

Seminararbeit

**Stefan Müllers HPSG-Grammatik für das Deutsche:
Grundbegriffe, Dominanzstrukturen, Konstituentenreihenfolge**

Thema und Zeitpunkt des Seminars:

Syntaxtheorien und computerlinguistische Praxis
Sommersemester 2000

Dozenten:

Prof. Dr. M. Hess
lic. phil. G. Schneider
lic. phil. S. Clematide

Institut für Informatik der Universität Zürich
Abteilung Computerlinguistik

Hans Estermann
Dohlenweg 28
8050 Zürich

Tel.: 01 / 301 19 72
Mail: estermann@dplanet.ch

18. Mai 2000

Inhalt

Einleitung	1
1 Grundbegriffe der HPSG.....	2
1.1 Aufbau von Merkmalstrukturen	4
1.2 Kopfmerkmale und Kopfmerkmalprinzip.....	6
1.3 Subkategorisierungsmerkmal (SUBCAT) und Subkategorisierungsprinzip	8
2 Dominanzstrukturen	10
2.1 Flache Dominanzstrukturen.....	10
2.2 Problemfälle flacher Strukturen	16
2.3 Binär verzweigende Dominanzstrukturen.....	17
3 Konstituentenreihenfolge im Deutschen	20
3.1 Modell des deutschen Satzes	20
3.2 Wortstellungsphänomene im Deutschen.....	21
3.2.1 Topikalisierung.....	22
3.2.2 Konstituentenreihenfolge im Mittelfeld.....	22
3.2.3 Extraposition.....	23
3.3 Ausblick	23
4 Schlussbemerkung	24
5 Bibliografie.....	25
Anhang: Aufbau des Babel-Systems.....	26
Morphologische Aspekte	26
WWW-Interface	27

Einleitung

Babel ist ein Sprachanalysesystem für das Deutsche, welches Eingabesätzen syntaktische und semantische Strukturen zuordnet. Programmiert wurde das System von Stefan Müller, der als wissenschaftlicher Mitarbeiter am Deutschen Forschungszentrum für Künstliche Intelligenz (<http://www.dfki.de>) in Saarbrücken arbeitet und im Rahmen des Projekts *Verbmobil* (<http://verbmobil.dfki.de>) eine HPSG-Grammatik für das Deutsche entwickelt. Das Babel-System dient zur praktischen Überprüfung dieser Theorie. Es ist in Prolog geschrieben, läuft auf unterschiedlichen Prolog-Versionen (SWI-Prolog, SICStus Prolog, HU-Prolog, Quintus Prolog) und wird seit Beginn der Implementation im August 1993 kontinuierlich weiterentwickelt und verbessert. Gegenwärtig umfasst der Quellcode des System rund 15'000 Zeilen (Stand Oktober 1999). Babel verfügt über ein Web-Interface und kann online getestet werden (<http://www.dfki.de/~stefan/Babel/>). Zusätzlich verfügt das System über ein "öffentliches" Lexikon. Dieses kann online erweitert und mit Einträgen für Verben, Nomina und Adjektive ergänzt werden (<http://www.dfki.de/~stefan/Babel/Interaktiv/NetGram/>). Ein paar weitere Informationen zum Babel-System finden sich im Anhang.

In dieser Arbeit werden wir uns primär mit der linguistischen Theorie beschäftigen, auf der das System basiert, einer Version der Head-Driven Phrase Structure Grammar (HPSG) für das Deutsche. Müller dokumentiert diese Grammatik in einem rund 500-seitigen Buch (Müller 1999). Darin werden vor allem syntaktische Besonderheiten des Deutschen, insbesondere Phänomene der Wortstellung, näher behandelt: für das Parsen deutscher Sätze ist vor allem die verhältnismässig freie Anordnung der Konstituenten problematisch, dies im Gegensatz zum Englischen, wo die Anordnung der einzelnen Konstituenten stärker normiert ist.

Im ersten Kapitel der Arbeit werden wir uns – im Rahmen einer kleinen Repetition - mit den Grundbausteinen der HPSG-Theorie auseinandersetzen. Im zweiten Kapitel soll gezeigt werden, welches Dominanzschema Müller zur Beschreibung von Kopf-Komplement-Strukturen im Deutschen vorschlägt, wie sich dieses von seinen Vorgängermodellen unterscheidet und welche Probleme für das Deutsche damit verbunden sind. Im dritten Kapitel werden wir uns kurz mit Phänomenen der Konstituentenreihenfolge im Deutschen beschäftigen.

1 Grundbegriffe der HPSG

Die wichtigsten Merkmale der HPSG lassen sich in einem kurzen einleitenden Überblick folgendermassen zusammenfassen (vgl. Bussmann 1990, S. 303f. / Liu 1999, S. 18f.):

HPSG ist ein Grammatikformalismus aus der Familie der Unifikationsgrammatiken. Sie übernimmt Elemente aus der GPSG (Generalized Phrase Structure Grammar), der FUG (Functional Unification Grammar), der LFG (Lexical Functional Grammar) und der Categorical Grammar. HPSG ist ein Grammatik-Formalismus, der sich an der Sprachoberfläche orientiert, d.h. es gibt keine theoretische Unterscheidung zwischen Oberflächen- und Tiefenstruktur.

Die HPSG integriert in ihrer Theorie sowohl syntaktische als auch semantische Aspekte von Sprache. In HPSG werden linguistische Objekte – wie in allen Unifikationsgrammatiken – durch komplexe Merkmalstrukturen dargestellt, wobei diese Merkmalstrukturen alle wesentlichen Informationen für die Repräsentation phonologischer, syntaktischer und semantischer Informationen enthalten. Auch Grammatikregeln werden in HPSG in Form von Merkmalstrukturen notiert, wobei diese als Beschränkungen (*Constraints*) für die Wohlgeformtheit eines sprachlichen Ausdrucks aufzufassen sind. Es wird angestrebt, eine möglichst kleine Anzahl solcher Grammatikschemata zu verwenden. Insgesamt ist die Grammatik der HPSG stark lexikalisiert, d.h. der grösste Teil der syntaktischen und semantischen Informationen ist bereits in den Merkmalstrukturen der Lexikoneinträge enthalten.

Die beiden grundlegenden Konzepte der HPSG (Merkmalstrukturen und Unifikation) lassen sich in Form einer Kurzdefinition folgendermassen umschreiben:

- **Merkmalstrukturen** sind eine Menge von Attribut-Wert-Paaren, wobei der Wert eines Attributs ein atomares Symbol oder wiederum eine komplexe Struktur sein kann. Wenn zwei Attribute einer Merkmalstruktur denselben Wert haben, dann werden sie als koreferent bezeichnet. Man spricht dann auch von Strukturteilung (*structure sharing*).
- Die **Unifikation** von Merkmalstrukturen geschieht nach folgendem Prinzip: wenn sich die Inhalte von zwei Merkmalstrukturen nicht widersprechen (d.h. wenn einem bestimmten Merkmal unverträgliche Werte zugewiesen werden), dann ist das Ergebnis einer Unifikation die Vereinigungsmenge der Informationen aus beiden unifizierten Merkmalstrukturen. Wenn es einen inhaltlichen Widerspruch gibt, dann schlägt die Unifikation fehl.

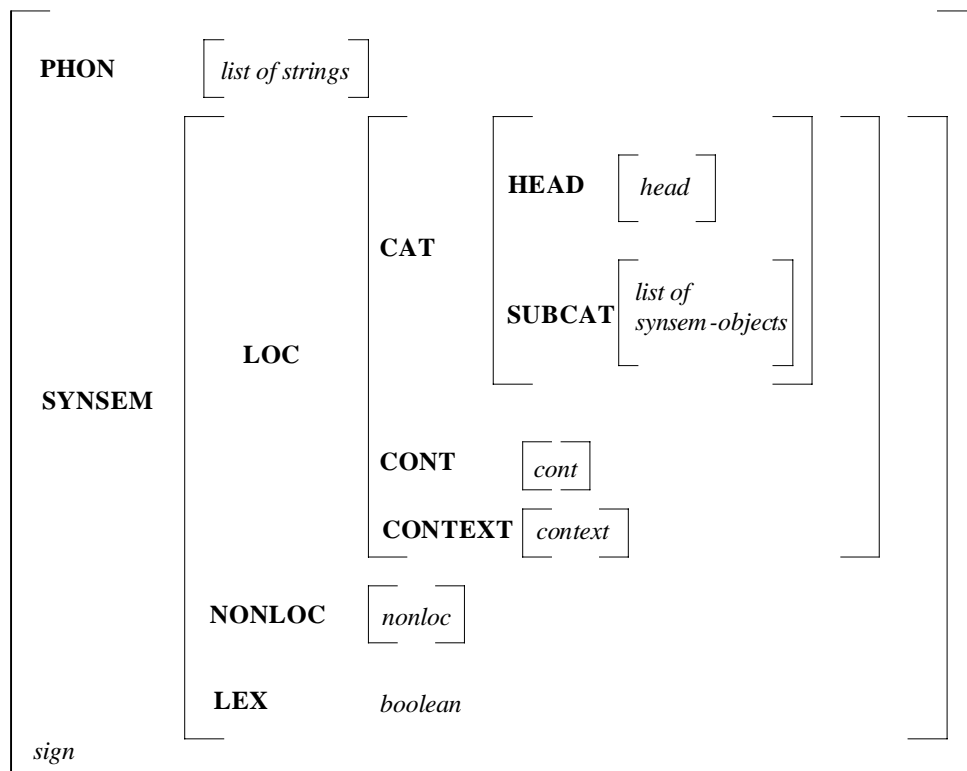
Als wichtigste Bestandteile einer HPSG-Theorie können zusammenfassend nochmals folgende drei Komponenten unterschieden werden (vgl. Schneider 1999):

- **Lexikon-Einträge:** Diese werden in Form von Merkmalstrukturen notiert und enthalten einen grossen Teil der syntaktischen und semantischen Information.
- **Grammatik-Prinzipien:** Hierbei handelt es sich um universalgrammatische, d.h. für alle Sprachen geltende Prinzipien, die die Wohlgeformtheit eines sprachlichen Ausdrucks festlegen. Gilt eines dieser Prinzipien nur einzelsprachlich, muss dies vom Verfasser der Grammatik entsprechend gekennzeichnet werden.
- **Grammatik-Regeln:** Diese beschreiben einzelsprachliche grammatische Phänomene und werden ebenfalls in Form von Merkmalstrukturen notiert. Ein mögliches Ziel der HPSG ist es, diese einzelsprachlichen Regeln zu Gunsten allgemeinerer universaler Prinzipien zu eliminieren.

Ein sprachlicher Ausdruck gilt dann als wohlgeformt, wenn er allen Grammatik-Prinzipien entspricht und durch mindestens einen Lexikon-Eintrag oder eine Grammatik-Regel abgeleitet werden kann.

