Schwerpunkt 'Semantische Verarbeitung' unter M. Pinkal entwickelt. – Das Dialogsystem wurde nicht via Internet öffentlich gemacht, da es zuletzt nicht als fertig und „toll“ genug betrachtet wurde, wie Projekt-Koordinatorin Julia E. Heine <heine@CoLi.Uni-SB.DE> mitteilte.

²⁹ Unter http://www.lpa.co.uk/ind_demo5.html findet man eine online-Version dieses Fossils im WWW.

³⁰ Alle folgenden Beispiele in Schreibmaschinenschrift stammen aus effektiv geführten Interaktionen meinerseits. Die gesprochenen Systemantworten wurden von mir verschriftlicht und sind jeweils kursiv ausgezeichnet.

Eine Sonderbehandlung bei der Beantwortung erfahren Präsuppositionsverletzungen von definiten Kennzeichnungen, die möglichst benutzerfreundlich zurückgemeldet werden. Wenn einer Kennzeichnung mehrere Objekte der Datenbank entsprechen, wird über die Verletzung mit einer Aufzählung aller möglichen Denotationen informiert:

Wann findet das Seminar statt?

Es gibt mehrere, nämlich das Seminar X, das Seminar Y,...

Bei „welche“-Fragen wird stillschweigend korrigiert:

Welche Vorlesung findet am Freitag statt?

Die Vorlesung X, die Vorlesung Y, ...

Das System versucht sich kooperativ zu verhalten, wenn bei einer Kennzeichnung die Existenzvoraussetzung verletzt ist. Beispiel:

Leitet Barry das Seminar Akustische Phonetik?

Meinen Sie vielleicht die Vorlesung Akustische Phonetik?

Ja.

*Ja.*³¹

Solche Rückfragen sind nur möglich, wenn der Gegenstandsbereich genügend strukturiert ist. Das könnte beispielsweise durch einen Thesaurus mit Informationen zu Hypernymie und Hyponymie oder durch eine dezidierte Bereichsontologie geschehen. Im Saarbrücker System werden allen Substantiven Gattungen zugeordnet, Gattungsfehler von Benutzenden sind somit leicht korrigierbar.

Die Verwendung von anaphorischen Bezügen mit Pronomen wird sowohl bei Anfragen und in den Antworten unterstützt.

Hält Professor Uszkoreit eine Vorlesung?

Ja.

Wann hält er sie?

Er hält sie am Mittwoch von 9 bis 11 Uhr.

3.2.2 Die Behandlung von Erweiterungsellipsen

Zunächst sollen einfache Erweiterungsellipsen betrachtet werden. Der folgende Dialog zeigt, wie eine Entscheidungsfrage durch ein simples „Wo?“ zu einer Konstituentenfrage erweitert werden kann. Man beachte, dass elliptische Konstituentenfragen immer elliptische Antwortellipsen auslösen.

Hält Professor Barry die Vorlesung Akustische Phonetik?

Ja.

Wo?

Im Phonetikstudio.

³¹ Entscheidungsfragen werden immer nur mit „Ja“ oder „Nein“ beantwortet. Eine Form von kooperativer Überbeantwortung war geplant, ist aber nicht realisiert worden.

Selbstverständlich können Erweiterungsellipsen auch zur Bildung von Entscheidungsfragen gebraucht werden.

Hält Professor Barry die Vorlesung Einführung in die Allgemeine Sprachwissenschaft?

Ja.

Im Phonetikstudio?

Nein.

Schwieriger wird die Sache, wenn mehr als eine elliptische Nachfrage gestellt wird.

a) Hält Professor Barry die Vorlesung Akustische Phonetik?

b) *Ja.*

c) Wo?

d) *Im Phonetikstudio.*

e) Wann?

f) *Am Dienstag von 14 bis 16 Uhr.*

Bei Frage (c) ist klar, dass die ausgetextete Formulierung folgendermassen lauten muss:

Wo hält Professor Barry die Vorlesung Akustische Phonetik?

Bei der Frage (e) gibt es allerdings zwei Möglichkeiten.³² Entweder wird die Antwort der vorgängigen Frage in die Nachfrage assimiliert:

e') Wann hält Professor Barry die Vorlesung Akustische Phonetik im Phonetikstudio?

oder die vorgängige Frage wird zur Doppelfrage erweitert:

e'') Wann hält Professor Barry wo die Vorlesung Akustische Phonetik?

Die Entscheidung fällt leicht, wenn man weiss, dass das System auf elliptische Konstituentenfragen immer nur die erfragten Konstituenten als Antwortellipse zurückgibt. Da in der Antwort (f) sowohl Zeit als Ort angegeben werden, muss intern die Frage (e) wie (e'') behandelt worden sein. Dieser Sachverhalt wird bestätigt, wenn die erste Nachfrage keine Konstituentenfrage ist, sondern eine Entscheidungsfrage:

Hält Professor Barry die Vorlesung Akustische Phonetik?

Ja.

Im Phonetikstudio?

Ja.

Wann?

Am Dienstag von 14 bis 16 Uhr.

In diesem Fall wird in der Antwort nur die Zeit erwähnt. Man könnte sich auf den Standpunkt stellen, dass es irrelevant ist, ob intern eine Variante wie (e') oder (e'') verwendet wird. Wenn man folgendes Beispiel betrachtet, sieht man jedoch sofort, dass es durchaus einen Unterschied macht:

³² Im Ellipsenmodul des UIS stellt sich diese Frage nicht, da die Antworten bei der Auflösung der Ellipse nicht miteinbezogen werden.

- a) Wo hält Barry die Akustische Phonetik?
- b) *Im Phonetikstudio.*
- c) Und die Einführung in die Allgemeine Sprachwissenschaft?
- d) *Im Seminarraum der Computerlinguistik.*

Die Antwort (d) kann nicht zustande kommen, wenn (c) als

- (c') Hält Barry die Einführung in die Allgemeine Sprachwissenschaft im Phonetikstudio?

verstanden wird. (c) muss als elliptische Konstituentenfrage der folgenden Form betrachtet werden:

- (c'') Wo hält Barry die Einführung in die Allgemeine Sprachwissenschaft?

Leider kann man nicht festlegen, dass immer die Variante (x'') die korrekte ist, d.h. dass w-Fragewörter wie „wann“, „wo“ usw. bei Nachfragen immer erhalten bleiben müssen. Wenn in ('c) „Und“ durch „Auch“ ersetzt wird, sieht die Sache nämlich anders aus, und die Antwort ('d), welche vom System gegeben wird, ist sicher unbefriedigend.

- a) Wo hält Barry die Akustische Phonetik?
- b) *Im Phonetikstudio.*
- c) Auch die Einführung in die Allgemeine Sprachwissenschaft?
- d) *Im Seminarraum der Computerlinguistik.*

Hier muss ('c) im Sinn von ('c') als Entscheidungsfrage verstanden werden, die mit „Nein“ zu beantworten ist.

- (c') Hält Barry auch die Einführung in die Allgemeine Sprachwissenschaft im Phonetikstudio?

Die verwendete Strategie, w-Fragewörter in iterierten elliptischen Nachfragen zu belassen, ist auch im folgenden Fall unintuitiv:

- a) Hält Barry eine Vorlesung?
- b) *Ja.*
- c) Wann?
- d) *Am Dienstag von 9 bis 11 Uhr.*
- e) Im Phonetikstudio?
- f) *Am Dienstag von 14 bis 16 Uhr.*

(e) wird kaum als echte Informationsfrage (e') zu verstehen sein, sondern als Bestätigungsfrage (e'').

- (e') Wann hält Barry eine Vorlesung im Phonetikstudio?

- (e'') Hält Barry eine Vorlesung am Dienstag von 9 bis 11 Uhr im Phonetikstudio?

3.2.3 Die Behandlung von Ersetzungsellipsen

Für die Behandlung von Fragen mit Ersetzungsellipsen müssen bestimmte Konstituenten der Vorfrage mit den Konstituenten der Ellipse ausgetauscht werden. Im fol-

genden Beispiel wird zunächst das Objekt und dann das Subjekt durch eine Nominalphrase der Ellipse ersetzt.

Hält Barry die Einführung in die Allgemeine Sprachwissenschaft?

Ja.

Und die Vorlesung Grammatiktheorie?

Nein.

Und Professor Uszkoreit?

Ja.

Die entscheidende Frage ist dabei, wie merkt das System, dass es das erste Mal das Subjekt und das zweite Mal das Objekt ersetzen muss? Immerhin wäre eine Frage wie

Hält Barry Professor Uszkoreit?

nicht bloss syntaktisch, sondern auch semantisch passabel. Eine allgemeine und intuitiv zugängliche Begründung könnte lauten: Es wäre eine Stilblüte, wenn das Verb „halten“ zunächst im Sinn von „etwas vortragen“ gebraucht würde und dann elliptisch im Sinn von „etwas oder jemanden fassen/festhalten“.

Eine technischere Lösung spricht (Alshawi 1992: 173) an:

„Sortal restrictions encode constraints on the sorts of objects that can fill argument positions of specified relations. [...] Sortal restrictions have been used, in some form or other, for disambiguation in natural language systems for a long time.“

Für das Verb „halten“ könnte man festlegen, dass das Subjekt etwas Menschliches und das Objekt etwas informationsvermittelndes Sächliches enthalten muss. Wenn gleichzeitig bekannt ist, dass „Professor Uszkoreit“ etwas Menschliches bezeichnet, und im Gegensatz dazu die „Vorlesung Grammatiktheorie“ ein informationsvermittelndes Sächliches, wäre das Finden der korrekten Ersetzung bei der Ellipsenbehandlung ein Kinderspiel. Einschränkungen dieser Art haben allerdings weniger damit zu tun, was Sprache bedeuten kann, sondern wie sie gebraucht werden soll (Alshawi 1992: 173):

„Since sortal restrictions largely depend on what is possible in the world or in some domain of discourse, they fall into the realm of pragmatics rather than semantics.“

3.2.3.1 Verbal-Lexikon

Wenn man im Vollformen-Verblexikon des Saarbrücker Dialogsystems nachschaut, wie die Verbform „hält“ repräsentiert ist, findet man PROLOG-Code wie er in der *Abbildung 2* zu sehen ist³³:

³³ Die lexikalischen Regeln sind wie die syntaktischen Regeln im *Definite Clause Grammar* Formalismus von Prolog notiert. Der abgebildete Eintrag wurde nicht von Hand geschrieben, sondern von einem Vorkompilierer erzeugt.

Der Lexikoneintrag ist eine Zusammenstellung von lexikalischen, morphologischen, syntaktischen, sortalen sowie semantischen Informationen, welche das System für sein Funktionieren benötigt. So beinhaltet die 2. Zeile, dass „hält“ als verbaler Kopf weder Partizip (`part(no)`), Infinitiv (`inf(no)`) noch Passivform (`pass(no)`) ist, mithin aber finit und zwar pluralisch (`pl(no)`). Eher lexikon-unüblich ist der Vermerk, dass „hält“ satzinitial (`init(_11936)`) vorkommen kann. Auf der 3. Zeile beginnt die Subkategorisierungs-Information, welche genau zwei nominale Komplemente (`np(...)`) zulässt. Die syntaktischen Funktionen sind implizit über die Reihenfolge ausgedrückt. Das Subjekt wird nicht nur in Kasus (`case(yes, no, no, no)`) und Numerus (`pl(no)`) bestimmt, sondern auch in seiner ontologischen Sorte:

```
sort(human(yes), local(no), temporal(no), course(no))
```

Dabei zeigt sich ein geschlossenes System, dass mit den Merkmalen MENSCH (`human/1`), ORT (`local/1`), ZEIT (`temporal/1`) und VERANSTALTUNG

```
v(
  head(v, part(no), inf(no), pass(no), init(_11936), pl(no),
    [np(case(yes, no, no, no), pl(no),
      sort(human(yes), local(no), temporal(no), course(no))),
    np(case(no, no, no, yes), pl(_11988),
      sort(human(no), local(no), temporal(no), course(yes)))]),
    ((_11996^_11997)^_11994)^wff(halten, [_11996, _11997, _11994]),
    v([haelt])) -->
  [haelt].
```

Abbildung 2: Vollformen-Lexikoneintrag zu „hält“

(`course/1`) auskommt. Als Subjekt eines Satzes mit dem Verb „hält“ darf also nur eine Nominalphrase auftreten, welche dem Merkmalsbündel

```
{+MENSCH, -ORT, -ZEIT, -VERANSTALTUNG}
```

genügt. Das Akkusativ-Objekt dagegen hat folgende Sortenbedingungen zu erfüllen:

```
{-MENSCH, -ORT, -ZEIT, +VERANSTALTUNG}
```

Die Merkmale müssen nicht alle vollständig spezifiziert sein. In PROLOG verwendet man Variablen, um diesen Sachverhalt auszudrücken. Die Verbform „ist“ hat für das Subjekt und das Objekt beispielsweise folgenden identischen Eintrag,

```
sort(human(_11982), local(no), temporal(no), course(_11988))
```

was als Merkmalsbündel etwa folgendermassen notiert wird:

```
{?MENSCH, -ORT, -ZEIT, ?VERANSTALTUNG}
```

Wenn man mit disjunkten Merkmale arbeitet, dürfen allerdings nicht zwei Variablen gleichzeitig positiv spezifiziert werden, bzw. zu *yes* instantiiert werden.

Fragen wie (a) und (c) sind zulässig, (e) kann aufgrund der Sortenbeschränkung nicht syntaktisch analysiert werden.

- a) Ist Pinkal ein Professor?
- b) *Ja.*
- c) Ist Akustische Phonetik eine Vorlesung?
- d) *Ja*
- e) Ist Pinkal eine Vorlesung?
- f) *Ich kann Sie leider nicht verstehen, bitte formulieren Sie Ihre Frage neu.*

Nur am Rande soll noch erwähnt werden, dass die für den Aufbau der Logischen Form notwendige kompositionelle Semantikinformaton aus lambda-abstrahierten „wohlgeformten Formeln“ (*well-formed-formula* = wff) besteht:

$((_996^{_997})^{_994})^{\text{wff}(\text{halten}, [_996, _997, _994])}$

Die Identität der Verbalbedeutung von Vollformen wie „hält“, „halten“, „gehalten“ drückt in den jeweiligen Lexikoneinträgen ein Ausdruck folgender Form aus:

$\text{wff}(\text{halten}, [\text{EVENT}, \text{AGENS}, \text{THEMA}])$ ³⁴

Lambda-Abstraktionen steuern dabei, wie die syntaktischen Komplemente beim Aufbau der Satzbedeutung in die Argumentpositionen von *halten* eingesetzt werden. Wenn beispielsweise Aktiv/Passiv-Varianten wie

Hält Barry die Akustische Phonetik?

