

Koppelung merkmalsbasierter Grammatikformalismen mit terminologischen Repräsentationsformalismen

Josef Ingenerf

Zusammenfassung:

Der vorliegende Aufsatz beschreibt die Konzeption einer benutzeranpaßbaren semantischen Sprachanalyse für die zunehmend relevante Dokumentation medizinsprachlich formulierter Sachverhalte. Bei den zu verarbeitenden Sprachausschnitt handelt es sich syntaktisch gesehen um verblose Nominalsyntaxen. Zu ihrer kompositionellen Semantikanalyse wird eine geeignete Koppelung eines merkmalsbasierten Grammatikformalismus mit einem terminologischen Repräsentationsformalismus vorgestellt. Auf der Basis einer formalen Terminologie als eine Menge von Konzeptdefinitionen wird dazu ein Meta-Modell eingeführt, in dem eher abstrakter Konzepte als semantische Kategorien ausgezeichnet und relational angeordnet werden. Ein solches Meta-Modell impliziert die potentiellen Beteiligungen der von den Kategorien subsumierten Konzepte am Aufbau komplexerer Strukturen. Daraus werden automatisch semantische Kontext-Informationen für die zu analysierenden sprachlichen Ausdrücke abgeleitet. Als Ausprägungen geeigneter Merkmale in einem merkmalsbasierten Grammatikformalismus läßt sich diese Information mit dem Standardverfahren - der Unifikation von Merkmalsstrukturen - auswerten und damit die grammatikalische Analyse im Rahmen des Chart-Parsing-Verfahrens steuern.

1. Einleitung: Dokumentation fachsprachlich formulierter Sachverhalte

Dokumentation als ein zentrales Aufgabengebiet in der Medizin beinhaltet definitionsgemäß die Sammlung und Erfassung, Erschließung und Ordnung sowie Speicherung und Wiedernutzbarmachung (unter evtl. neuen Gesichtspunkten) von Information, insbesondere von sprachlich repräsentierter Information. Die medizinische Fachsprache ist noch vor Bildern und Meßdaten das vorherrschende Ausdrucksmittel, in dem zu dokumentierende Sachverhalte vorliegen. Von besondere Wichtigkeit sind Beschreibungen diagnostischer und therapeutischer Sachverhalte. Eigenschaften einer solchen Fachsprache wie Mehrdeutigkeit, begriffliche Unschärfe sowie zahlreiche synonymen Umschreibungsmöglichkeiten machen sie für Dokumentationszwecke nur bedingt brauchbar. Durch die Nutzung von begrifflichen Ordnungssystemen, in denen die für die Dokumentation zugelassenen Begriffe taxonomisch angeordnet sind, erreicht man eine kontrollierte, d.h. vergleichbare inhaltliche Erschließung der zu dokumentierenden Sachverhalte. Die beabsichtigte Nutzung der in einer Datenbank gespeicherten standardisierten, d.h. formalen Rekonstruktionen der Klartextdiagnosen stützt sich dann auf das zugrundegelegte Ordnungssystem. Zum Beispiel werden alle "Neubildungen am Urogenitalsystem" gesucht und auch die "Myome an der Niere" gefunden. Eine rechnergestützte Vollautomatisierung einer solchen Dokumentation gewinnt - insbesondere vor dem Hintergrund des neuen Gesundheitsstrukturgesetzes 1993 - zunehmend an Bedeutung [Seelos 1993]. Ausführungen zu der ganzen Breite medizin-linguistischer Ansätze finden sich in [Schröder 1993] und [Ingenerf 1994].

Dabei steht für eine automatische semantische Sprachanalyse mit den medizinischen Abschlußdiagnosen ein geeigneter Input zur Verfügung. Das betrifft sowohl ihre Relevanz zur Charakterisierung einer Krankengeschichte als auch ihren knappen sprachlichen 'Nominalsyntaxmen'-Stil, in dem sie üblicherweise abgefaßt werden. Die folgende Abbildung skizziert die wichtigsten syntaktischen Konstruktionen.

<u>Nominale Lexeme</u>	<i>Wortbildung</i>
Präfigierung	("epi_cutan", "Ober_bauch", "Post_infarctsyndrom")
Suffigierung	("Nephr_itis", "My_om")
Komposition	("Magen_schleim_haut_entzündung")
<u>Nominale Syntagmen</u>	<i>NP im weiteren Sinne</i>
Nominalphrasen	<i>NP im engeren Sinne</i>
Adjektivisches Attribut	("akute Nephritis")
Adverbielles Attribut	("stark blutende Wunde")
Nominelles Attribut	("Morbus Hodgkin")
Nominalgruppen	<i>ein dominierender Kern</i>
Genitivattribut	("Entzündung der Niere", "Ablatio retinae")
Präpositionalattribut	("Beschwerden nach Nephritis")
Nominalkomplexe	<i>nebengeordnete Kerne</i>
Konjunktion	("Abschürfung an Hand und Fuß")

Abb.1 "Syntaktische Konstruktionen auf Nominalsyntaxmenenebene"

Diese kompakte, nüchterne Form der sprachlichen Beschreibung von Sachverhalten ist symptomatisch für Fachsprachen [Scheffé 1981]. Sie haben i.a. eine substantivierte Form passivischer Sätze und das Subjekt der entsprechenden Aktiv-Form bleibt ungenannt.

z.B. Der Patient hat/leidet unter

vordere Kreuzbandruptur im rechten Kniegelenk

Das Zellbild zeigt (bei Gutachten in der Pathologie)

Adenokarzinom des Ovars mit Stenosierung des Sigmas und intraabdominaler Metastasierung

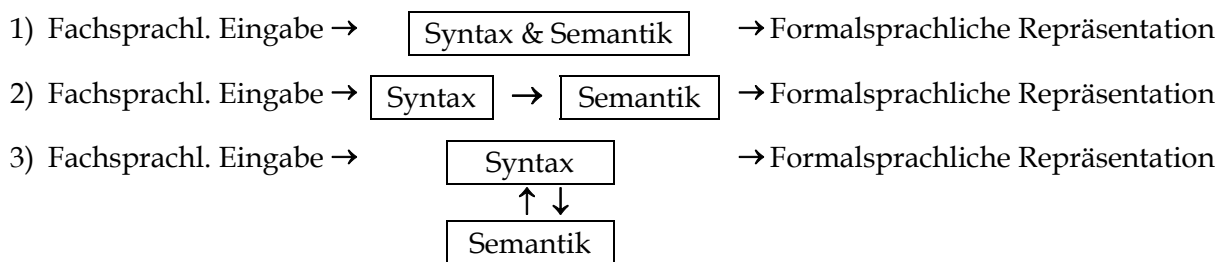
Das in der medizinischen Fachsprache wichtigste und gleichzeitig kompakteste Konstrukt ist die Komposition. Die Semantikanalyse von Komposita ist ein bisher wenig befriedigend gelöstes Problem, denn sie liefern keine morfo-syntaktische Information über die Art der begrifflichen Beziehung zwischen den Teilbegriffen, wie sie durch die Präpositionen in der jeweils paraphrasierten Version in Abb.2 ausgedrückt werden. Die Lösung erfordert also eine geeignete Modellierung von Begriffsbildung.

Kopfhautentzündung	→ Entzündung <i>der</i> (Haut <i>am</i> Kopf)
Tracheobronchitis	→ Entzündung <i>an</i> (Trachea <i>und</i> Bronchus)
Schleimhautreizung	→ Reizung <i>der</i> (Schleim <i>erzeugenden</i> Haut)
Leberzellkarzinom	→ (Zelle <i>in</i> Leber) <i>charakterisiert</i> Karzinom
M. bronchooesophageus	→ Muskel <i>zwischen</i> Luft- <i>und</i> Speiseröhre
Pyeloureterolyse	→ (Herauslösung <i>des</i> Harnleiters) <i>aus</i> Nierenbecken
Fibromyom	→ gutartige Geschwulst <i>aus</i> Binde- <i>und</i> Muskelgewebe
	→ <i>Kombination</i> (Fibrom, Myom)
<u>aber:</u> Myom und Fibrom → zwei Tumore	

Abb. 2 "Umschreibungen der Bedeutung (Paraphrasen) von Komposita"

2. Kompositionelle, semantische Sprachanalyse

Die vorliegende Arbeit beschreibt eine automatisierte Inhaltserschließung fachsprachlich formulierter Sachverhalte mittels einer kompositionellen, semantischen Sprachanalyse. Auf Architekturebene unterscheidet Görz [1988] folgende Sprachverarbeitungs-Modelle:



Die folgende Abbildung skizziert den vorliegenden Ansatz, der sich auf der Architekturebene dem dritten Sprachverarbeitungsmodell zuordnen lässt.