1.1 Aufbau von Merkmalstrukturen

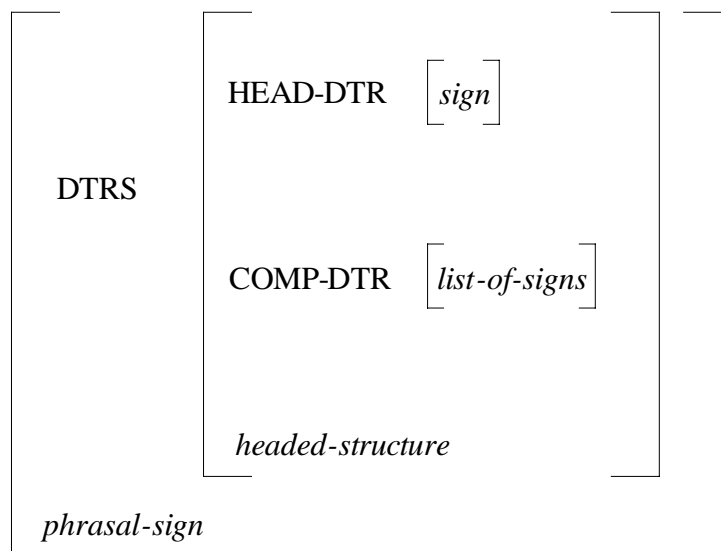
In HPSG werden alle linguistischen Objekte – Wort, Phrase oder Satz – durch Merkmalsstrukturen repräsentiert. Gemäss Müller lässt sich die Merkmalsstruktur einer linguistischen Einheit, die in Anlehnung an Ferdinand de Saussure *Zeichen (sign)* genannt wird, durch folgende Struktur darstellen (vgl. Müller 1999, S. 2):



- Der Wert von PHONOLOGY repräsentiert die phonologische Struktur eines Zeichens.
- Unter dem Attribut SYNSEM werden syntaktische und semantische Merkmale eines Zeichens zusammengefasst. Dabei gilt es zu unterscheiden zwischen lokalen Merkmalen (LOC) und nicht-lokalen Merkmalen (NONLOC).
- Unter dem Pfad SYNSEM|LOC werden Informationen zur syntaktischen Kategorie eines Zeichens (CAT), aber auch semantische Merkmale repräsentiert (CONT[ENT], CONTEXT): im Falle von CONTENT sind dies kontext-unabhängige, im Falle von CONTEXT kontext-abhängige semantische Informationen. Auf die Repräsentation semantischer Aspekte werden wir hier nicht eingehen.
- Unter dem Pfad SYNSEM|LOC|CAT wird zusätzlich zwischen den Merkmalen HEAD und SUBCAT unterschieden.
- Der Wert von HEAD ist eine Merkmalsstruktur, in der jene syntaktischen Informationen enthalten sind, die von einem Zeichen an seine Projektionen weitervererbt werden (also an jene phrasalen Einheiten, deren Kopf ein lexikalisches Zeichen ist).

- Unter dem Merkmal SUBCAT werden die Valenzpartner eines Zeichens festgelegt: Der Wert von SUBCAT besteht aus einer Liste von Elementen, mit denen ein Zeichen kombiniert werden muss, damit eine gesättigte, d.h. vollständige Phrase entsteht.
- Der Wert von LEX ist ein binäres Merkmal, das festlegt, ob ein Zeichen eine lexikalische Einheit ist (+) oder durch Kombination aus mehreren solchen entstanden ist (-).

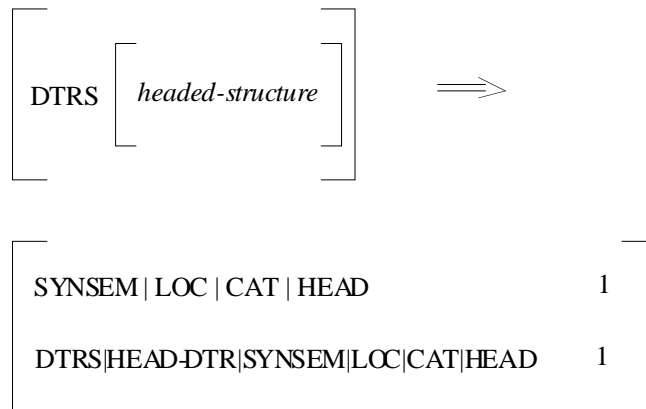
Der Begriff *Zeichen* bzw. *sign* zur Beschreibung linguistischer Objekte gilt in HPSG als Oberbegriff und umfasst neben den lexikalischen Zeichen (*lexical sign*) auch zusammengesetzte phrasale Zeichen (*phrasal sign*). Bei phrasalen Zeichen gibt es zusätzlich zu den oben genannten ein DAUGHTERS-Merkmal, welches Informationen über die Töchter enthält, aus denen sich ein phrasales Zeichen zusammensetzt. Dabei wird bei Kopf-Komplement-Strukturen z.B. zwischen Kopf-Tochter (HEAD-DTR) und Komplement-Töchtern (COMP-DTRS) unterschieden.



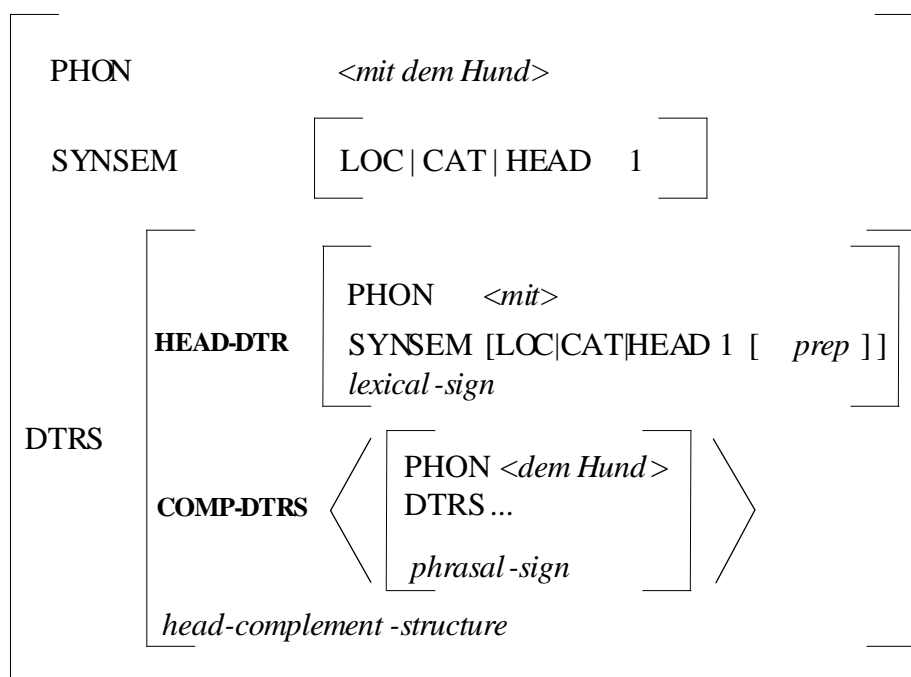
1.2 Kopfmerkmale und Kopfmerkmalprinzip

Jede Phrase enthält ein bestimmtes Wort, welches viele der syntaktischen Eigenschaften der gesamten Phrase bestimmt und deshalb als *lexical head* (lexikalischer Kopf) der Phrase bezeichnet wird. In HPSG gilt für Köpfe folgendes universalgrammatisches Prinzip (vgl. Müller 1999, S. 4):

Head Feature Principle (Kopfmerkmalprinzip)



In Worten ausgedrückt bedeutet dies: wenn ein Zeichen eine Kopftochter (HEAD-DTR) hat, dann teilt dieses seine Kopfmerkmale mit jenen der Kopftochter. Diese Koreferenz oder Strukturteilung (*structure-sharing*) wird in der Darstellung oben durch die 1 signalisiert. Bei der folgenden Präpositionalphrase (*mit dem Hund*) teilt das Zeichen z.B. seine Kopfmerkmale mit jenen des lexikalischen Kopfes der Phrase (*mit*):



Die Merkmale, die von einem lexikalischen Kopf an seine Projektionen weitervererbt werden, sind abhängig von der grammatischen Kategorie eines lexikalischen Zeichens: je nach Typ (*noun*, *verb*, *adj*, *prep*, *det* oder *adv*) werden unterschiedliche Merkmale an die übergeordnete Phrase weitervererbt. Das folgende Beispiel zeigt z.B. die Kopfmerkmale des Verbs. Bei Verben unterscheiden sich die Kopfmerkmale je nachdem, ob es sich um ein finites Verb (*fin-verb*) oder ein infinites Verb (*non-fin-verb*) handelt (vgl. Müller 1999, S. 5f.):

VFORM	<i>fin</i>	VFORM	<i>non-fin</i>
INITIAL	<i>boolean</i>	INITIAL	<i>boolean</i>
AUX	<i>boolean</i>	AUX	<i>boolean</i>
SUBJ	< >	SUBJ	[<i>list of synsem-objects</i>]
		AUXF	<i>auxf</i>
		FLIP	<i>boolean</i>
		ERG	[<i>list of synsem-objects</i>]
<i>fin-verb</i>			
		<i>non-fin-verb</i>	

Beide Verbarten (finites und infinites Verb) erben bestimmte Merkmale von einem Supertyp *verb*:

- VFORM legt fest, ob es sich um eine finite Verbform (*fin*), um eine Infinitivform (*bse*), um eine Infinitivform mit *zu* (*inf*) oder um eine Partizipform (*ppp*) handelt.
- INITIAL ist ein binäres Merkmal, das bei Verberststellung den Wert +, andernfalls den Wert – hat.
- AUX ist ebenfalls ein binäres Merkmal, das bei Hilfsverben den Wert +, andernfalls den Wert – hat.
- SUBJ ist ein Merkmal, das als Wert eine Liste hat. Bei finiten Verben ist diese Liste leer (vgl. Kap. 1.3 für weitere Informationen zum SUBJ-Merkmal).

Infinite Verben haben zu den oben genannten noch zusätzliche Merkmale: das Merkmal AUXF besitzt z.B. bei Partizipien (*ppp*) den Wert *haben* oder *sein*, bei anderen infiniten Verbformen (*bse* oder *inf*) den Wert *none*. Die beiden letzten Merkmale, auf die wir nicht näher eingehen, dienen zur Beschreibung von Wortstellungsphänomenen (FLIP) oder von Passivkonstruktionen (ERG).

1.3 Subkategorisierungsmerkmal (SUBCAT) und Subkategorisierungsprinzip

Die möglichen Valenzpartner eines lexikalischen oder phrasalen Zeichens werden in HPSG unter dem Merkmal SUBCAT in einer Liste festgelegt. In dieser Liste sind die syntaktischen und semantischen Merkmale anderer Zeichen definiert, mit denen ein Zeichen kombiniert werden muss, um eine vollständige Phrase zu bilden. Bei Verben würde eine derartige Liste z.B. bei intransitiven Verben nur ein Element, bei transitiven Verben zwei Elemente und bei ditransitiven Verben drei Elemente enthalten. Die Reihenfolge der einzelnen Elemente in der SUBCAT-Liste sagt nichts über die effektive Reihenfolge der einzelnen Komplemente aus, wie sie in Sätzen real vorkommen. Ein phrasales Zeichen gilt dann als vollständig oder gesättigt, wenn die SUBCAT-Liste keine Elemente enthält.