Wird die Akustische Phonetik von Barry gehalten?

dieselbe Semantik haben sollen, reicht dazu ein minimaler Unterschied bei der Abstraktion. So ist die Semantikinformaton der passivisch verwendeten Verbform „gehalten“ im Lexikon folgendermassen vermerkt:

$((_259^{_260})^{_257})^{\text{wff}(\text{halten}, [_259, _257, _260])}$

Bei genauem Hinschauen und Vergleich mit dem Lexikon-Eintrag von „hält“ in Abbildung 2 auf Seite 32 sieht man, dass zwei Variablen bei der Lambda-Abstraktion ausgetauscht sind. Die Subkategorisierungsliste der Verbform „gehalten“ besteht aus:

```
[np(case(yes,no,no,no),pl(_201),
  sort(human(no),local(no),temporal(no),course(_229))),
  pp(von,case(no,no,yes,no),
  sort(human(yes),local(no),temporal(no),course(no)))]
```

³⁴ *EVENT* drückt den Ereignis-Charakter von „halten“ aus. Thematische Rollen, wie ich sie hier mit *AGENS* und *THEMA* eingefügt habe, tauchen im System selbst nirgends explizit auf. Die Rollen sind – wie im weiteren Text erklärt – implizit über die Stellung in der Argumentliste, Sorteninformation und Lambda-Abstraktion kodiert.

```

n(
  head(n, bar(no, no, no, no, _3884), mc(_3915),
    distr(rc(yes), det(no), qtty(no), quantifier(no),
    exact(no), prenomgen(no), wh(no)),
    sort(human(yes), local(no), temporal(no), course(no),
    pn(yes), addressing(no), first_name(no),
    family_name(yes))),
  agr(case(_3855, no, _3856, _3857), gender(yes, no, no),
    pl(no), strong(_3858)),
  [* , * , * ]36,
  [iota(yes), rep(_3859^wff(ident, [_3859, s_dozent, barr])),
  sem_sort(dozent), anaphor(_3989)],
  n([barr]) --> [barr] .

```

Abbildung 3: Vollformen-Lexikoneintrag zum Eigennamen „Barry“

Wenn nun bei der kompositionellen Semantikkonstruktion das syntaktische Subjekt auf den Lambda-Term angewendet wird, erscheint es nicht mehr an der *AGENS*-Argumentstelle, sondern an der *THEMA*-Stelle. Dafür wird das zweite Komplement, die Präpositionalphrase mit „von“, durch die Lambda-Konversion bei der *AGENS*-Stelle eingesetzt.

Zum Abschluss dieses Exkurses über das Verbal-Lexikon halte ich nochmals fest:

- Im Lexikon ist bei jeder Verbform für jedes Komplement eine Sorteninformation kodiert, die folgende disjunkten Merkmale umfasst: MENSCH, ORT, ZEIT und VERANSTALTUNG.
- Durch die Merkmale werden Sorten differenziert wie Menschen, Orte, Zeiten, Veranstaltungen und der ganze Rest.³⁵

3.2.3.2 Nominal-Lexikon

Soll die Sortenbeschränkung in den Subkategorisierungs-Slots der Verben sinnvoll benutzt werden, braucht es auf der Seite der Nominalphrasen, welche die Slots füllen sollen, ebenfalls die passende Information.

Als Beispiel dient der Lexikoneintrag für den Eigennamen „Barry“ in *Abbildung 3*. Es gibt zwar keine exakte Kopie der vierstelligen Sortenrelation *sort/4* aus dem Verbal-Lexikon, aber die ersten vier Argumente der achtstelligen Relation *sort/8* sind trotzdem passgenau mit denselben Merkmalen gefüllt.

Die Behandlung von nominalen Ersetzungselipsen wäre somit auf der Syntaxebene nach einem einfachen Rezept möglich:

³⁵ Zur Restsorte gehören beispielsweise Studienscheine.

³⁶ Hinter dieser Sternliste verbirgt sich ein Subkategorisierungsschema für Nomen, das von folgendem Schema ausgeht: [Apposition, Genitivattribut, Präpositionalphrase]. Gesternte Positionen dürfen nicht gefüllt werden. In *Abbildung 4* auf Seite 38 sieht man ein Beispiel, wo Apposition und Genitivattribut spezifiziert sind.

1. Nehme die Nominalphrase der elliptischen Frage.
2. Bestimme ihre Sorte und den Kasus.
3. Nehme die syntaktische Struktur der Vorfrage oder scheitere.
4. Bilde eine neue Frage, indem die elliptische Nominalphrase an der Subkategorierungs-Stelle mit der passenden Sorte und dem passenden Kasus in der Vorfrage eingesetzt wird oder scheitere.

Mehrdeutigkeiten könnten höchstens noch auftauchen, wenn ein Verb mehreren Komplementen dieselbe Sorte zuweist und zusätzlich die Kasusmarkierung der elliptischen Phrase nicht genügend differenziert wäre.³⁷

Das im System tatsächlich angewandte Vorgehen weicht in den Punkten 3 und 4 ab: Erstens wird keine Syntaxstruktur der Ellipse aufgebaut, sondern direkt die semantische. Zweitens steht für die Auflösung der Vorfrage kein Baum als syntaktische Struktur zur Verfügung, sondern eine flache Liste der Konstituenten in der Reihenfolge ihres Auftretens. Diese Reihenfolge müssen auch Nachfrage-Ellipsen genau einhalten.³⁸

Folgende Doppelersetzung ist beispielsweise gestattet:

Hält Barry die Akustische Phonetik?

Ja.

Und Uszkoreit die Grammatiktheorie?

Ja.

Im Gegensatz dazu löst die vertauschte Reihenfolge der Komplemente einen Syntaxfehler aus:

Hält Barry die Akustische Phonetik?

Ja.

Und die Grammatiktheorie Uszkoreit?

Ich kann Sie leider nicht verstehen, formulieren Sie ...

Obwohl sie für mich bei kontrastierender Intonation von „Uszkoreit“ kein Problem darstellt, ist das Vertauschen von Subjekt und Objekt in der Nachfrage syntaktisch verboten. Das heisst nicht, dass eine Ellipse wie

Und die Grammatiktheorie Uszkoreit?

grundsätzlich ausgeschlossen ist. Bei geeignetem Antezedenssatz wird sie korrekt verarbeitet:

³⁷ Eine solches Vorgehen wäre für das UIS die einfachste Möglichkeit gewesen, wenn es denn solche Sorten wie im Saarbrücker Dialogsystem gegeben hätte.

³⁸ Die strenge Einhaltung der Reihenfolge der Satzglieder in Ellipse und kontrollierendem Ausdruck (syntaktische Parallelitätsbedingung) ist von der angelsächsischen Linguistik mehrfach gefordert worden. Im Deutschen lässt sie sich nicht halten. Interessanterweise ignoriert Lappin (1996) die Parallelitätsbedingung ebenfalls. – Vergleiche Abschnitt 3.4 .

Welche Vorlesung hält Barry?

Barry hält die Vorlesung Akustische Phonetik, die Vorlesung Einführung in die Allgemeine ...

Und die Grammatiktheorie Uszkoreit?

Ja.

Das Prinzip der parallelen Reihenfolge in Frage und elliptischer Nachfrage lässt allfällige Mehrdeutigkeiten bei Verben mit sortengleichen Komplementen nicht zu. Das einzige solche Verb im System neben dem schon angesprochenen „sein“ ist „voraussetzen“. Es hat sowohl für Subjekt wie Objekt folgende Merkmale:

{-MENSCH, -ORT, -ZEIT, +VERANSTALTUNG}

Im folgenden verwende ich „voraussetzen“, obwohl es inhaltlich ungenügend als reflexive Relation definiert ist.³⁹ Das Verb „sein“ ist aber zum Demonstrieren noch ungeeigneter, da die Antworten ‘benutzerfreundlich’ gekürzt sind.

Welche Veranstaltung ist die Grammatiktheorie?

Die Vorlesung Grammatiktheorie.

Die folgenden Beispiele mit „voraussetzen“ mögen pragmatisch verquer erscheinen; sie sind aber die einzige Möglichkeit gewisse Eigenheiten der Ellipsenbehandlung aufzuzeigen.

Welche Veranstaltung setzt die Grammatiktheorie voraus?

Die Grammatiktheorie setzt die Vorlesung Grammatiktheorie voraus.

In vollständigen Antworten verwendet das System immer die Satzglied-Reihenfolge Subjekt-Verb-Objekt. Da nicht erfragte Komplemente in der Antwort immer wörtlich zurückgegeben werden, in diesem Fall „die Grammatiktheorie“, lässt sich folgende syntaktische Struktur erschliessen:

[Welche Veranstaltung]_{Objekt} setzt [die Grammatiktheorie]_{Subjekt} voraus?

Wie werden nun Doppelersetzungen aufgelöst?

a) Welche Veranstaltung setzt die Grammatiktheorie voraus?

b) *Die Grammatiktheorie setzt die Vorlesung Grammatiktheorie voraus.*

c) Und die Textgenerierung die Vorlesung Semantik?

d) *Ich kann dazu keine Angaben finden.*

Aufgrund der parallelen Verarbeitung muss (c) die Interpretation (c') haben:

(c') Setzt [die Vorlesung Semantik]_{Subjekt} [die Textgenerierung]_{Objekt} voraus?

(c'') Setzt [die Textgenerierung]_{Subjekt} [die Vorlesung Semantik]_{Objekt} voraus?

³⁹ Den Grund für diesen Defekt erkläre ich bei der Darstellung der Konzeptualisierungsmöglichkeiten im Dialogsystem. Diese ist wie in Abbildung 5 auf Seite 40 gezeigt ausserordentlich restriktiv und lässt gar keinen sinnvollen Gebrauch von „voraussetzen“ zu.

Gefühlsmässig würde man vermutlich auf (c'') tippen.⁴⁰ Die Interpretation (c'') wird vom System zwingend gewählt, wenn die Antezedens-Frage wie ('c) lautet:

('c) Setzt [die Grammatiktheorie]_{Subjekt} [die Vorlesung Grammatiktheorie]_{Objekt} voraus?

Es fragt sich, ob die strenge Parallelität sinnvoll ist bei Konstruktionswechseln von Konstituentenfrage zu Entscheidungsfragen. Oder ob nicht besser einfach die unmarkierte normale Satzstellung angenommen werden sollte.

Bei einfachen Ellipsen in Fragen aus Verben mit sortengleichen Komplementen wird das erste passende Komplement ersetzt, auch wenn die Datenbank damit nichts anfangen kann.

Setzt die Akustische Phonetik die Grammatiktheorie voraus?

Ich kann dazu keine Angaben finden.

Und die Akustische Phonetik?

Ich kann dazu keine Angaben finden.

Als minimaler Kontrast:

Setzt die Akustische Phonetik die Grammatiktheorie voraus?

Ich kann dazu keine Angaben finden.

Und die Grammatiktheorie?

Ja.

Das folgende Beispiel demonstriert, dass die Ersetzung im System tatsächlich nicht über die syntaktische Rolle verläuft, sondern über die Semantik. Die Ersetzung des Subjekts ersetzt auch gleich das Objekt mit, weil es die gleiche Bedeutung hat:

Setzt die Grammatiktheorie die Grammatiktheorie voraus?

Ja.

Und die Textgenerierung?

Ja.

Zurück zum Thema Nominal-Lexikon: Gewöhnliche Substantive haben etwas andere Lexikoninformation, wie *Abbildung 4* mit dem Vollformen-Eintrag zu „Vorlesung“ zeigt.

⁴⁰ Dieses Gefühl könnte damit zu tun haben, dass in Nachfrage-Ellipsen der Wechsel von einer Konstituentenfrage zu einer Entscheidungsfrage nicht besonders akzeptabel ist, und die Nachfrage nicht als Parallelkonstruktion zur Vorfrage, sondern zur Systemantwort aufgefasst wird.


```

n(
  head(n,bar(no,no,no,no,_3900),mc(_3929),
    distr(rc(no),det(no),qty(no),quantifier(no),
      exact(no),prenomgen(no),wh(no)),
    sort(human(no),local(no),temporal(no),course(yes),
      pn(no),addressing(no),first_name(no),family_name(no))),
  agr(case(_3859,_3860,_3861,_3862),gender(no,yes,no),
    pl(no),strong(_3863)),
  [apposition(
    agr(case(_3859,_3860,_3861,_3862),
      gender(no,yes,no),pl(no),strong(_3863)),
    app_noun(
      agr(case(yes,no,no,no),gender(_3870,_3871,_3872),
        pl(no),strong(_3873)),
      sort(human(no),local(no),temporal(no),course(yes),
        pn(_3868),addressing(no),first_name(no),family_name(no)),
      _3869)),
  np_gen(postnom(_3874),
    sort(human(_3875),local(_3876),temporal(_3877),course(no),
      pn(_3878),addressing(_3879),first_name(_3880),
      family_name(_3881)),
    _3882),*],
  [iota(no),rep(_3883^wff(g_vorlesung,[_3883])),sem_sort(_3978),
  anaphor(no)],
  n([vorlesung]))-->[vorlesung].

```

Abbildung 4: Vollformen-Lexikoneintrag zu ^aVorlesung

Interessant ist der Subkategorisierungsrahmen bei Nomen (9.-22. Zeile), der aus maximal 3 Einträgen besteht.⁴¹ Im Fall von „Vorlesung“ besagt er, dass eine Apposition (*apposition/1*) der Sorte VERANSTALTUNG zulässig ist. Weiter kann ein postnominales Genitivattribut (*np_gen(post_nom/1)*) zu „Vorlesung“ treten, sofern es keine VERANSTALTUNG ist. Bei Verben enthält der Subkategorisierungsrahmen normalerweise die Komplemente, die miteinander auftreten. Bei Nomen gibt es Beschränkungen darüber, welche Erweiterungen zugleich erscheinen dürfen:

- a) die Vorlesung Barrys
- b) die Vorlesung Akustische Phonetik
- c) *die Vorlesung Barrys Akustische Phonetik
- d) *die Vorlesung Akustische Phonetik Barrys
- e) Barrys Vorlesung Akustische Phonetik

Das Dialogsystem lässt immer nur eine einzige Erweiterung zu. (a-b) werden akzeptiert, Fälle wie (c-d) in Übereinstimmung mit der Grammatik abgelehnt. (e) wäre grammatisch korrekt. Das Dialogsystem kann wie erwartet nichts damit anfangen, da der Lexikoneintrag von „Vorlesung“ gar keine pränominalen Genitivattribute vorsieht.