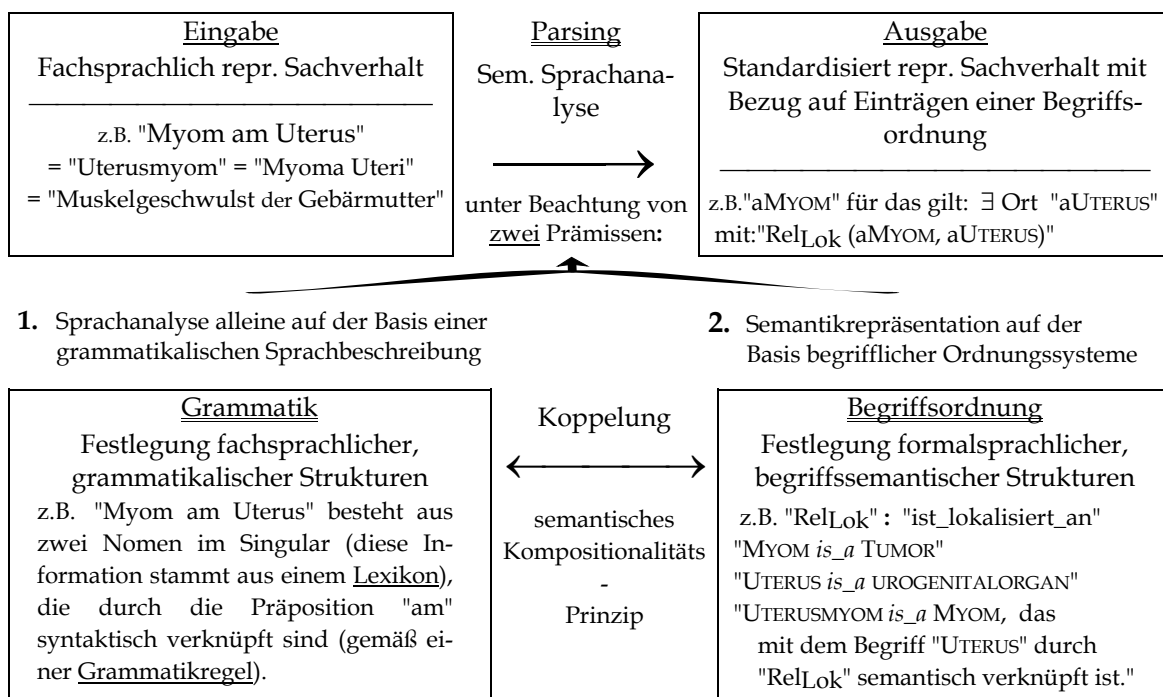


Abb. 3 "Kompositionelle semantische Sprachanalyse"

Im vorliegenden Ansatz gemäß des dritten Architekturtyps sind mit Abb. 3 die Syntax- und Semantikanalyse zwar miteinander gekoppelt, aber trotzdem unabhängig voneinander. Damit ist eine Berücksichtigung semantischer, kontextueller Informationen bereits während der grammatikalischen Analyse möglich. Andererseits ist das System auf andere semantische Gegenstandsbereiche übertragbar. Es wird weder eine semantische Repräsentation nach einer syntaktischen Ableitung erzeugt noch semantisch motivierte nicht-terminale, domainspezifische Kategorien verwendet.

Eine solche semantische Sprachanalyse eines fachsprachlichen Ausdrucks ist also auf linguistisches Wissen über die grammatikalischen Strukturen der Fachsprache und auf terminologisches Wissen über Begriffsbedeutungen angewiesen. Eine Automatisierung (1. Prämisse) hat die Konsequenz, daß Einträge der Begriffsordnung ausschließlich über sprachliche Ausdrücke identifiziert werden können. Falls es sich um lexikalisierte Ausdrücke (Worte bzw. Teilworte) handelt, genügt ein Zugriff auf die Bedeutung im Lexikon (z.B. "Myom", "Uterus"). In allen anderen Fällen ist man auf eine linguistische Analyse der Semantik des sprachlichen Ausdrucks (z.B. "Myom am Uterus") angewiesen, die dem Kompositionalitätsprinzip genügt, da darüberhinaus keine Information zur Verfügung steht, einen Begriff zu identifizieren (siehe unten).

Zur formalen Beschreibung der fachsprachlichen Strukturen aus Abb.1 werden merkmalsbasierte Grammatikformalismen verwendet [Shieber 1986]. Das sogenannte Chart-Parsing-Verfahren (siehe Kilbury [1984]) wird zur algorithmischen Realisierung der grammatikalischen Sprachanalyse eingesetzt. Die Anwendung und Erweiterung des Chart-Parsing-Verfahrens zur Analyse basierend auf Merkmalsgrammatiken wird ausführlich in Gazdar und Mellish [1989] beschrieben.

Die zweite Prämisse in Abb.3 fordert eine Semantikrepräsentation auf der Basis begrifflicher Ordnungssysteme. Nimmt man das Kompositionalitätsprinzip ernst, so darf zur Etablierung einer entsprechenden begrifflichen Ordnungsrelation (Ober-/Unterbegriffsrelation) ausschließlich terminologisches Wissen für Begriffsbedeutungen genutzt werden [Rahmstorf 1989]. Als konsequenter Vertreter dieser definitorischen Auffassung von Begriffsbildung entwickelte Brachman (u.a. in Reaktion auf die Kritik Wood's an der mangelnden Fundierung semantischer Netze) den terminologischen Repräsentationsformalismus KL-ONE [Brachman, Schmolze 1985]. Dieser Formalismus besitzt eine modelltheoretische Semantik, die es gestattet, die Bedeutung definitorisch eingeführter Begriffe kompositionell zu 'berechnen'. Die zweite ganz wesentliche Eigenschaft dieser Ansätze betrifft gerade die epistemologisch motivierte Trennung einer intensionalen und einer extensionalen Beschreibungsebene. Definitionen von Konzepten (z.B. "FRAKTUR") dienen hier zur Informationsanreicherung sowie zur Disambiguierung lexikalischer Mehrdeutigkeit. Die extrahierte Bedeutung einer Sachverhaltsaussage wird auf einer extensionalen Beschreibungsebene als sogenannte Objektbeschreibung über Instanzen der definierten Konzepte

(z.B. "aFRAKTUR") aus einer zugrundeliegenden Terminologie (auch T-Box genannt) formuliert und in einer Objektdatenbank (auch A-Box genannt) abgespeichert. Die folgende Abbildung faßt die beschriebenen Zusammenhänge zusammen.

3. Verb-zentrierte Ansätze und ein alternativer Ansatz

Als Ausgangspunkt für die folgenden Ausführungen seien zunächst die oben angesprochenen merkmalsbasierten Grammatikformalismen eingeführt. Aus der Menge der existierenden theoretischen Präzisierungen wird ein solcher Formalismus mit Bezug auf die Arbeit von Nebel und Smolka [1990] definiert. Ihre Beschreibung als 'Attributive Description Formalisms' stützt sich auf eine Analogiebetrachtung dieser merkmalsbasierten Grammatikformalismen im Verhältnis zu den terminologischen Repräsentationsformalismen, die im nächsten Kapitel eingeführt werden. Der Unterschied besteht hauptsächlich in der Verwendung von funktionalen Merkmalen im Gegensatz zu mehrwertigen Rollen sowie verschiedener Zielsetzungen, die mit diesen Formalismen verfolgt werden. Die Analogie bezieht sich auf die jeweilige Formalisierbarkeit als Untermenge der Prädikatenlogik mit einer modelltheoretischen Tarski-Semantik, die für komplexe Strukturen nach dem Kompositionitätsprinzip auf der Basis elementarer Strukturen berechnet werden kann. Sie formulieren jeweils partielle Information über Objekte (einerseits Zeichenketten als linguistische Objekte und andererseits Objekte der Wirklichkeit als Wissenselemente) und erben jeweils Struktureigenschaften boolescher Verbäede (hierzu und zur Semantik der folgenden Beschreibungssprache, siehe Ingenerf [1993]).

Merkmalsstruktur-Beschreibungssprache: Syntax

Gegeben: Menge von atomaren Merkmalswerten (Atom), incl. Top 'T' und Bottom '⊥'
 Menge von Merkmalsnamen (Feature-Namen, Feature)
 Menge von Merkmalsstrukturen ('directed acyclic graphs', Dag)

$\langle \text{Dag} \rangle ::= \text{T} \quad | \quad \perp \quad - \text{'unter- bzw. überspezifizierte Merkmalsstruktur'}$
 $\quad | \quad \langle \text{Feature} : \text{Atom} \rangle \quad - \text{'atomare Merkmalsstruktur'}$
 $\quad | \quad \langle \text{Feature} : \text{Dag} \rangle \quad - \text{'komplexe Merkmalsstruktur'}$
 $\quad | \quad \langle \text{Dag} \rangle \sqcap \langle \text{Dag} \rangle \quad \text{'Konjunktion'}$
 $\langle \text{Pfad} \rangle ::= \langle \text{Feature} \rangle \quad \text{'Pfadinitialisierung'}$
 $\quad | \quad \langle \text{Feature} \rangle : \langle \text{Pfad} \rangle \quad \text{'Pfadverkettung'}$
 $\langle \text{Dag} \rangle ::= \langle \text{Pfad} \rangle \doteq \langle \text{Pfad} \rangle \quad - \text{'reentrant-fähiger Dag' (Agreement, Value-Sharing)}$

Analog zu den kontextfreien Grammatiken dienen Merkmalsgrammatiken als formale Charakterisierungen von korrekt aufgebauten sprachlichen Ausdrücken. Allerdings gehen sie über die üblichen 'string-rewriting rules' in Chomsky-Grammatiken, d.h. Relationen zwischen Zeichenketten hinaus. Parallel zu diesem Konkatenieren von wohlgeformten Zeichenketten zu größeren wohlgeformten Zeichenketten, werden mit den Zeichenketten as-

soziierte Merkmalsstrukturen in Relation gesetzt und gemäß der annotierten Merkmalsgleichungen ausgewertet. Das folgende Beispiel verdeutlicht diesen Zusammenhang.