Das Subkategorisierungsmerkmal ermöglicht es uns, die aus Phrasenstruktur-Regeln bekannten Grammatiksymbole als komplexe Merkmalstrukturen zu beschreiben. In HPSG werden diese traditionellen Symbole häufig als Abkürzungen innerhalb von Merkmalstrukturen verwendet; sie stehen dann stellvertretend für die komplexere Darstellungsweise. Durch entsprechende Instanzierung der LEX- und SUBCAT-Merkmale können zum Beispiel die unterschiedlichen Komplexitätsstufen von Kategorien aus der X-Bar-Theorie dargestellt werden. Die verschiedenen Bar-Level von N lassen sich beispielsweise durch folgende Merkmalstrukturen beschreiben (vgl. dazu auch Müller 1999, S.8):

Symbol	Merkmalstruktur (Teilspezifikation)
N	$\left[\begin{array}{l} \text{LOC CAT} \left[\begin{array}{l} \text{HEAD} \quad [\text{noun}] \\ \text{SUBCAT} \quad \langle \text{DET} \rangle \end{array} \right] \\ \text{LEX} \quad + \end{array} \right]$
\bar{N}	$\left[\begin{array}{l} \text{LOC CAT} \left[\begin{array}{l} \text{HEAD} \quad [\text{noun}] \\ \text{SUBCAT} \quad \langle \text{DET} \rangle \end{array} \right] \\ \text{LEX} \quad - \end{array} \right]$
$\bar{\bar{N}}$	$\left[\begin{array}{l} \text{LOC CAT} \left[\begin{array}{l} \text{HEAD} \quad [\text{noun}] \\ \text{SUBCAT} \quad \langle \rangle \end{array} \right] \\ \text{LEX} \quad - \end{array} \right]$

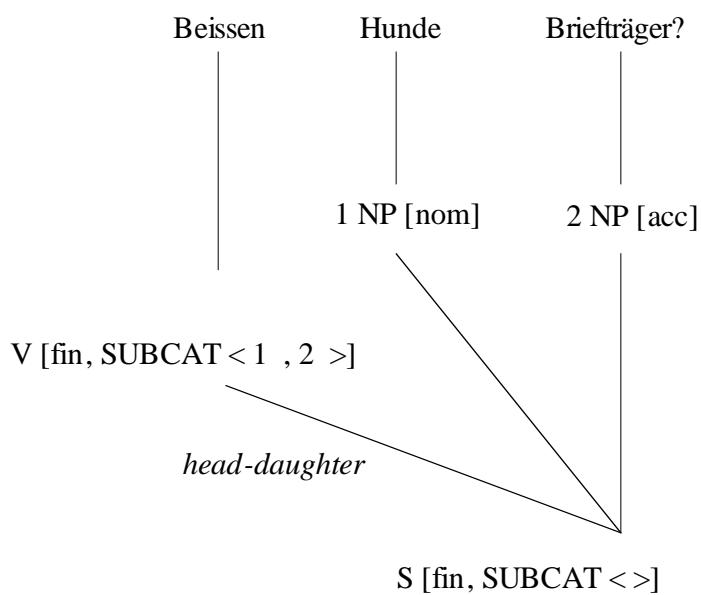
Beim ersten Symbol N (Bar-Level 0) handelt es sich um ein lexikalisches Zeichen, weshalb es im Gegensatz zu den übergeordneten Stufen (Bar-Level 1 und 2) als einziges das Merkmal LEX + besitzt. Bei der bereits vollständigen Kategorie N'' (Bar-Level 2) steht unter SUBCAT die leere Liste – dies im Gegensatz zu den noch unvollständigen Stufen N und N'.

Ein wichtiges universalgrammatisches Prinzip der HPSG ist das Subkategorisierungsprinzip (*Subcat Principle*). Müller formuliert dieses Prinzip folgendermassen (vgl. Müller 1999, S. 10):

Subkategorisierungsprinzip (Subcat Principle)

In einem phrasalen Zeichen, dessen DTRS-Wert vom Typ headed-structure ist, ist der SUBCAT-Wert der Kopftochter die Verknüpfung der Subcat-Liste des phrasalen Zeichens mit der Liste der SYNSEM-Werte der Elemente in der COMP-DTRS-Liste.

Das Subkategorisierungsprinzip lässt sich wiederum mit einem konkreten Beispiel verdeutlichen:



NP[nom] und NP[acc] sind Komplement-Töchter von S und stehen hier stellvertretend für komplexere SYNSEM-Ojekte. Das Subkategorisierungsprinzip besagt, dass sich die SUBCAT-Liste der Kopftochter aus der Verknüpfung der folgenden Elemente zusammensetzt:

- den Merkmalen der Komplement-Töchter NP [nom] und NP [akk] und
- den SUBCAT-Elemente der übergeordneten Phrase (also hier S mit der leeren Liste).

Einen besonderen Status bei der Subkategorisierung hat das Subjekt: Müller zählt zu den Subjekten einerseits NP's im Nominativ (sofern sie nicht prädikativ gebraucht werden), andererseits aber auch Subjektsätze (z.B. *Dass der Hund ständig bellt, geht ihm auf die Nerven*). In Sätzen mit finitem Verb ist das Subjekt Bestandteil der SUBCAT-Liste des Verbs; Sätze sind demnach als vollständige Projektionen des Verbs aufzufassen. Bei Konstruktionen mit infinitem Verb (z.B. *Ich versuche, den Hund zu streicheln*) ist dies nicht der Fall: hier gibt es zusätzlich ein Kopfmerkmal SUBJ, das als Wert eine Liste mit maximal einem Element hat (vgl. Müller 1999, S. 11f.).

2 Dominanzstrukturen

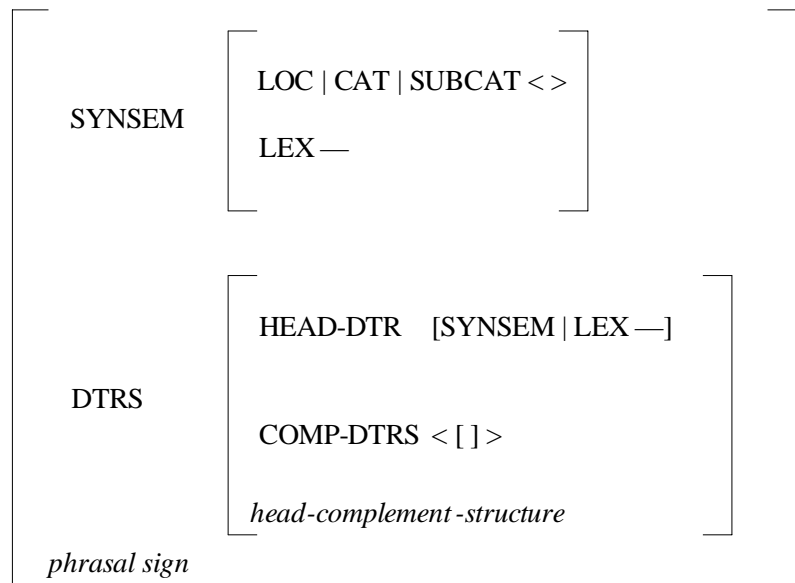
Jener Teil von Grammatiken, der unabhängig von Einzelsprachen gültig ist, wird bekanntermassen als Universalgrammatik bezeichnet. Zwei wichtige universalgrammatische Prinzipien (Kopfmerkmalprinzip und Subkategorisierungsprinzip) wurden bereits vorgestellt. Ein weiteres Prinzip, welches in HPSG verwendet wird, ist das Immediate Dominance Principle (IDP). Dieses enthält wiederum mehrere Dominanzschemata: Dominanzschemata sind eine abstrakte Beschreibung von möglichen (lokalen) Phrasenstrukturbäumen. Von diesen Schemata muss mindestens eines auf eine bestimmte natürlichsprachige Konstruktion zutreffen, damit sie als wohlgeformt gilt. Die ID-Schemata entsprechen den Grammatik-Regeln, Phrasenstruktur-Regeln oder X-Bar-Schemata aus anderen Grammatiktheorien. Das Immediate Dominance Principle beinhaltet grundsätzlich alle möglichen Schemata, die in natürlichen Sprachen vorkommen können. Für eine Einzelsprache wie Deutsch sind dementsprechend nur einige dieser Schemata anwendbar bzw. notwendig.

Bei Pollard und Sag (1994) werden insgesamt drei verschiedene ID-Schemata zur Beschreibung von Kopf-Komplement-Strukturen vorgeschlagen. Diese werden in Kapitel 2.1 nochmals ausführlich vorgestellt. Anschliessend soll gezeigt werden, warum Müller die Verwendung dieser Schemata für das Deutsche ablehnt, welche Lösung er vorschlägt und welche weitergehenden Probleme (insbesondere für Phänomene der Wortstellung im Deutschen) sich dadurch ergeben.

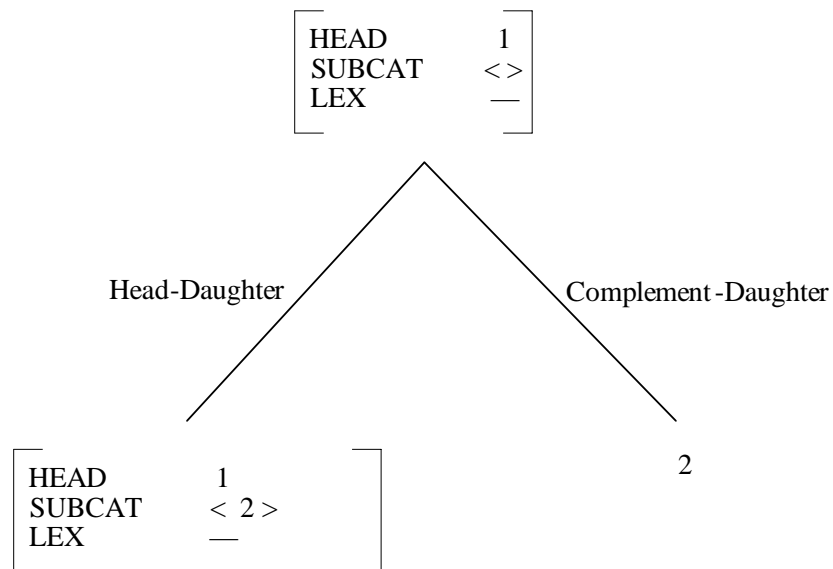
2.1 Flache Dominanzstrukturen

Die drei nachfolgenden Kopf-Komplement-Schemata stehen bei Pollard und Sag (1994) neben einer Reihe von weiteren Regeln (z.B. zur Erklärung von Kopf-Adjunkt-Strukturen). Gemäss dem ID-Prinzip gilt eine Phrase dann als wohlgeformt, wenn sie eine Instanz von mindestens einem dieser Schemata darstellt (vgl. Pollard und Sag 1994, S. 37ff. und Müller 1999, S. 15ff.):

Kopf-Komplement-Schema 1: Eine mögliche wohlgeformte Phrase ist „eine bereits gesättigte Phrase, deren Kopftochter ein phrasales Zeichen ist, und die genau eine Komplementtochter hat.“ (Müller 1999, S. 15)