⁴¹ Vergleiche auch Fussnote 36 auf S. 34.

Erstaunlicherweise – und ziemlich gegen mein Sprachgefühl – behandelt das System nun Ersetzungsellipsen von postnominalen Genitivattributen⁴²:

Findet eine Vorlesung Barrys am Dienstag statt?

Ja.

Und Uszkoreits?

Nein.

Wenn man dasselbe Spiel mit einer Erweiterungsellipse versucht, was grammatisch sinnvoller wäre, scheitert es allerdings.

Findet am Dienstag eine Vorlesung statt?

Ja.

Barrys?

Ich kann Sie leider nicht verstehen, formulieren Sie ...

Zum Schluss noch etwas zur Semantik. Die Bedeutungsinformation des Appellativs „Vorlesung“

```
rep(_3883^wff(g_vorlesung, [_3883]))
```

drückt aus, dass es die Gattung *Vorlesung* (`g_vorlesung`) bezeichnet. Über die im Nominallexikon vorhandenen Gattungen sind in einem Gattungsmodul noch strukturierende Relationen definiert.⁴³ Wie im Abschnitt 3.2.1 schon demonstriert, ermöglicht diese Strukturierung, gewisse Kategorienfehler zu korrigieren.⁴⁴

Intern werden die Gattungsverbindungen vor allem bei der Umsetzung in die Datenbankabfrage benötigt. Bei dieser wird jede Anfrage zunächst in eine Instanz desselben Konzepts (im Sinn von *frame*) gebracht. In *Abbildung 5* ist das „Universalkonzept“ (`concept/17`) dargestellt, das 17 Rollen (*slots*) umfasst. Universal ist es allerdings nur in dem Sinn, dass es für die Bearbeitung aller pragmatisch sinnvollen Anfragen hinreichen soll. Ein etwas reichhaltigerer Anwendungsbereich würde sicher keine so simple Konzeptualisierung mehr erlauben.

⁴² Für mich ist nur die Lesart nachvollziehbar, wo die Ellipse „Uszkoreits“ als pränominales Genitivattribut eines elliptischen Nomens betrachtet wird. Das wäre dann eine Form von N-Ellipse, die ich im Abschnitt 2.2.7 nicht besonders erwähnt habe.

⁴³ Dies ergibt eine ontologische Strukturierung bzw. eine Art semantischer Thesaurus.

⁴⁴ Für Fehler bei den ontologischen Sorten können selbstverständlich keine Korrekturvorschläge gemacht werden, da dann meist schon die syntaktische Analyse scheitert.

```

concept (
  s_dozent (Do),           % Dozent
  s_student (Stud),       % Student
  s_veranst (Ver),        % Veranstaltung
  s_zeit (Zeit),          % Zeit
  s_tag (Tag),            % Wochentag
  s_vonU (VonU),          % Veranstaltungsbeginn
  s_bisU (BisU),          % Veranstaltungsende
  s_ort (Ort),            % Veranstaltungsort
  s_semester (Zeitsem),   % Sommer-/Wintersemester
  s_woche (Woche),        % Semesterwoche
  s_studsem (Studsem),   % ?
  s_hfach (HF),           % Hauptfach
  s_schein (Schein),      % Testatschein
  s_nfach (NF),           % Nebenfach
  s_beginn (Beginn),      % Beginn in Woche
  s_gebiet (Gebiet),      % Studiengebiet
  s_studteil (Studteil)  % ?
)

```

Abbildung 5: Das Konzept concept⁴⁵

3.2.3.3 Zusammenspiel von Ellipsen- und Anaphernauflösung

Durch die Sorteninformation im Nominal- und Verbal-Lexikon wird auch die Auflösung der anaphorischen Bezüge von Personalpronomen vereinfacht. Sie kann nach folgendem Schema stattfinden:

- Bestimme die syntaktische Funktion eines anaphorischen Ausdrucks im Satz.
- Weise der Anapher die Sorte zu, welche beim Verb für die betreffende syntaktische Funktion vermerkt ist.
- Suche im vorangehenden Text nach Nominalphrasen mit passender Sorte, Genus und Numerus.

Ein Beispiel:

Hält Vestre den Programmierkurs LISP?

Ja.

Wann wird er von ihm gehalten?

Er wird von ihm am Montag von 11 bis 13 Uhr gehalten.

Weil Konjunktionen, Nebensätze und Reflexivpronomen fehlen, muss das Antezedenselement immer in einem vorhergehenden Satz sein. Der Suchraum für die Antezedenten von Fragen ist auf das vorangehende Frage-Antwort-Paar beschränkt:

a) Wann hält Barry eine Vorlesung?

b) *Barry hält eine Vorlesung am Dienstag von 9 bis 11 Uhr.*

c) Und ein Seminar?

⁴⁵ Zur besseren Verständlichkeit habe ich zu jedem Konzeptbestandteil eine Beschreibung beigelegt, sofern mir die Funktion klar war.

- d) *Am Donnerstag von 14 bis 16 Uhr.*
- e) *Wo hält er sie?*
- f) *Es gibt keinen Referenten.*

Die benutzerunfreundliche Fehlermeldung (f) will besagen, dass das System für „sie“ in (e) keinen Antezedenten finden kann. Es sieht den Bezug zu „eine Vorlesung“ in (a-b) nicht. Mit „er“ gibt es in (e) keine Probleme, weil „Barry“ in (c-d) implizit erwähnt wird. Folgendes Beispiel belegt es:

- Wann hält Barry eine Vorlesung?
- Barry hält eine Vorlesung am Dienstag von 9 bis 11 Uhr.*
- Und ein Seminar?
- Am Donnerstag von 14 bis 16 Uhr.*
- Wo hält er es?
- Er hält es im Phonetikstudio.*

Etwas irritierend erscheint vielleicht folgendes Beispiel:

- a) Wann findet die Übung zur Vorlesung Einführung in die Semantik statt?
- b) *Die Übung zur Vorlesung Einführung in die Semantik findet am Donnerstag von 11 bis 13 Uhr statt.*
- c) Und die Vorlesung?
- d) *Am Dienstag von 11 bis 13 Uhr.*

Man könnte meinen, dass hier von der Komponente Ellipsenbehandlung eine besondere Leistung vollbracht wird. Wenn eine Frage wie (c') gestellt wird, meldet das System normalerweise (d') als Präsuppositionsverletzung zurück.

- (c') Wann findet die Vorlesung statt?
- (d') *Es gibt mehrere Vorlesungen, nämlich ...*

Da es aber mit (d) antwortet, liegt der Gedanke nahe, das System habe (c) als (c'') behandelt, da diese Vorlesung tatsächlich am Dienstag von 11 bis 13 Uhr stattfindet.

- (c'') Wann findet die Vorlesung Einführung in die Semantik statt?

Man könnte fälschlicherweise denken, dass beim Ersetzen von „Übung zur Vorlesung Einführung in die Semantik“ durch „Vorlesung“ letzteres aufgrund des gemeinsamen Kerns irgendwie zu „Vorlesung Einführung in die Semantik“ gematcht wurde.

Tatsächlich behandelt das System (c) als (c'), die Präsuppositionsverletzung wird jedoch verhindert, weil definite Nominalphrasen auch eine Lesart als anaphorische Ausdrücke besitzen. Durch die Anfrage (a) wird ein Diskursreferent i eingeführt, auf den in (c) ein anaphorischer Bezug gemacht wird:

- a) Wann findet die Übung zur [Vorlesung Einführung in die Semantik] _{i} statt?
- b) Und [die Vorlesung] _{i} ?

Die anaphorische Lesart von definiten Nominalphrasen erzeugt im Zusammenhang mit Appositionen manchmal unerwünschte Mehrdeutigkeiten. Folgendes Beispiel zeigt den Effekt:

- a) Setzt die Akustische Phonetik die Grammatiktheorie voraus?
- b) *Ich kann dazu keine Angaben finden.*
- c) Und die Vorlesung Grammatiktheorie?
- d) Ja.

Nach dem bis anhin gültigen Prinzip, dass bei einfachen Ersetzungsellipsen bei Verben mit sortengleichen Komplementen immer die erste Konstituente ersetzt wird, muss (c) die Interpretation (c') haben:

(c') Setzt [die Akustische Phonetik]_{Subjekt} [die Vorlesung Grammatiktheorie]_{Objekt} voraus?

Die Bejahung in der Systemantwort (d) wird dann aber unverständlich, da jede Veranstaltung nur sich selbst als Voraussetzung hat. Die Lösung des Rätsels hat damit zu tun, wie die Syntaxkomponente unterschiedliche Lesarten erzeugt. Der Parser ist so eingestellt, dass die 'wahrscheinlichste' Lesart zuerst geliefert wird. Zu dieser Lesart wird sofort eine Antwort gesucht. Wenn das System keine Angaben dazu findet, wird die nächste Lesart erzeugt. Im obigen Fall führt diese Strategie dazu, dass (c) nicht als einfache Ellipse, sondern als Doppelellipse mit „die Vorlesung“ als anaphorisches Element verstanden wird:

- (a) Setzt [die Akustische Phonetik]_{Subjekt} [die Grammatiktheorie]ⁱ_{Objekt} voraus?
- (c) Und [die Vorlesung]ⁱ_{Subjekt} [Grammatiktheorie]_{Objekt}?

Damit erhält die Frage (c) die Interpretation (c'), für die das System eine befriedigende Antwort findet.

(c') Setzt [die Vorlesung]ⁱ_{Subjekt} [Grammatiktheorie]_{Objekt} voraus?

3.2.3.4 Fazit Ersetzungsellipsen

Das Wichtigste zur Auflösung von Ersetzungsellipsen nochmals zusammengefasst:

- Substantive haben in ihrem Lexikoneintrag Sorteninformation kodiert, die mit der Sorteninformation in den Subkategorisierungsrahmen der Verben zusammenpasst.
- Die Behandlung von einzelnen Ersetzungsellipsen wird durch die passende Sorteninformation gesteuert.
- Die Auflösung von Ersetzungsellipsen wird durch die Forderung der parallelen Reihenfolge von Komplementen in Frage und elliptischer Nachfrage eingeschränkt.

3.2.4 Mischung von Erweiterungs- und Ersetzungsellipsen

Die Unterscheidung in Erweiterungs- und Ersetzungsellipsen stellt keine Klassifikation von elliptischen Äußerungen dar. Sie umreißt die zwei unterschiedlichen Arten, mit denen die Konstituenten einer elliptischen Anfrage zu behandeln sind. Im Dialogsystem ist die Auflösung von Mischungen von Erweiterungs- und Ersetzungsellipse nicht realisiert⁴⁶:

Wann hält Barry eine Vorlesung?

Barry hält eine Vorlesung am Dienstag von 9 bis 11 Uhr.

Und wann Uszkoreit?

Ich kann Sie leider nicht verstehen, ...

Mischformen sind keinesfalls exotischer in der Kommunikation als reine Formen. Ein praktisches System muss Ellipsen, die das Fragewort wiederholen, auf jeden Fall verarbeiten können.

3.2.5 Zur Architektur der Ellipsenbehandlung

In den bisherigen Abschnitten wurde die Leistungsfähigkeit des Systems mehr von aussen betrachtet. Nun soll der gesamte Aufbau skizziert werden. In *Abbildung 6* sind die Verarbeitungsschritte beim Beantworten einer einzelnen Frage abgebildet.

3.2.5.1 Syntaxanalyse und Semantikkonstruktion

Der Aufbau der kompositionellen semantischen Struktur verläuft parallel mit der syntaktischen Analyse. Das Ellipsenmodul arbeitet nicht als Unterkomponente des normalen Syntax/Semantik-Moduls, sondern als gleichwertige Alternativ-Komponente, die ebenfalls aus einer Tokenfolge eine (flache) syntaktische und vollständige semantische Struktur aufbaut. Das Ellipsenmodul wird aufgerufen, wenn das normale Modul nicht durchkommt.

⁴⁶ Man findet im Code Ansätze zur Realisierung.

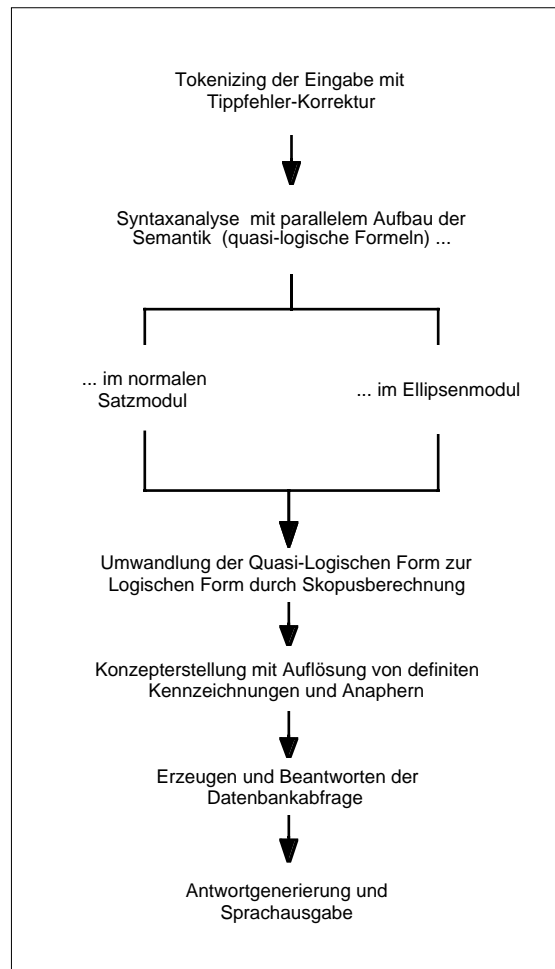


Abbildung 6: Schema der Anfrage-Beantwortung im Saarbrücker System

3.2.5.2 Syntaxanalyse und Semantikkonstruktion

Der Aufbau der kompositionellen semantischen Struktur verläuft parallel mit der syntaktischen Analyse.⁴⁷ Das Ellipsenmodul arbeitet als gleichwertige Alternativ-Komponente zum normalen Syntax/Semantik-Modul und baut ebenfalls aus einer Tokenfolge eine – allerdings flache – syntaktische und semantische Struktur auf. Das Ellipsenmodul wird aufgerufen, wenn das normale Modul nicht durchkommt.