Grammatik (-Regeln)	Lexikon (Terminale Merkmalsstrukturen)
Die Namen der Variablen bzw. Nonterminals sind 'Syntactic sugar' für den Benutzer.	
S → NP VP	T → "liebt" (mit dem Terminal "liebt")
-----	T:cat = verb
S:cat = satz	T:agree: Numerus = sg
VP:cat = vp	T:agree: person = dritte
NP:cat = VP:subjekt:cat	T:subjekt:cat = np
NP:agree = VP:agree	T:objekt:cat = np
S:agree = VP:agree	T:sem: pred = '^liebt'
VP:subjekt:sem = NP:sem	T:sem: arg1 = T:subjekt:sem
S:sem = VP:sem	T:sem: arg2 = T:objekt:sem
VP → V NP	T → "Norbert" (mit dem Terminal "Norbert")
-----	T:cat = np
VP:cat = vp	T:agree: Numerus = sg
V:cat = verb	T:agree: person = dritte
NP:cat = V:objekt:cat	T:sem = '^Norbert'
VP:agree = V:agree	T → "Uli" (mit dem Terminal "Uli")
VP:subjekt = V:subjekt	T:cat = np
V:objekt:sem = NP:sem	T:agree: Numerus = sg
VP:sem = V:sem	T:agree: person = dritte
	T:sem = '^Uli'
	T → "Menschen" (mit dem Terminal "Menschen")
	T:cat = np
	T:agree: Numerus = pl

Abb. 4 "Kompositionelle Semantik über Verbvalenzkomplemente"

Jeweils der kontextfreie Teil einer Regel (z.B. S → NP VP) legt die Relation zwischen Zeichenketten fest. Die NP-Konstituente ist links von der VP-Konstituente lokalisiert ('linear precedence (LP)') und beide sind unmittelbar der Satz-Konstituente untergeordnet ('immediate dominance (ID)'). Durch diese Relationen ist die Konstituentenstruktur einer grammatikalisch wohlgeformten Zeichenkette festgelegt.

Die einer solchen kontextfreien Regel annotierten Merkmalsgleichungen drücken die Anwendbarkeitsbedingungen der Regel aus und können als Anweisungen verstanden werden, die in ihnen benannten Teilstrukturen (Konstituenten) durch ihre Unifikation zu ersetzen. Anders ausgedrückt: die Anwendbarkeit einer Grammatikregel hängt ab von der Lösbarkeit der annotierten Gleichungen, d.h. der Unifizierbarkeit der beteiligten Merkmalsstrukturen gemäß den Gleichungen. Die als Lösung erzeugte kombinierte Merkmalsstruktur für die 'Mutterkategorie' enthält keine Information, die über die in den Teilstrukturen enthaltene Information hinausgeht (Monotonie)

Im obigen <u>Beispiel</u> ergibt sich nach Analyse des Satzes "Norbert liebt Uli." als erzeugte Merkmalsstruktur beim Startsymbol "S:Dag"	mit: Dag:sem =	<table border="0" style="border-collapse: collapse;"> <tr> <td style="padding-right: 10px;"><i>pred</i></td> <td style="padding-right: 10px;">=</td> <td>'^liebt'</td> </tr> <tr> <td style="padding-right: 10px;"><i>arg1</i></td> <td style="padding-right: 10px;">=</td> <td>'^Norbert'</td> </tr> <tr> <td style="padding-right: 10px;"><i>arg2</i></td> <td style="padding-right: 10px;">=</td> <td>'^Uli'</td> </tr> </table>	<i>pred</i>	=	'^liebt'	<i>arg1</i>	=	'^Norbert'	<i>arg2</i>	=	'^Uli'
<i>pred</i>	=	'^liebt'									
<i>arg1</i>	=	'^Norbert'									
<i>arg2</i>	=	'^Uli'									

Die Analyse des Satzes "Menschen liebt Uli" scheitert, da eine Unifikation der Merkmalsgleichungen nicht gelingt bzw. das Gleichungssystem (die Gleichung '*NP:agree* = *VP:agree*' der ersten Regel) wegen widersprüchlicher '*numerus*'-Merkmalswerte nicht lösbar ist. Dagegen gelingt die Analyse des Satzes "Uli liebt Menschen", da die zweite Regel eine solche Gleichung nicht enthält. Man erkennt im obigen Beispiel auch die Möglichkeit, morphologische (z.B. '*numerus*', '*person*'), syntaktische (z.B. '*cat*', '*objekt*', '*subjekt*') und semantische (z.B. '*pred*', '*arg1*', '*arg2*') Merkmale in uniformer Art und Weise auszudrücken. Komplexe Merkmale lassen sich dann zur Formulierung prägnanter, ausdrucksstarker Merkmalsgleichungen nutzen (zum Beispiel die Merkmale '*agree*' und '*sem*'). Aufgrund des deklarativen Charakters merkmalsbasierter Grammatikformalismen ist man nicht festgelegt bzgl. der Art und Weise der algorithmischen Strukturanalyse (Parsing), d.h. der Bestimmung der Merkmalsstruktur zu einer Zeichenkette gemäß Lexikon und Grammatikregeln.

Die in Abbildung 4 skizzierten Verb-zentrierten Ansätze nutzen die Valenzeigenschaften des Verbs im Zusammenhang mit funktional-semantischen Kategorien (Subjekt, Objekt, ...). Mit Fillmore [1968] spricht man auch von Kasusrahmen, die für jedes Verb gemäß seiner Valenz festlegen, welche semantischen Partner an welchen syntaktischen Positionen erlaubt sind. Solche impliziten semantischen Relationen zwischen Teilstrukturen haben in Nominalsyntaxen i.a. keine Entsprechungen auf der Satzoberfläche [Grewendorf 1979]. Besonders deutlich wird das bei den Komposita, deren potentielle Beteiligung am Aufbau komplexerer Ausdrücke nicht einfach durch im Lexikon vordefinierte Zugriffe auf die bekannten Partner festgelegt werden können. Diese semantischen Partner als auch die verknüpfende Relation sind im vorhinein nicht bekannt.

Aus diesem Grunde wird im folgenden Kapitel ein alternativer Ansatz vorgestellt. Auf der Grundlage einer konzeptuellen Meta-Modellierung lassen sich für alle Ausdrücke geeignete semantische Kategorien sowie deren potentielle Beteiligung am Aufbau komplexerer Ausdrücke (Rollen) benutzeranpaßbar bestimmen. Als semantische Merkmale im merkmalsbasierten Grammatikformalismus lassen sie sich dann im Rahmen einer grammatikalischen Ableitung daten-gesteuert auswerten und '*steuern*' damit die semantische Sprachanalyse. Die dazu notwendige Erweiterung des Chart-Parsing-Verfahrens wird im Kapitel 6 erläutert.

Für die folgenden Überlegungen wird mit Bezug auf Abb. 1 von einer Grammatik ausgegangen, von der aus Übersichtlichkeitsgründen hier nur der '*monadische*' Kern aufgeführt wird. Es wird also nur die syntaktische Kategorie (Wortart) als eines von mehreren, aber wichtigstes Merkmal berücksichtigt. Der Merkmalsausprägung (z.B. *syn:cat* = *np*) dient zugleich als Name der gesamten Merkmalsstruktur (z.B. NP).

NP	→	N	"ein Noun ist (bereits) eine Nominalphrase"
BS	→	Bindestr	"künstlicher Bindestrich zur Kennung von Komposita"
NC	→	N_wt	"ein <u>W</u> ortteil ist (bereits) ein Kompositum"
NC	→	NC BS NC	" <u>B</u> sp.: "Fuss-Pilz"
NC	→	NC BS KONJ NC	" <u>B</u> sp.: "Magen- und Darminfektion"
NP	→	NC_be	"vollständige Komposita von <u>B</u> eginn bis <u>E</u> nde sind NP's"
NP	→	Prefix BS ADJ_wt NP	" <u>B</u> sp.: "post-infektiöse Beschwerden"
NP	→	ADJ NP	" <u>B</u> sp.: "funktionelle Beschwerden"
NP	→	NP ADJ	" <u>B</u> sp.: "Arteria renalis"
NG	→	NP	"eine Nominalphrase ist (bereits) eine Nominalgruppe"
NG	→	NP GEN NG	" <u>B</u> sp.: "Entzündung der großen Arterie der Lunge"
NG	→	NG PRAEP NS	" <u>B</u> sp.: "Abschürfung an Hand und Fuss"
NS	→	NG	"eine Nominalgruppe ist (bereits) eine Nominalsyntaxme"
NS	→	NG KONJ NS	" <u>B</u> sp.: "Abschürfung an Hand und Fuss-pilz"

Abb. 5: "Ausschnitt aus monadischem Teil der hier syntax-orientierten Grammatikregeln"

Analog zu den morphosyntaktischen Merkmalen, wie sie in Abb. 4 zur Beschreibung der abzuleitenden Merkmalsstrukturen zur Charakterisierung von syntaktisch korrekten Ausdrücken benutzt wurden, werden jetzt geeignete semantische Merkmale ergänzt.

Beispiel 1 (syntaktische Mehrdeutigkeit):

"Abschürfung an (Hand und Fuß) versus (Abschürfung an Hand) und Pilz

Lösung 1 (Angabe von geeigneten semantischen Kategorien):

"Nosologie (Krankheit) RelOrt (Topographie UND Topo) versus (Nos RelOrt Topo) UND Nos

Bemerkung 1 (= syntaktische Kategorie auf syntaktischer Ebene):

Die Relation bzw. Konjunktion als Verknüpfung der Teilausdrücke sind **explizit** gegeben.

Beispiel 2 (unbekannte Relation):

"Fuß-pilz" versus "Fuß-Virus" (letzteres ist kein semantisch wohlgeformter Ausdruck)

Lösung 2 (Angabe potentieller Relationen mit Ausdrücken anderer Kategorien)

(Nos RelOrt Topo) versus (Topo Rel? Ätiologie) "ätiolog. Ausdruck ist kein erlaubter Kontext"

Bemerkung 2 (= syntaktische Relationen, ausgedrückt durch kontextfreie Grammatikregeln):

Die verknüpfende Relation ist nur **implizit** durch die beteiligten Kategorien gegeben.

Beispiel 3 (Relationierung oder Konjunktion):

"virale Broch-itis" versus "Tracheo-bronch-itis"

Lösung 3 (gleiche Kategorie heißt Konjunktion, sonst Relationierung wie bei Beispiel 2)

"(Nos RelOrt Topo) RelUrsache Ätiologie) versus Nos RelOrt (Topo UND Topo)

Bemerkung 3: siehe Beispiele in Abb. 2

Beispiel 4 (lexikalische Mehrdeutigkeit):

"Rippen-bruch1" versus "Rippen-bruch2"

Lösung 4 (Konzeptdefinitionen, über Disjunktheit der Konzepte "Knochen" und "Weichteile")

"(Nos RelOrt Topo1) versus -

mit: Rippen definiert Unterkonzept zu Knochen und Bruch1 definiert als Krankheit an

Knochen (Fraktur) und Bruch2 definiert als Krankheit an Weichteilen (Hernie).