Als Strukturbaum lässt sich das folgendermassen darstellen:

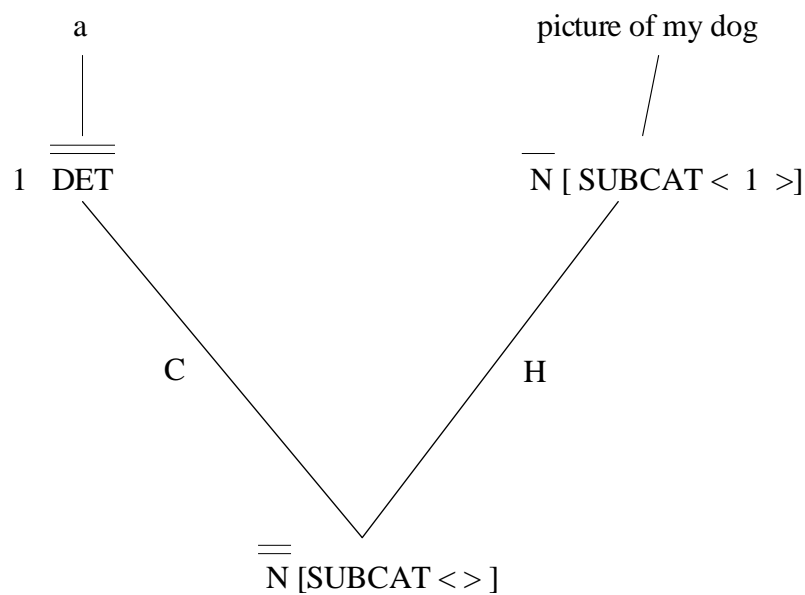
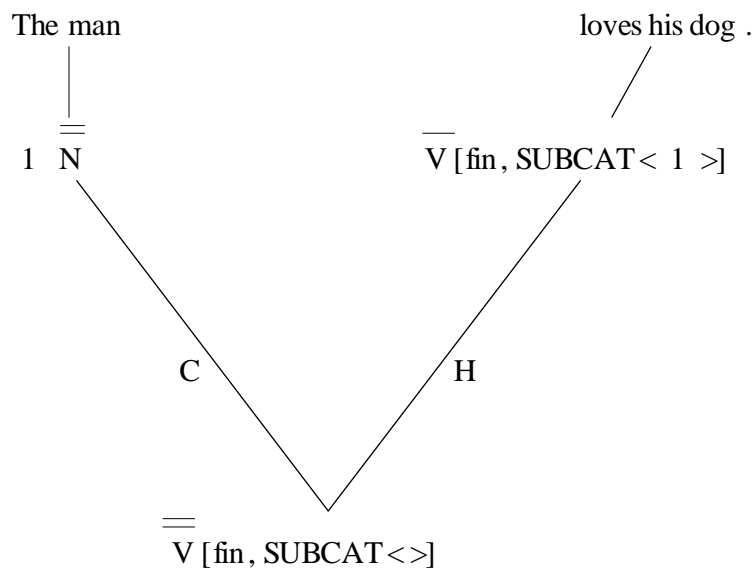


Schema 1 entspricht folgender X-Bar-Regel:

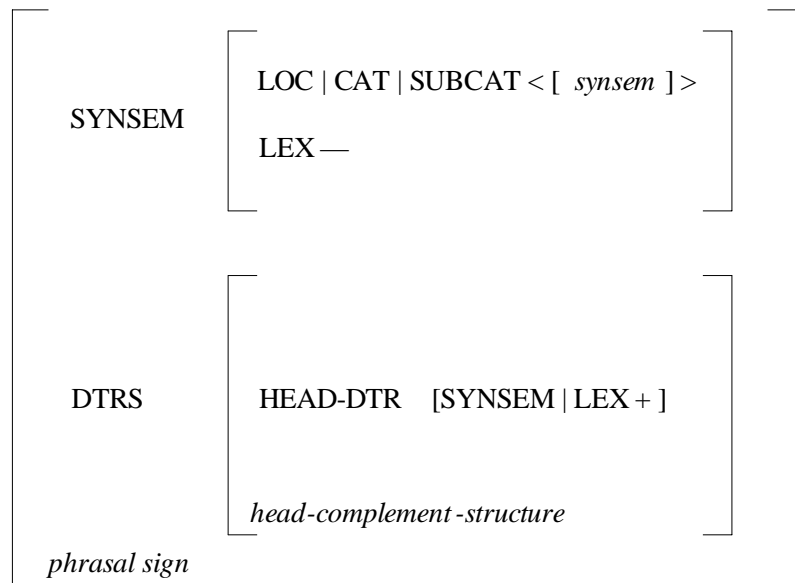
$X'' \rightarrow \text{specifier}'' X'$

Alle drei Darstellungen (Merkmalstruktur, Syntaxbaum und X-Bar-Regel) besagen dasselbe: eine vollständige Phrase (Bar-Level 2) besteht aus einem noch nicht ganz vollständigen, phrasalen Kopf (Bar-Level 1) und genau einem Komplement. Das hinzugefügte Komplement wird auch als Spezifikator (*specifier*) bezeichnet. Bei Konstruktionen im Englischen, auf die Schema 1 zutrifft, gilt eine Linearisierungsregel, die besagt, dass der Spezifikator auf der linken Seite des Kopfes steht, also z.B. das Subjekt vor dem finiten Verb.

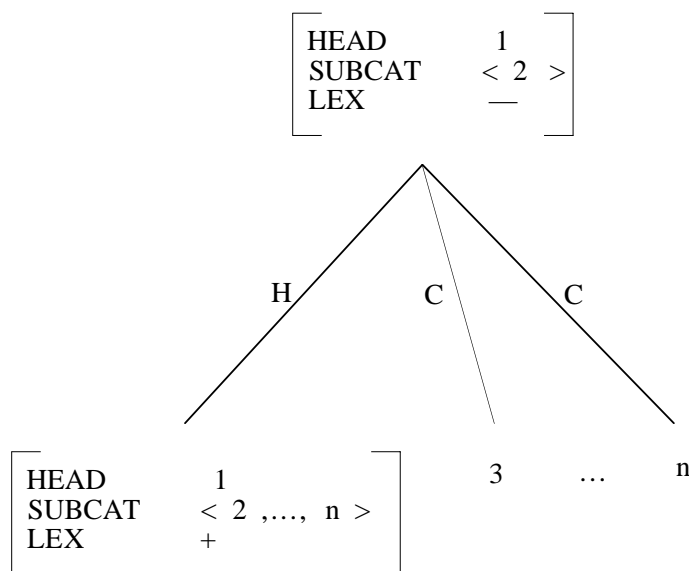
Mögliche Beispiele aus dem Englischen für Schema 1 sind folgende Phrasen:



Kopf-Komplement-Schema 2: Eine mögliche wohlgeformte Phrase ist „eine fast gesättigte Phrase, deren Subcat-Liste nur noch ein Element enthält, mit einem DTRS-Wert vom Typ *head-complement-structure*, in der die Kopftochter ein lexikalisches Zeichen ist.“ (Müller 1999, S. 15)



Im Gegensatz zu Regel 1 ergibt dies eine flache Syntaxstruktur, da ein lexikalischer Kopf mehr als zwei Elemente in seiner SUBCAT-Liste enthalten kann, z.B. bei ditransitiven Verben:

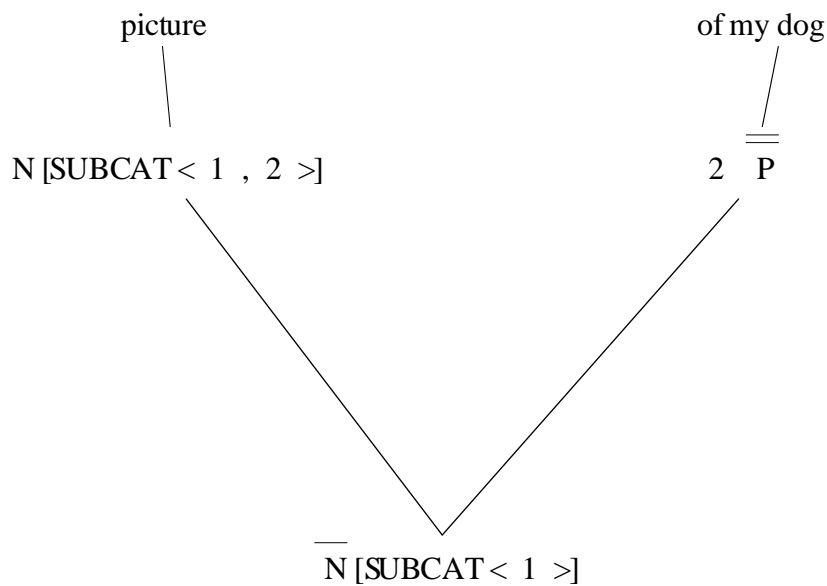
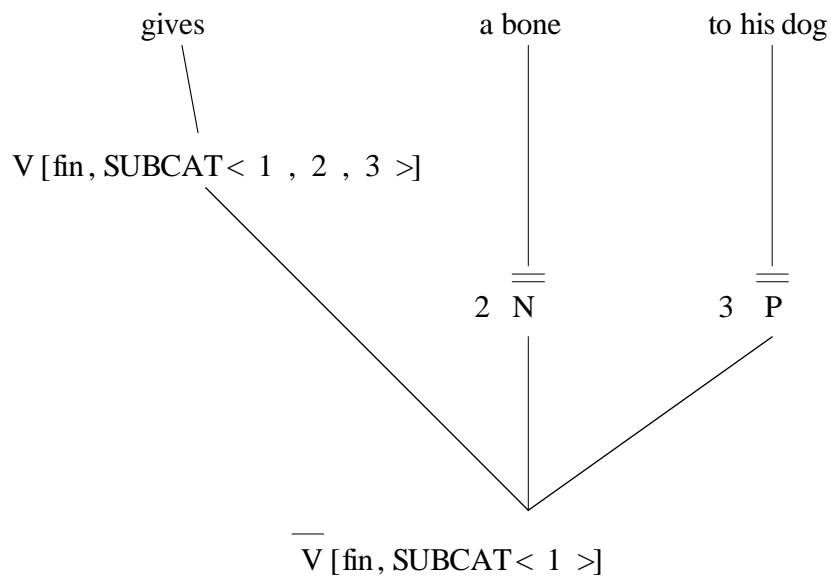


Schema 2 korrespondiert mit folgender X-Bar-Regel:

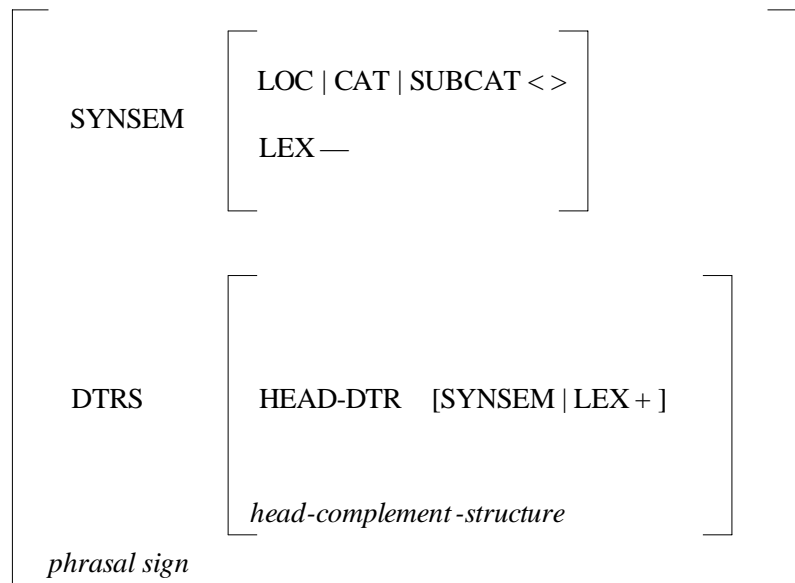
$$X' \rightarrow X (\text{complement}')^*$$

Auch hier drücken alle drei Darstellungsweisen das Gleiche aus: ein lexikalischer Kopf (Bar-Level 0) wird mit all seinen Komplementen (bis auf das erste Element der SUBCAT-Liste) zu einer noch nicht vollständigen Phrase (Bar-Level 1) ergänzt. Zur Vollständigkeit fehlt noch der Spezifikator. Bei Konstruktionen, die durch Schema 2 abgeleitet werden können, gilt im Englischen eine LP-Regel, die besagt, dass die Komplemente auf der rechten Seite des lexikalischen Kopfes stehen. Wie sich diese Komplemente dann jeweils untereinander anordnen lassen, ist wiederum eine andere Frage.