Auf der syntaktischen Seite arbeiten beide Komponenten mit *DCG*-Regeln, die im *top-down* Verfahren angewendet werden. Das Ellipsenmodul kann dabei nicht von der Vorarbeit des normalen Parsings profitieren und kleinere Teilbäume übernehmen.⁴⁸

⁴⁷ Die Lexikoneinträge haben wie gezeigt die notwendigen semantischen Strukturen definiert.

⁴⁸ Im UIS kann das Ellipsenmodul die Ergebnisse des normalen Parsings vollständig wiederverwerten.

Folgende elliptischen Syntagmen werden von der Ellipsengrammatik zugelassen:⁴⁹

$$\text{ellipse} ::= [„und“ [„auch“]] \text{ (NP|PP)+} \\ | [„und“] \text{ (NP|PP)+ [„auch“]}$$

Die Konjunktion „und“ und das Adverb „auch“ werden beim Parsen verwendet, dass keine Syntagmen wie „und auch Uszkoreit auch“ durchschlüpfen. Weiter triggern sie manchmal im Zusammenhang mit w-Fragewörtern zwischen Ersetzungs- und Erweiterungsellipsen. Die Restriktionen sind teilweise domänenspezifisch und etwas arbiträr. So dürfen w-Fragewörter nicht mit „auch“ in einer Ellipse auftreten:

Hält Barry die Akustische Phonetik?

Ja.

Und wer auch?

Das System streikt bei der Nachfrage, obwohl es durchaus eine Welt geben könnte, wo nicht jede Vorlesung von höchstens einem Dozent gehalten wird. Im UIS konnten solche spezifischen Einschränkungen deshalb nicht übernommen werden.

Die Nominal- und Präpositionalphrasen erscheinen in einer Konstituentenliste. Die Hauptarbeit des Ellipsenmoduls besteht darin, aus der Konstituentenliste mit Hilfe des Kontexts eine passende und vollständige Satzbedeutung zu konstruieren.

Beide Komponenten – das normale Syntaxmodul und das Ellipsenmodul – liefern als Resultat eine semantischen Struktur, die aus quasi-logischen Formeln besteht: Die **Quasi-Logische Form (QLF)**. Die Bezeichnung „Quasi-Logische Form“ begreift man am einfachsten, wenn man von der **Logischen Form (LF)** ausgeht. Unter der LF eines Satzes wird in der Semantik die Gesamtheit seiner Wahrheitsbedingungen verstanden, die meist in Form von prädikatenlogischen Ausdrücken repräsentiert sind. Die Quasi-Logische Form eines Satzes umfasst *quasi* die Gesamtheit seiner Wahrheitsbedingungen. Es fehlt insbesondere noch Skopusinformation.⁵⁰ Ein klassisches Beispiel zur Illustration: Auf der Ebene der QLF hat die Frage⁵¹

Halten alle Professoren eine Vorlesung?

nur *eine* Repräsentation, die grob etwa so aussehen könnte:

⁴⁹ Eine starke Beschränkung ist, dass wie in der übrigen Grammatik des Systems keine Negation möglich ist.

⁵⁰ Im *Core Language Engine (CLE)* wird ebenfalls eine quasi-logische Form verwendet. Trotz einiger Übereinstimmungen zwischen dem *CLE* Ansatz und dem Saarbrücker Dialogsystem gibt es im Detail doch viele Unterschiede. So bleiben im Dialogsystem Pronomen auch in der Logischen Form stehen und werden erst bei der Auswertung aufgelöst, während es für den *CLE*-Ansatz charakteristisch ist, dass Anaphern und Ellipsen beim Übergang von Quasi-Logischer zur Logischen Form gemeinsam aufgelöst werden. Vergleiche (Eijck Alshawi 1992: 11).

⁵¹ Das Problem, wie Fragen im Gegensatz zu Aussagesätzen zu repräsentieren sind, klammere ich aus.

halten($\forall x.$ **professor**(x), $\exists y.$ **vorlesung**(y))

Auf der ausbuchstabierteren Ebene der LF sind zwei Repräsentationen nötig,

$$\forall x.(\mathbf{professor}(x) \rightarrow \exists y.(\mathbf{vorlesung}(y) \wedge \mathbf{halten}(x, y)))$$

$$\exists y.(\mathbf{vorlesung}(y) \wedge \forall x.(\mathbf{professor}(x) \rightarrow \mathbf{halten}(x,y)))$$

welche folgenden Lesarten entsprechen:

Gilt für alle Professoren, dass sie irgendeine Vorlesung halten?

Gibt es eine Vorlesung, die alle Professoren halten?

Die konkrete quasi-logische Formel des Ausdrucks

Hält Professor Barry eine Vorlesung?

sieht im Dialogsystem so aus:

```
wff(halten,
  [E,
    term(iota, X,
      wff(and, [wff(g_professor, [X]),
                wff(ident, [X, s_dozent, barr])])]),
    term(exists, Y, wff(g_vorlesung, [Y]))])
```

Der Nutzen der QLF liegt in ihrer Kompaktheit und zwar in zweifacher Hinsicht: Erstens stellt sie eine verdichtete Repräsentation der LF dar, weil aus der QLF immer ihre eventuell bedeutungsunterschiedlichen Logischen Formen berechnet werden können. Zweitens ist die Bedeutung der Komplemente kompakt innerhalb des Prädikats dargestellt: Im obigen Beispiel das Subjekt mit dem Jota-Quantor für definite Kennzeichnung und das existenz-quantifizierte Objekt. Die Auflösung von Frage-Ellipsen kann so auf der Ebene der QLF noch lokal erfolgen.

Der einfachste Fall sind Ersetzungsellipsen, wo in der QLF eine Teilformel durch eine andere Teilformel ersetzt wird. Bei der Auflösung der Nachfrage

Und Professor Uszkoreit?

wird zunächst die quasi-logische Formel für die Nominalphrase „Professor Uszkoreit“ berechnet:

```
term(iota, X,
  wff(and, [wff(g_professor, [X]),
            wff(ident, [X, s_dozent, uszkorei])])]),
```

Danach ergibt sich die QLF der Nachfrage durch eine simple Term-Substitution in der QLF der Vorfrage (Substituiertes fett):

```
wff(halten,
  [E,
    term(iota, X,
      wff(and, [wff(g_professor, [X]),
                wff(ident, [X, s_dozent, uszkorei])])]),
    term(exists, Y, wff(g_vorlesung, [Y]))])
```

Etwas aufwendiger sind Erweiterungsellipsen. Bei einer Nachfrage wie „Im Phonetikstudio?“ wird zunächst die quasi-logische Formel für die Präpositionalphrase „im Phonetikstudio“ berechnet:

$$E^{\wedge} \text{wff}(\text{in}, [E, \text{term}(\text{iota}, W, \text{wff}(\text{ident}, [W, s_{\text{ort}}, \text{phstud}])))])$$

Danach wird die QLF der Vorfrage entsprechend erweitert (Erweitertes fett):

```
wff(
  and,
  [wff(halten,
    [E,
      term(iota, X, wff(and, [wff(g_professor, [X]),
        wff(ident, [X, s_dozent, barr])])),
      term(exists, Y, wff(g_vorlesung, [Y]))]),
    wff(and, [wff(in, [E, term(iota, W, wff(ident, [W, s_ort, phstud])])])])])])
```

Wie das System entscheidet, dass es ersetzen oder erweitern muss, und was es ersetzen bzw. erweitern muss, bespreche ich im nächsten Abschnitt.

3.2.5.3 Diskurgedächtnis und Ellipsenbehandlung

Für die Behandlung von Frageellipsen und für die Auflösung von anaphorischen Bezügen braucht es Informationen, die über die Analyse einer einzelnen Äußerung hinaus erhalten bleiben. Um den notwendigen Kontext für den Dialog aufrechtzuerhalten, werden gewisse Daten im sogenannten Dialog- oder Diskurgedächtnis gespeichert. Es enthält folgende Informationen:

- **Diskursreferenten:** Es handelt sich um eine Liste der NP/PPs der jeweiligen Vorfrage, auf die man sich in aktuellen Fragen anaphorisch beziehen kann. Repräsentiert sind Sorte, syntaktische Funktion und QLF sowie bei Nominalphrasen noch Genus⁵².
- **Ellipseninfo:** Es handelt sich um eine Liste der NP/PPs der Vorfrage in exakt derselben Reihenfolge, wie sie in der Vorfrage erschienen sind. Repräsentiert sind Sorte, syntaktische Funktion und quasi-logische Formel sowie bei Nominalphrasen noch Kasus und bei Präpositionalphrasen noch die Modifikationsfunktion (Ort, Zeit) und die Präposition bei Präpositionalobjekten.
- **Vorfrage-Semantik:** Es handelt sich um die QLF der Vorfrage.

Normales Parsingmodul und Ellipsenmodul aktualisieren jeweils alle drei Gedächtnisteile. Die Diskursreferenten sind für die Anaphernbehandlung notwendig, Ellipseninfo und Vorfrage-Semantik für die Ellipsenauflösung.

Der Algorithmus soll nun spezifiziert werden: e_0, \dots, e_n seien die Konstituenten der elliptischen Nachfrage E und a_0, \dots, a_m die Bestandteile des Antezedensfrage A aus der Ellipseninfo. Fett formatierten Zeichen wie e_i , a_k , E , A bezeichnen die Be-

⁵² Numerus wird entgegen den Erwartungen nicht repräsentiert.

deutung in quasi-logischer Form. A ist dabei eine durch die Vorfrage-Semantik gegebene Grösse.

Fall Ersetzungsellipse: Für jedes e_i muss ein verträgliches a_j gefunden werden, wobei für alle Paare (e_i, a_j) und (e_{i+1}, a_k) die Bedingung $j > k$ gelten muss. Wenn das gelingt, ergibt sich die QLF von E aus der Gleichung:

$$E = A\{a_i / e_0, \dots, a_k / e_n\}$$

$A\{a_i / e_0, \dots, a_k / e_n\}$ bezeichnet die Substitution, die so gelesen wird: Ersetze in A alle Ausdrücke a_i bis a_k durch die entsprechenden e_0 bis e_n . Die neue Ellipseninfo A' wird durch folgende Gleichung gegeben: $A' = A\{a_i / e_0, \dots, a_k / e_n\}$.

Fall Erweiterungsellipse: Eine Erweiterungsfunktion *erweitere* kombiniert die Bedeutungen von e_0, \dots, e_n mit A :

$$E = \text{erweitere}(A, (e_0, \dots, e_n))$$

Die neue Ellipseninfo A' ergibt sich durch: $A' = a_0, \dots, a_m, e_0, \dots, e_n$.

Zuerst wird immer versucht Ellipsen im Antezedenssatz zu ersetzen. Nur wenn das scheitert, wird erweitert. Der verwendete Algorithmus ist sehr allgemein, und beinhaltet wenig linguistisches Wissen. Die Steuerungsmechanismen müssen in der Verträglichkeitsrelation und in den Datenstrukturen kodiert sein, letztere neigen deshalb zu Unübersichtlichkeit. Der rein semantische Ansatz bei der Substitution kann in seltenen Fällen zur Falle werden, wenn zwei verschiedene Satzglieder identische quasi-logische Formeln haben.

Aus der Spezifikation des Algorithmus sieht man nun sofort, warum als Nachfrage zum folgenden Dialog

- a) Hält Barry die Akustische Phonetik?
- b) Ja.
- c) Wo?
- d) Im Phonetikstudio.
- e) Und Uszkoreit die Grammatiktheorie wo?

die natürlichere Alternative (e') nicht zugelassen wird:

- (e') Und wo Uszkoreit die Grammatiktheorie?

3.2.6 Schlusskommentar

Die Beschränkung auf einen engen Diskursbereich mit wenigen Sorten sowie die eingeschränkte Sprache nehmen viel an potentieller Desambiguierungsarbeit vorweg. Die Auflösung von Ellipsen findet so einschränkende sortale Bedingungen bei Verben und Nominalen, dass Mehrdeutigkeiten kaum entstehen können. Wenig Spielraum lässt der Zwang, in Frage und elliptischer Nachfrage die gleiche lineare

Abfolge von Konstituenten einhalten zu müssen; es können unnatürliche Beschränkungen entstehen. Der Übergang von Konstituentenfragen zu Entscheidungsfragen ist manchmal problematisch. Iterierte elliptische Nachfragen können zu unterschiedlichen Vorstellungen führen, was alles erfragt wird.

Das System ist ein Musterbeispiel dafür, was eine kontrollierte Untersprache (*controled sublanguage*) durch ihre Reduktion ermöglicht. Ganz nach dem Motto: *Was man nicht kann fragen, darf man auch nicht sagen.*

3.3 Core Language Engine (CLE)

3.3.1 Überblick

Am CLE-Projekt⁵³ wurde zwischen 1986 und 1989 mit folgender Zielsetzung und Methodik gearbeitet:

„The Core Language Engine (CLE) is a general purpose device for mapping between natural language sentences and logical form representations of their meaning. [...] We adopted a modular staged design for the CLE in which explicit intermediate levels of linguistic representation are used as an interface between successive phases of analysis.“ (Alshawi 1992:1)

In *Abbildung 7* sind die Stufen der Verarbeitung eines Satzes schematisch dargestellt. Als linguistisch gut fundiertes und möglichst domänenunabhängiges Kernmodul – mit Analyse- und Generierungsfunktion – sollte die CLE in unterschiedliche Anwendungen eingesetzt werden: Natürlich-sprachliche Datenbankabfrage, Antwortgenerierung, Übersetzung usw. Die Pfeile in *Abbildung 7* deuten allerdings nur diejenige Laufrichtung der *engine* an, welche im Zusammenhang mit der Ellipsenauflösung interessiert.

In der CLE wird unterschieden zwischen kompositionell und kontextuell zu interpretierenden Satzbestandteilen. Diese Unterscheidung lässt sich aber nicht exakt auf einer Repräsentationsebene festmachen. Das folgende Zitat sollte trotzdem die Grundidee der unterschiedlichen Repräsentationsebenen erhellen.

„To summarize, we can view a QLF as giving a semantic analysis for an utterance, an RQLF as a representation of its interpretation in context and an LF as a representation of its truth conditions. (Alshawi 1992: 189)

⁵³ Das ganze System ist im Überblick in (Alshawi 1992) dargestellt, die Weiterentwicklung im Nachfolgeprojekt CLARE in (Alshawi, Carter et al. 1992b). Hinter dem Projekt stehen das *SRI International Artificial Intelligence Center*, das *Cambridge University Computer Laboratory* und private Firmen. CLE ist in PROLOG implementiert und nicht frei erhältlich. Information im WWW dazu: <<http://kidwelly.cam.sri.com/ccsrc/cle.html>>.