Bemerkung 4: (= 'selectional restrictions' in Abb. 5, siehe Merkmal 'agree'):

Konzeptdefinitionen dienen auch der Informationsanreicherung,

z.B. "Tuberkulose definiert als Granulom an der Lunge verursacht durch Bakterien".

Beispiel 5 (Relationierung oder Konjunktion, lexikalische Mehrdeutigkeit):

"Kopf-haut1-entzündung" versus "Hirn-haut2-entzündung"

Lösung 5 (Verfeinerung der sem. Kategorien und entspr. Relationen, Konzeptdefinitionen)

"Nos RelOrt1 (Topo1 RelOrt2 Topo2) versus Nos RelOrt1 (Topo1 UND Topo3)

Bemerkung 5 (= 'selectional restrictions' wie in Beispiel 4, Verfeinerung der Kategorien):

Der mehrdeutige systemtopographisch gebrauchte Ausdruck "Haut1" (Topo1) als Außen-

haut (Kutis) and Regionen (Topo2, z.B. "Kopf") und Haut2 (Topo1) als Schleimhaut (Tunica) an Organen (Topo3, z.B. "Hirn"), die Konzeptdefinitionen erlauben für das Konzept "Haut" im Sinne einer Kutis nur Regionen, im Sinne einer Tunica nur Organe als Partner, siehe Kapitel 4 und 5.

In den folgenden Kapiteln wird eine Konzeptbeschreibungssprache und eine darauf aufbauende Meta-Modellierung skizziert, mit denen die oben zitierten Lösungsansätze verwirklicht werden. Die benutzten Beispiele werden dazu wieder aufgegriffen.

4. Motivation einer Meta-Modellierung

Ein Konzeptrepräsentationsformalismus stellt eine formale Sprache bereit, mit der auf der Basis von gegebenen atomaren Konzepten sowie Rollen zwischen Konzepten komplexere Konzepte konstruiert werden können. Für die weiteren Ausführungen wird eine solche Konzeptbeschreibungssprache sowie eine daran gekoppelte Objektbeschreibungssprache eingeführt. Dabei wird statt der bisherigen halbformalen Notation eine formale Notation mit " \wedge Konzept" statt "KONZEPT" sowie " \forall Konzept" statt "aKONZEPT" zugrundegelegt (bzgl. der Präfixe " \wedge " und " \forall " siehe Frost [1986, S.385]). In von Luck [1991] werden weitere Sprachkonstrukte sowie deren modelltheoretische Interpretation vorgestellt. Bezüglich der Ausdrucksmächtigkeit und Komplexität der darauf basierenden Inferenzalgorithmen existiert eine Fülle von Arbeiten und Ergebnissen, auf die an dieser Stelle ebenfalls nicht weiter eingegangen wird (siehe Nebel [1990]).

Formale Terminologie: Syntax

$\wedge\langle\text{Konzeptname}\rangle \sqsubseteq \langle\text{Konzeptterm}\rangle$ "notwendige Angaben"

$\wedge\langle\text{Konzeptname}\rangle = \langle\text{Konzeptterm}\rangle$ "notwendige und hinreichende Angaben"

Konzeptbeschreibungssprache: Syntax

Gegeben: : - Menge von Konzepten(-namen), incl. Top 'T' und Bottom '1' bzgl. der implizit ...
 - Menge von Rollen(-namen) ... induzierten Verbandsordnung

$\langle\text{Konzeptterm}\rangle ::= \wedge\langle\text{Konzeptname}\rangle$ 'atomares Konzept'
 | $\langle\text{Konzeptterm}\rangle \sqcap \langle\text{Konzeptterm}\rangle$ 'Konjunktion (als Schnittmenge)'
 | $\forall \langle\text{Rolle}\rangle \bullet \langle\text{Konzeptterm}\rangle$ 'Allquantifizierte Wertrestriktion'
 | $\geq 1 \langle\text{Rolle}\rangle$ 'Kardinalitätsrestriktion'
 | $\leq 0 \langle\text{Rolle}\rangle$ 'Kardinalitätsrestriktion'

Schreibweise: $(\forall^1 R \bullet \wedge K) = (\forall R \bullet \wedge K) \sqcap (\geq 1 R)$

Objektbeschreibungssprache: Syntax

Gegeben: - Menge von Konzepten(-namen)
 - Menge von Rollen(-namen)

$\langle\text{Objektterm}\rangle ::= \forall\langle\text{Konzeptname}\rangle$ 'atomarer Objektterm'
 | $\forall\langle\text{Konzeptname}\rangle \text{ mit } \langle\text{Rolle}\rangle \bullet \langle\text{Objektterm}\rangle$ 'relationale Ergänzung'
 | $\langle\text{Objektterm}\rangle \text{ und } \langle\text{Objektterm}\rangle$ 'Konjunktion (von Aussagen)'

Ein sehr wesentlicher Nutzen einer solchen Konzeptrepräsentation ist die Deduktion impliziter Information, z.B. das Classifying als Anordnung eines definierten Konzeptes an die richtige Stelle eines bereits etablierten Verbandes, z.B:

$\wedge\text{Myom} \sqsubseteq \wedge\text{Geschwulst} \sqcap \forall^1 \text{ntAllg} \bullet \wedge\text{Muskelgewebe}$

$\wedge\text{LeioMyom} \sqsubseteq \wedge\text{Geschwulst} \sqcap \forall^1 \text{ntAllg} \bullet \wedge\text{GlattesMuskelgewebe}$

dann gilt mit: $\wedge\text{GlattesMuskelgewebe} \sqsubseteq \wedge\text{Muskelgewebe}$ auch: $\wedge\text{LeioMyom} \sqsubseteq \wedge\text{Myom}$

Eine Terminologie als Menge definierter Konzepte impliziert nun alleine aufgrund der internen syntaktischen Konzeptdefinitionen eine Verbandsordnung. Hier werden dann anwendungsabhängig einige eher abstrakte Konzepte als semantische Kategorien ausgezeichnet.

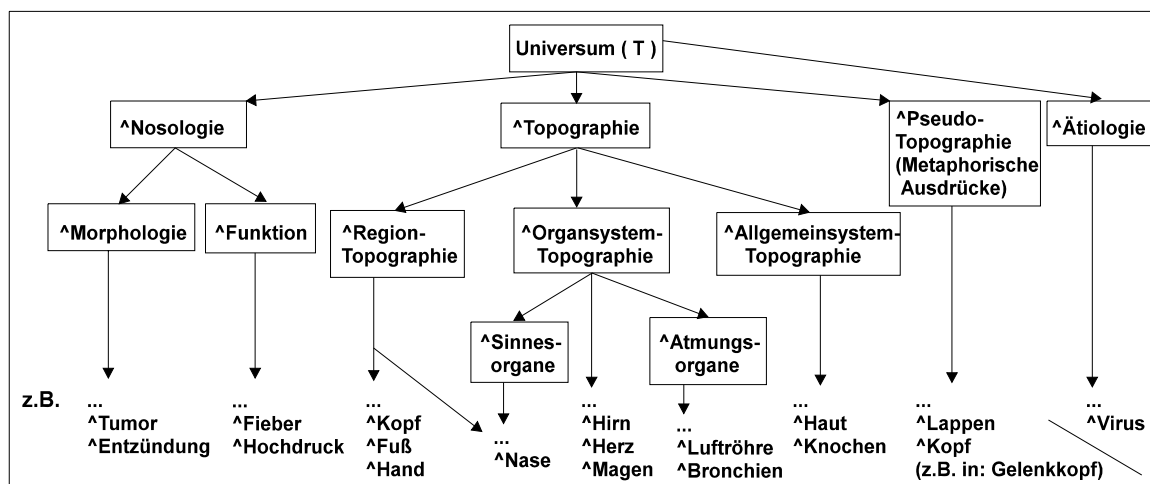


Abb. 6 "Oberer Anteil einer typischen medizinischen Konzepthierarchie (Taxonomie)"

Die selektierten Kategorien werden nun ihrerseits auf einer Meta-Ebene in einem gerichteten Graphen angeordnet. Der Mediziner kann z.B. mit '①' angeben, daß er von Krankheiten an Lokalisationen bzw. wegen Ursachen redet, d.h. damit wird die Rolle zwischen den Kategorien implizit mitgeliefert. Ein solches Modell kann er verfeinern ('②'). Eine Namensgebung der damit implizit festgelegten Rollen ist im übrigen sehr schwierig. Hier haben sich die Anfangsbuchstaben der beteiligten Kategorien bewährt.

