

Der Einsatz von Wortklassenagenten für die automatische Sprachverarbeitung

Teil II - Die vier Verarbeitungsstufen

Hermann Helbig
FernUniversität Hagen
Praktische Informatik VII/Künstliche Intelligenz
58084 Hagen

e-mail:
hermann.helbig@fernuni-hagen.de

Abstract

Im Teil I dieser Veröffentlichung wurde ein wortklassenorientiertes Modell zur Verarbeitung natürlicher Sprache vorgeschlagen, dem die These zugrundeliegt, daß dem Wort die zentrale Rolle beim Sprachverstehen zukommt. Das genannte Modell wurde in Form einer sogenannten Wortagentenmaschine realisiert. Im Mittelpunkt dieses Beitrags stehen die vier Verarbeitungsstufen dieser Wortagentenmaschine für die automatische Sprachverarbeitung. Die beiden ersten Verarbeitungsstufen sind für die Bildung von elementaren und komplexen *Kernen* zuständig. In der dritten Stufe werden die Unterordnung der bereits gebildeten Kerne unter das Verb und die damit verbundene Sättigung der von demselben ausgehenden Valenzen behandelt. Die vierte Verarbeitungsstufe hat die Aufgabe, Modalangaben und Satzrelatoren syntaktisch zu analysieren und semantisch zu repräsentieren. Die Wortklassenagenten selbst werden in einem gesonderten Beitrag beschrieben.

Inhaltsverzeichnis

1	Vorbemerkungen	3
2	Die vier Verarbeitungsstufen der WCAM	6
2.1	Die Bildung elementarer Kerne	6
2.2	Die Bildung komplexer Kerne	13
2.3	Die Sättigung der Verbvalenzen	17
2.4	Modalitäten und Satzrelatoren	20
3	Zusammenfassung und Ausblick	22
4	Literaturverzeichnis	24
5	Abbildungsverzeichnis	25
6	Tabellenverzeichnis	25

1 Vorbemerkungen

Die grammatischen und kognitiven Grundlagen des Wortagentenmodells sowie die darin verwendeten linguistischen Prinzipien werden im ersten Teil [Helbig, Mertens 94] ausführlich behandelt. Außerdem werden darin das Schichtenmodell für die Verarbeitung natürlicher Sprache und der prinzipielle Aufbau des Wortklassenagentenmodells vorgestellt. In dem vorliegenden zweiten Teil stehen die vier Verarbeitungsstufen der Wortklassenagentenmaschine im Mittelpunkt.

Das Wortagentenmodell basiert auf den Grundannahmen, daß dem Wort die zentrale Rolle beim Sprachverstehen zukommt und daß sich Wörter mit gleicher grammatischer Funktion zu einer Klasse zusammenfassen und damit bei der Verarbeitung natürlicher Sprache einheitlich behandeln lassen. Die grammatische Funktion einer Wortklasse wird mit Hilfe von Wortklassenagenten (abgekürzt: WCA) dargestellt. Jeder der Agenten ist i.a. in einen OPEN-Act und einen COMPLETE-Act geteilt. Diese Zweiteilung spiegelt zwei grundsätzliche "Wirkungsweisen" eines Wortes wieder, die bei der Verarbeitung natürlicher Sprache berücksichtigt werden müssen: in der ersten Phase eröffnet ein in die Analyse einbezogenes Wort Erwartungen (Valenzen) und in der zweiten Phase können die eröffneten Valenzen dann durch andere Wörter bzw. Konstituenten gesättigt werden. Die Wortklassenagenten werden von einem übergeordneten Monitor, der Wortklassenagentenmaschine (abgekürzt: WCAM), überwacht und zentral gesteuert. Diese Maschine ist eine Art nichtdeterministischer Automat mit den drei verschiedenen Zuständen OPEN, CLOSED und COMPLETE.¹

Im Zustand OPEN wird ein neues Wort in die Analyse einbezogen und der OPEN-Act des aktuell analysierten Wortes aufgerufen. Danach wird ein Repräsentant RSLT für das Resultat dieses WCA-Aufrufs gebildet. Das Ergebnis RSLT wird mit OP markiert, wenn die von ihm ausgehenden Valenzen ungesättigt sind, und es wird mit CL markiert, wenn es einen vollständig gesättigten Kern² darstellt (wie das z.B. bei Eigennamen oder nicht als Graduatoren wirkenden Adverbien der Fall ist).

Befindet sich die WCAM im Zustand CLOSED, so ist eine Konstituente gerade zu Ende analysiert und ein vollständiger Kern der genannten Art gebildet worden. Der semantische Repräsentant für diesen Kern wird mit CL markiert und als Filler für die Valenzen anderer Wörter zur Verfügung gestellt.

Im Zustand COMPLETE wird die Sättigung von Valenzen ausgeführt, indem der COMPLETE-Act desjenigen Wortklassenagenten aufgerufen wird, der von der WCAM (evtl. nach Entscheidung über Mehrdeutigkeiten) die höchste Priorität erhalten hat. Beim Aufbau elementarer Kerne ist das immer der COMPLETE-Act des Toplevel-Elements im zentralen Arbeitsgedächtnis (abgekürzt: ZAG).

¹Vgl. dazu auch das Übergangendiagramm für die Zustände der Wortklassenagentenmaschine in Abbildung 1.

²Allgemein verstehen wir unter *Kernen* abgeschlossene semantische Bedeutungsstrukturen, von denen keine Valenzen ausgehen. Der Terminus *Kerne* widerspiegelt die Auffassung, daß sich während des Sprachverstehens nach und nach abgeschlossene mentale Repräsentationen (eben die Kerne) herausbilden, die durch Herstellung von Beziehungen zu anderen Kernen oder funktionelle Komposition zu höher aggregierten Strukturen zusammengesetzt werden.

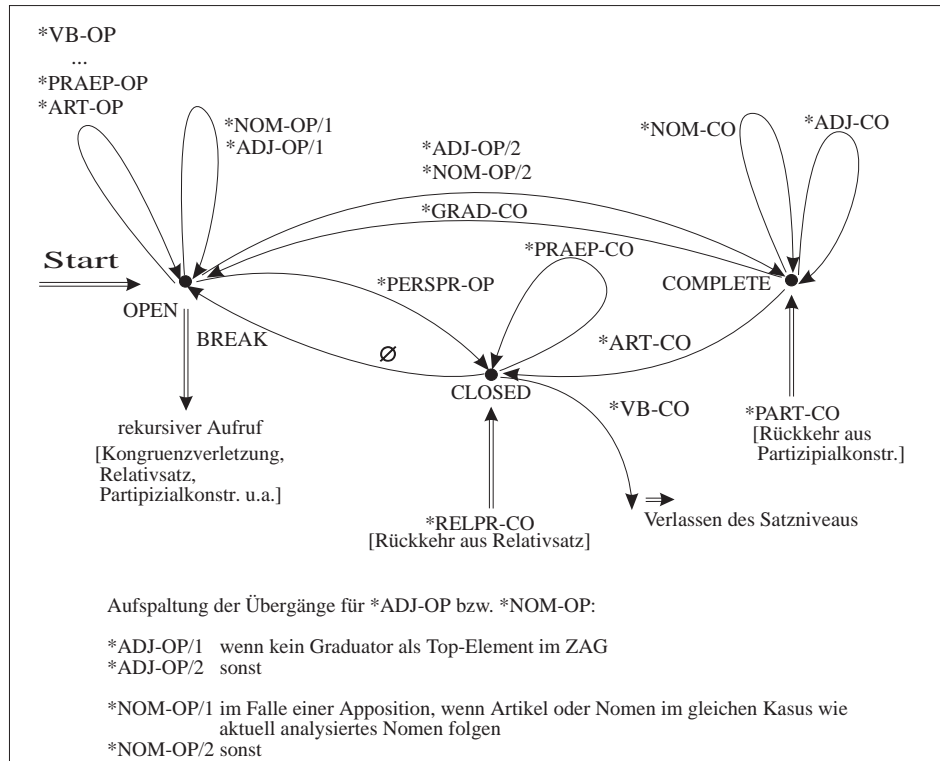


Abbildung 1: Übergangendiagramm für die Zustände der WCAM

Die WCAM verarbeitet natürlichsprachliche Ausdrücke in vier aufeinanderfolgenden Stufen. Die beiden ersten Stufen sind für die Bildung der elementaren bzw. komplexen Kerne des gerade analysierten Satzes zuständig (vgl. Abschnitte 2.1 und 2.2). In der dritten Stufe werden die bereits gebildeten Kerne als Argumente dem Verb als Hauptrelator des Satzes untergeordnet (vgl. Abschnitt 2.3) und damit die von ihm ausgehenden Valenzen gesättigt. Die Modalangaben, die eine Stellungnahme des Sprechers zur Gültigkeit eines geäußerten Satzes ausdrücken, und die Satzrelatoren, die die propositionalen Kerne eines Satzes miteinander verbinden, werden in der vierten Verarbeitungsstufe behandelt (vgl. Abschnitt 2.4). In Abbildung 2 werden die einzelnen Komponenten und wichtige Arbeitsstrukturen der WCAM dargestellt.