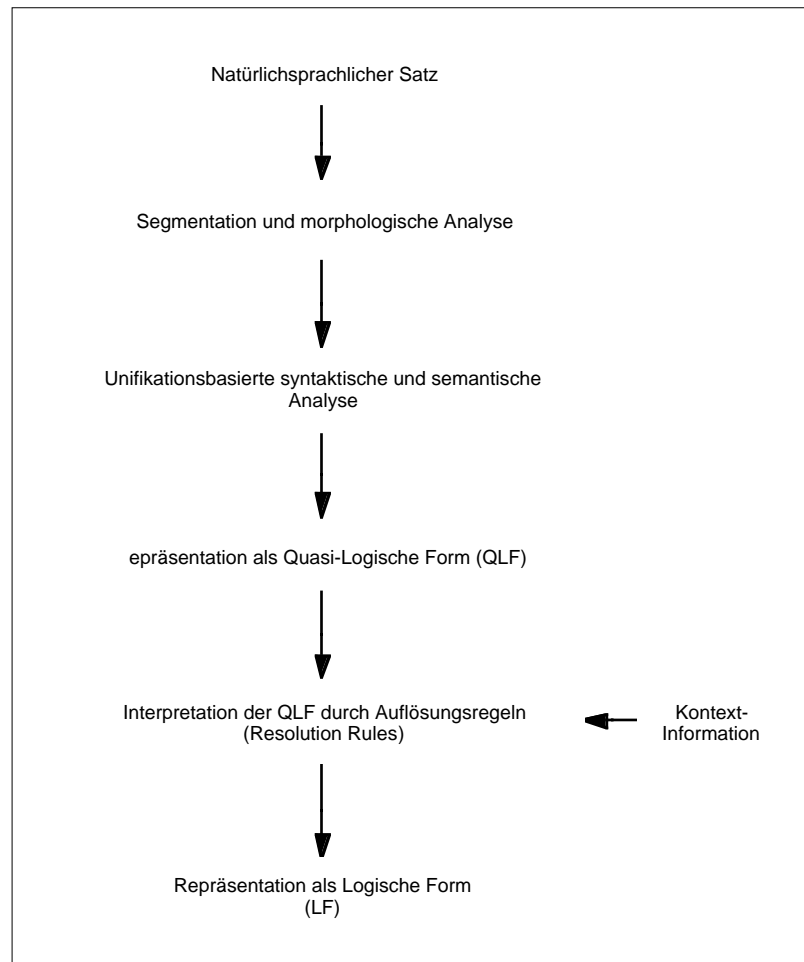


Abbildung 7: Schema der Satzanalyse in der CLE

Am einfachsten zu verstehen ist die Logische Form (LF), welche in (Alshawi 1992: 11ff.) mittels Prädikatenlogik erster Stufe dargestellt wird. Die LF ist zusätzlich erweitert mit Lambda-Abstraktion zur Bildung komplexer Funktionen sowie mit Lambda-Applikation. Weiter werden noch verallgemeinerte Quantoren und verschiedene Operatoren gebraucht, um die Semantik der natürlichen Sprache ausdrücken zu können.

Die Ausdrücke der Quasi-Logische-Form (QLF) stellen nach (Alshawi 1992: 33ff.) eine Erweiterung der Sprache der LF dar. Als zusätzliche Terme kommen quantifizierte Ausdrücke, wo der Skopus des Quantors noch nicht festgelegt ist, aber auch anaphorische Terme. Wichtig bezüglich Ellipsen sind sogenannte *a_forms* (steht für anaphorische Formeln), welche implizite, nicht aufgelöste Beziehungen darstellen. Dazu zählen nebst anderen Dingen auch Ellipsen wie *Did Mary?, an expensive one, more horses than John.*

Auf der Ebene der *resolved quasi logical form* (RQLF) stehen nach (Alshawi 1992: 187ff.) Auflösungsregeln (*resolution rules*) zur Verfügung, welche die speziellen Terme der QLF in die Sprache der LF zu übersetzen erlauben. Dabei werden auch satzübergreifende Bezüge mitverwendet. Die QLF-Terme, welche satzübergreifende Ellipsen wie etwa Nachfrage-Ellipsen (*follow-on questions*) repräsentieren, können dadurch aufgelöst werden.

Die CLE verfügt somit über einen ausgefeilten linguistisch-semantischen Apparat, um die Bedeutung eines Satzes innerhalb eines sprachlichen Kontexts zu gewinnen. Dies ermöglicht es, die manchmal als eher pragmatisch eingestufte Leistungen wie Ellipsenbehandlung in der ‘Kernsprache’ abzuhandeln:

„Use of the Core Language Engine as a general purpose natural language processor has enabled us to place the interpretation of elliptical input firmly within the linguistic processor and not within a general dialogue manager.“ (Lewin and Pulman 1995: 1)

3.3.2 Ellipsenbehandlung

Wie muss man sich die Ellipsenbehandlung nun konkret vorstellen? Ich möchte es am Beispiel folgender Nachfrage-Ellipse durchspielen:⁵⁴

- a) Was Mary nominated by Bill?
- b) Susan?

„Susan?“ wird von der Syntaxkomponente als elliptischer Satz erkannt und als anaphorische Formel in der QLF repräsentiert:

`a_form(<t=ell,p=vp>,P,[P,susan])`.

In den spitzen Klammern sind Merkmalspaare für den Typ (*t*) und die syntaktische Kategorie (*p*) der `a_form` vermerkt: Es handelt sich um eine Ellipse (`ell`) der Kategorie Satz (`vp`). `[P,susan]` repräsentiert die Bedeutung von „Susan“.

Der Antezedens-Satz (a) muss als RQLF-Interpretation vorliegen, um daraus und aus der QLF von (b) eine explizite Substitution zu berechnen. Es ist ein wichtiges Merkmal der CLE, dass die RQLF-Interpretation des Antezedens-Satzes verwendet wird. RQLFs sind im Gegensatz zu QLFs determiniert bezüglich Skopus, Bindungen und impliziten Relationen. Für elliptische Konstruktionen wird gemeinhin gefordert, dass sie bei struktureller Mehrdeutigkeit dieselbe Lesart wie die kontrollierende

⁵⁴ Vergleiche (Alshawi 1992: 251ff.). – Trotz vieler guter Erklärungen im Buch ist es doch recht schwierig, das Vorgehen im Detail zu verstehen. Der Report des Nachfolgeprojekts CLARE (Alshawi, Carter et al. 1992b) enthält mehr Material, allerdings sind auch die zugrundeliegenden Strukturen nochmals komplexer. Für das Beispiel verwende ich deshalb die Notation aus dem CLE.

Konstruktion haben. Dieser semantische Parallelismus ergibt sich recht natürlich, wenn man sich auf die RQLF bezieht.

Wie geht nun die Berechnung der Substitution vor sich? Die CLE macht dazu eine Anleihe bei (Dalrymple, Shieber et al. 1991), welche die Auflösung von Ellipsen als eine Anwendung von Unifikation höherer Ordnung betrachten. Die grundlegende Idee lässt sich recht einfach skizzieren.⁵⁵ Den Aspekt der Frage, Aktionsart, Tempus usw. blende ich dabei aus, um die Sache nicht unnötig zu komplizieren. Die ‘logische Form’ von (a) wird dann wie folgt repräsentiert:

nominate(bill, mary)

Die Ellipse „Susan?“ versteht man zunächst so, dass irgendeine Eigenschaft bezüglich *Susan* erfragt ist. Das kann mit der Prädikatsvariable P ausgedrückt werden:

$P(\text{susan})$

In der Antezedensfrage muss nun das gleiche Prädikat P ausgesagt sein, allerdings von einer andern Person. Wenn man von den zwei Möglichkeiten ‘bill’ und ‘mary’ letztere wählt, kann folgende Gleichung aufgestellt werden:

$P(\text{mary}) = \text{nominate}(\text{bill}, \text{mary})$

Lösungen zu solchen Gleichungen mit Prädikatsvariablen lassen sich durch Unifikation höherer Ordnung (*higher-order unification*) berechnen. Das gewünschte Resultat sieht folgendermassen aus:⁵⁶

$P = \lambda x. \text{nominate}(\text{bill}, x)$

Die Auflösung der Ellipse, d.h. die Bestimmung ihrer Bedeutung geschieht dann durch Lambda-Applikation, bzw. Konversion:

$\lambda x. \text{nominate}(\text{bill}, x) \text{ susan} \Rightarrow \text{nominate}(\text{bill}, \text{susan})$

In der CLE wird nicht direkt das Verfahren von (Dalrymple, Shieber et al. 1991) angewendet, sondern eine PROLOG-gerechte, eingeschränkere Lösung, welche die zu substituierenden Terme explizit berechnet. Im Beispielfall wäre es folgende Substitution:

[susan/marry]

Diese Substitution wird durch eine Regel auf die RQLF der Antezedensfrage angewendet und ergibt die RQLF der Ellipsenfrage. Für den Beispielfall heisst das im wesentlichen, dass alle marry-Terme durch susan-Terme ersetzt werden. Die

⁵⁵ Der Ansatz von (Dalrymple, Shieber et al. 1991) ist um einiges komplexer und verknüpft beweistheoretische Mittel mit der Ellipsenauflösung. (Alshawi, Carter et al. 1992b: 142) gibt eine genauere und trotzdem verständliche Zusammenfassung.

⁵⁶ Es werden auch unerwünschte Lösungen berechnet, im Beispielfall wäre $\lambda x. \text{nominate}(\text{bill}, \text{mary})$ ein weiteres mögliches Prädikat. Lösungen mit solchen leeren Variablenbindungen werden im Ansatz von (Dalrymple, Shieber et al. 1991) ausgefiltert.

RQLF der Ellipsenfrage lässt sich dann mittels Standardverfahren in die LF überführen.

Bei Ellipsen, die nicht satzübergreifend sind, wird nach analogen Auflösungsregeln vorgegangen:

„Roughly speaking, these [= satzinterne Auflösungsregeln] resolve the ellipsis *a_form* to a formula resulting from substituting arguments into a subformula of the RQLF interpretation being derived for the sentence.“ (Alshawi 1992: 258)

Ich möchte nochmals betonen, dass in der CLE die Ellipsenauflösung auf der Ebene der semantischen Repräsentation erfolgt: Genau ausgedrückt zwischen QLF und RQLF. Einzelne syntaktische Kategorien, die in der QLF noch vorhanden sind, werden für die Steuerung der Auflösung verwendet. Ein Vorteil dieses Ansatzes besteht darin, dass die Behandlung Ellipsen, nominalen Anaphern und Bindungsrelationen miteinander geschehen kann. Die CLE bietet für gewisse semantisch vertrackte Sätze recht saubere Lösungen an.

So lassen sich die unterschiedlichen Lesarten folgender Ellipsen durch eine kleine Variation der Substitutionspaare erzeugen.

John loves his mother, and Peter does too.

In der sogenannten *sloppy* Lesart (a), wo sich das elliptische Possessivpronomen auf das Nomen der Ellipse bezieht, wird über den Index *k* gebunden. In der *strict* Lesart (b) wird durch Referenzidentität gebunden.⁵⁷

a) John_i loves his_i mother, and Peter_k does ~~love his_k mother~~ too.

b) John_i loves his_i mother, and Peter_k does ~~love his_i mother~~ too.

In deutschen Nachfrage-Ellipsen wie (c) scheint mir allerdings die *sloppy* Lesart fast die einzig mögliche zu sein.⁵⁸

c) *F*: Liebt Hans_i seine_i Mutter? *A*: Ja. *F*: Und Peter auch?

3.4 HPSG-Ansatz

In (Lappin und Shih 1996) wird ein syntaktischer Ansatz zur Auflösung einer Reihe von Ellipsenkonstruktionen für das Englische vorgestellt. Der Algorithmus wurde zunächst für eine sogenannte Slot-Grammatik (McCord Bernth et al. 1992) entwickelt und später von Lappin(1998) in einer HPSG-Teil-Implementation verwendet.⁵⁹ Die Ellipsenbehandlung und die normale Grammatikregeln wurden dabei uniform in

⁵⁷ Die gezeigte Analyse hängt natürlich davon ab, dass das elliptische Material tatsächlich so ausbuchstabiert wird.

⁵⁸ Diese wird bei einem syntaktische Ansatz wie im UIS auch erzeugt.

⁵⁹ Dokumentation und PROLOG-Quell-Code zum HPSG ELLIP-Projekt ist erhältlich im WWW unter http://semantics.soas.ac.uk/ellip/hpsg_ellip.html.

ProFIT (Erbach 1995) spezifiziert. Der Ansatz umfasst folgende Konstruktions-typen:

- VP-Ellipsen (*VP-Ellipsis*):
John completed his paper before he expected to.
- Normales *gapping*
Sam teaches in London, and Lucy in Boston
- Pseudo-Gapping (*pseudo-gapping*)
John sent the flowers to Lucy before he did the chocolates.
- *Stripping* bzw. *bare-ellipsis* mit Komplementen und Adjunkten
Bill wrote reviews for the journal last year, and articles this year.

Der detaillierte Rekonstruktions-Algorithmus aus (Lappin und Shih 1996: 688) lässt sich etwa folgendermassen wiedergeben:

1. Übernehme den Verbalkopf des Antezedenssatzes als Verbalkopf der Ellipse.
2. Fülle die Subkategorisierungs-Slots des Ellipsenverbs mit den passenden NPs und PPs der Ellipse.
3. Fülle die restlichen Subkategorisierungs-Slots des Ellipsenverbs mit den entsprechenden Phrasen des Antezedenssatzes.
4. Übernehme diejenigen Adjunkte des Antezedenssatzes in die Ellipse, für die es in der Ellipse keine entsprechenden Adjunkte gibt.
5. Übernehme die restlichen Bestandteile der Ellipse, welche in Schritt 4 nicht eingesetzt wurden, als Adjunkte.