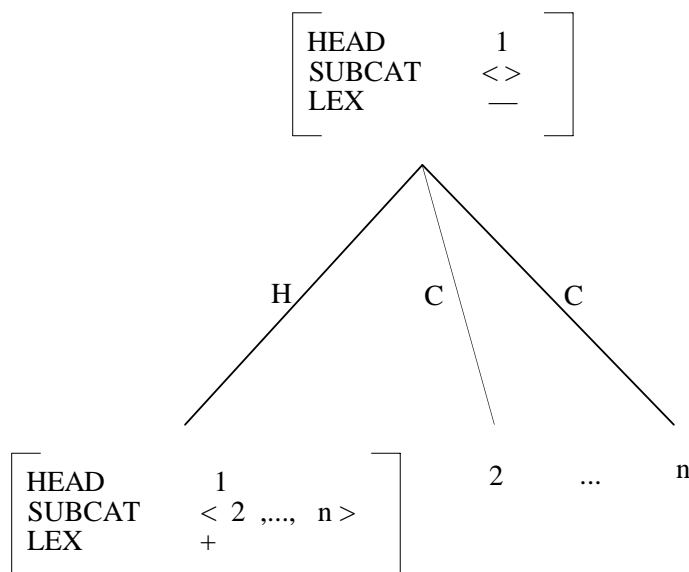
Mögliche Beispiele für Schema 2 aus dem Englischen sind folgende Phrasen:



Kopf-Komplement-Schema 3: Eine mögliche wohlgeformte Phrase ist „eine gesättigte Phrase mit einem DTRS-Wert vom Typ *head-complement-structure*, in der die Kopftochter ein lexikalisches Zeichen ist.“ (Müller 1999, S. 15)



Auch hier ergeben sich in der Baumdarstellung (wie bei Schema 2) flache Syntaxstrukturen mit mehreren Ästen:



Da hier die Zwischenstufe X' direkt "übersprungen" wird, könnte Schema 3 durch folgende zwei X-Bar-Regeln zusammengefasst werden:



Durch Schema 3 wird ein lexikalischer Kopf mit all seinen Komplementen zu einer gesättigten Phrase kombiniert. Damit werden im Englischen bestimmte Spezialfälle der Wortstellung beschrieben, beispielsweise invertierte Sätze mit Erststellung des finiten Verbs (z.B. *Can your dog bark?*). Dies ist insofern problematisch, da sich dadurch auch falsche Strukturen (z.B. **Gives the man a bone to the dog*) ableiten lassen, sofern dies nicht durch ein entsprechendes Kopfmerkmal ausgeschlossen wird, z.B. durch ein binäres Merkmal INV im Lexikon, das Verberststellung nur bei Hilfs- und Modalverben erlaubt.

2.2 Problemfälle flacher Strukturen

Alle drei Kopf-Komplement-Schemata, die oben besprochen wurden, lassen sich grundsätzlich auch zur Beschreibung deutscher Konstruktionen verwenden. Schema 1 kann zur Beschreibung von Nominalphrasen und Sätzen angewendet werden:

$N'' \rightarrow \text{DET}'' N'$	[<i>das [Halsband von Struppi]</i>]
$V'' \rightarrow N'' V'$	[<i>Der Hund [beisst den Mann]</i>].

Durch Schema 2 können N'- und V'-Konstruktionen abgeleitet werden:

$N' \rightarrow N P''$	[<i>Halsband [von Struppi]</i>]
$V' \rightarrow V N'' N''$	[<i>gibt [dem Hund einen Knochen]</i>]

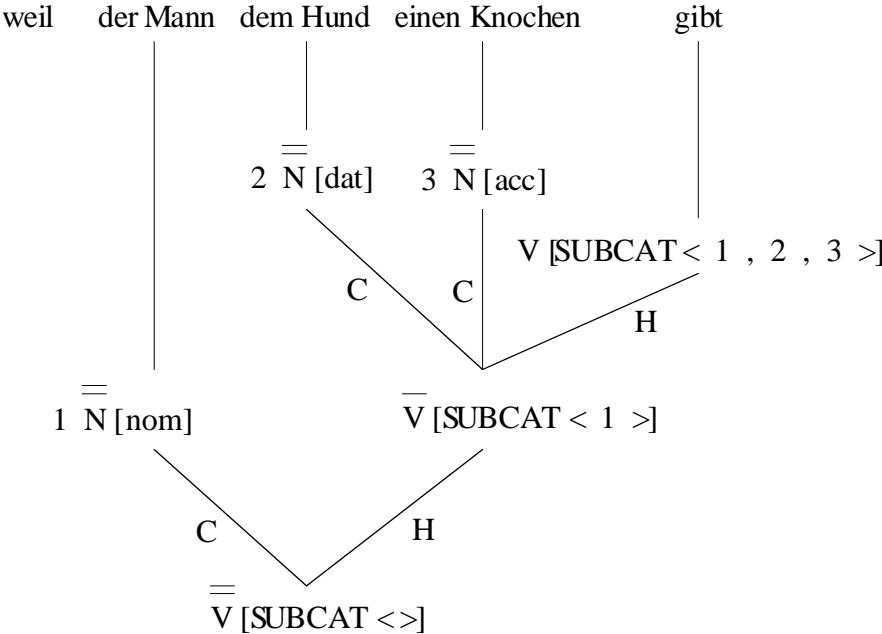
Durch Schema 3 lassen sich Sätze, Adjektivphrasen und Präpositionalphrasen beschreiben:

$S \rightarrow V (\text{Complement}')^*$	[<i>Gibt [der Mann dem Hund einen Knochen]</i>]?
$A' \rightarrow A \text{ Complement}''$	der [[<i>auf seinen Hund</i>] wütende] Mann
$P' \rightarrow P N''$	[<i>mit [meinem Hund]</i>]

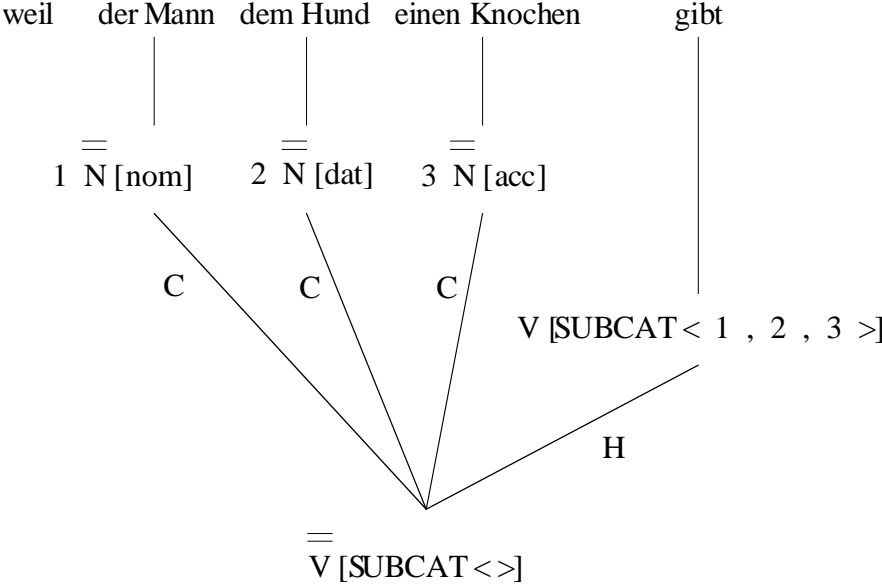
Müller verzichtet jedoch auf die Verwendung der Kopf-Komplement-Schemata, wie sie bei Pollard und Sag vorgeschlagen werden. Seine Kritik äussert sich vor allem in folgenden Einwänden:

Kopf-Komplement-Regeln sind Beschreibungen möglicher lokaler Phrasenstrukturen, sagen jedoch grundsätzlich nichts aus über die Oberflächenreihenfolge der einzelnen Komplemente. Bei den drei vorgestellten Regeln sind jedoch implizit auch Phänomene der Wortstellung berücksichtigt, die eigentlich nur für eine SVO-Sprache wie das Englische relevant sind: so gibt es z.B. unterschiedliche Regeln für Konstruktionen mit linksgestelltem Komplement (Schema 1), mit rechtsgestellten Komplementen (Schema 2) oder mit Erststellung des finiten Verbs (Schema 3). Nach diesem Muster könnten für jede Einzelsprache weitere Schemata formuliert werden. Damit würde aber gleichzeitig die universalgrammatische Berechtigung des IDP in Frage gestellt, wenn dieses bloss die (potentiell unendliche) Aufzählung beliebiger Regeln umfasst.