Eingesetzt wird die Wortklassenagentenmaschine im Rahmen des natürlichsprachlichen Literaturrecherchesystems LINAS (**L**iteraturrecherche in **N**atürlicher **S**prache), das im Lehrgebiet Praktische Informatik VII/Künstliche Intelligenz an der FernUniversität Hagen entwickelt wird und im ersten Teil (vgl. [Helbig, Mertens 94]) bereits ausführlich vorgestellt wurde.

Die wichtigsten Komponenten des Auskunftssystems LINAS umfassen eine Lexikographenwerkbank (LEXBENCH) und ein multivalent nutzbares Computerlexikon (COLEX), die Analyse mit Hilfe von autonomen und parallel arbeitenden Wortklassenagenten (WCA), die Bedeutungsdarstellung in Form von mehrschichtigen, erweiterten semantischen Netzen (MESNET) und die Transformation von Ausdrücken der semantischen Repräsentationssprache in die Datenbankanfragesprache.

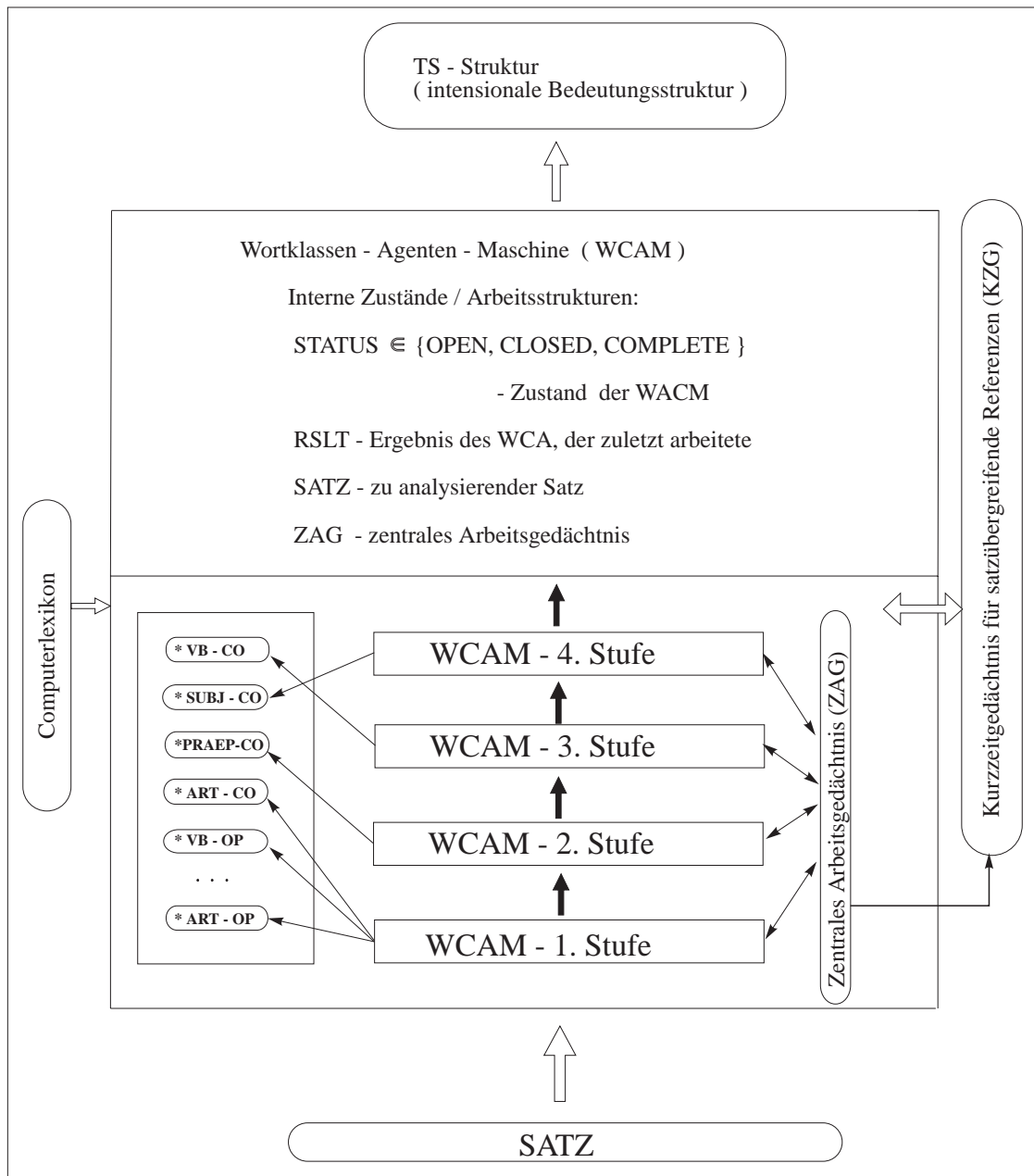


Abbildung 2: Die Komponenten der WCAM

2 Die vier Verarbeitungsstufen der WCAM

Nachdem im ersten Teil (vgl. [Helbig, Mertens 94]) das Schichtenmodell der Verarbeitung vorgestellt wurde, soll in den nachfolgenden Abschnitten die funktionelle Realisierung der diesem Modell zugrundeliegenden vier Verarbeitungsstufen behandelt werden.³

2.1 Die Bildung elementarer Kerne

In der **ersten Verarbeitungsstufe** der WCAM werden die Wörter eines Satzes eingelesen und einer lexikalisch-morphologischen Analyse (abgekürzt: LXMA) unterzogen. Dabei werden die flektierten Wörter auf ihre Grundformen zurückgeführt (Lemmatisierung) und mit entsprechenden morphosyntaktischen Merkmalen (Grammatemen) versehen. Da wir in diesem Zusammenhang nicht näher auf die LXMA eingehen können, kann für das Verständnis der WCAM angenommen werden, daß diese Grammateme wie in einem vollständigen Wortformenlexikon einfach bei jeder Wortform im Lexikon eingetragen seien. In der Praxis - und das trifft auch auf das vorgestellte System zu - wird bei Sprachen wie dem Deutschen oder Russischen zumindest eine flexivische Analyse eingesetzt, die Genus, Kasus, Numerus (für Artikel, Adjektiv, Nomen und Pronomen), die Steigerungsstufen (bei Adjektiven, Adverbien) bzw. Person, Numerus, Genus verbi, Tempus und Modus (für Verben bzw. Hilfsverben) automatisch ermittelt.

Die wesentliche Leistung der WCAM erster Stufe besteht in der Bildung elementarer Kerne. Dieser Vorgang soll anhand eines Beispiels erläutert werden, wobei auf das Übergangendiagramm (vgl. Abbildung 1) und die funktionelle Beschreibung (vgl. Abbildung 3) Bezug genommen wird. Der nachstehend angegebene Beispielsatz 1 wird auch für die Diskussion der weiteren Analyseschritte verwendet:

- (1) Der unbekannte Schriftsteller schrieb ein sehr interessantes Buch über die Mammothöhle in Kentucky.

Die einzelnen Analyseschritte der WCAM 1. Stufe für diesen Satz sind in Tabelle 1 angegeben. In den Spalten dieser Tabelle sind für jeden Analyseschritt folgende Angaben enthalten: der Ausgangszustand der WCAM vor Ausführung des Schrittes (das ist gleichzeitig auch der Endzustand des vorhergehenden Schrittes), das neu in die Analyse einbezogene Wort, der jeweils aktive Wortklassenagent (differenziert nach OPEN-Act und COMPLETE-Act) und die Wirkung der zugehörigen Aktion. Die Analyse beginnt gemäß Übergangendiagramm (Abbildung 1) mit STATUS=OPEN.

³In den Abbildungen 3, 7, 8, 10 und 11 sind die zugehörigen Funktionsschemata der entsprechenden Verarbeitungsstufen der WCAM dargestellt. Sie sind in Form von Strukturschemata (vgl. [Winograd 83]) mit eingelagertem Pseudocode beschrieben und in COMMON LISP implementiert. Die darin verwendeten Pfeilsymbole haben folgende Bedeutungen: $a \Rightarrow b$: "a wird in der Arbeitsstruktur b gespeichert" (*Eintragung*), $x \Leftarrow y$: "der Variablen x wird der Wert y zugewiesen" (*Zuweisung*) und $[p] \mapsto q$: "falls p zutrifft, dann wird q ausgeführt" (*Fallunterscheidung*).