Die Schritte 1 bis 5 rekonstruieren aus einem elliptischen Gebilde einen vollständigen Satz. Interessante, aber von Lappin nicht weiter ausgeführte Punkte sind dabei, wie man „passende NPs“ oder „entsprechende Adjunkte“ erkennt, d.h. wie man mehrdeutig rekonstruierbaren Ellipsen zur intendierten Struktur verhilft. Im Englischen ist das Problem etwas weniger akut als im Deutschen, aber trotzdem vorhanden. Lappin(1996: 689) bringt mit dem Beispiel für *Stripping*

John gives Mary flowers, and chocolates too.

einen mehrfach rekonstruierbaren Satz. Mit einem Konditional wird die Mehrdeutigkeit jedoch elegant überspielt. Sinngemäss heisst es: Falls *chocolates* bei Schritt 2 in den Subkategorisierungs-Slot des direkten Objekts eingefüllt wird, dann füllen in Schritt 3 *John* und *Mary* die Position des Subjekt bzw. des indirekten Objekts auf.

Der Vorteil des syntaktischen Ansatzes von (Lappin und Shih 1996) ist seine Einfachheit. Durch die getrennte Analyse von Antezedenssatz und Ellipse taugt er für die Rekonstruktion von satzübergreifenden Ellipsen. Er entspricht in den Grundzü-

gen dem Verfahren, dass ich im UIS für die Behandlung der Frageellipsen implementiert habe. Für das Deutsche entstehen durch die freiere Wortstellung und den häufigen Fall von Ellipsen ohne finites Verb (*bare ellipsis*) oft strukturelle Mehrdeutigkeiten. Ob ein elliptisches Nomen Subjekt und direktes Objekt ist, kann auf Grund der Kasusmarkierung im Deutschen oft nicht entschieden werden. Der Ansatz von Lappin bietet dazu keine Hilfe an.

Eine interessante Weiterentwicklung des Ansatzes haben Lappin und Gregory in (Lappin und Gregory 1998) vorgenommen, wo eine Implementation beschrieben ist, welche die ausgelassenen Elemente *in situ* rekonstruiert. Für Ellipsen, die nicht satzübergreifend sind, scheint mir eine stärkere Einbettung der Ellipsenbehandlung in die 'normale' Syntax erstrebenswert. Es besteht allerdings die Gefahr, dass es zwischen den normalen Syntaxregeln und den elliptischen zu unliebsamen Inferenzen kommt.⁶⁰

3.5 UIS

Die Architektur des UIS und die Grundidee des *passage retrieval* habe ich in der Einleitung dargestellt. Es soll nun präsentiert werden, wie die Parsing-Komponente mit der Behandlung der Frageellipsen im Ellipsenmodul zusammenspielt und wie die Ellipsenbehandlung implementiert wurde.

3.5.1 Die F-Struktur als Zielformat

Der Weg von der schriftlichen Eingabe eines Satzes bis zur Konstruktion seiner F-Struktur ist in *Abbildung 9* schematisch dargestellt. In der Mitte der Aktivität steht die normale Syntaxanalyse, welche einen Kernbereich der deutschen Sprache verarbeiten kann.

⁶⁰ Im UIS-System, wo verschiedene Leute beteiligt sind, sprechen auch software-entwicklungstechnische Gründe für eine strikte Modularisierung.

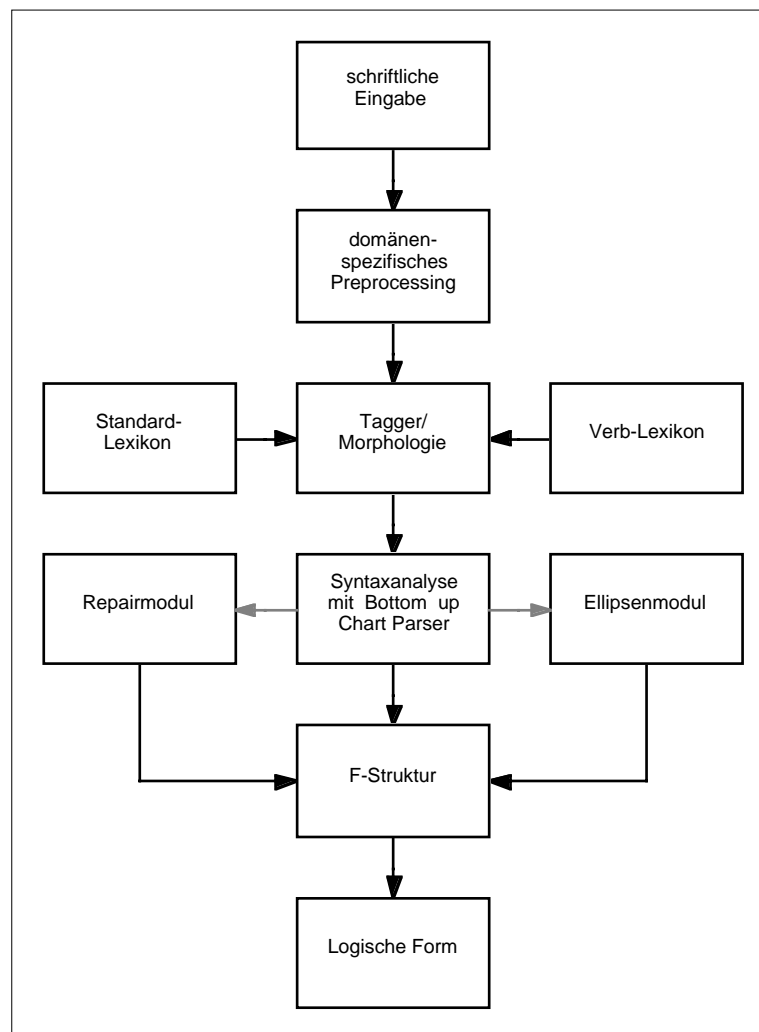


Abbildung 8: Konstruktion der F-Struktur im UIS

In natürlich-sprachlichen Abfragesystemen muss immer mit nicht ganz wohlgeformten Eingaben gerechnet werden. Auch wenn keine normale Syntaxanalyse möglich ist, sollten solche Eingaben einigermassen vernünftig umgesetzt werden – das System sollte sich robust verhalten. Das Repairmodul ist als Komponente des UIS gedacht, die das Scheitern einer normalen Syntaxanalyse auffängt, und auch fehlerhaften Input zu F-Strukturen verarbeiten kann. Das Ellipsenmodul ist eine weitere Komponente, die einspringt, wenn normales Parsen keine vollständige Satzstruktur aufbauen kann. Es macht sich dabei zwei Eigenschaften des Parsers geschickt zunutze:

- Als *bottom-up* Parser arbeitet er sich von den kleinsten lexikalischen zu den übergreifenden phrasalen Bestandteilen durch. Syntaktische Bruchstücke erhal-

ten damit soviel syntaktische Struktur, wie es grammatische Regeln für sie gibt.⁶¹

- Als *chart*-Parser merkt er sich alle gefundenen Strukturen, indem er sie in der *chart*-Tabelle ablegt. Wenn der *chart*-Parser scheitert, darf man sicher sein, dass alle möglichen Strukturen aufgebaut wurden.

Was sind nun F-Strukturen? Die Abkürzung ‘F’ steht für funktional und hat ihren Ursprung in der Lexikalisch-Funktionalen Grammatik (*Lexical Functional Grammar*), kurz LFG, welche anfangs der 80er-Jahre von R.M. Kaplan und J. Bresnan entwickelt wurde. F-Strukturen enthalten im Gegensatz zur reinen Konstituentenstruktur (C-Struktur) eine explizite Repräsentation der grammatischen Relationen⁶². Darunter versteht man insbesondere das Subjekt und die verschiedenen Objekte. F-Strukturen werden normalerweise – wie in unifiktions-basierten Grammatiken üblich – aus Merkmalstrukturen definiert.

Um die F-Strukturen, welche für das UIS verwendet werden, einfach beschreiben zu können, habe ich die Merkmalstrukturen mit ihren Substrukturen in eine kanonische Reihenfolge gebracht⁶³. Dies erlaubt es, die möglichen F-Strukturen kompakt in EBNF zu notieren.⁶⁴

```
Satz ::= Satzart Subjekt Objekte Praedikat SatzAdv SatzModi .
Satzart ::= "decl"
          | "jn_frage"
          | "e_frage"
          | "n_satz"
          | "relat" .
Subjekt ::= [NP] .
Objekte ::= Objekt1 Objekt2 PObjekt .
Objekt1 ::= [NP] .
Objekt2 ::= [NP] .
PObjekt ::= [PP] .
SatzAdv ::= ["AdverbLemma"] .
SatzModi ::= [PP] .
```

Abbildung 9: Beschreibung der F-Struktur des Satzes im UIS

⁶¹ Die Grammatik ist unifiktions-basiert und lexikalisch-funktional ausgerichtet, sie enthält Phrasenstruktur- und ID/LP-Regeln.

⁶² Manchmal spricht man auch von Satzfunktionen.

⁶³ In einer gewissen Korrespondenz zur Implementation in PROLOG.

⁶⁴ Eine Darstellung in Diagrammform aller verwendeten F-Strukturen im UIS sind im Anhang A aufgeführt.

decl	pos (0, 8)
--- subj	pos (3, 4)
----- spec_typ (INDEF)	
----- spec_lem (man)	
----- num (SG)	
----- pers (3)	
----- gender (_975)	
----- head (pron)	pos (3, 4)
----- modi ([])	
----- modi ([])	
--- obj	pos (0, 2)
----- spec_typ (DEF)	pos (0, 1)
----- spec_lem (das)	
----- num (SG)	
----- pers (3)	
----- gender (NEUTR)	
----- head (Vor les~ung\s#ver zeich-nis)	pos (1, 2)
----- modi (_905)	
----- modi (_901)	
--- obj2 ([])	
--- pobj ([])	
--- pred	pos (2, 3)
----- head (be komm~en)	pos (2, 3)
----- perg ([])	
----- modal ([])	
----- tense (PRAES)	
----- subcat (A000)	
--- modi (_97963)	
--- modi	pos (4, 7)
----- spec (in)	pos (4, 5)
----- head	pos (5, 7)
----- spec_typ (DEF)	pos (5, 6)
----- spec_lem (die)	
----- num (SG)	
----- pers (3)	
----- gender (FEM)	
----- head (Kanzl~ei)	pos (6, 7)
----- modi (_781)	
----- modi (_777)	

Abbildung 10: F-Struktur eines Satzes im UIS

Abbildung 9 zeigt die Definition der F-Struktur von Sätzen. Man sieht, dass die Grammatik zum Zeitpunkt der Implementation der Ellipsenbehandlung gewisse Einschränkungen aufgewiesen hat. So ist beispielsweise nur ein Satzadverb (SatzAdv) und ein Satzmodifikator (SatzModi) zugelassen.

In Abbildung 10 ist die F-Struktur des Satzes „Man bekommt das Vorlesungsverzeichnis in der Kanzlei“ im *pretty-printing* aufgelistet. Zusätzlich zur F-Struktur-Information ist noch eine Positionsangabe (pos) vermerkt, die die lineare Erstreckung der einzelnen Komponenten angibt. Bezugseinheit sind die *Token* des Satzes. Implementiert ist eine F-Struktur als einzelner, komplex geschachtelter PROLOG-Term.

3.5.2 Ellipsenauflösung als F-Struktur-Rekonstruktion

Wie werden Nachfrageellipsen im UIS aufgelöst? Für Menschen ist das Verständnis von F2 als Nachfrage zu F1 im mündlichen Verkehr meist schon durch den wieder-aufnehmenden Tonfall offensichtlich. Wie bringt man ein NLP-System mit schriftlichem Input wie das UIS dazu, dass es die elliptische Frage F2 zu „Wo bekommt man das Studienreglement?“ vervollständigt?

F1: Wo bekommt man das Vorlesungsverzeichnis? ...

F2: Und das Studienreglement?

Nun, für die Frage F1 kann der Parser eine vollständige F-Struktur berechnen. Bei der Analyse von F2 kann die normale Syntaxkomponente keine Satzstruktur aufbauen und scheitert. Immerhin hat der *Bottom-up*-Parser dabei feststellen und als *Chart* Parser auch festhalten können, dass eine Konjunktion und eine NP vorliegt. Die F-Struktur der Vorfrage und die syntaktische Analyse der elliptischen Nachfrage reichen aus, um eine vollständige Frage – in Form ihrer F-Struktur – zu rekonstruieren. Der eingeschränkte Kommunikationshorizont eines Anfragesystems machen es sinnvoll und kooperativ, aus einer Frage-NP mithilfe der Vorfrage eine vollständige Frage zu ergänzen. Ob dabei die intendierte Frage entsteht, wenn man nur eine einzige F-Struktur als Lösung rekonstruieren darf, ist ein anderes Thema.

Die Ellipsenbehandlung auf der Ebene der F-Struktur profitiert von deren grösseren Abstraktheit im Vergleich zur Konstituentenstruktur: Oberflächenmerkmale wie lineare Abfolge der Satzglieder, Verbstellung, Subjekt-Finitum-Kongruenz sind aufgelöst bzw. normalisiert und müssen beim Rekonstruieren nicht miteinbezogen werden.

3.5.2.1 Erklärungen zur Implementation

Im folgenden soll nun meine Implementation des Ellipsenmoduls diskutiert werden. Anhang B enthält den vollständigen und recht sorgfältig kommentierten Programm-Code.

Als Frageellipsen sind nur Konstituentenfolgen zugelassen, die bestimmten Regeln folgen. Diese lassen sich als kleine Ellipsengrammatik darstellen:

$$e ::= [\text{Konjunktion}] (\text{NP}|\text{PP}) [\text{NP}|\text{PP}] [\text{Adverb}] (<\text{Q}>|<\text{S}>)$$

$$| [\text{Konjunktion}] [\text{Adverb}] (\text{NP}|\text{PP}) [\text{NP}|\text{PP}] (<\text{Q}>|<\text{S}>)$$

Im Gegensatz zum normalen Parser, der *bottom-up* arbeitet, werden die elliptischen Konstituenten im Ellipsenmodul *top-down* zusammengesetzt.⁶⁵ Für jede der beiden Zeilen der Ellipsengrammatik gibt es eine Klausel des Prädikats `el_behandle_`

⁶⁵ Dabei wird mit den Zwischenpositionen der Token im Sinne von *Definite Clause Grammar* gearbeitet.

ellipse/3, das die Auflösung komplett steuert. Als Input nimmt sie die Position nach dem letzten Token der Ellipseingabe und die F-Struktur der Vorfrage. Die Position des letzten Tokens braucht es, damit wirklich die ganze Ellipse abgearbeitet wird. Die Klausel für die erste Ellipsenregel sieht folgendermassen aus:

```
el_behandle_ellipse(To,DG_FS,FS):-
    el_ellipsen_ende(To,Token),
    el_ellipsen_konjunktion(0,From,KonjLemma),
    To1 is To - 1,
    el_ellipsen_adverb(To1,To2,AdvLemma),
    el_ellipsen_konstituenten(From,To2,DG_FS,Konstituenten,Switch),
    el_ersetze_konstituenten(Switch,DG_FS,Konstituenten,FS0),
    el_korrekt_satzadverb(FS0,KonjLemma,AdvLemma,FS1),
    el_korrekte_satzart(Token,FS1,FS).
```

el_ellipsen_ende/2 wird dazu verwendet, um am Schluss die korrekte Satzart der Ellipse einzustellen (el_korrekte_satzart/3). el_ellipsen_konjunktion/3 und el_ellipsen_adverb/3 extrahieren allfällige Konjunktionen bzw. Adverbien, die bei der Bestimmung des korrekten Satzadverbs in el_korrekt_satzadverb/4 notwendig sind. Der Kern der Ellipsenbehandlung steckt in den beiden Prädikaten el_ellipsen_konstituenten/5 und el_ersetze_konstituenten/4.