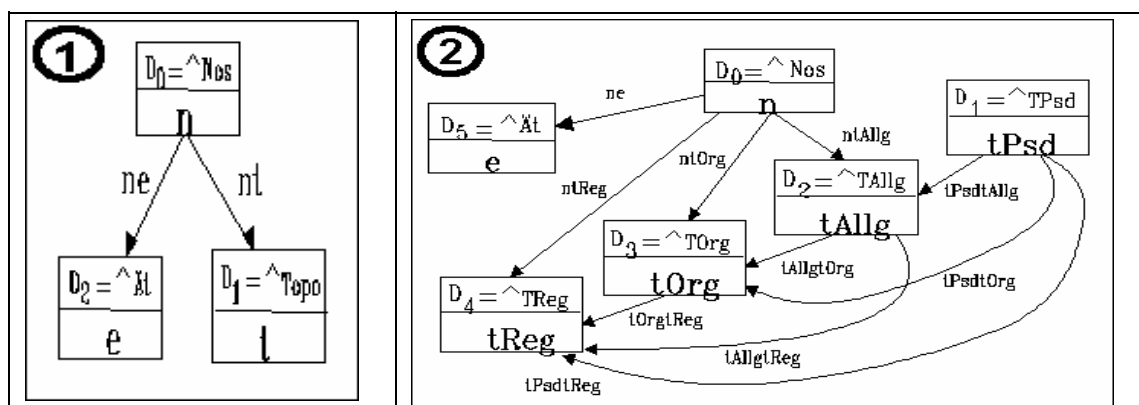


Abb. 7 "Grobes vs. feines Metamodell, d.h. gerichteter Graph mit Kategorien & Rollen"

Im Gegensatz zu den festgelegten Kategorien der SNOMED-Nomenklatur (Topographie, Morphologie, Ätiologie, Funktion usw.) [Wingert 1984] erlaubt das hier vorgestellte Meta-Modell die Auszeichnung allgemeiner Konzepte als semantische Kategorien auf der Basis einer formalen Terminologie und setzt sie in Relation zueinander, z.B. die Verfeinerung der Topographie-Kategorie in Allgemeinsysteme (z.B. "Haut"), Organe (z.B. "Hirn") und Regionen (z.B. "Kopf"). Damit kann eine gewünschte begriffliche Generativität eingestellt werden. Die Wichtigkeit einer inhaltlichen Modellierung zeigt sich mit Lang [1991] auch im LILOG-Projekt: "How to interrelate word knowledge and world knowledge?".

Jeder zu erschließende bedeutungstragende Ausdruck kann als Objektterm und damit als Instanz eines Konzeptes rekonstruiert werden. Durch die Subsumptionsbeziehung dieses

Konzeptes zu einer der ausgezeichneten Konzepte in Abb. 8 kann dem Ausdruck eine semantische Kategorie zugeordnet werden. Mit dem Meta-Modell aus Abb. 9 lassen sich dann die entsprechenden Kontextinformationen zuordnen, d.h. bzgl. welcher Rollen ist der Ausdruck mit welchen 'Partnern' semantisch verknüpfbar. Gehören zwei Ausdrücke derselben Kategorie an, so werden sie konjunktiv verknüpft.

Lösung 1: mit Meta-Modell '①' folgen die Objektterme mit Angabe der Rolle "nt" durch "an":
 "VAbschürfung **mit** nt . (VHand **und** VFuß)" versus "(VAbschürfung **mit** nt . VHand) **und** VPilz)"

Lösung 2: mit Meta-Modell '①' folgt der Objektterm mit impliziter Festlegung der Rolle "nt":
 "VPilz **mit** nt . (VFuß)"

Lösung 3: mit Meta-Modell '①' folgen die Objektterme, wobei die Rollen "nt" und "ne" implizit gegeben sind und die Konjunktion "**und**" Terme gleicher Kategorie verknüpft:
 "(VEntz. **mit** nt . VBronchien) **mit** ne . (Virus)" versus "VEntz. **mit** nt . (VLuftröhre **und** VBronchien)"

Lösung 4: mit Meta-Modell '②' folgt der Objektterm mittels folgender Konzeptdefinitionen:

$$\text{^Rippen} \sqsubseteq \text{^Knochen} \sqcap \forall^1 \text{tAllgtReg} . \text{^Thorax}$$

$$\text{^Fraktur} \sqsubseteq \text{^Bruch} \sqcap \forall^1 \text{ntAllg} . \text{^Knochen}$$

$$\text{^Hernie} \sqsubseteq \text{^Bruch} \sqcap \forall^1 \text{ntAllg} . \text{^Weichteile}$$

Durch Instanziierung dieser definierten Konzepte wird das Homonym "Bruch" disambiguiert. Eine fälschliche Objekttermverknüpfung "VHernie **mit** ntAllg . VRippen" scheitert. Die Konzeptdefinition von "Hernie" verlangt für alle Partner bzgl. der Rolle "ntAllg", daß sie Instanz eines Konzeptes sind, das von "Weichteile" subsumiert wird. Der zu verknüpfende Partner ist allerdings Instanz des Konzeptes "Rippen". Nach Einsetzen der Konzeptdefinitionen scheitert die Verknüpfung rein formal erst durch die erforderliche implizite Disjunktheitsannahme zweier 'elementarer' Bruderkonzepte.

$$\text{disjoint}(\text{^Weichteile}, \text{^Knochen}), \text{ und damit: } \text{^Weichteile} \sqcap \text{^Knochen} = \perp$$

$$\text{"VFraktur mit ntAllg . VRippen"}$$

Lösung 5: mit Meta-Modell '②' folgt der Objektterm mittels folgender Konzeptdefinitionen:

$$\text{^Kutis} \sqsubseteq \text{^Haut} \sqcap (\geq 1 \text{tAllgtReg}) \sqcap (\leq 0 \text{tAllgtOrg})$$

$$\text{^Tunica} \sqsubseteq \text{^Haut} \sqcap (\geq 1 \text{tAllgtOrg}) \sqcap (\leq 0 \text{tAllgtReg})$$

$$\text{"Entzündung mit ntAllg . (VKutis mit tAllgtReg . VKopf)" versus}$$

$$\text{"Entzündung mit ntOrg . (VTunica mit tOrgtReg . VHirn)"}$$

Bemerkung: mit dem Meta-Modell '①' gehörten die Terme "Haut", "Kopf" und "Hirn" alle der semantischen Kategorie "Topo" an, so daß sie fälschlicherweise kon-

junktiv verknüpft würden; die obigen Definitionen wären nicht formulierbar.

Der automatische Transfer der obigen Informationen als Merkmalsausprägungen geeigneter Merkmale des Grammatikformalismus wird im folgenden Kapitel skizziert. Diese werden im Rahmen des Standardauswertungsverfahrens - nämlich der Unifikation von Merkmalsstrukturen - ausgewertet und steuern damit die grammatikalische Ableitung.

5. Steuerung der grammtikalischen Ableitung

Eine Fachsprache stellt nun ihrerseits Mittel bereit, komplexe Ausdrücke aus einfacheren zu bilden. Eine Bedeutungsextraktion aus fachsprachlichen Strukturen stützt sich gemäß dem Kompositionalitätsprinzip auf eine morpho-syntaktische Beschreibungsebene zur

a) Kontrolle der Bedeutungsextraktion (Gewährleistung korrekter Fachsprache), d.h.

- Morphologie: korrekte Wortbildungen (z.B. "Entzündungs"),

- Syntax: syntaktisch wohlgeformte Ausdrücke (z.B. "Entzündung nach virale")

und für einen

b) Beitrag zur Bedeutungsextraktion über Inhaltsworte (bzw. Lexeme) hinausgehend.

- Morphologie: z.B. "Beine" versus "Bein", "Nieren-becken" versus "Becken-niere"

- Syntax: z.B. "Beschwerden nach Infektion" vs. "Infektion vor Beschwerden" vs. "postinfektiöse Beschwerden"

Semantisches Kompositionalitätsprinzip

Die Bedeutung von komplexen, nicht lexikalisierten Ausdrücken kann alleine aus der Bedeutung der Teilausdrücke sowie ihrer syntaktischen Verknüpfung ermittelt werden.

Lexikalischer Zugriff (bzw. Lexikalische Analyse)

(i.a. inclusive: Morphologische Analyse)

Syntaktische Analyse

Sei als Eingabe ein nicht-lexikalischer Ausdruck als Sachverhaltsaussage gegeben. Andernfalls wird im Lexikon auf die Bedeutung zugegriffen, falls es sich um ein bedeutungstragendes Lexem handelt (z.B. "Bronchien"). Oder es handelt sich um ein Funktionswort (z.B. die Präposition "an") ohne selbständige Bedeutung. Für nicht-lexikalische Ausdrücke ist über die Teilbedeutungen hinaus eine Berücksichtigung der syntaktischen Struktur nötig.

Für den Ausdruck "Bronch-itis" kann gemäß dem Meta-Modell '①' ermittelt werden, daß die Teilausdrücke "Bronchien" der semantischen Kategorie "n" und "Entzündung" der Kategorie "t" angehören. Darüberhinaus lassen sich aus den Relationen zu dominanten (Vater-) sowie abhängigen (Sohn-) Kategorien eines solchen Modellgraphen die entsprechenden Kontextinformationen ableiten, d.h. bzgl. welcher Rollen der Ausdruck mit welchen 'Partnern' semantisch verknüpfbar ist. Zusätzlich erfordert das Kompositionalitätsprinzip einen Kongruenztest. Und zwar muß der semantisch dominante Partner (hier "Entzündung")

bzgl. der semantischen Rolle (hier "nt") gleichzeitig die syntaktisch dominante Teilstruktur sein. In einem Kompositum trägt z.B. das letzte Wortteil die Gesamtbedeutung; in diesem Falle also die "-itis". Ohne einen solchen Kongruenztest könnte man nicht wohlgeformte Ausdrücke (z.B. "Entzündungsbronchien") nicht vermeiden sowie die Bedeutungsunterschiede in den Ausdrücken "Nieren-becken" versus "Becken-niere" bzw. "Kopf-gelenk" versus "Gelenk-kopf" nicht differenzieren.

Dieser skizzierten Idee wird nun durch eine geeignete Koppelung der terminologischen Hintergrundrepräsentation inklusive des Meta-Modells mit dem merkmalsbasierten Grammatikformalismus inklusive des ableitenden Chart-Parsers entsprochen. Ein Großteil der semantischen Auswertung kann nach einem Transfer der Information in Ausprägungen geeigneter Merkmale bereits durch Unifikation von Merkmalsstrukturen geleistet werden.

A) Lexikalische Analyse und Initialisierung der Merkmalsstrukturen

- Zugriff auf morphosyntaktische Merkmale, insbesondere die syntaktische Kategorie (weniger wichtige morphologische Merkmale werden hier weggelassen) sowie die
- Ableitung semantischer Merkmale, insbesondere die lexikalische Bedeutung sowie die semantische Kategorie und die Kontextinformationen gemäß Meta-Modell.