WCAM - 1. Stufe [Bildung elementarer Kerne]	
Zweck: - lexikalisch - morphologische Analyse der Wortformen - Aufruf der OPEN - Acts der WCA - Bildung elementarer Kerne	
Hintergrundwissen: Computerlexikon	
globale Struktur: zu Beginn leeres ZAG Input: Satz der natürlichen Sprache	
Ergebnis: ZAG mit (soweit als möglich) zusammengefaßten und mit 'CL markierten elementaren Kernen bzw. noch nicht gesättigten, mit 'OP markierten Lemmata	
innere Zustände/ Arbeitsstruktur:	SENT - noch zu analysierender Satzrest * - LXMA - Ergebnis für aktuell analysiertes Wort W_{act} ** - LXMA - Ergebnis für das auf W_{act} folgende Wort
<p>(0) Öffnen einer linken Satzklammer (Eintragen von BREAK in ZAG) (1) SENT \leftarrow gesamter zu analysierender Satz (2) STATUS \leftarrow 'OPEN; ** \leftarrow LXMA - Ergebnis für erstes Wort, Verkürzung von SENT um dieses Wort [Ist ** Konjunktion] \mapsto Konjunktion wird als Relator vor Satzklammer in ZAG gesetzt ** \leftarrow LXMA des ersten Wortes von SENT; Verkürzen von SENT Wiederhole bis Satzende erreicht (d.h. ** = \emptyset) bzw. Teilsatzende erkannt ist (z.B. ** ein KOMMA) :</p>	
<p>(3) * \leftarrow ** ; Verkürzen von SENT um ein Wort (4) ** \leftarrow LXMA - Ergebnis für erstes Element von SENT (5) Aufruf des OPEN-Act von WCA(*) RSLT \leftarrow Ergebnis des Aufrufs (6) [Unterbrechung durch WCA(*) erkannt ?] \mapsto</p> <div style="border: 1px solid black; padding: 5px; margin: 5px auto; width: fit-content;"> <ul style="list-style-type: none"> - rekursiver Aufruf der WCAM für eingeschobenen Teilsatz - Einbau des Analyseergebnisses für eingeschobenen Teilsatz in RSLT - Ergebnis \Rightarrow ZAG </div> <p>Wiederhole bis EXIT signalisiert wird:</p>	
<p>(7) Fallunterscheidung auf Grund des Wertes von STATUS : [STATUS = ' OPEN] \mapsto</p> <div style="border: 1px solid black; padding: 5px; margin: 5px auto; width: fit-content;"> <ul style="list-style-type: none"> - Eintragen von RSLT mit Markierung 'OP in ZAG - EXIT </div> <p> [STATUS = ' COMPLETE] \mapsto</p> <div style="border: 1px solid black; padding: 5px; margin: 5px auto; width: fit-content;"> <ul style="list-style-type: none"> - Aufruf des COMPLETE-Act des WCA des Top-Elements von ZAG mit den Parametern: TPH \leftarrow RSLT TPL \leftarrow Top - Element von ZAG - Entfernen von TPL aus ZAG - RSLT \leftarrow Ergebnis des COMPLETE - Acts </div> <p> [STATUS = ' CLOSED] \mapsto</p> <div style="border: 1px solid black; padding: 5px; margin: 5px auto; width: fit-content;"> <ul style="list-style-type: none"> - Eintragung von RSLT mit Markierung 'CL in ZAG - STATUS \leftarrow 'OPEN - EXIT </div>	
<p>(8) Schließen der rechten Satzklammer (Eintragen eines BREAK in ZAG) SATZ \leftarrow SENT (Verkürzen des Satzes um gerade analysierten Teilsatz)</p>	
<p>Prozesse:</p> <div style="border: 1px solid black; padding: 5px; margin: 5px auto; width: fit-content;"> <p>OPEN-Acts : OPEN-Acts aller WCA</p> </div> <div style="border: 1px solid black; padding: 5px; margin: 5px auto; width: fit-content;"> <p>COMPLETE-Acts: *ADJ-CO, *ART-CO, *NUM-CO, *ORD-CO, *GRAD-CO</p> </div>	

Abbildung 3: Die erste Verarbeitungsstufe der WCAM

	STATUS	neu einbezogenes Wort	WCA	Aktivität
1	OPEN	der	*ART-OP	Eröffnung einer Nominalphrase
2	OPEN	unbekannt	*ADJ-OP	Verschärfung der Erwartungen bezüglich NP; z.B. treten neue Kasusanforderungen hinzu
3	OPEN	Schriftsteller	*NOM-OP	Erzeugung eines Repräsentanten K1 für den Begriff <Schriftsteller>
4	COMPLETE	-	*ADJ-CO	Komplettierung der Erwartungen, die durch (2) ausgelöst wurden; Anreicherung von K1 durch Eigenschaft <unbekannt>
5	COMPLETE	-	*ART-CO	Komplettierungen der Erwartungen aus (1)
6	CLOSED	-	-	Abschluß des elementaren Kerns K1
7	OPEN	schreiben	*VB-OP	Eröffnung der vom Verb ausgehenden Valenzen
8	OPEN	ein	*ART-OP	analog zu (1)
9	OPEN	sehr	*GRAD-OP	Eröffnung einer Adjektivgruppe
10	OPEN	interessant	*ADJ-OP	Filler für Erwartungen von (9) erkannt
11	COMPLETE	-	*GRAD-CO	Erzeugung einer Repräsentation für graduierte Eigenschaft P1 <sehr interessant>, STATUS \Leftarrow OPEN
12	OPEN	Buch	*NOM-OP	Erzeugung eines Repräsentanten K2 für den Begriff <Buch>
13	COMPLETE	-	*ADJ-CO	Erwartungen, die von noch geöffneter Konstituente P1 ausgehen, werden erfüllt; K2 wird durch graduierte Eigenschaft P1 erweitert
14	COMPLETE	-	*ART-CO	Erwartungen von (8) erfüllt
15	CLOSED	-	-	Abschluß des elementaren Kerns K2
16	OPEN	über	*PRAEP-OP	Eröffnung der Erwartungen der Präposition
17	OPEN	die	*ART-OP	analog zu (1) und (8)
18	OPEN	Mammuthöhle	*NOM-OP	Erzeugung eines Repräsentanten K3 für den Begriff <Mammuthöhle>
19	COMPLETE	-	*ART-CO	Erwartungen von (17) erfüllt
20	CLOSED	-	-	Abschluß des elementaren Kerns K3
21	OPEN	in	*PRAEP-OP	analog zu (16)
22	OPEN	Kentucky	*NAME-OP	Erzeugung eines Repräsentanten K4 für Individualbegriff <Kentucky>
23	CLOSED	-	-	Abschluß des elementaren Kerns K4

Tabelle 1: Darstellung der Analyseschritte der WCAM erster Stufe

Für das Verständnis der in den Abbildungen und Beispielen verwendeten Wortklassenagenten und semantischen Darstellungsmittel sei auf die Tabellen 2 und 3 verwiesen. Die an den ZAG-Elementen eingetragenen Bezeichnungen DET und INDET sind Werte des Merkmals REF (Referenzhinweis) und bedeuten, daß das Denotat der betreffenden Kerne *genau* festgelegt (determiniert) bzw. *nicht* festgelegt (indeterminiert) ist (vgl. [Helbig, Herold 94]).

Nach dem dritten Analyseschritt hat das ZAG den in Abbildung 4 dargestellten Inhalt (von den grammatischen und semantischen Merkmalen der ZAG-Elemente sind nur die Wortklassen angedeutet).

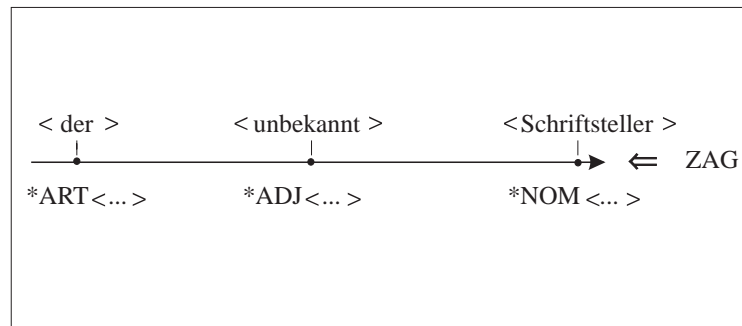


Abbildung 4: Der Inhalt des ZAG nach dem dritten Analyseschritt

Am Ende von Analyseschritt 6 besteht das ZAG aufgrund der bereits bis dahin möglichen Zusammenfassungen nur noch aus dem elementaren Kern K1 (vgl. Abbildung 5).

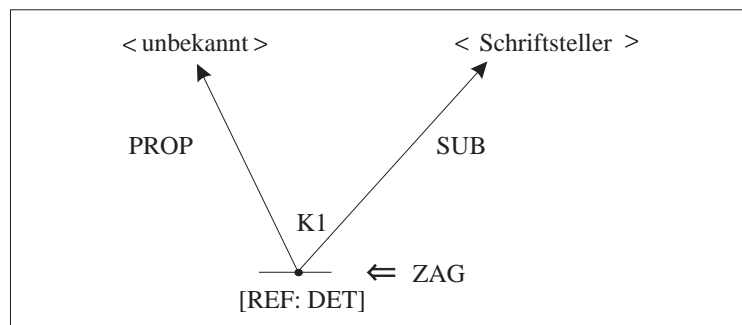


Abbildung 5: Der Inhalt des ZAG nach dem sechsten Analyseschritt

Nach Abschluß der Arbeit der WCAM erster Stufe (d.h. am Satz- bzw. Teilsatzende; vgl. Analyseschritt 23 in Tabelle 1) enthält das zentrale Arbeitsgedächtnis schließlich die in Abbildung 6 aufgeführten Elemente, wobei die elementaren Kerne K1 bis K4 bereits eine semantische Repräsentation besitzen.