Mit den vorgestellten Schemata lassen sich darüber hinaus für identische Konstruktionen unterschiedliche Syntaxstrukturen ableiten. Dies macht wenig Sinn, solange dafür keine semantische Motivation vorliegt. Das Problem der Mehrdeutigkeit können wir beim folgenden Beispiel nachvollziehen, bei dem zwei Syntaxstrukturen möglich sind. Die erste Analyse ergibt sich aus einer Kombination der Regeln 1 und 2:



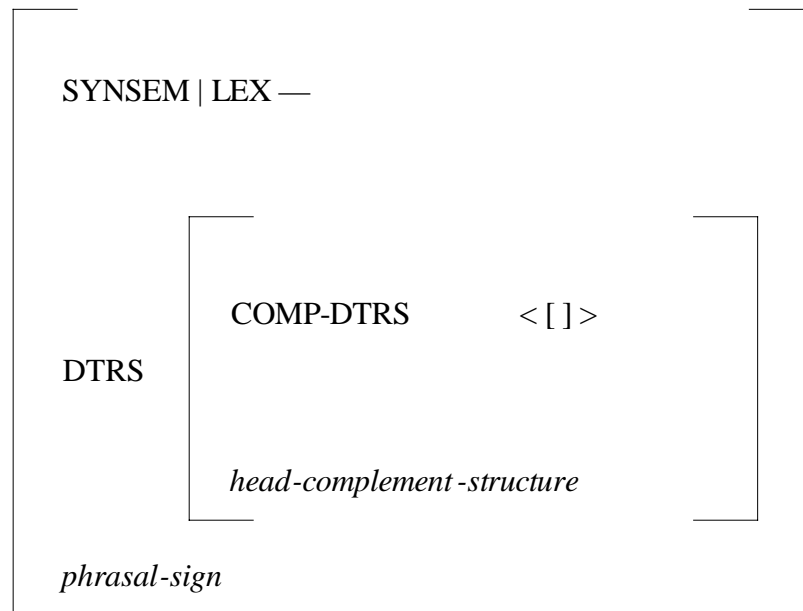
Die Phrase kann aber auch direkt durch Schema 3 abgeleitet werden, was folgende Struktur ergibt:



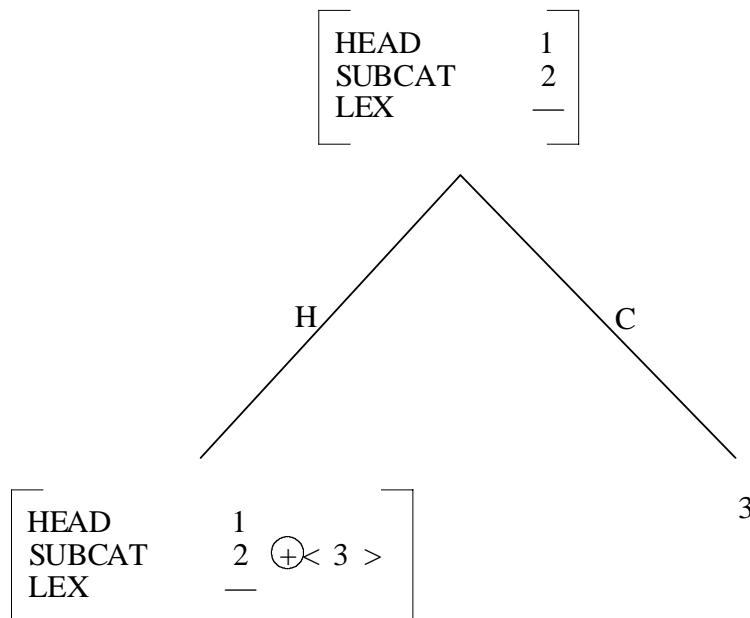
2.3 Binär verzweigende Dominanzstrukturen

Müllers Lösung für die oben beschriebenen Probleme sieht so aus, dass er für die Beschreibung von Kopf-Komplement-Strukturen nur ein einziges Schema formuliert (vgl. Müller 1999, S. 21):

Generalisiertes Kopf-Komplement-Schema (vereinfachte Darstellung):



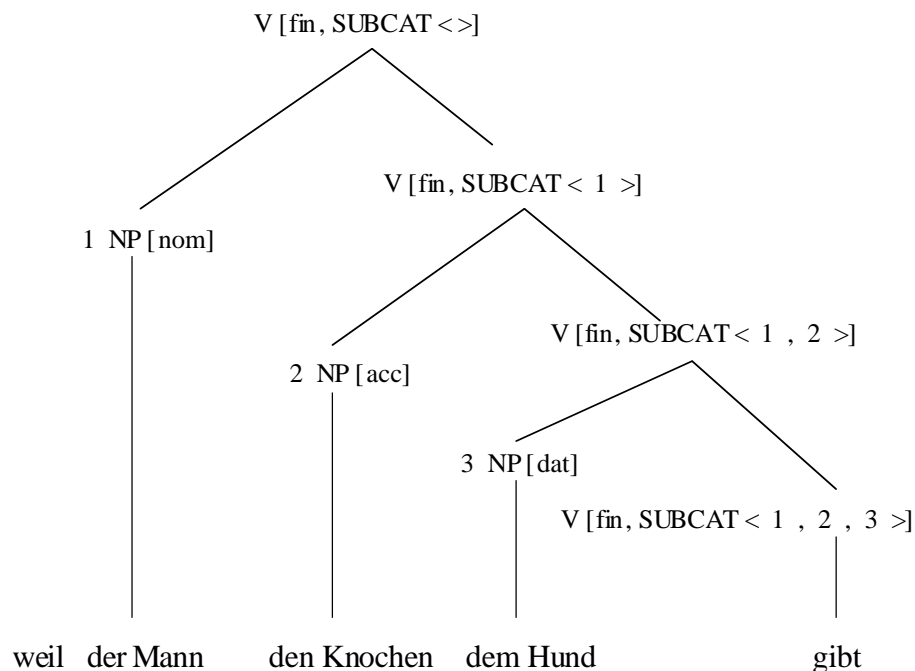
Dies ergibt folgende Syntaxstruktur:



Müllers Schema besagt, dass es in Kopf-Komplement-Strukturen neben der Kopftochter immer genau eine Komplementtochter gibt. Dies ergibt binär verzweigende Strukturen, bei denen immer zwei Elemente, ein Kopf und ein Komplement, zu einer übergeordneten Einheit kombiniert werden.

Der \oplus -Operator im Strukturbaum ist ein Listenverkettungsoperator, der besagt, dass sich die SUBCAT-Liste der Kopftochter gemäss dem Subkategorisierungsprinzip aus den SUBCAT-Elementen der Mutter ($\boxed{2}$) und den SYNSEM-Werten der Komplementtochter ($\boxed{3}$) zusammensetzt.

Beim nächsten Beispiel ergibt die Anwendung dieser Regel folgende Konstituenten-Struktur:



Müllers Dominanzschema hat gegenüber den Vorgänger-Modellen einleuchtende Vorteile: es wird eine stärkere Verallgemeinerung vorgenommen und somit eher dem universalgrammatischen Postulat nach wenigen und eindeutigen Beschreibungsmitteln entsprochen. Ausserdem werden bei der syntaktischen Ableitung keine mehrdeutigen Strukturen erzeugt.

Trotzdem ergeben sich bei der Anwendung aufs Deutsche Probleme: Deutsch ist eine Sprache mit relativ freier Anordnung der Konstituenten, was sich rein intuitiv durch die Verwendung flacher Dominanzschemata besser erklären liesse. Bei binären Phrasenstrukturen ist die Beschreibung der Oberflächenreihenfolge jedoch nicht mehr ohne weiteres möglich. Konstituenten sind sie nicht mehr beliebig permutierbar, da sie nicht vom unmittelbar gleichen Mutterknoten dominiert werden.

Auf einige Phänomene deutscher Wortstellung werden wir im nächsten Kapitel kurz eingehen.

3 Konstituentenreihenfolge im Deutschen

Die Dominanzschemata, die wir im vorhergehenden Kapitel besprochen haben, sind Beschreibungen möglicher Phrasenstrukturen, sagen aber prinzipiell nichts aus über die Oberflächenreihenfolge der Konstituenten. In welcher Reihenfolge Konstituenten auftreten dürfen, wird durch die *Linear-Precedence-Rules* (LP-Regeln) gesteuert. In der Standard-HPSG-Theorie werden diese LP-Regeln auf Konstituenten innerhalb von flachen Dominanzstrukturen angewendet. Da LP-Regeln jedoch nur für unmittelbare Töchter eines Zeichens formuliert werden können, führt dies bei binär verzweigenden Dominanzstrukturen, wie sie Müller propagiert, zu Schwierigkeiten: bei binär verzweigenden Strukturen gibt es neben dem Kopf nur genau eine Komplementtochter.

Wie bereits mehrmals erwähnt, ist Deutsch eine Sprache mit relativ freier Konstituentenanordnung, Komplemente sind - mit gewissen Einschränkungen - beliebig permutierbar. Dieses Phänomen wird auch als *Scrambling* bezeichnet. In diesem Kapitel soll einleitend gezeigt werden, in welche grundlegenden Bestandteile oder Felder deutsche Sätze eingeteilt werden können. Anschliessend werden wir auf der Grundlage dieser Typologie einige *Scrambling*-Phänomene des Deutschen mit Beispielen dokumentieren. Auf eine Darstellung der konkreten "Implementierung" dieser Phänomene in Müllers HPSG-Theorie werden wir dabei verzichten, da dies den Rahmen dieser Arbeit sprengen würde.

3.1 Modell des deutschen Satzes

Deutsche Sätze können, je nach Stellung des finiten Verbs, in drei Klassen eingeteilt werden:

- Sätze mit Verbendstellung: *Ich glaube, dass der Hund den Mann gebissen **hat**.*
- Sätze mit Verberststellung: ***Hat** der Hund den Mann gebissen?*
- Sätze mit Verbzweitstellung: *Der Hund **hat** den Mann gebissen.*

Wie aus den oben gezeigten Beispielen deutlich wird, bildet das finite Verb mit seinen Zusätzen nur beim ersten Beispiel eine Einheit. Beim zweiten und dritten Beispiel bilden das finite Verb und seine Verbkomplemente keine Einheit, d.h. sie sind diskontinuierlich. Der Satz wird durch die einzelnen Verbeile eingerahmt, was auch als *Satzklammer* bezeichnet wird. Durch diese Typologie können deutsche Sätze in insgesamt fünf verschiedene Felder eingeteilt werden (vgl. Müller 1999, S. 83):

- Vorfeld
- Linke Satzklammer (finites Verb oder *complementizer*)
- Mittelfeld

- Rechte Satzklammer (wird auch als Verbalkomplex oder *verb cluster* bezeichnet)
- Nachfeld

In der folgenden Tabelle wird ersichtlich, wie sich diese Typologie auf einige konkrete Beispiele anwenden lässt:

Vorfeld	Linke Satzklammer	Mittelfeld	Rechte Satzklammer	Nachfeld
Der Hund	schläft.			
Der Hund	hat		geschlafen.	
Der Hund	beisst	den Mann.		
Der Hund	greift	den Mann	an.	
Der Hund	hat	den Mann	gebissen.	
Der Hund	hat	den Mann als er vorbeiging	gebissen.	
Der Hund	hat	den Mann heute morgen	gebissen.	
Der Hund	hat	den Mann zu beissen	versucht.	
Der Hund	hat		versucht	den Mann zu beissen.
	Schlafen	Hunde?		
	Schlaf!			
	Greif	den Mann	an!	
	Hat	er doch heute diesen Mann	gebissen.	
	weil	der Hund den Mann	gebissen hat	ohne auf mich zu hören.
	weil	der Hund immer wieder	zubeissen wollen wird	ohne dabei gestört zu werden.