Den Aufruf des Unterziels

```
el_ellipsen_konstituenten(From,To2,DG_FS,Konstituenten,Switch)
```

möchte ich etwas genauer betrachten. Wie erwähnt, speichert der normale *bottom-up*-Parser beim Versuch, die Ellipse als Satz zu analysieren, alle erkannten Teilstücke in die *Chart*. Mithilfe von `From`, `To2` lässt sich nun überprüfen, ob zwischen diesen Positionen geeignete Phrasen gefunden wurden. Falls ja, müssen diese Phrasen auch noch mit passenden Satzfunktionen in der F-Struktur der Vorfrage verträglich sein. Diese wird darum via Argument `DG_FS` eingespiesen. Allfällig passende Phrasen werden dann in ihrer Satzfunktion bestimmt und an die Variable `Konstituenten` gebunden zurückgegeben.⁶⁶

el_ersetze_konstituenten substituiert die als passend ausgemachten Konstituenten in die F-Struktur der Antezedensfrage und erzeugt so die vervollständigte F-Struktur der Ellipse. Der Termunifikations-Mechanismus von PROLOG kann bei

⁶⁶ Das Argument `Switch` ist reiner Komfort, der beim späteren Ersetzen mühsames *Pattern-Matching* in die verschachtelten Datenstrukturen erspart. – NPs und VPs werden einfach als Listen ihrer Merkmale repräsentiert und haben leider keinen identifizierenden Hauptfunktork. Das würde die Lesbarkeit des Codes massiv erhöhen und zudem funktorspezifisches *pretty-printing* dieser PROLOG-Terme via Systemausgabe beim *Debugging* erlauben.

diesen Verträglichkeitstests und allfälligen Instantiierungen optimal ausgenützt werden.

Das Prinzip ist also recht einfach und funktioniert auch, aber der Teufel hält sich bekanntlich in den Details versteckt.

3.5.3 Umgang mit Mehrdeutigkeiten

Das schon oben verwendete Beispiel einer Nachfrage-Ellipse ist für Menschen eindeutig zu interpretieren: F2 muss „Wo bekommt man das Studienreglement?“ bedeuten.

F1: Wo bekommt man das Vorlesungsverzeichnis? ...

F2: Und das Studienreglement?

Wenn man vom Aspekt der Bedeutung abstrahiert, zeigt sich aber, dass aus morphosyntaktischen Überlegungen die Variante „Wo bekommt das Studienreglement das Vorlesungsverzeichnis?“ genau so plausibel ist. Tatsächlich ist die ‘falsche’, d.h. nicht intendierte zweite Lesart, diejenige, welche aus rein morphosyntaktischen Gründen rekonstruiert würde.

Es ist nämlich sinnvoll, beim Suchen der verträglichen Satzfunktionen einer bestimmten Präferenzrelation nachzufolgen. Für Nominalphrasen lautet die Reihenfolge:

Subjekt < direktes Objekt/Prädikatsergänzung < indirektes Objekt/Objekt II

Alleiniges Kriterium stellt dabei die Kasusmarkierung dar. Da der Kasus im Deutschen morphologisch oft nicht eindeutig markiert ist, können z.B. intendierte direkte Objekte in die Subjektsposition geraten. In vielen Fällen liefert die Präferenzrelation jedoch gute Resultate. Für die Behandlung von *Gapping*-Konstruktionen – aus der Perspektive der Ellipsenauflösung auch ‘Doppelellipsen’ genannt – muss zusätzlich die lineare Abfolge in der Ellipse berücksichtigt werden, d.h. was in der elliptischen Frage zuerst kommt, wird zuerst eingesetzt bei der Rekonstruktion.

Im folgenden Beispiel wird F2 korrekt zu „Brauchen die Auditoren ein Reglement?“ aufgelöst und nicht zu „Braucht ein Reglement die Auditoren?“.

F1: Brauchen alle Studenten ein Vorlesungsverzeichnis? A: ...

F2: Und die Auditoren ein Reglement?

Wenn anstelle von F2 die Frage F2’ gestellt wird, entsteht – nach diesen Regeln – die syntaktisch korrekte, aber nicht besonders intuitive Frage „Was braucht die Auditoren?“.

F2': Und was die Auditoren?

Bevor ich meinen Lösungsvorschlag für eine verbesserte Behandlung von Mehrdeutigkeiten vorstelle, soll noch die Zuordnung von Präpositionalphrasen besprochen werden.

Präpositionalphrasen können ebenfalls an mehreren Orten im Satz auftreten: Als Präpositionalobjekte, Prädikatsergänzungen sowie als Satzmodifikatoren.⁶⁷ Da Präpositionalobjekte und Prädikatsergänzungen nicht gekoppelt auftreten und die F-Struktur so restringiert ist, dass nur ein Satzmodifikator erscheinen darf, ist das Problem weniger akut.⁶⁸ Trotzdem ist eine Festlegung der Reihenfolge sinnvoll:

Präpositionalobjekt/Prädikatsergänzung < Satzmodifikator

Als Entscheidungskriterium dienen die Präpositionen, die bei Verben mit Präpositionalobjekt meist fixiert und im Lexikoneintrag vermerkt sind.

Wenn mehrere Satzmodifikatoren zugelassen werden, was absolut notwendig ist für eine allgemeinere Syntaxkomponente, entsteht eine systematische Quelle von syntaktischer Mehrdeutigkeit. Bei Nachfrage-Ellipsen kann aufgrund syntaktischer Gegebenheiten nicht entschieden werden, ob wie in F2 das „heute morgen“ genauer spezifiziert wird zu „heute morgen um 8 Uhr“ oder ob wie in F3 „heute morgen“ durch „am Dienstag“ zu ersetzen ist.

F1: Kommt Gerda heute morgen? A: ...

F2: Um 8 Uhr?

F3: Oder am Dienstag?

Die Anforderung, dass das Ellipsenmodul als einzige oder erste Lösung die intendierte liefert, kann allein mit syntaktischen Kriterien unmöglich erfüllt werden. Die Mehrdeutigkeit müsste vom Ellipsenmodul als solche weitergegeben und erst auf einer Stufe der semantischen Verarbeitung disambiguiert werden. Ganz ähnlich wie in der CLE, wo Quasi-Logische Formen verschiedene potentielle Logische Formen enthalten.

Ellipsen können auch von denselben Problemen struktureller Mehrdeutigkeit befallen werden wie vollständige Sätze. Das sogenannte Problem des *PP-Attachment* (PP-Anbindung) tritt im Ellipsenmodul ebenfalls auf. In F2 kann es sich um eine Doppelellipse (F2') handeln mit separater NP und PP. „im Wald“ wäre dann als Modifikator des Satzes zu analysieren. „den Mann im Wald“ kann in F2 auch als komplexe NP aufgefasst werden, was einer Einzel-Ellipse (F2'') entsprechen würde.

F1: Sieht sie das Kind? A: ...

⁶⁷ Satzmodifikatoren werden oft als Adjunkte bezeichnet. Im Gegensatz zu Komplementen können sie in beliebiger Zahl auftauchen.

⁶⁸ Das nicht ellipsenspezifische Problem, dass Präpositionalphrasen Bestandteil komplexer Nominalphrasen oder Satzmodifikatoren sein können, diskutiere ich weiter unten.

F2: Und den Mann im Wald?

F2': Sieht sie [_{NP}den Mann] [_{PP}im Wald]?

F2'': Sieht sie [_{NP}den Mann [_{PP}im Wald]]?

Durch Vertauschen der Klauseln des Prädikats `el_ellipsen_konstituenten/5` können global gewisse Lesarten bevorzugt werden. Das Problem bleibt aber bestehen.

Das Saarbrücker Dialogsystem kann durch die starke Einschränkung der Sorten in den einzelnen Satzgliedern Mehrdeutigkeiten eliminieren und liefert damit erstaunlich gute Resultate. Im UIS, das einen grösseren sprachlichen Abdeckungsgrad (*coverage*) anstrebt, können die Verbal- und die Substantivbedeutungen nicht einfach willkürlich beschnitten werden. Sortenbeschränkungen haben mit dem Anwendungsbereich von Sprache zu tun und gehören zur Pragmatik eines Systems. Die Syntaxkomponente des UIS – und dazu gehört auch das Ellipsenmodul – soll gegenüber den Anwendungsbereichen möglichst neutral sein.

Mit der verwendeten kasusbezogenen Präferenzrelation für NPs liefert die Auflösung von Einzel-Ellipsen als direkten Objekten aber so viele unbefriedigende und unintuitive Resultate, dass etwas getan werden musste. Da sortale Beschränkungen aus den obenerwähnten Gründen schlecht möglich sind, habe ich einen andern Ansatz entwickelt, der von sortalen Präferenzen ausgeht. Als Motivation möchte ich das bereits bekannte Beispiel wiederaufnehmen:

F1: Wo bekommt man das Vorlesungsverzeichnis? ...

F2: Und das Studienreglement?

Zwischen „Studienreglement“ und „Vorlesungsverzeichnis“ besteht ein stärkerer inhaltlicher Bezug als zwischen „man“ und „Studienreglement“. Diese Affinität könnte zur Steuerung der Substitution des elliptischen Ausdrucks ausgenutzt werden. Zwei Probleme tauchen dabei auf:

- Wie lässt sich die Affinität repräsentieren? Ein aufwendiger Ansatz besteht im Erstellen einer ontologischen Strukturierung des Lexikons. Aber auch ein guter Thesaurus könnte dazu dienen.⁶⁹
- Ist die stärkere Affinität zwischen dem zu Substituierenden regelhaft? Die sinnvolle Interpretation einer Äusserung hängt letztlich immer von der Situation ab. Wenn man eine elliptische Nachfrage oder Koordinationsfrage stellt, muss aber aus kommunikativer Sicht ein thematischer Zusammenhang bestehen. Im besten

⁶⁹ Ein Thesaurus braucht es im UIS sowieso, da beim *passage retrieval* nicht vorausgesetzt werden kann, dass die Benutzenden genau die im Text vorkommenden Wörter in der Frage verwenden.

Fall kann die Verwendung der Affinität diesen thematischen Zusammenhang ausnützen und eine valable Heuristik abgeben.

Im Ellipsenmodul habe ich diese Ideen zu implementieren versucht gemäss folgender heuristischer Regel:

Präferiere die Ersetzung von ontologisch Nahem.

Zur Berechnung der ontologischen Nähe von Nominalphrasen, verwende ich einen Minithesaurus, der die Lexeme in hypernymisch strukturierte Synonymklassen einteilt.⁷⁰ Der einfache Ellipsenbehandlungs-Algorithmus, der ganz auf den Backtracking-Mechanismus von PROLOG als Suchstrategie setzen konnte, musste abgeändert werden. Konkret werden beim Ausfindigmachen von passenden Satzfunktionen durch das Prädikat `el_kompatible_np_konstituenten/4` alle verträglichen Lösungen auf einmal berechnet (`find_all`). In der Lösungsliste, die gemäss der NP-Präferenz geordnet ist, wird dann das erste ontologisch nahe Ersetzungspaar ausgewählt. Falls keine ontologische Präferenz auszumachen ist, wird die normale kasusgesteuerte NP Präferenz benutzt.

Für die Verwendung mit einem grösseren Thesaurus muss die Suchstrategie allenfalls noch verfeinert werden. Immerhin liefert das Ellipsenmodul mit dieser Heuristik in seiner gegenwärtigen Form intuitiv ansprechendere Resultate.

3.5.4 Behandlung der Satznegation

Bei der Besprechung der *Gapping* und *Stripping*-Phänomene in Abschnitt 2.2.4 habe ich aufgezeigt, dass sich die Satznegation in Ellipsen etwas anders verhält als normale Satzadverbien. Im Ellipsenmodul sind diese Erscheinungen berücksichtigt. Die Nachfrage F2 wird als F2' interpretiert:

F1: Brauchen die Professoren das Vorlesungsverzeichnis nicht? A:...

F2: Aber die Studenten?

F2': Aber brauchen die Studenten das Vorlesungsverzeichnis?

Die Regel dahinter lautet: Ignoriere die Satznegation der Vorfrage, wenn eine Nachfrage mit restriktiver nebenordnenden Konjunktion wie „aber“ oder „doch“ eingeleitet wird.⁷¹

⁷⁰ Wenn der Thesaurus umfangreich ist, kann die Berechnung von Nähe zwischen Lexemen, die letztlich doch nicht zusammenhängen, sehr aufwendig werden. Die Suchtiefe sollte dann durch einen *threshold* begrenzt werden. Für solche Zwecke ist im Prädikat `el_ontologie_ein/1`, das den Gebrauch der heuristischen Regel steuert, das Argument gedacht. – Koordinierte NPs sind eine zusätzliche Quelle von Komplexität, da nicht zwei Lexeme, sondern eigentlich zwei Lexemlisten untersucht werden müssen.

⁷¹ Vergleiche Grammatikduden §654 (Drosdowski 1984) für die Unterteilung der Konjunktionen.

Bei kopulativen Konjunktionen wie „und“, „auch“ usw. bleibt die Satznegation bestehen, ausser ein konzessives Adverb wie „doch“, „dennoch“, „trotzdem“ usw. erscheint. Im folgenden Beispiel versteht das Ellipsenmodul F2 als „Und liest sie die NZZ nicht?“ und F2' als „Und liest sie die NZZ trotzdem?“.

F1: Liest sie den Blick nicht? A:...

F2: Und die NZZ?

F2': Und die NZZ trotzdem?

Die allfällige Elimination der Satznegation wird durch das Prädikat `el_korrektes_satzadverb/4` gesteuert.