Bemerkenswert ist die Tatsache, daß die Zusammenfassung elementarer Kerne im großen und ganzen deterministisch verläuft. Eine Unterbrechung dieses regulären Prozesses, die durch ein BREAK im ZAG markiert wird, hat im Deutschen vor allem folgende sprachliche Erscheinungen als Ursachen (vgl. Programmschritt 6 in Abbildung 3):

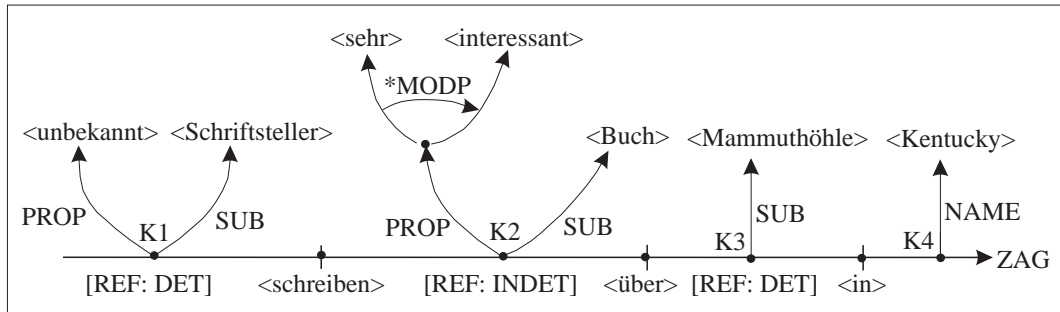


Abbildung 6: Der Inhalt des ZAG nach Abschluß der Verarbeitung in der ersten Stufe

- (a) Verletzung von Kongruenzanforderungen zwischen Artikel, Adjektiv und Nomen:
BREAK zwischen "der" und "klein":
Der kleinen Kindern gut schmeckende Brotaufstrich
- (b) Vom Normalsatz abweichende Aufeinanderfolge von Wortklassen-Agenten:
*ART+*ART:
Der ein teures Rennrad besitzende Schüler
*ART+*PRAEP:
Ein mit Teppichen handelnder Reisender
*PRAEP+*PRAEP:
Er kam *mit vor* Nässe triefenden Haaren
Die BREAKs liegen in den genannten Beispielen zwischen den kursiv gedruckten Wörtern.
- (c) Eröffnung eines eingeschobenen Nebensatzes (BREAK durch *KOMMA):
Der Junge, der das Fenster zerschlagen hatte,
Da nahm der Junge, als er das Fenster zerschlagen hatte,
- (d) Konstituenten-Koordination von Linksattributen:
syndetisch:
Das schwere und farblose Mineral
asyndetisch:
Das schwere, farblose Mineral

Bei den Beispielsätzen unter den aufgeführten Punkten (a) und (b) erfolgt ein Abstieg in ein untergeordnetes Satzniveau (Partizipialkonstruktion), wobei nach Verlassen desselben der gleiche Zustand erreicht ist, wie nach Abarbeitung des COMPLETE-Act eines Adjektivs (vgl. *ADJ-CO Übergang in Abbildung 1).

Eingeschobene Relativsätze und eingeschobene Nebensätze allgemein (vgl. Beispielsätze unter Punkt c) bewirken ebenfalls einen rekursiven Abstieg in ein untergeordnetes Satzniveau. Nach ihrer erfolgreichen Bearbeitung wird das Ergebnis der Analyse in das ZAG eingefügt. Im Fall eines Relativsatzes wird das Ergebnis in die semantische Repräsentation desjenigen Kerns, der als Antezedent des Relativpronomens ermittelt wurde, integriert. Anschließend wird die Analyse so fortgesetzt, als wäre der Nebensatz nicht vorhanden gewesen. Im Falle eines Relativsatzes kehrt die Analyse in den

WCA	Wortart bzw. sprachliche Erscheinung (Beispiel)
*ADJ	Adjektive (Der <i>teure</i> Monitor steht im Schaufenster.)
*ADV	Adverbien (Wie <i>schnell</i> arbeitet der Drucker?)
*ART	Artikel (Der <i>teure</i> Monitor steht im Schaufenster.)
*GRAD	Graduatoren (Ein <i>sehr</i> schneller Rechner ...)
*HV	Hilfsverben (Der Händler <i>hat</i> ihn verkauft.)
*KOMP	Komparative (Das Auto ist <i>teurer</i> als ...)
*KOORD	Koordinierende Konjunktionen (<i>und, oder, aber, ...</i>)
*MV	Modale Hilfsverben (<i>Kann</i> sie den Computer liefern?)
*NAME	Namen von Individuen (Gibt es Bücher von <i>Woods</i> ?)
*NEG	Negatoren außer <i>kein</i> (Peter sah das Auto <i>nicht</i> .)
*NEGATTR	Negatoren mit attributivem Gebrauch (<i>Kein</i> Auto ist schneller.)
*NOM	Nomen (Der <i>teure</i> <i>Monitor</i> steht im Schaufenster.)
*NUM	Zahlen und Zahlwörter (Peter ist <i>60</i> kg schwer.)
*ORD	Ordnungszahlen (Am <i>24.</i> Dezember ist ...)
*PART	Partizipien (Die in Rußland <i>produzierten</i> Güter ...)
*PERSPR	Personalpronomen (Wer hat <i>es</i> veröffentlicht?)
*PRAEP	Präpositionen (Der Autor schrieb ein Buch <i>über</i> ...)
*RELPR	Relativpronomen (Der Schüler, <i>der</i> gestern ...)
*SUBJ	subordinierende Konjunktionen (<i>weil, nachdem, ...</i>)
*SUPL	Superlative (Das <i>teuerste</i> Fahrrad kostet ...)
*VB	Verben (Der Autor <i>schrieb</i> ein Buch über ...)

Tabelle 2: Kurzcharakteristik der in den Funktionsschemata und Beispielen verwendeten Wortklassenagenten (WCA)

Relation	Charakteristik	Beispiel
AFF	affiziertes Objekt	Er zerschlug <i>die Vase</i> .
AGT	Handlungsträger	<i>Peter</i> malt ein Bild.
CAUS	Verursacher	<i>Weil die Bremsen nicht repariert wurden</i> , ist er verunglückt.
INSTR	Instrument	Er fuhr <i>mit dem Auto</i> nach Paris.
LOK	Ortsangabe	Die Post befindet sich <i>in der City</i> .
MODL	Modalitätsangabe	Weil die Bremsen <i>nicht</i> repariert wurden, ist er verunglückt.
NAME	Individuenname	Er wohnt in <i>München</i> .
PROP	Eigenschaftszuordnung	Das Auto <i>ist teuer</i> .
SETOF	Gesamtheit	Alle <i>Autos</i> haben einen Motor.
SUB	Begriffliche Unterordnung	Ein Auto <i>ist ein Fahrzeug</i> .
SUBA	Unterordnung von Handlungen	<i>Peter fliegt</i> nach Ibiza.
TEMP	Zeitangabe	Er wurde <i>1920</i> geboren.
ZEXP	Zustandsträger	<i>Peter</i> wohnt in München.
Funktion	Charakteristik	Beispiel
*MODP	Eigenschaftsspezifikation	Das Buch ist <i>sehr</i> interessant.
*NON	Negation	Die Firma produziert <i>keinen</i> neuen Computer.

Tabelle 3: Kurzcharakteristik der in den Beispielen verwendeten semantischen Relationen und Funktionen

gleichen Zustand zurück, der bei Abschluß der letzten Konstituente vor dem Relativsatzkomma vorgelegen hat (das ist bei eingeschobenen Relativsätzen der CLOSED-Zustand in Abbildung 1).

Eine weitere Unterbrechung des normalen "Expand/Reduce"-Mechanismus innerhalb nominaler Kerne kann durch eine Konstituentenkoordination ausgelöst werden (vgl. Punkt d). Das schwierigste Problem liegt hier in der Unterscheidung von Satz- und Konstituentenkoordination (vgl. dazu [Helbig et al. 94]). Wenn letztere vorliegt, wird sofort nach Beendigung der Analyse der zweiten Konstituente in der Koordination eine Zusammenfassung der koordinierten Kerne durch *KOORD-CO durchgeführt und die Analyse so fortgesetzt, als wäre nur ein Kern vorhanden gewesen.