B) Grammatikregeln aus Abb. x mit angepassten Merkmalsgleichungen (incl. Kongruenztest).

C) Verknüpfung zweier Konstituenten und Erweiterung des Chart-Parser-Prinzips

- Auswertung unifizierter Kontextterme und Ableitung der u.U. impliziten Rolle 2'er Terme,
- Verknüpfung 2'er Objektterme als Bedeutungsrepräsentation des komplexeren Ausdrucks,
- Weitergabe geeigneter semantischer Merkmale an die Mutterkonstituente

Zunächst werden unter A) im Rahmen der lexikalischen Analyse einige Merkmale bereits im Lexikon vordefiniert bzw. über Grammatikregeln abgeleitet, insbesondere

- $X : \text{syn} : \text{cat} =$ syntaktische Kategorie eines Ausdrucks und
- $X : \text{sem} : \text{erg} =$ Objektterm als Bedeutung eines Ausdrucks

bzw. gemäß der Konzeptdefinitionen und dem Meta-Modell (im folgenden $\textcircled{1}$) abgeleitet:

- $X : \text{sem} : \text{cat} =$ semantische Kategorie für bedeutungstragende Ausdrücke (Sorte)
- $X : \text{sem} : \text{rel}_u =$ Kontextterme für bed. Ausdrücke ('Partner' ungleicher Kategorie)
- $X : \text{sem} : \text{rel}_g =$ Kontextterme für bed. Ausdrücke ('Partner' gleicher Kategorie)
- $X : \text{sem} : \text{rel}_r$ bzw. $X : \text{sem} : \text{rel}_l =$ Kontextterme für Funktionsworte ('Partner' gleicher Kat.)

Interessant ist nun die Strategie, mit der die Kontextterme aufgebaut werden. Gemäß dem Meta-Modell werden alle potentiell möglichen semantischen 'Partner' als Kontextterme $\text{kt}(_, _, _)$ aufbereitet, die als Merkmalsausprägungen geeigneter Merkmale aufgefaßt werden. Ähnliche Kontext-Informationen nutzte bereits Heydthausen [1988] in den von ihm entwickelten 'relationalen Grammatiken'. Dabei unterscheiden wir drei Fälle:

- Relationen zwischen Ausdrücken ungleicher semantischer Kategorie (laut Meta-Modell):

- der abhängige Partner mit Objektterm "Ot" bzgl. einer semantischen Rolle ("rollej") erwartet den dominanten Partner an der syntaktisch dominanten Position,
 $X : sem : rel_u : rollej = kt(_, Ot, \mathbf{Var})$
 $X : sem : head = \mathbf{Var}$
- der dominante Partner mit Objektterm "Ot'" bzgl. einer semantischen Rolle ("rollej") erwartet den abhängigen Partner und spezifiziert explizit die eigene semantische Kategorie,
 $X : sem : rel_u : rollej = kt(Ot', _, \mathbf{SemCat})$
- Relationen zwischen Ausdrücken gleicher sem. Kategorie 'semCat' (über Funktionsworte):
 - jeder Ausdruck mit Objektterm "Ot" kann bzgl. einer semantischen Rolle ("rolleg",) (z.B. "nach") bzw. einer Konjunktion Argument an beiden Argumentpositionen sein,
 $X : sem : rel_g : semCat = kt(_, Ot, _) \text{ bzw. } X : sem : rel_g : junktor = kt(_, Ot, _)$
 - für Funktionsworte (etablieren Relationen zwischen Ausdrücken gleicher sem. Kategorie):
 - Funktionsworte für gerichtete Rollen (z.B. "nach") verlangen Argumentpositionen (1,2),
 $X : sem : rel_l : semCat = kt(rolleg, _, 1) \text{ bzw. } X : sem : rel_r : semCat = kt(rolleg, _, 2)$
 - Konjunktionen (und auch ungerichtete Rollen) verlangen keine Argumentpositionen,
 $X : sem : rel_l : junktor = kt(konj, _, _) \text{ bzw. } X : sem : rel_r : junktor = kt(konj, _, _)$

Beispiel:beachte:

"Entzündung" →	<pre>syn: cat: noun erg: Ot = [·^VEntzündung'] cat: Cat = [n] sem: rel_u: nt: kt (Ot, _, Cat) ne: kt (Ot, _, Cat) rel_g: n: kt (_, Ot, _) junktor: kt (_, Ot, _)</pre>	<p>Mit Meta-Modell '①' gibt es für die semantische Kategorie "n" die abhängigen Partner "t" und "e" bzgl. der Rollen "nt" und "ne". Letztere fungieren unter "rel_u" als Merkmalsnamen.</p>
----------------	--	---

"Bronchien" →	<pre>syn: cat: noun erg: Ot = [·^VBronchien'] cat: t sem: rel_u: nt: kt (_, Ot, Var) rel_g: t: kt (_, Ot, _) junktor: kt (_, Ot, _) head: Var</pre>	<p>Mit Meta-Modell '①' gibt es für die semantische Kategorie "t" nur den dominanten Partner "n" bzgl. der Rolle "nt". Letztere fungiert unter "rel_u" als Merkmalsname.</p>
---------------	---	---

"und" →	<pre>syn: cat: konj relator: Rel = [konj] sem: rel_l: junktor: kt (Rel, _, _) rel_r: junktor: kt (Rel, _, _)</pre>
---------	---

"nach" →	<pre>syn: cat: praep relator: Rel = [zeitl_folg] sem: rel_l: n: kt (Rel, _, 1) rel_r: n: kt (Rel, _, 2) erlaubt: gleiche_semantik</pre>
----------	--

Bemerkung:

Das Funktionswort "vor" erhält die gleiche Merkmalsstruktur. Nur die Argumentpositionen (Pos = 1 bzw. 2) in "kt(Rel, _, Pos)" werden vertauscht.

Nach der Spezifikation bzw. Initialisierung semantischer Merkmale im Rahmen der lexikalischen Analyse werden nun unter dem obigen Punkt B) die Grammatikregeln um geeignete Merkmalsgleichungen erweitert. Dabei differiert die Vorgehensweise für drei Regeltypen:

- Abschlußregeln, auch Kettenproduktionen genannt (bloßes Hochschieben der Information):

NP → N	NS → NG
NP : <i>sem</i> = N : <i>sem</i> , ...	NS : <i>sem</i> = NG : <i>sem</i> , ...
NG → NP	NP → NC
NG : <i>sem</i> = NP : <i>sem</i> , ...	NP : <i>sem</i> = NC : <i>sem</i> , ...

- binäre Regeln, die zwei 'Partner' (Konstituenten) ungleicher sem. Kategorie verknüpfen:

NP0 → ADJ NP1	NC0 → NC1 BS NC2 ("Kompositaregel 1")
ADJ : <i>sem : rel_u</i> = NP1 : <i>sem : rel_u</i> ,	NC1 : <i>sem : rel_u</i> = NC2 : <i>sem : rel_u</i> ,
ADJ : <i>sem : head</i> = NP1 : <i>sem : cat</i> ,	NC1 : <i>sem : head</i> = NC2 : <i>sem : cat</i> ,
NP0 : <i>test : rel_u</i> = NP1 : <i>sem : rel_u</i> , ...	NC0 : <i>test : rel_u</i> = NC2 : <i>sem : rel_u</i> , ...
NG0 → NP GEN NG1	NG0 → □ NG1 PRAEP NS ("Präpositionsregel 1")
NP : <i>sem : rel_u</i> = NG1 : <i>sem : rel_u</i> ,	NG1 : <i>sem : rel_u</i> = NS : <i>sem : rel_u</i> ,
NG1 : <i>sem : head</i> = NP : <i>sem : cat</i> ,	NS : <i>sem : head</i> = NG1 : <i>sem : cat</i> ,
NG0 : <i>test : rel_u</i> = NP : <i>sem : rel_u</i> , ...	NG0 : <i>test : rel_u</i> = NG1 : <i>sem : rel_u</i> ,
	PRAEP : <i>erlaubt</i> = ungleiche_semantik , ...

- binäre Regeln, die zwei 'Partner' (Konstituenten) gleicher sem. Kategorie verknüpfen

NC0 → NC1 BS NC2 ("Kompositaregel 2")	NG0 → NG1 PRAEP NG2 ("Präpositionsregel 2")
NC1 : <i>sem : cat</i> = NC2 : <i>sem : cat</i> ,	NG1 : <i>sem : cat</i> = NG2 : <i>sem : cat</i> ,
NC1 : <i>sem : rel_g</i> = BS : <i>sem : rel_l</i> ,	NG1 : <i>sem : rel_g</i> = PRAEP : <i>sem : rel_l</i> ,
NC2 : <i>sem : rel_g</i> = BS : <i>sem : rel_r</i> ,	NG2 : <i>sem : rel_g</i> = PRAEP : <i>sem : rel_r</i> ,
NC0 : <i>test</i> = BS : <i>sem</i> , ...	NG0 : <i>test</i> = PRAEP : <i>sem</i> ,
NS0 → NG KONJ NS1	PRAEP : <i>erlaubt</i> = gleiche_semantik , ...
NG : <i>sem : cat</i> = NS1 : <i>sem : cat</i> ,	
NG : <i>sem : rel_g</i> = KONJ : <i>sem : rel_l</i> ,	
NS1 : <i>sem : rel_g</i> = KONJ : <i>sem : rel_r</i> ,	
NS0 : <i>test</i> = KONJ : <i>sem</i> , ...	

Bem.: Durch eine Präposition wird i.a. nur eine der beiden 'Präpositionsregeln' gemäß dem 'erlaubt'-Merkmal selektiert (z.B. "nach" für gleiche und "an" für ungleiche Semantik). Für Komposita werden beide 'Kompositaregeln' aktiviert und erst durch gleiche oder ungleiche Partner ausdifferenziert.