3.5.5 Mögliche Erweiterungen der Ellipsenauflösung

So wie das Ellipsenmodul jetzt implementiert ist, behandelt es die wichtigsten Fälle von Frage-Ellipsen: *Stripping* und *Gapping*.

Die Behandlung von *Sluicing* würde es den Benutzenden erlauben, indirekte Fragen zu stellen. Dieses Manko ist nicht so schwerwiegend, da wie im Abschnitt 2.1 über Mensch-Maschine-Kommunikation erwähnt, die Benutzenden sich dem Computer gegenüber relativ leicht das Stellen von direkten Fragen angewöhnen.

VP-Ellipsen sind in Frageellipsen eher selten im Deutschen ebenso wie die nicht-phrasale Koordinationsreduktion. Ihre Behandlung könnte zu unerwünschten Interferenzen mit dem Repairmodul führen, da in solchen Ellipsen finite Verben vorhanden sind.

N-Ellipsen sind häufiger, und deren Behandlung wäre wünschenswert, aber vermutlich auch aufwendig. Satz-Ellipsen erscheinen dem UIS-Parser nicht eigentlich als unvollständig, da er blosse Nebensätze akzeptiert. Für die semantische Weiterverarbeitung ist die Auflösung des elliptischen Hauptsatzes allerdings Bedingung.

Nebst diesen möglichen Erweiterungen lässt sich die Behandlung von *Stripping* und *Gapping* noch optimieren. Wenn auf der F-Struktur Aktiv und Passiv in ein Diathesen-Merkmal aufgelöst wird, ist das miteinzubeziehen. Die PP „Von wem“ in F2 muss bei der Ellipsenauflösung zu einer NP in Subjektsfunktion gemacht werden.

F1: Werden die Reglemente verschickt? A: ...

F2: Von wem?

Wenn die F-Struktur mehrere Satzmodifikatoren enthalten kann, muss das vom Ellipsenmodul integriert werden können:

F1: Bekommt man die Studienbestätigung in der Kanzlei? A:...

F2: Wann?

Die spezielle Behandlung der Satznegation bei restriktiven Konjunktionen bzw. konzessiven Adverbien funktioniert erst mit „nicht“. Verschmolzene Satznegation via „kein“ oder „nie“ wird nicht aufgelöst.

Die Beschränkung auf die Auflösung von maximal zwei elliptischen Konstituenten ist nicht syntaktisch begründet. Im Deutschen kann es *Gapping* mit drei Konstituenten geben. Auch Nachfrage-Ellipsen mit drei Konstituenten sind konstruierbar, wie folgendes Beispiel zeigt:

F: Gibt sie ihrer Tochter jeden Tag Wein? A: ...

F: Und seit wann ihrem Sohn Whisky?

Solche eher seltenen Fälle könnten ohne konzeptionelle Änderungen ins Ellipsenmodul eingebaut werden.

3.5.6 Fazit

Ich habe für das UIS ein Ellipsenmodul implementiert, das die wichtigsten Formen von Nachfrage-Ellipsen auf der syntaktischen F-Struktur behandelt. Durch die Benutzung von ontologisch/thesaurischer Information in einer heuristischen Präferenzregel konnte die Auflösung zur intendierten Frage verbessert werden. Inwiefern dieses Verfahren auch mit grossen ontologisch/thesaurischen strukturierten Domänen befriedigend funktioniert, muss noch geprüft werden.

Bei der Implementation des Ellipsenmoduls habe ich grossen Wert auf die Wartbarkeit gelegt und einfache Lösungen bevorzugt. Das Ellipsenmodul ist stark abhängig vom normalen Parser und dessen Datenstrukturen. Änderungen in diesem Bereich haben direkten Einfluss auf die Leistung des Ellipsenmoduls. Mit Datenkapselung und Modularisierung lässt sich dieser Einfluss lokal einschränken, aber nicht verhindern. So wie viele linguistischen Analyseansätze von Ellipsen untrennbar von der Grammatiktheorie sind, auf der sie aufsetzen, bleibt auch die Ellipsenbehandlung ein vom Parser abhängiges Produkt.

4 Schluss

Ich habe in meiner Arbeit Ellipsen aus unterschiedlichen Perspektiven beleuchtet. Sie haben sich dabei als ein schillerndes Phänomen erwiesen, dass syntaktische, semantische, pragmatische und phonologische Fragen tangiert. Für den engeren Bereich der syntaktisch kontrollierten Ellipsen gibt es – zumindest im angelsächsischen Raum – seit dreissig Jahren detaillierte Versuche, die exakten grammatischen Bedingungen für Elliptizität zu spezifizieren. Allgemein anerkannte Ergebnisse sind aus diesen Bemühungen leider nicht entstanden, da die Eigenheiten der betrachteten Einzelsprachen und verwendeten Sprachtheorien zu stark mitspielen. Ersteres gilt insbesondere für die deutsche Sprache, die viele Freiheiten bezüglich Ellipsen gestattet.

Ich habe aufgezeigt, dass verschiedene Arten von satzübergreifenden Frage-Ellipsen analog zu bekannten satzinternen elliptischen Konstruktionstypen einzuordnen sind. Für die gängigsten Konstruktionstypen *Gapping* und *Stripping* habe ich im Rahmen des UIS-Projekts ein Modul zur Behandlung solcher Frage-Ellipsen entworfen und implementiert. Für *passage retrieval* wie es im UIS-Projekt vorgesehen ist, schränkt sich der sinnvolle Kontext der Ellipsenauflösung auf allfällige Vorfragen ein. Der gewählte Ansatz rekonstruiert das elliptische Material dabei auf der syntaktischen Ebene der F-Struktur mithilfe der vollständigen F-Struktur der Vorfrage.

Wenn nur die Information der F-Struktur zur Verfügung steht, kann der Anspruch an das Ellipsenmodul, die intendierte Lesart als einzige oder zumindest als präferierteste Lösung zu liefern, aufgrund der vielen morphosyntaktischen Mehrdeutigkeiten nicht eingelöst werden. Rigide sortale Einschränkungen bei Verben und Substantiven, wie sie in Systemen mit stark eingeschränktem Anwendungsbereich möglich sind, können in dieser Form nicht auf das UIS übertragen werden. Um das Problem der strukturellen Mehrdeutigkeit etwas eindämmen zu können, habe ich einen Ansatz entwickelt, der nicht mit Einschränkungen, sondern mit semantischer Affinität von Lexemgruppen arbeitet. Im Minimalfall kann ein anwendungsbereichsspezifischer Thesaurus mit Synonymklassen und rudimentärem Hypernymiekonzept für die Bestimmung der Affinität verwendet werden.

Inwiefern die Auflösung von Nachfrage-Ellipsen von den zukünftigen Benutzenden des UIS eingesetzt wird, lässt sich zum jetzigen Zeitpunkt nicht sagen. Als

Abfragesystem unterliegt es vermutlich etwas anderen Kommunikationsbedingungen als echte Dialogsysteme.

Der von mir entwickelte Ansatz der Ellipsenauflösung auf der Ebene der F-Struktur könnte mit geeigneten Modifikationen auch zur Behandlung von gewissen satzinternen Ellipsen eingesetzt werden.

5 Literatur

- ALSHAWI, H. (Hg.): The Core Language Engine. Cambridge, Mass. 1992. (= ACL-MIT Press series in natural language processing)
- ALSHAWI, H./ CARTER, D./ CROUCH, R. ET AL.: CLARE: A Contextual Reasoning and Cooperative Response Framework for the Core Language Engine. Bericht Nr.CRC033, SRI International, Cambridge Computer Science Research Centre, December 1992b.
- CHAO, W.: On ellipsis. New York 1988. (= Outstanding Dissertations in Linguistics)
- CHOMSKY, N.: Syntactic Structures. Den Haag 1957.
- DALRYMPLE, M./ SHIEBER, S.M. / PEREIRA, F.C.N.: Ellipsis and Higher-Order Unification. In: Linguistics and Philosophy 14. (1991) S. 399-452.
- DENTLER, S.: Verb und Ellipse im heutigen Deutsch. Universität Göteborg 1990. (= Göteborger Germanistische Forschung 31)
- DROSDOWSKI, G. (Hg.): Duden "Grammatik der deutschen Gegenwartssprache". 4., völlig neu bearbeitete und erweiterte Auflage. Mannheim/Wien/Zürich 1984. (= Der Duden in 10 Bänden; Band 4)
- EIJCK, J.V. / ALSHAWI, H.: Logical Forms. In: H. Alshawi (Hg.): The Core Language Engine. Cambridge, Mass. 1992. S. 11-39. (= ACL-MIT Press series in natural language processing)
- EIKMEYER, H.-J.: Ellipsen und Analysestrategien in inkrementellen Sprachverarbeitungsmodellen. In: R. Meyer-Hermann und H. Rieser (Hg.): Ellipsen und fragmentarische Ausdrücke. Tübingen 1985. S. 1-25. (= Linguistische Arbeiten 148)
- ERBACH, G.: ProFIT: Prolog with Features, Inheritance and Templates. In: Seventh Conference of the European Chapter of the Association for Computational Linguistics. Dublin 1995. S. 180-187.
- FAUSER, A. / RATHKE, C.: Studie zum Stand der Forschung über natürlichsprachliche Frage/Antwort-Systeme. Karlsruhe 1981. (= Bundesministerium für Forschung und Technologie. Forschungsbericht ID 81-006)
- GRICE, H.P.: Logic and Conversation. In: P. Cole und J. Morgan (Hg.): Syntax and Semantics: Speech Acts. Bd. 3 New York 1975. S. 41-58.
- HANKAMER, J.: 1971. In: Linguistic Inquiry 4. (1973) S. 17-68.

- ISSATSCHENKO, A.V.: Kontextbedingte Ellipse und Pronominalisierung. In: W. Dresslerl (Hg.): Textlinguistik. Darmstadt 1978. S. 79-92. (= Wege der Forschung 427)
- JÖNSSON, A.: Dialogue Management for Natural Language Interfaces. Linköping 1993. (Diss.) (= Linköping Studies in Science and Technology. No. 312)
- KINDT, W.: Grammatische Prinzipien sogenannter Ellipsen und ein neues Syntaxmodell. In: R. Meyer-Hermann und H. Rieser (Hg.): Ellipsen und fragmentarische Ausdrücke. 2 Bände. Bd. 1 Tübingen 1985. S. 161-290. (= Linguistische Arbeiten 148)
- KLEIN, W.: Ellipse. In: J. Jacobs, A. v. Stechow und D. Wunderlich (Hg.): Syntax. Ein internationales Handbuch. Bd. 1 Berlin/New York 1991. S. 764-799.
- KRAUSE, J.: Mensch-Maschine-Interaktion in natürlicher Sprache. Evaluierungsstudien zu praxisorientierten Frage-Antwort-Systemen und ihre Methodik. Tübingen 1982. (= Sprache und Information Bd.1)
- KÜPPER, D./ RÖSNER, D. / STRIEGL-SCHERER, A.: Natürlichsprachliche Zugangssysteme mit Deutsch als Interaktionssprache - eine vergleichende Studie. In: Sprache und Datenverarbeitung 16.1. (1992) S. 29-66.
- LAPPIN, S. / GREGORY, H.: A Computational Model of Ellipsis Resolution. In: http://semantics.soas.ac.uk/ellip/papers/comp_ellip.ps.gz . (1998) S. 1-17.
- LAPPIN, S. / SHIH, H.-H.: A Generalized Reconstruction Algorithm for Ellipsis Resolution. In: Proceedings of COLING-96 2. (1996) S. 687-692.
- LEWIN, I. / PULMAN, S.G.: Inference in the Resolution of Ellipsis. Bericht SRI International, 1995.
- MAST, M.: Ein Dialogmodul für ein Spracherkennungs- und Dialogsystem. Erlangen 1993. (= DISKI 50)
- MAY, R.: Logical Form. Cambridge, Mass. 1985. (= Linguistic Inquiry Monographs 12)
- MCCORD, M./ BERNTH, A./ LAPPIN, S. ET AL.: Natural Language Processing within a Slot Grammar Framework. In: International Journal on Artificial Intelligence Tools 1. (1992) S. 229-277.
- MOLLÀ ALIOD, D./ BERRI, J. / HESS, M.: A Real World Implementation of Answer Extraction. In: ? . (forthcoming) S. ?
- ORTNER, H.: Die Ellipse. Ein Problem der Sprachtheorie und der Grammatikschreibung. Tübingen 1987. (= Reihe Germanistische Linguistik Bd.80)
- PINKAL, M.: Semantik. In: G. Görz (Hg.): Einführung in die künstliche Intelligenz. 2. Auflage. Bonn 1995. S. 431-491.
- RATH, R.: Strukturelle Aspekte und kommunikative Funktion sprachlicher Verkürzungen. In: Grazer Linguistische Studien 10. (1979) S. 217-239.

- RAYNER, M. / CARTER, D.: Adapting the Core Language Engine to French and Spanish. In: Proceedings of NLP-IA-96. New Brunswick 1996. S. 224-232.
- ROSS, J.: Guess who? In: Papers from the Fifth Regional Meeting of the Chicago Linguistic Society . (1969) S. 252-286.
- ROSS, J.: Gapping and the order of constituents. In: M. Bierwisch und K. E. Hei-dolph (Hg.): Progress in linguistics. The Hague 1970. S. 249-259.
- SAG, I.: Deletion and Logical Form. Massachusetts 1976. (Diss.)
- SCHUH, H.-M.: Ellipse - Text - Kommunikation. Bonn 1974. (= Romanische Versu-che und Vorarbeiten 48)
- SCHWABE, K.: Syntax und Semantik situativer Ellipsen. Tübingen 1994. (= Studien zur deutschen Grammatik Bd. 48)
- STECHOW, A.V. / STERNEFELD, W.: Bausteine syntaktischen Wissens. Ein Lehr-buch der generativen Grammatik. Opladen 1988.
- STROHNER, H. / RICKHEIT, G.: Funktionale Vollständigkeit sprachlicher Äusserun-gen: Probleme, Experimente und Perspektiven. In: R. Meyer-Hermann und H. Rieser (Hg.): Ellipsen und fragmentarische Ausdrücke. Tübingen 1985. S. 250-268. (= Linguistische Arbeiten 148)
- WARREN, D.H. / PEIRERA, F.C.N.: An Efficient Easily Adaptable System for Inter-pretating Natural Language Queries. In: American Journal of Computational Linguistics 8.3-4. (1982) S. 110-122.

6 Anhänge