2.2 Die Bildung komplexer Kerne

Die **zweite Stufe** der WCAM (vgl. Abbildungen 7 und 8) ist durch die Gegenüberstellung von (z.T. miteinander konkurrierenden) offenen Valenzen (Slots) und von geschlossenen Kernen geprägt, die als potentielle Filler für diese Slots in Frage kommen. Die Träger der Valenzen werden in einer Verweisliste OList von offenen Elementen und die Filler in einer Verweisliste CList von geschlossenen Kernen zusammengefaßt (vgl. Abbildung 9). Zu Beginn werden die Position und die Merkmale des Hauptrelators des Satzes aus den Elementen von OList ermittelt. Im Beispielsatz 1 trägt hierzu nur das Vollverb "schreiben" bei. Damit ist auch eine Zuordnung der Konstituenten zu Vorbereich und Nachbereich des Verbs gegeben. Diese Information ist wichtig, da im Deutschen im Hauptsatz vor dem finiten Verb nur eine Konstituente, d.h. vor dem Hauptrelator nur ein (elementarer oder komplexer) Kern stehen darf, wodurch eine wichtige Disambiguierungshilfe gegeben ist. In Nebensätzen gilt das nicht, da dort (im Deutschen) das finite Verb ans Satzende tritt, und damit natürlich i.a. nicht alle Kerne im Vorbereich des Verbs zu einem komplexen Kern zusammenzufassen sind.

Zur Charakterisierung des Hauptrelators können mehrere Wörter der Oberflächenstruktur des Satzes beitragen. Die Ermittlung des Hauptrelators bzw. seiner Merkmale aus den Elementen von OList (genauer aus den Oberflächenbestandteilen Hauptverb, Hilfsverb(en) und abtrennbarer Verbbestandteil) wird in der derzeitigen WCAM-Version von einer speziellen Funktion VBGROUP geleistet. Sie soll aber in Zukunft in die Wortklassenagenten *VB bzw. *HV hineinverlagert werden.

Die schwierigste Aufgabe in der zweiten Stufe der WCAM besteht in der Entscheidung, ob die abgeschlossenen elementaren Kerne einer Kette vom Typ <Kern> <Präposition> <Kern> ... jeweils mit Hilfe des durch die Präposition beschriebenen Relators einem davorstehenden Kern oder dem Verb (dem Hauptrelator des Satzes) unterzuordnen sind. Diese Phase der Analyse wird primär durch die Art der Elemente von CList und deren Anordnung im ZAG bestimmt. Für das Verständnis ist es noch wichtig anzumerken, daß jeder sicher erkannte Genitiv einer Nominalgruppe bei Abschluß des COMPLETE-Acts von *ART durch eine "künstliche" Präposition GENATTR ersetzt wird, die etwa die Funktion des englischen "of" besitzt. Dadurch lassen sich Genitivattribute ähnlich wie präpositionale Fügungen behandeln. Das allgemeine Vorgehen soll anhand der Analysesituation in unserem Beispielsatz 1 bei Eintritt in die WCAM

WCAM - 2. Stufe [Bildung komplexer Kerne]
Zweck: - Auflösung von Pronomenreferenzen (Bereitstellung der potentiellen Antezedenten) - Bereitstellung des Hauptrelators des Satzes - Disambiguierung der Unterordnung der elementaren Kerne - Aggregation komplexer Kerne innerhalb eines Satzklammerpaares
Hintergrundwissen: Kurzzeitgedächtnis KZG für die Analyseergebnisse der n letzten (Teil-) Sätze
globale Struktur: ZAG - als Ergebnis der Arbeit von WCAM - 1. Stufe
Resultat: ZAG mit höher aggregierten (nominalen) Kernen, aufgelösten Referenzen und zusammengefaßtem Hauptrelator. Übergabewerte an WCAM 3 - Stufe: CL - letzter Stand der aggregierten CList OP - letzter Stand der verkürzten OList VB - Hauptrelator
lokale Zustände/Arbeitsstrukturen: CList - Verweisliste auf die abgeschlossenen, mit CL markierten Konstituenten aus ZAG (Reihenfolge wie in ZAG) OList - Verweisliste auf die offenen, mit OP markierten Konstituenten aus ZAG (Reihenfolge wie in ZAG) VB - Hauptrelator
(0) Aufbau von CList und OList (entsprechend Definition oben) (1) Ermittlung des Hauptrelators: Zusammenfassen der Verbbestandteile zu einem Element VB in ZAG und OList - VB \Leftarrow VBGROUP (OList) Wiederhole für jedes Element NE aus CList (von rechts nach links):
<p>(2) [NE = <Pronomen>] \mapsto Ersetzen von NE durch potentiellen Antezedenten NE \Leftarrow ANTEZED (KZG)</p> <p>(3) [NE = <Komparativergänzung>] \mapsto Substitution von NE in Leerstelle der Komparativrepräsentation</p> <p>(4) [NE = <Superlativergänzung>] \mapsto Substitution von NE in Leerstelle der Superlativrepräsentation</p> <p>[Fortsetzung]</p>

Abbildung 7: Die zweite Verarbeitungsstufe der WCAM (Teil 1)

[Fortsetzung]

(5) [Geht dem NE in ZAG unmittelbar ein (präpositionaler) Relator R und eine weitere geschlossene Konstituente AN voraus] \mapsto

Disambiguierung der Unterordnung von NE anhand des Ergebnisses von Kette (AN, R, NE, VB, CL)

[Fall 1: NE eindeutig unter VB unterzuordnen] \mapsto
Markierung von NE als Verbkonstituente (Aktant bzw. freie Fügung)

[Fall 2: NE eindeutig unter AN unterzuordnen] \mapsto
- Aufruf von *PRAEP-CO (AN, R, NE) \Rightarrow RSLT
(Zusammenfassen von AN, R, NE zu neuer Konstituente RSLT)
- Tilgen von NE aus CList und ZAG von R aus OList
- RSLT nimmt Stelle von AN ein

[Fall 3: NE unter andere geschlossene Konstituente aus CL unterzuordnen] \mapsto
- Setzen einer Verzweigungsanforderung

(6) [Steht zwischen NE und der nächsten vorangehenden geschlossenen Konstituente AN ein *KOORD] \mapsto

[Kongruieren AN und NE in Kasus (evtl. nach vorangehender Präposition) und gemeinsamer Einordnungsinstanz] \mapsto
Fallunterscheidung anhand des Ergebnisses (ja/nein) :

ja: - Aufruf von *KOORD-CO(AN, NE) \Rightarrow RSLT
(Zusammenfassen von AN, NE zu Gesamtheit)
- Tilgen von NE aus ZAG und CL und von *KOORD aus OList
- RSLT nimmt Stelle von AN ein

nein: - Setzen einer Verzweigungsanforderung

(7) [Wurde bei vorher untersuchter CL-Konstituente NE' mit Verzweigungsanforderung gesetzt] \mapsto
- Aufruf von { *KOORD-CO bzw. *PRAEP-CO } entsprechend Relator R' \Rightarrow RSLT
(Zusammenfassen von NE, R', NE' zu neuer Konstituente RSIT)
- Tilgen von NE' aus ZAG und CL, von R' aus OList
- RSLT wird neues, aktuell zu untersuchendes NE

Prozesse: COMPLETE-Acts : *PRAEP-CO, *KOMP-CO, *SUPL-CO

VBGROUP: - Ermitteln des Hauptrelators
(Zusammenfassen der im Deutschen getrennten Verbbestandteile)
- Kombination { *HV, *MV, *PRAEP, *ADV, *VB } \rightarrow *VB + Grammateme

KETTE : - Disambiguierung der Unterordnung der (meist nominalen) Kerne unter vorangehenden Kern oder unter Verb

ANTEZED: - Ermittlung der potentiellen Antezedenten für referenzauslösendes Pronomen (Inferenzkomponente)

Abbildung 8: Die zweite Verarbeitungsstufe der WCAM (Teil 2)

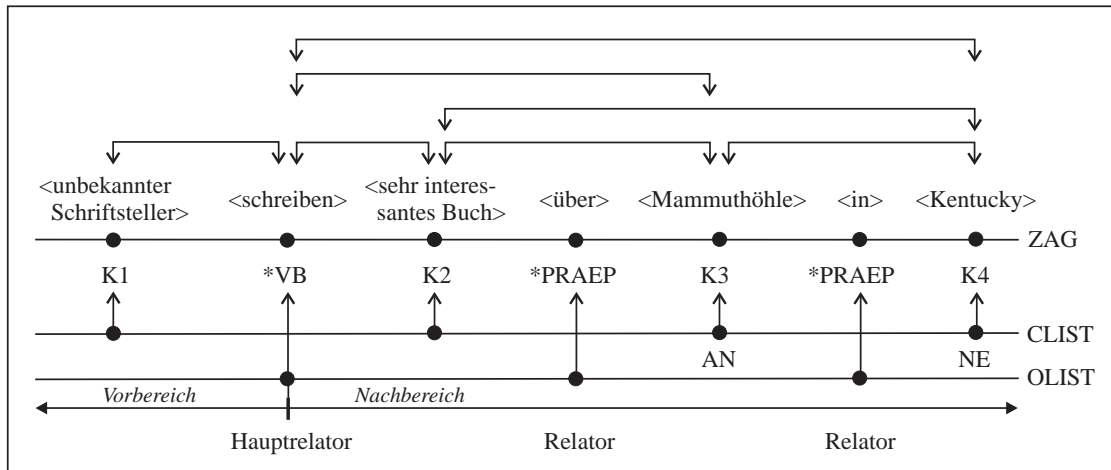


Abbildung 9: Die Verweislisten der WCAM

2. Stufe veranschaulicht werden. Die Analysesituation einschließlich der syntaktisch möglichen Zuordnungen der einzelnen Konstituenten zum Verb bzw. zu anderen Konstituenten des Satzes sind in Abbildung 9 dargestellt.