Abb. 8 "Semantische Merkmale in Merkmalsgleichungen der Grammatikregeln aus Abb. 4"

Die 'Ausstattung' des Grammatikformalismus mit semantischen Merkmalen (die morpho-syntaktischen Merkmale wurden hier weggelassen, siehe Abb. 8) erlaubt nun eine Steuerung der grammatikalischen Ableitung in dem Sinne, daß nur bedeutungstragende Konstituenten mit 'sortengerechten' Objekttermen als Ergebnisrepräsentation gemäß dem Meta-Modell ableitbar sind. Die Kontextterme haben dabei denselben Effekt wie die Kasusrahmen bei Verb-zentrierten Ansätzen gemäß Abb. 8.

Ich komme damit zum letzten Punkt C) der Koppelung der terminologischen und grammatikalischen Beschreibungsformalismen. Um für diese Ausarbeitung ein vertretbares Maß an Ausführlichkeit nicht zu überschreiten, wird anhand des einfachen Beispiels

"Bronch-itis" die zugrundeliegende Idee skizziert. Eine Segmentierung auf morphologischer Ebene habe bereits die Wortteile "Bronchien" und "Entzündung" erkannt und verknüpft sie auf syntaktischer Ebene durch einen Bindestrich ("BS"). Damit wird die obige Kompositaregel 1 (NC0 → NC1 BS NC2) mit den entsprechenden Merkmalsgleichungen anwendbar. Es handelt sich um den interessanteren Fall der Verknüpfung zweier Konstituenten mit ungleicher Kategorie.

Betrachtet man die Merkmalsstrukturen der beiden bedeutungstragenden Konstituenten, so ist bei beiden unter "*sem:rel_u:nt*" ein Kontextterm spezifiziert, nämlich "*kt (^VEntzündung', _, n)*" für die Konstituente "NC2" und "*kt (_, ^VBronchien', Var)*" für die Konstituente "NG1". Gemäß der Merkmalsgleichung "*NC1:sem:rel_u = NC2:sem:rel_u*" und werden sie und - falls vorhanden - auch alle weiteren potentiellen Kontextterme unifiziert. Für die gemeinsame Rolle "nt" werden also gerade die komplementären Kontextterme unter dem gemeinsamen Merkmal "*nt*" durch beide Partner vervollständigt ('dichter Kontextterm', hier "*kt (^VEntzündung', ^VBronchien', Var = n)*"). Die zweite Gleichung "*NC1:sem:head = NC2:sem:cat*" fordert, daß die semantische Kategorie des syntaktischen 'Head' (hier die Konstituente "NC2"; d.h. der letzte Wortteil trägt in Komposita die Bedeutung; in anderen Regeln sind andere Konstituenten syntaktisch dominant) gleich der laut Meta-Modell dominanten Kategorie bzgl. der Rolle "nt" ist. Diese Bedingung wird im Kontextterm des abhängigen Partners der jeweiligen Rolle (hier: die Konstituente "NC1") über 'Variablen-Sharing' ausgedrückt. Nach der Unifikation der beiden Kontextterme fordert die Gleichung die (erfolgreiche) Unifizierbarkeit von "*Var = n*" und "*n*". Mit dieser Strategie wird also dem oben beschriebenen Kompositionalitätsprinzip im allgemeinen und dem skizzierten Kongruenztest im besonderen entsprochen. Mit der letzten Gleichung "*NC0:test:rel_u = NC2:sem:rel_u*" wird die gesamte Merkmalsstruktur an die Mutterkonstituente zur weiteren Auswertung weitergereicht. An dieser Stelle ist beachtenswert, daß die Kontextterme in ihrer Aussagekraft über die üblicherweise nur als atomare Werte zugelassenen Merkmalsausprägungen in merkmalsbasierten Grammatikformalismen hinausgehen.

Zur Algorithmisierung der grammatikalischen Ableitung wird hier das Chart-Parsing-Verfahren verwendet. Die Nutzung der semantischen Informationen, insbesondere der Kontextinformationen in Kongruenz zu den ableitenden Regeln, begründet die Favorisierung einer Bottom-up-Parsingstrategie. Zum Abschluß dieses Kapitels wird nun eine notwendige Modifikation des Standardmechanismus des Chart-Parsings, genauer der 'Fundamental Rule' für die Expansion einer passiven und aktiven Kante, die jeweils Konstituenten mit selbständiger Bedeutung repräsentieren, angedeutet. Diese Modifikation beinhaltet die Auswertung der unifizierten Kontextterme und damit die Ableitung der u.U. impliziten Rolle zwischen zwei Termen, die Verknüpfung zweier Objektterme als Bedeu-

tungsrepräsentation des komplexeren Ausdruckes sowie die Weitergabe geeigneter semantischer Merkmale an die Mutterkonstituente.

Sind die zwei zu verknüpfenden Objektterme semantisch laut Meta-Modell verträglich, so muß es einen dichten Kontextterm bzgl. einer gemeinsamen Rolle (als Grammatik-Merkmal !) unter dem Merkmal "*test:rel_u*" der Mutterkonstituente (im Beispiel: "NC0") geben, d.h. alle drei Argumente des Kontextterms sind dicht bzw. instantiiert (Grundterm). Das Merkmal (im Beispiel: "*nt*") liefert also die gesuchte implizite Rolle. Bei nicht verträglichen Objekttermen würde die Unifikation scheitern, oder es gibt keine dichte Rolle. Bei Verknüpfung zweier Terme gleicher Kategorie werden die Rolle sowie die Argumente über den Funktor bzw. die Konjunktion - Konstituente abgegriffen. Auf diesen Fall gehe ich hier nicht näher ein.

Fazit: die Lösungsskizzen der Beispiele 1, 2 und 3 der Kap. 3, 4 werden hiermit abgedeckt.

Die Objektterme der beiden bedeutungstragenden Konstituenten sowie die verknüpfende Rolle stehen also zur Verfügung. Die beiden Objektterme stellen, wie bereits formuliert, Instanzen von Konzepten dar, die u.U. in einer Terminologie definiert wurden. Diese Konzeptdefinitionen können nun im Rahmen einer Objekttermverknüpfung zur Disambiguierung lexikalischer Mehrdeutigkeit genutzt werden (z.B. "Rippen-bruch" im letzten Kapitel). Damit geht dieser Ansatz über die Unifikation als einzige auswertende und informationsanreichernde Operation in merkmalsbasierten Grammatikformalismen hinaus.

Fazit: die Lösungsskizzen der Beispiele 4 und 5 der Kap. 3 und 4 werden hiermit abgedeckt.

Als letztes wird der hergeleitete Objektterm als selbständige Bedeutung der Mutterkonstituente unter das Merkmal "*NC0 : sem: erg*" repräsentiert. Die semantische Kategorie und die Kontextterme werden i.a. von der syntaktisch dominanten Konstituente (im Beispiel: "*NC : sem*") übernommen. Die Auswertung der unifizierten Kontextterme geschah unter dem Merkmal "*test:rel_u*", welches nicht kopiert wird. Auf Besonderheiten, wie die Berücksichtigung eines Phänomens 'Rollenkonsumierung', wird hier nicht weiter eingegangen. Die so präparierte Mutterkonstituente geht nun wiederum ihrerseits ein in das Standard Bottom-up-Chart-Parsing. Das ganze wird solange iteriert, bis der Ausdruck gemäß der Grammatik abgeleitet ist bzw. auf das Startsymbol (hier eine Konstituente mit 'ns' als syntaktischer Kategorie und "n" = 'Wurzel des Meta-Modells' als semantische Kategorie) reduziert wurde. Zu beachten ist, daß die Monotonieeigenschaft des Chart-Parsing-Verfahrens durch diesen 'Eingriff' gezielt verletzt wird, was u.a. eine Reihenfolgeabhängigkeit der Regelanwendung impliziert.

Endergebnis:

"Bronch-itis" →	<i>syn:</i>	<i>cat:</i>	ns
		<i>erg:</i>	Ot = [^v Entzündung mit nt • ^v Bronchien]
		<i>cat:</i>	Cat = [n]
	<i>sem:</i>	<i>rel_u:</i>	ne: kt (Ot, _, Cat)
		<i>rel_g:</i>	n: kt (_, Ot, _)
		<i>junktor:</i>	kt (_, Ot, _)

Der vorliegende Ansatz wurde als Prototyp mit dem Namen 'SMART' (Sprachverarbeitungssystem in der Medizin: Begriffliche Analyse und Repräsentation auf der Basis formaler Terminologien) auf einem PC unter Verwendung des Werkzeugs ARITY-PROLOG [Arity Corporation 1988 und 1989] implementiert. Dem Benutzer steht mit einer Auswahl an Parsing-Strategien und verschiedenen Kontroll- und Steuerungsmechanismen eine flexible und transparente Entwicklungsumgebung für eine semantische Sprachanalyse auf der Basis formaler Terminologien zur Verfügung.