3.2 Wortstellungsphänomene im Deutschen

Die unmarkierte Satzstellung bei Müller sind Sätze mit Verberststellung, bei denen das finite Verb die linke Satzklammer bildet und alle Konstituenten im Mittelfeld gruppiert sind. Dies trifft auf Konstruktionen wie Alternativfragen oder Imperativesätze zu:

- (1) *Hat der Mann dem Hund den Knochen gegeben?*
- (2) *Gib mir den verdammten Knochen!*

Von dieser Grundkonstellation aus sind verschiedene Variationen denkbar: einerseits lassen sich die Konstituenten im Mittelfeld untereinander permutieren, andererseits können Konstituenten verschoben werden, und zwar entweder in das Vorfeld oder in das Nachfeld des Satzes. Einige dieser Möglichkeiten werden anschliessend mit Beispielen erklärt.

3.2.1 Topikalisierung

Sätze mit Verbzweitstellung werden als abgeleitete Konstruktionen von Sätzen mit Verberststellung betrachtet, indem ein Komplement ins Vorfeld verschoben wird. Dabei können praktisch alle Komplemente des Verbs im Vorfeld auftauchen:

- (3) *Der Mann hat dem Hund den Knochen gegeben.*
Dem Hund hat der Mann den Knochen gegeben.
Den Knochen hat der Mann dem Hund gegeben.

Es können auch Konstituenten ins Vorfeld verschoben werden, die nicht unmittelbares Komplement des Verbs sind: so können z.B. Komplemente von Nominal- oder Adjektivphrasen an den Anfang gesetzt werden:

- (4) *[Von meinem Struppi]_i, habe ich gestern [ein Foto _i] gemacht.*
[Auf meinen Struppi]_i, bin ich wirklich sehr [_i stolz].

Topikalisierung ist auch für Adjunkte möglich:

- (5) *Gestern hat Struppi im Wald nach einem Knochen gegraben.*
Im Wald hat Struppi gestern nach einem Knochen gegraben.

3.2.2 Konstituentenreihenfolge im Mittelfeld

Die Anordnung der Konstituenten im Mittelfeld unterliegt im Deutschen nur wenigen, allenfalls stilistisch oder semantisch begründeten Restriktionen, so sind z.B. bei folgendem Beispiel (mit dem indirekten Verb *geben*) grundsätzlich alle Permutationen der Komplemente zulässig:

- (6) *Gab der Mann dem Hund den Knochen?*
Gab der Mann den Knochen dem Hund?
Gab dem Hund der Mann den Knochen?
Gab dem Hund den Knochen der Mann?
Gab den Knochen der Mann dem Hund?
Gab den Knochen dem Hund der Mann?

Komplemente von Nominal- oder Adjektivphrasen können auch im Mittelfeld getrennt von ihren Köpfen auftreten, analog zu den Beispielen bei (4), wo diese Elemente ins Vorfeld verschoben wurden:

- (7) *Deshalb habe ich [von meinem Struppi]_i gestern [ein Foto -_i] gemacht.
Deshalb bin ich [auf meinen Struppi]_i wirklich sehr [_i stolz].*

3.2.3 Extraposition

Verschiedene Satzglieder können im Deutschen aus dem Mittelfeld ins Nachfeld des Satzes verschoben werden. Dies gilt z.B. für Satz- und Infinitivkomplemente, sowie Präpositional-, Adverbial- oder Nominalphrasen. Sowohl Komplemente als auch Adjunkte können extrapониert werden.

3.3 Ausblick

Die Phänomene der Wortstellung im Deutschen, welchen Restriktionen sie unterliegen und wie das in HPSG repräsentiert werden kann, all dies wird in Müller (1999) ausführlich analysiert: in Kapitel 9 werden Topikalisierungsphänomene behandelt und durch den NONLOC-Mechanismus erklärt. Kapitel 11 und 12 beschäftigen sich mit der Anordnung der Konstituenten im Mittelfeld des Satzes und zeigen Lösungen auf, wie sich trotz binärer Dominanzstrukturen die freie Wortstellung im Deutschen erklären lässt. Kapitel 13 widmet sich der Extraposition von Satzgliedern.

Da dies im Rahmen einer einführenden Seminararbeit nicht im Detail erklärt werden kann, haben wir uns in diesem Kapitel auf einige Stichworte beschränkt, die das grundsätzliche Problem skizzieren. In der endgültigen Version dieser Arbeit, die noch um einen zweiten Teil ergänzt werden sollte, werden wir uns damit ausführlicher auseinandersetzen können.

4 Schlussbemerkung

Müller präsentiert mit seiner HPSG-Theorie ein umfassendes Grammatik-Modell, das Grundlagen der HPSG mit Aspekten deutscher Syntax und Semantik verbindet. Mit dem Babel-System beweist er gleichzeitig die praktische Anwendbarkeit seiner Theorie.

Müllers Grammatik stellt den interessierten Laien jedoch vor fast unlösbare Probleme:

- Die Grundelemente der HPSG-Theorie von Pollard und Sag, auf die sich Müller im Wesentlichen stützt, werden zwar in den ersten sieben Kapiteln seines Buches erklärt. Für das Verständnis des Nachfolgenden sind gründliche HPSG-Vorkenntnisse jedoch beinahe unabdingbar.
- Die restlichen 400 Seiten seines Buches widmen sich der Erklärung deutscher Wortstellungsphänomene und deren Repräsentation in HPSG, wobei sich Müller wiederum auf Erklärungsansätze zahlreicher Autoren zu Phänomenen deutscher Syntax bezieht. Auch hier würden entsprechende Vorkenntnisse dem Leser das Verständnis erleichtern.

Es bleibt die Frage, ob es sinnvoll und möglich ist, eine derart umfassende und schwierige Syntaxtheorie, wie sie uns Stefan Müller vorgibt, im Rahmen einer zweistündigen Seminarsitzung und einer kurzen einführenden Arbeit zu behandeln. Dies wird dem enormen Umfang des Themas wahrscheinlich nicht gerecht.

5 Bibliografie

Borsley, Robert D. (1997): Syntax-Theorie. Ein zusammengefasster Zugang. Tübingen. (Konzepte der Sprach- und Literaturwissenschaft, Bd. 55)

Bussmann, Hadumod (1990): Lexikon der Sprachwissenschaft. Stuttgart.

Liu, Gang (1998): Eine unifikationsbasierte Grammatik für das moderne Chinesisch, dargestellt in der HPSG. Diss. Universität Konstanz. <http://www.uni-konstanz.de/ZE/Bib/diss/1/liu/liu.htm> (17.5.2000)

Müller, Stefan (1996): The Babel-System – An HPSG Prolog Implementation. <http://www.dfki.de/~stefan/Pub/babel.html> (17.5.2000)

Müller, Stefan (1997): Scrambling in German – Extraction into the *Mittelfeld*. <http://www.dfki.de/~stefan/Pub/scrambling.html> (17.5.2000)

Müller, Stefan (1999): Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche. Tübingen. (Linguistische Arbeiten, Bd. 394)

Pollard, Carl J. / Sag, Ivan A. (1994): Head-Driven Phrase Structure Grammar. Chicago.

Schneider, Gerold (1999/2000): Formale Grammatiken und Syntaxanalyse. Vorlesungsskript. <http://www.ifi.unizh.ch/CL/gschneid/SyntaxVorlesung> (17.5.2000)

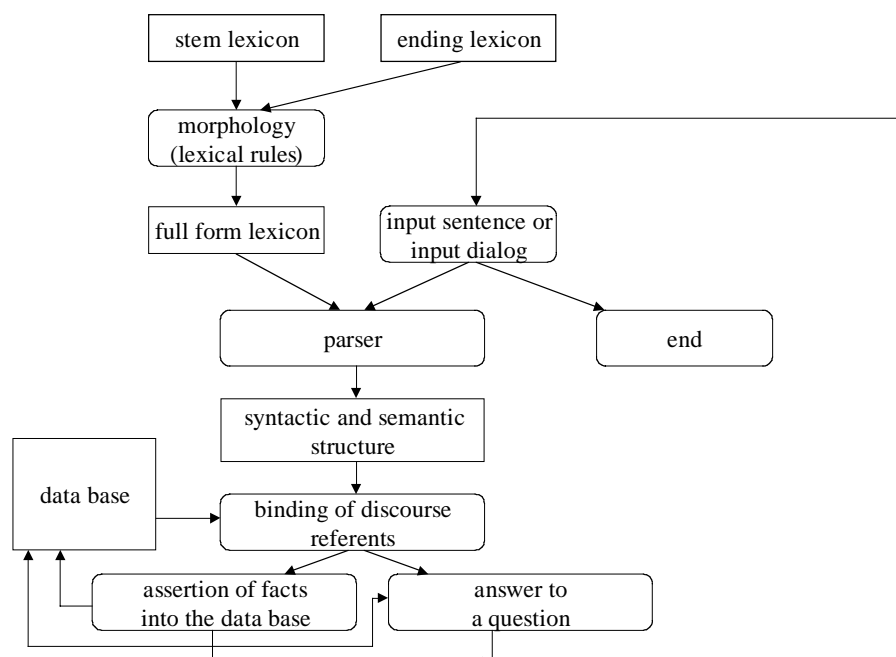
Volk, Martin (1998/1999): Formale Grammatiken und Syntaxanalyse. Vorlesungsskript. <http://www.ifi.unizh.ch/CL/volk/SyntaxVorl> (17.5.2000)

Anhang: Aufbau des Babel-Systems

Stefan Müllers Babel-System ist ein modular aufgebautes Prolog-Programm. Es können Module für folgende Bereiche unterschieden werden (vgl. Müller 1996, S. 2):

- Morphologie
- Parsen (syntaktische und semantische Analyse)
- Hinzufügen von Fakten zur Datenbasis und Beantworten von Fragen
- WWW-Interface

Der Aufbau des Systems lässt sich mit folgender Darstellung beschreiben (vgl. Müller 1996, S. 3):



Morphologische Aspekte

Das Lexikon des Systems umfasst zwei Komponenten: ein Wortstamm-Lexikon (*stem lexicon*) und ein Endungs-Lexikon (*ending lexicon*). Die meisten linguistischen Fakten (Informationen über Subkategorisierungsregeln, syntaktische und semantische Informationen) sind im Stamm-Lexikon integriert. Das *stem lexicon* umfasst Einträge für Verben, Nomina, Adjektive und Determinatoren.