Die elementaren Kerne, die in CLIST vorkommen, werden von rechts nach links - also entgegengesetzt zur Reihenfolge ihres Auftretens im Satz - analysiert und (soweit zulässig) zusammengefaßt. In diesem Prozeß sind folgende Hauptfälle zu unterscheiden (dabei wird mit NE das augenblicklich untersuchte Element von CLIST bezeichnet und mit AN das vor NE in CLIST stehende Element, falls ein solches existiert; andernfalls ist AN leer):

- (a) Wenn NE eine referenzauslösende Konstituente ist (Pronomen, Proadverb), dann wird es durch den wahrscheinlichsten Antezedenten ersetzt, auf den es referiert. Anschließend wird mit dem neuen NE so verfahren wie unter (b) und (c). Die übrigen Elemente aus der Liste der potentiellen Antezedenten werden in der Repräsentation von NE als Reserve-Stack gehalten und rücken an die Stelle des alten NE, wenn dieses beim Überprüfen von Valenzanforderungen (z.B. des Verbs) ein negatives Ergebnis bringt (vgl. Abschnitt 2.3).
- (b) Es liegt eine Komparativ- oder Superlativanforderung vor, die an zwei Dingen erkennbar ist:
 - an einer im ZAG vorausgehenden, bereits eröffneten Komparativ- bzw. Superlativrepräsentation, die noch eine Leerstelle besitzt;
 - an einer spezifischen unmittelbar vor NE stehenden Präposition (im Falle des Komparativs ist das ein "als" und im Falle des Superlativs eine der Präpositionen "von", "in", "unter", "aus" oder GENATTR); im letztgenannten Fall wird die bereits ermittelte semantische Repräsentation von NE in die Leerstelle der Komparativ- oder Superlativkonstruktion eingefügt. Anschließend werden NE aus CLIST bzw. ZAG und die Präposition aus OLIST bzw. ZAG entfernt.

Im Gegensatz zu anderen präpositionalen Fügungen (vgl. Punkt c) können beim Komparativ - und mit etwas geringerer Akzeptanz auch beim Superlativ - die beiden zusammengehörigen Konstituenten durchaus auf verschiedenen Seiten des Hauptrelators stehen:

- (2) Weil er ein *schöneres* Bild besitzt *als sein Freund*, ...
- (3) Weil er *am schnellsten* rechnete *unter all seinen Mitschülern*, ...

Der elliptische Charakter von Komparativ- und Superlativkonstruktionen wird in dieser Phase der Verarbeitung noch nicht berücksichtigt. Dieses Problem ist Gegenstand einer tieferen semantischen Interpretation.

(c) Wenn unmittelbar vor NE in CList ein weiterer elementarer Kern AN steht und zwischen NE und AN im ZAG eine Präposition P vorkommt, wird nach den folgenden vier Unterfällen verfahren (vgl. dazu auch die Fälle 1 bis 3 unter Programmschritt 5 in Abbildung 8):

- Fall 1: NE ist eindeutig VB unterzuordnen (z.B. läßt sich in dem Satz "Der unbekannte Schriftsteller berichtete in Kentucky über die Mammuthöhle." der Kern *Mammuthöhle* nur dem Verb unterordnen);
- Fall 2: NE ist eindeutig AN unterzuordnen (z.B. führt das Ersetzen von *in Kentucky* in Abbildung 9 durch *des Staates Kentucky* zur Unterordnung von K4 unter K3, falls VB keine Genitivreaktion besitzt und P=GENATTR gilt);
- Fall 3: NE ist einer anderen Konstituente aus CList unterzuordnen (z.B. wird in dem Satz "Der unbekannte Schriftsteller aus Kentucky mit dem schwarzen Hut schläft." der Kern *schwarzer Hut* dem Kern *unbekannter Schriftsteller* untergeordnet);
- Fall 4: Keine Entscheidung möglich (z.B. kann der Kern K4 aus Abbildung 9 sowohl VB als auch dem Kern K3 untergeordnet werden). In diesem Fall wird derzeit durch Rückfrage an den Nutzer entschieden, welcher der Fälle 1 bis 3 tatsächlich vorliegt.

Die Kriterien, die von der WCAM zur automatischen Entscheidung der Fälle 1 bis 3 eingesetzt werden, sind in [Helbig et al. 94] beschrieben. Darin findet sich auch eine ausführliche Diskussion der Disambiguierungsprobleme mit weiteren Beispielen für die hier vorgestellten Fälle.⁴

⁴Zur Disambiguierungsproblematik vgl. z.B. [Schmitz, Quantz 93] und [Hindle, Rooth 93]. Linguistische Prinzipien, die zur Auflösung struktureller Mehrdeutigkeiten beitragen, werden z.B. in [Allen 87], [Hemforth et al. 92]) und [Helbig, Mertens 94] vorgestellt.

2.3 Die Sättigung der Verbvalenzen

Die in der zweiten Verarbeitungsstufe bereitgestellten, nicht weiter aggregierbaren Kerne werden in der **dritten Stufe** als Argumente des Hauptrelators des Satzes dem Verb VB untergeordnet. Die nach der zweiten Stufe in CList verbleibenden Kerne können aus linguistischer Sicht in einem grammatischen Satz drei verschiedene Rollen spielen:

- (a) als Filler für die *obligatorischen* Valenzen von VB, d.h. zur Sättigung derjenigen Erwartungen, die vom Verb ausgehen und die unbedingt durch abgeschlossene Kerne erfüllt werden müssen, damit der Satz grammatisch vollständig ist. Bei einem Verb wie "schlagen" (in der Bedeutung I: "mit einem Gegenstand schlagen") und dem Kasusrahmen <schlagen_I> [((AGT x) (AFF y)) ((INSTR z))] hätte man als obligatorische Valenzen einen Handlungsträger AGT und ein affiziertes Objekt AFF sowie die fakultative Valenz Instrument, ausgedrückt durch INSTR. Dazu treten noch die Selektionsrestriktionen für x, y und z, auf die hier nicht näher eingegangen werden soll.
- (b) als Filler einer der *fakultativen* Valenzen von VB, d.h. zur Sättigung von Erwartungen, die zwar eng mit der Verbbedeutung verknüpft sind, die aber in der Oberflächenstruktur des Satzes nicht notwendigerweise durch eine entsprechende Teilphrase beschrieben sein müssen. Wenn z.B. in einem Satz, der das Verb "schlagen" in der Bedeutung I enthält, die Kasusrollen bis auf die Instrumentalbestimmung INSTR belegt sind, dann ist er grammatisch vollständig.
- (c) als *freie Fügung*, die weitestgehend unabhängig von der Verbbedeutung praktisch zu fast allen oder zumindest zu einer großen Klasse von Verben (genauer zu den von den Verben ausgedrückten propositionalen Kernen) als Bedeutungsergänzung hinzutreten kann. Typische Beispiele hierfür sind lokale und temporale Angaben.

Die unter (a) und (b) genannten Filler für Verbvalenzen werden nach Tesnière auch als Aktanten (actants) bezeichnet. Ihre Substitution in die entsprechenden Slots des Verbkasusrahmens wird vom COMPLETE-Act des Verbs (*VB-CO) in der dritten Verarbeitungsstufe der WCAM durchgeführt (vgl. dazu die Programmschritte 1 bis 4 in Abbildung 10). Die unter (c) genannten Konstituenten heißen nach Tesnière Umstandsbestimmungen (circonstants). Sie umfassen temporale, lokale und modale Bestimmungen im engeren Sinne. Während die semantische Einbindung der Filler für obligatorische und fakultative Valenzen in die Bedeutungsstruktur des Satzes durch den im Lexikon verankerten Kasusrahmen des Verbs VB bestimmt wird, unterliegen die freien Fügungen einer besonderen semantischen Deutung (vgl. Programmschritt 5 in Abbildung 10). Diese ergibt sich aus der Verbbedeutung, der Art der Präposition, falls eine solche vor dem Kern NE im ZAG steht, sowie aus der semantischen Sorte von NE.