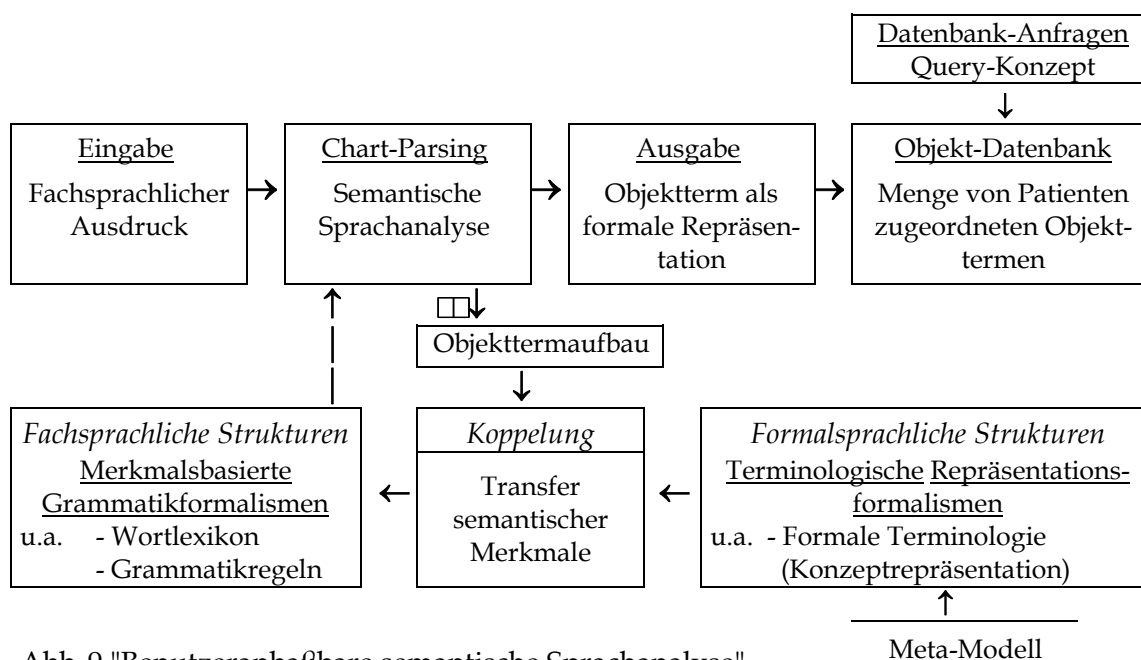


Abb. 9 "Benutzeranbaßbare semantische Sprachanalyse"

7. Ausblick

Dieses Methodenspektrum hat sich bereits im LILOG-Projekt bewährt [Herzog, Rollinger 1991] und wurde hier im wesentlichen durch die Idee einer Meta-Modellierung ergänzt. An dieser Stelle seien nur in Stichworten zwei verwandte Arbeiten mit einer allerdings jeweils verschiedenen Zielsetzung erwähnt. Zum einen sind das Pereira und Pollack [1991] mit der Absicht "Account for the influences of context on interpretation, while preserving compositionality to the extent possible."; und zum anderen Backofen et al. [1991] mit der Idee "partial consistency checks during the unification of feature terms; final consistency checks by classifying concept terms". In Gazdar und Mellish [1989, S.293] wird ebenfalls die

Tragfähigkeit der Arbeitshypothese "Computational Semantics as feature instantiation" bezweifelt; d.h. die Ableitbarkeit semantischer Strukturen entlang grammatikalischer Strukturen.

Literatur:

- ARITY CORPORATION (1988). *Arity/Prolog Language Reference Manual*. Concord, Ma..
- ARITY CORPORATION (1989). *Arity/Expert Language Reference Manual, Taxonomy Language und Classifier*. Concord, Ma..
- BACKOFEN, R., TROST, H., USZKOREIT, H. (1991). *Linking Typed Feature Formalisms and Terminological Knowledge Representation Languages in Natural Language Front-Ends*. In: Brauer, W., Hernández, D. (Hrsg.). *Verteilte Künstliche Intelligenz und kooperatives Arbeiten*. Proc. des 4. Int. GI-Kongreß 'Wissensbasierte Systeme', München, Okt.1991, Informatik-Fachberichte 291. Berlin: Springer., 375-383.
- BAUD, R.H., RASSINOX, A.-M., SCHERRER, J.-R. (1992). *Natural Language Processing and Semantical Representation of Medical Texts*. *Methods of Information in Medicine*, 31, 117-125.
- BRACHMAN, R.J., SCHMOLZE, J.G. (1985). *An Overview of the KL-ONE Knowledge Representation System*. *Cognitive Science* 9, 171-216.
- DOYLE, J., PATIL, R.S. (1991). *Two theses of knowledge representation: language restrictions, taxonomic classification, and the utility of representation services*. *Artificial Intelligence* 48, 261-297.
- FILLMORE, C. (1968). *The Case for Case*. In: Bach, E., Harms, R. (Eds.). *Universals in Linguistic Theory*. New York: Holt, 1-88.
- FROST, R.A. (1986). *Introduction to Knowledge Base Systems*. London: Collins.
- FUGMANN, R. (1975). *Die Grenzen des Thesaurus-Verfahrens bei der Wiedergabe von Begriffsbeziehungen*. *Nachr. Dok.* 26, 2-7.
- GAZDAR, G., MELLISH, C. (1989). *Natural Language Processing in PROLOG: An Introduction to Computational Linguistics*. Reading: Addison-Wesley.
- HAHN, U. (1990). *Lexikalisch verteiltes Text-Parsing: eine objektorientierte Spezifikation eines Wortexpertensystems auf der Grundlage des Aktorenmodells*. *Informatik-Fachberichte* 243, zugel. Diss., Uni. Konstanz, 1987. Berlin: Springer.
- HAIMOWITZ, I.J., PATIL, R.S., SZOLOVITS, P. (1988). *Representing Medical Knowledge in a Terminological Language is Difficult*. In: Greenes, R.A. (Ed.). *Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care*. New York: IEEE Computer Society Press.,101-105.
- HERZOG, O., ROLLINGER, C.-R. (Eds.) (1991). *Text Understanding in LILOG (Integrating Computational Linguistics and Artificial Intelligence: Final Report on the IBM Germany LILOG-Projekt)*, *Lecture Notes in Computer Science* 546. Berlin: Springer.
- INGENERF, J. (1993). *Benutzeranpaßbare semantische Sprachanalyse und Begriffsrepräsentation für die medizinische Dokumentation*. Dissertation an der Mathematisch-Naturwissenschaftlichen Fakultät, RWTH Aachen.
- KILBURY, J. (1985). *Chart Parsing and the Early Algorithm*. In: Klenk, U. (Hrsg.). *Kontextfreie Syntaxen und verwandte Systeme*. Tübingen: Niemeyer., 76-89.
- KLEIN, R. (1992). *TL (Terminologic Logic) Users Workshop, 24.-25. Okt. 1991 in Berlin*. *Künstliche Intelligenz*, 1, 47-48.
- LANG, E. (1991). *The LILOG-Ontology from a Linguistic Point of View*. In: Herzog, O., Rollinger, C.-R. (1991), 464-481.
- V.LUCK, K. (1991). *Hybride logikbasierte Systeme*. *Künstliche Intelligenz* 2, 27-31.
- NEBEL, B. (1990). *Terminological Reasoning is Inherently Intractable*. *Artificial Intelligence* 43, 235-249.

- PEREIRA, F.C.N., POLLACK, M.E. (1991). *Incremental interpretation*. *Artificial Intelligence* 50, 37-82.
- PIETRZYK, P.M. (1988). *Syntaktische Analyse zytopathologischer Beurteilungstexte*. In: Rienhoff, O., Piccolo, U., Schneider, B. (Eds.). *Expert Systems and Decision Support in Medicine, Lecture Notes in Medical Informatics* 36. Berlin: Springer., 443-448.
- RAHMSTORF, G. (1991). *Compositional semantics and concept representation*. In: Bock, H.-H., Ihm, P. (Eds.). *Classification, Data Analysis and Knowledge Organisation - Proc. of the 14th Annual Conf. of the GfKI in Marburg*. Berlin: Springer., 163-172.
- RECTOR, A.L., NOWLAN, W.A., KAY, S. (1992). *Conceptual knowledge: The core of medical information systems*. In: Lun K.C., Degoulet, P., Piemme, T.E., Rienhoff, O. (Eds.) (1992). *MEDINFO-92 - Proceedings of the 7th World Congress on Medical Informatics in Geneva Palexpo, Switzerland*. Amsterdam: North-Holland.
- ROSTEK, L. (1979). *Methoden des partiellen Parsing für das automatische Indexing - Syntaxgraphen zur Analyse von Sprachmustern*. In: Kuhlen, R. (Hrsg.). *Datenbasen, Datenbanken, Netzwerke, Praxis des Information Retrieval*, Bd.1. München: Saur.,251-281.
- SAGER, N., FRIEDMAN, C., LYMAN, M.S. (1987). *Medical Language Processing: Computer Management of Narrative Data*. Reading, Ma.: Addison Wesley.
- SCHEFE, P. (1981). *Zur Funktionalität der Wissenschaftssprache - Am Beispiel der Medizin*. In: Bundgarten, T. (Hrsg.). *Wissenschaftssprache*. München: Fink. 356-371.
- SHIEBER, S.M. (1986). *An Introduction to Unification-based Approaches to Grammar*. CSLI Lecture Notes, no.4.
- SCHMIEDEL, A. (1991). *Modeling the Medical Data of the Heart Catheter Report*. Interner Arbeitsbericht, Projektgruppe Medizin Informatik, Berlin.
- TROST, H. (1989). *Begriffsbildung im Bereich Natürlichsprachiger Systeme*. In: Retti,J., Leidlmair,K. (Hrsg.). 5. *Österreichische AI-Tagung*. Berlin: Springer. 261-269.
- WINGERT, F. (1984). *SNOMED - Systematische Nomenklatur der Medizin, 2 Bände*. New York: Springer.

Stichworte: kompositionelle Semantikanalyse, Konzeptrepräsentation, Meta-Modellierung

Josef Ingenerf studierte an der RWTH Aachen Informatik mit Nebenfach Operations Research. Nach dem Diplom (Arbeit über das automatische Indexieren medizinische Diagnosen auf der Basis der SNOMED-Nomenklatur) arbeitete er seit 1987 als wissenschaftlicher Angestellter am Institut für Medizinische Informatik und Biometrie am Klinikum der RWTH Aachen. Nach seiner Promotion (Arbeit über eine benutzeranpaßbare semantische Sprachanalyse und Begriffsrepräsentation für die medizinische Dokumentation) arbeitet er seit Herbst 1993 als wissenschaftlicher Assistent weiterhin am genannten Institut. Er leitet u.a. den informatischen Arbeitskreis 'Terminologie' im Rahmen des BMFT-Förderschwerpunktes MEDWIS (Medizinische Wissensbasen).