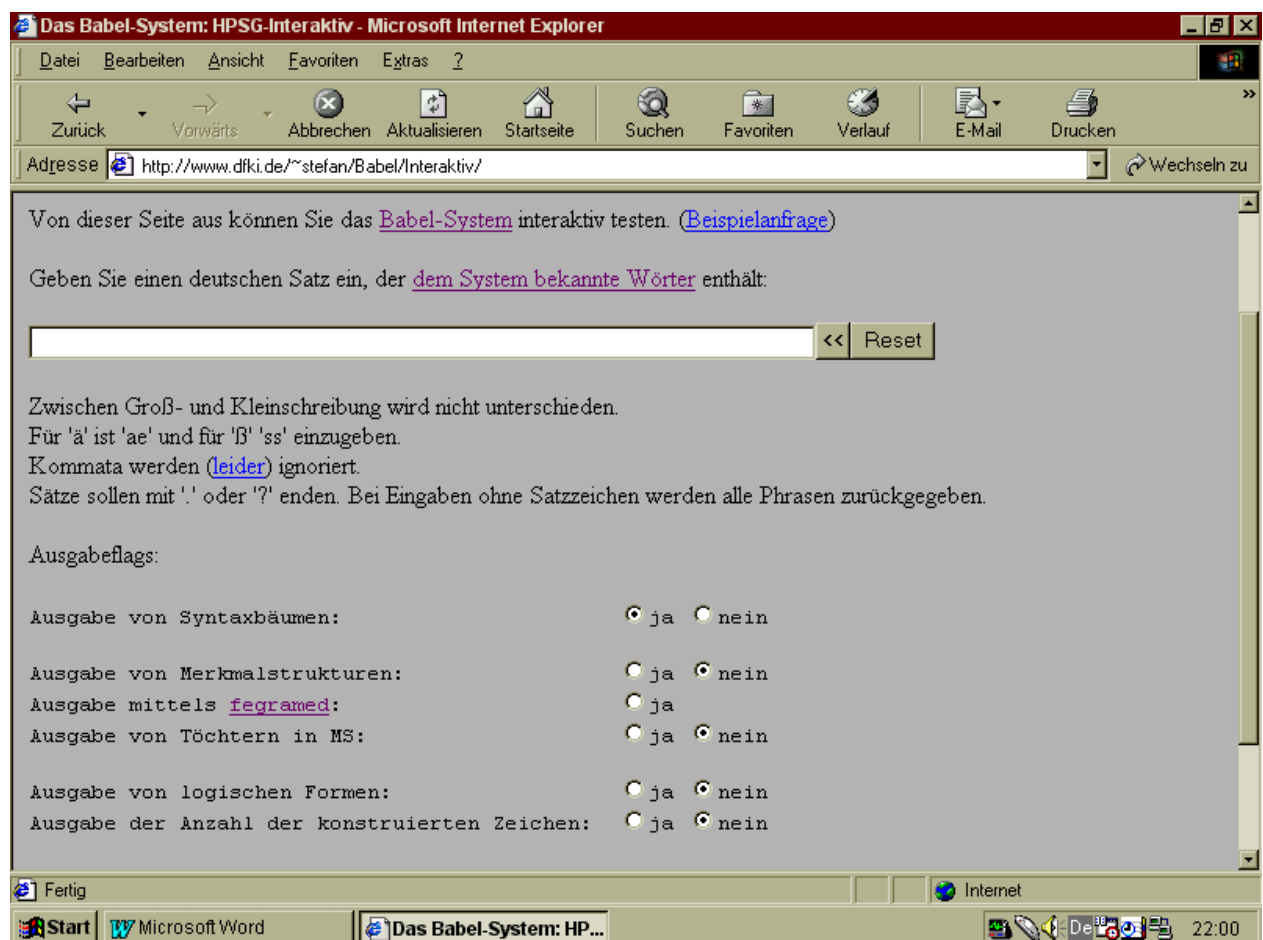
Durch lexikalische Wortbildungsregeln werden die Einträge von *stem lexicon* und *ending lexicon* zu den vollständig flektierten lexikalischen Einträgen im Vollformen-Lexikon (*full form lexicon*) kombiniert. Morphologische Informationen, die in der wortübergeordneten Syntax nicht mehr benötigt wer-

den, sind in den Einträgen (Merkmalstrukturen) des *full form lexicons* nicht mehr berücksichtigt. Die lexikalische Wortbildung geschieht beim Babel-System offline; d.h. beim Parsen der Benutzereingabe wird auf die bereits vollständigen Einträge des *full form lexicon* zugegriffen.

Lexikalische Regeln und Morphologie werden theoretisch ausführlicher behandelt in Müller (1999, Kapitel 7).

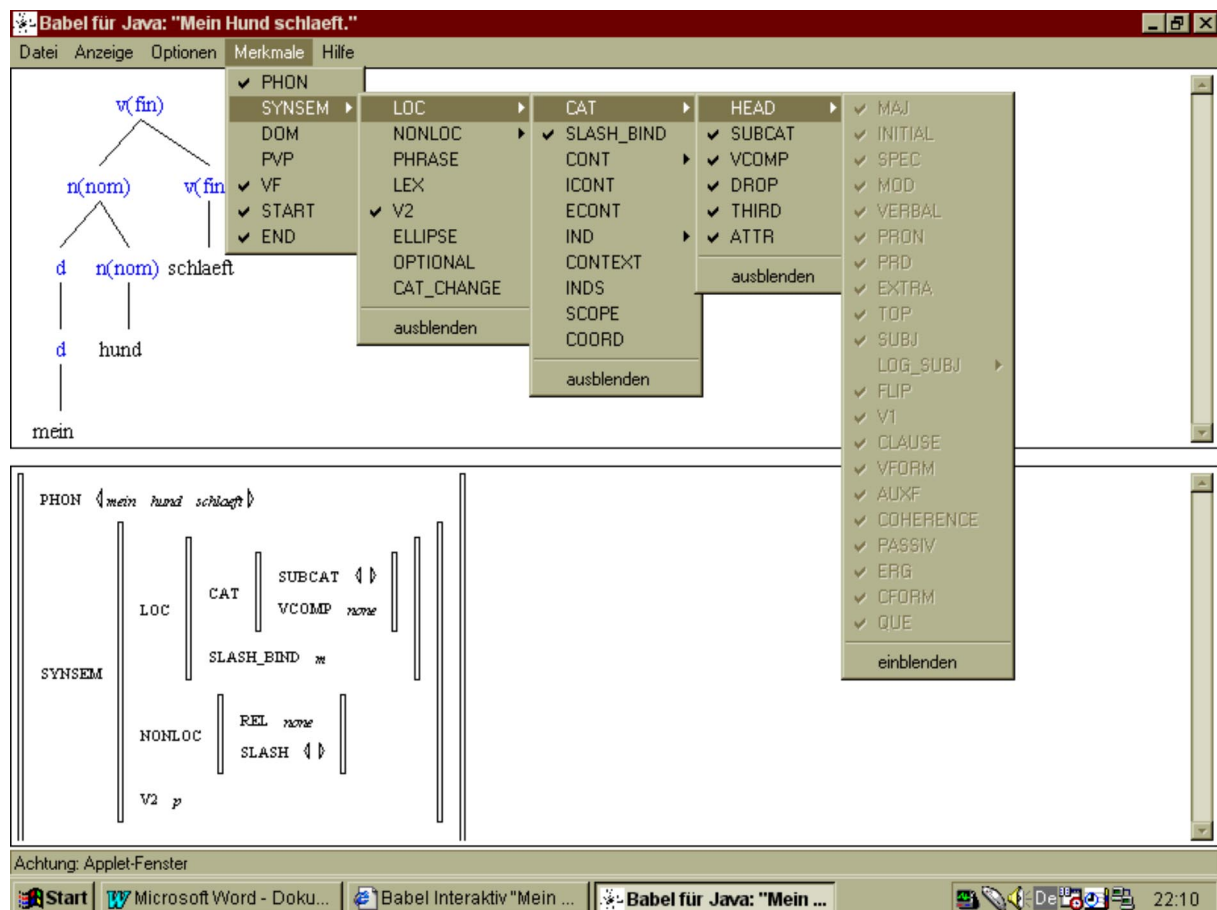
WWW-Interface

Das Babel-System verfügt über ein Interface im World-Wide-Web und kann online getestet werden. Der Benutzer kann zwischen zwei Möglichkeiten wählen, einer (relativ schnellen) HTML-Version (<http://www.dfki.de/~stefan/Babel/Interaktiv/>) sowie einer um interaktive Möglichkeiten ergänzten, jedoch ziemlich langsamen Java-Version (<http://www.dfki.de/~stefan/Babel/Interaktiv/Babajava/>). Der folgende Screenshot zeigt die Eingabeseite der HTML-Version:



Falls eine Analyse erfolgreich war, werden Syntaxbäume, Merkmalstrukturen oder eine semantische Repräsentation der Phrase ausgegeben. Wie auf dem Screenshot ersichtlich, kann der Benutzer bei der HTML-Version durch Flags auf der Eingabeseite auswählen, welche Repräsentation er bevorzugt.

Bei der Java-Version wird für die Ausgabe ein zusätzliches Applet-Fenster geöffnet, bei dem diverse Menü-Funktionen zur Verfügung stehen. Im Menüpunkt 'Anzeige' kann zwischen den unterschiedlichen Darstellungsformen gewechselt werden. Ausserdem können über den Menüpunkt 'Merkmale' einzelne Werte der Merkmalstruktur ein- oder ausgeblendet werden:



Wenn die Analyse misslingt, wird der Benutzer auf eine Seite umgeleitet, wo ihm dies ziemlich lapidar mitgeteilt wird. Falls der Eingabesatz dem System unbekannte Wörter enthält, wird das speziell erwähnt. Es wird ein Link auf eine entsprechende Seite angeboten, von wo aus das Lexikon online erweitert werden kann (http://www.dfki.de/~stefan/Babel/Interaktiv/NetGram/schreib_einen.html). Durch einen weiteren Link wird ausserdem auf eine Seite verwiesen, wo weitere, z.T. enorm nützliche Informationen zu lesen sind, warum für den Eingabesatz keine Analyse gefunden werden konnte (<http://www.dfki.de/~stefan/Babel/Interaktiv/warum.html>):

Das Babel-System: HPSG-Interaktiv - Warum wurde keine Analyse gefunden?

Es gibt mehrere mögliche Gründe dafür, daß keine Analyse gefunden werden konnte.

- *Ein Wort war nicht im Lexikon, obwohl der Stamm in der Stamm-Liste enthalten ist. Hier wird Ihnen oder mir ein Fehler bei der Flexion unterlaufen sein.*
- *Die im Eingabesatz enthaltenen Wörter wurden alle gefunden, eine Analyse jedoch nicht. Das liegt daran, dass zwar die Wörter im Lexikon enthalten sind, jedoch die Funktion, die sie im eingegebenen Satz innehaben, nicht beschrieben ist. Zum Beispiel braucht man für die beiden Vorkommen von *singen* in (1) zwei verschiedene Lexikoneinträge.*

- (1) a. *Maria singt.*
 b. *Maria singt ein Lied.*

Wenn Sie der Meinung sind, Ihr Satz müsste unbedingt vom System erkannt werden, tragen Sie ihn in die Wunschliste ein.

- *Sie haben einen Satz mit Extraktion aus dem eingebetteten dass-Satz eingegeben. Merke: Das ist Bayern-Deutsch! Wenn Sie so komische Sätze analysieren wollen, müssen Sie statt dass ein da_s verwenden. (Peter glaube ich, da_s Hans kennt.)*
- *Sie haben einen Fragesatz eingegeben, aber kein Fragezeichen. Der Punkt wird als Default-Wert angenommen.*
- *Das Programm funktioniert nicht. Merke: Jedes nichttriviale Programm hat mindestens einen Fehler. Wenn man einen Fehler behebt, baut man dabei zehn neue ein.*
- *Die Grammatik hat einen Fehler. Sprache ist komplexer als normale Menschen und Linguisten glauben.*
- *Sie haben Quatsch eingegeben.*