Die Zuordnung der Elemente von CList zu einer der unter (a) bis (c) genannten Rollen muß in der Reihenfolge, wie sie in den Programmschritten 1 bis 5 im funktionellen

WCAM - 3. Stufe [Verbkomplettierung]
Zweck : - Unterordnung der aggregierten Kerne unter Hauptrelator (komplettiertes Verb) - Disambiguierung der Verbbedeutungen
globaler Speicher: ZAG mit aggregierten Kernen aus WCAM - 2. Stufe
Übergabeparameter: CL, OP, VB (s. WCAM - 2. Stufe)
lokaler Parameter: STRULIST - Liste von Ergebnisstrukturen
Ergebnis :- Repräsentant für propositionale Kerne - Übergabewerte an WCAM - 4. Stufe: - NEG - Satznegations-Markierung - MOD - Modalität
Wiederhole für jede Verbbedeutung VB_i des Hauptrelators VB (aufruf jeweils von *VB-CO):
<p>(0) Zusammenfassen von Präposition und nachfolgendem CL-Kern in ZAG zu PP (1) Sättigung der obligatorischen Valenzen aus Kasusrahmen von VB_i - Streichen der jeweiligen zugeordneten Aktanten aus CL</p> <p>(2) [Gibt es mehr obligatorische Valenzen als Elemente von CL] \mapsto -Return. UNGRAMMATISCH oder ELLIPSE; -EXIT [CL leer] \mapsto -Einbau der gefundenen Struktur in STRULIST; -EXIT</p> <p>(3) Sättigung der fakultativen Valenzen von VB_i - Streichen der jeweils zugeordneten Aktanten aus CL</p> <p>(4) [CL leer] \mapsto -Einbau der gefundenen Struktur in STRULIST; -EXIT</p> <p>(5) Versuch der Deutung der Elemente von CL als freie Fügung - Streichen der entsprechenden zuordenbaren Elemente aus CL</p> <p>(6) [CL nicht leer] \mapsto Return: UNGRAMMATISCH</p>
(7) -Entfernen aller Elemente aus ZAG bis zum nächsten BREAK. -Auswahl der wahrscheinlichsten Struktur aus STRULIST. -Bildung eines Satzrepräsentanten für den propositionalen Kern des Satzes im ZAG.
Prozesse: COMPLETE-Acts: *VB-CO

Abbildung 10: Die dritte Verarbeitungsstufe der WCAM

Schema (vgl. Abbildung 10) zum Ausdruck kommt, durchgeführt werden. Das wird besonders in den Fällen deutlich, in denen z.B. eine lokale Angabe obligatorisch verlangt wird. Dies ist u.a. bei dem Verb "wohnen" mit der Tiefenstruktur <wohnen₁> [((ZEXP x) (LOK 1))] der Fall. Würde man z.B. freie Fügungen vor den Verbvalenzen behandeln, so würde in dem Satz "Peter wohnt in München." die Teilphrase "in München" nicht mehr als Filler für die obligatorische Valenz LOK zur Verfügung stehen.

Falls sich die Kerne aus CList in keine der drei Kategorien (a) bis (c) einordnen lassen oder überhaupt kein Verb VB im Satz bzw. Teilsatz vorkommt, dann ist der Satz entweder ungrammatisch oder eine elliptische Konstruktion. Auf das Problem der Ellipsen kann aber in diesem Zusammenhang nicht näher eingegangen werden.

2.4 Modalitäten und Satzrelatoren

In der **vierten Verarbeitungsstufe** der WCAM (vgl. Abbildung 11) werden zunächst die Modalangaben, die eine Stellungnahme des Sprechers bzw. des Textautors zur Gültigkeit des (Teil-) Satzes zum Ausdruck bringen, mit der semantischen Repräsentation des propositionalen Kerns des Satzes vereinigt. Zu den Modalangaben in diesem Sinne gehören neben der Negation alle weiteren Angaben, die den Wahrheitsgehalt des propositionalen Kerns des Satzes einschränken, nicht aber Modalangaben im Sinne der traditionellen Grammatik wie z.B. Art und Weise oder Umstandsbestimmungen. In dieser Hinsicht unterscheiden wir uns von den Auffassungen der Kasusgrammatik [Robinson 70] oder der Terminologie der traditionellen Grammatik (vgl. dazu z.B. [Jung 82, §208]).

Modalitäten können in der Oberflächenstruktur durch verschiedene sprachliche Mittel ausgedrückt werden (vgl. [Gerstenkorn 76]). Hierzu gehören:

- (a) lexikalisch durch modale Hilfsverben (z.B. "glauben", "dürfen", "sollen") oder Adverbien (z.B. "vielleicht", "möglicherweise");
- (b) morphologisch durch Konjunktivformen (z.B. "käme" vs. "kam", "ginge" vs. "ging");
- (c) grammatisch durch Matrixsätze (z.B. "Es ist nicht wahr, daß ...", "Es ist zweifelhaft, ob ...");
- (d) Tempusangaben (z.B. "gestern", "im Jahre 1984").

Betrachten wir zur Repräsentation von Modalitäten das nachfolgende Beispiel:

- (4) Weil die Bremsen nicht repariert wurden, ist der Fahrer verunglückt.

Für Beispielsatz 4 haben das zentrale Arbeitsgedächtnis (ZAG) bzw. der noch zu analysierende Satzrest (SATZ) bei Eintritt in die vierte Stufe der WCAM die in Abbildung 12 angegebene Struktur.

WCAM - 4. Stufe [Modalitäten und Satzverbindungen]
Zweck: - Verarbeitung der modalen Komponente (i.w.S.) - Generierung der semantischen Repräsentation für komplexe Satzgefüge bzw. Satzverbindungen
globale Struktur: ZAG mit RSLT als Top-Element (Ergebnis der WCAM 3. Stufe)
Übergabeparameter: NEG, MOD (s. WCAM - 2. Stufe)
Ergebnis: Semantische Repräsentation des Gesamtsatzes
<p>(1) Verbinden von Negationsinformation NEG und modaler Information MOD mit semantischer Struktur von RSLT</p> <p>(2) [Satz nicht leer] \mapsto rekursiver Aufruf der gesamten WCAM für verbleibenden Satzteil</p> <p>(3) [Befinden sich satzwertige Relatoren im ZAG] \mapsto</p> <p style="text-align: center;">Aufruf der entsprechenden Koplettierungsfunktionen von *KOORD bzw. *SUBJ</p> <p>(4) Ausgabe des einzigen verbliebenen Satzrepräsentanten im ZAG ($\hat{=}$ RSLT der letzten Komplettierungsaktion)</p>
<div style="border: 1px solid black; padding: 5px;"> Prozesse: COMPLETE - Acts: *KOORD-CO, *SUBJ-CO </div>

Abbildung 11: Die vierte Verarbeitungsstufe der WCAM

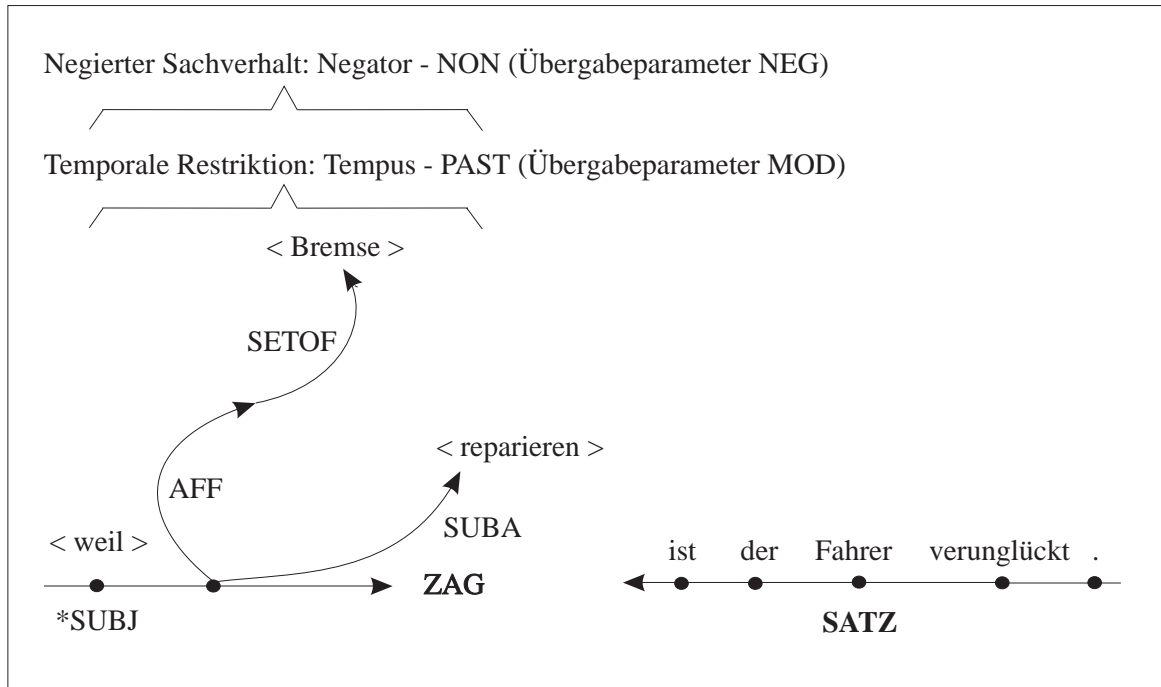


Abbildung 12: ZAG- und SATZ-Struktur bei Eintritt in die vierte Verarbeitungsstufe (vgl. Beispielsatz 4)

Wenn der Satz nur einen Hauptsatz ohne Modalitäten darstellt (vgl. Beispielsatz 5), dann ist die Arbeit der WCAM bereits mit der dritten Stufe beendet. Bezüglich der allgemeinen Behandlung der Darstellung von Modalitäten in der semantischen Repräsentation muß auf [Helbig 83] verwiesen werden.

(5) Der Fahrer ist verunglückt.

Falls nach der Analyse des ersten Teilsatzes die Warteschlange SATZ noch nicht leer ist, d.h. noch ein nachgestellter Haupt- oder Nebensatz folgt, so ist die Analyse für diesen Teil des Satzes fortzusetzen (rekursive Abarbeitung der WCAM). Nach Rückkehr aus dem untergeordneten Satzniveau wird dann mit Hilfe der Relatoren, die propositionale Kerne miteinander verbinden, die semantische Repräsentation für den Gesamtsatz aufgebaut. In unserem Beispielsatz 4 handelt es sich auf grammatischer Ebene um den Relator *SUBJ und auf semantischem Niveau um die Relation CAUS.

Zur Repräsentation des Gesamtsatzes wird der COMPLETE-Act des Wortklassenagenten der entsprechenden Konjunktion (*KOORD-CO) bzw. Subjunktion (*SUBJ-CO) einschließlich der semantischen Interpretation für das entsprechende Bindewort aktiviert (im Beispiel 4 ist das der WCA der Subjunktion "weil"). Die Bedeutungsdarstellung des gesamten Satzes hat für Beispielsatz 4 die in Abbildung 13 dargestellte Struktur.

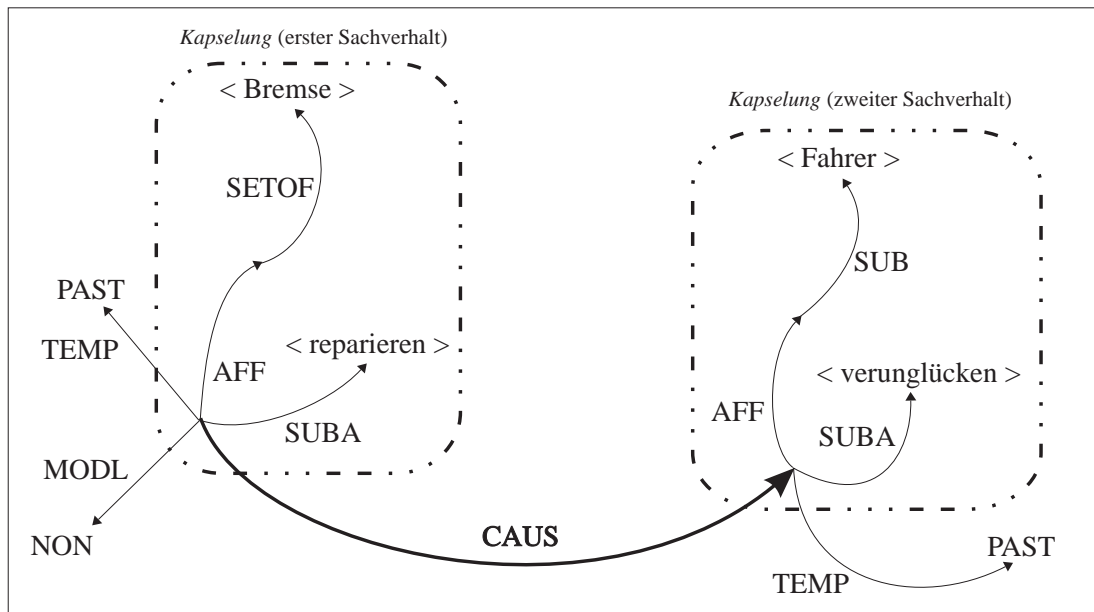


Abbildung 13: Ergebnisstruktur nach Abschluß der vierten Verarbeitungsstufe (vgl. Beispielsatz 4)

3 Zusammenfassung und Ausblick

Der Einsatz der Wortklassenagentenmaschine erfolgt im Rahmen des Literaturrecherchesystems LINAS an der FernUniversität Hagen. Zur Überführung eines natürlichsprachlichen Ausdrucks in eine semantische Tiefenstruktur wird das im ersten Teil (vgl. [Helbig, Mertens 94]) und in diesem Beitrag (Teil II) beschriebene wortklassenorientierte vierstufige Analysemodell eingesetzt, das sich auf ein merkmalsbasiertes Computerlexikon als Hintergrundwissen stützt (vgl. [Helbig, Schulz 94]).

Die Bedeutungsdarstellung einer natürlichsprachlichen Benutzeranfrage wird mit den Darstellungsmitteln für mehrschichtige, erweiterte semantische Netze durchgeführt (vgl. [Helbig, Herold 94]).

Ein wichtiger Forschungsgegenstand ist die Repräsentation der Wortklassenagenten als eigenständige Prozesse. Das Ziel besteht dabei in der Entwicklung von *autonomen* und *parallel arbeitenden* Wort- bzw. Wortklassenagenten, die analysespezifische Informationen per Intertask-Kommunikation austauschen und als gemeinschaftliches Ergebnis die semantische Repräsentation der natürlichsprachlichen Benutzereingabe erzeugen. Dazu muß das Wissen über die Ablaufsteuerung, das zur Zeit in der Wortklassenagentenmaschine repräsentiert ist, auf die einzelnen Agenten verteilt werden. Ein Anliegen der Arbeiten zu diesem Themengebiet wird die Untersuchung sein, ob die Analyse ohne einen übergeordneten Steuerungsprozeß – wie er in der WCAM verkörpert ist – auskommen kann.

4 Literaturverzeichnis

- [Allen 87] Allen, J.: Natural Language Understanding. Menlo Park, California: Benjamin/Cummings 1987
- [Gerstenkorn 76] Gerstenkorn, A.: Das "Modal"-System des heutigen Deutsch. München: Fink Verlag 1976
- [Helbig 83] Helbig, H.: Semantische Repräsentation von Wissen in einem Frage-Antwort-System. Dissertation. Berlin: Akademie der Wissenschaften 1983
- [Helbig et al. 94] Helbig, H., Mertens, A., Schulz, M.: Disambiguierung mit Wortklassenagenten. Informatik Berichte. FernUniversität Hagen 1994 (in Vorbereitung)
- [Helbig, Herold 94] Helbig, H., Herold, C.: Mehrschichtige erweiterte semantische Netze (MESNET). Informatik Berichte. FernUniversität Hagen 1994 (in Vorbereitung)
- [Helbig, Mertens 94] Helbig, H., Mertens, A.: Der Einsatz von Wortklassenagenten für die automatische Sprachverarbeitung. Teil I - Überblick über das Gesamtsystem. Informatik Berichte Nr. 158. FernUniversität Hagen 1994
- [Helbig, Schulz 94] Helbig, H., Schulz, M.: COLEX - Ein Computerlexikon für die automatische Sprachverarbeitung. Informatik Berichte. FernUniversität Hagen 1994 (in Vorbereitung)
- [Hemforth et al. 92] Hemforth, B., Konieczny L., Scheepers, C., Strube, G.: SOUL-Processing: Semantik-orientierte Prinzipien menschlicher Sprachverarbeitung. In: Görz, G. (ed.): KONVENS 92 - 1. Konferenz "Verarbeitung natürlicher Sprache", Nürnberg, 7.-9. Oktober 1992. Berlin etc.: Springer-Verlag 1992, S. 198-208
- [Hindle, Rooth 93] Hindle, D., Rooth, M.: Structural Ambiguity and Lexical Relations. In: Computational Linguistics 19, S. 103-120
- [Jung 82] Jung, W.: Grammatik der deutschen Sprache. Leipzig: VEB Bibliographisches Institut 1982
- [Robinson 70] Robinson, J. J.: Case, Category and Configuration. Journal of Linguistics 1970, S. 57-80
- [Schmitz, Quantz 93] Schmitz, B., Quantz, J. J.: Defaults in Machine Translation. KIT-Report 106. Technische Universität Berlin 1993
- [Winograd 83] Winograd, T.: Language as a Cognitive Process. Volume 1: Syntax. Reading, Massachusetts: Addison-Wesley 1983

5 Abbildungsverzeichnis

1	Übergangendiagramm für die Zustände der WCAM	4
2	Die Komponenten der WCAM	5
3	Die erste Verarbeitungsstufe der WCAM	7
4	Der Inhalt des ZAG nach dem dritten Analyseschritt	9
5	Der Inhalt des ZAG nach dem sechsten Analyseschritt	9
6	Der Inhalt des ZAG nach Abschluß der Verarbeitung in der ersten Stufe	10
7	Die zweite Verarbeitungsstufe der WCAM (Teil 1)	14
8	Die zweite Verarbeitungsstufe der WCAM (Teil 2)	15
9	Die Verweislisten der WCAM	16
10	Die dritte Verarbeitungsstufe der WCAM	19
11	Die vierte Verarbeitungsstufe der WCAM	21
12	ZAG- und SATZ-Struktur bei Eintritt in die vierte Verarbeitungsstufe .	22
13	Ergebnisstruktur nach Abschluß der vierten Verarbeitungsstufe	23

6 Tabellenverzeichnis

1	Darstellung der Analyseschritte der WCAM erster Stufe	8
2	Kurzcharakteristik der in den Funktionsschemata und Beispielen ver- wendeten Wortklassenagenten	11
3	Kurzcharakteristik der in den Beispielen verwendeten semantischen Re- lationen und Funktionen	12