

Ausgewählte Techniken der Maschinellen Übersetzung

Susanne J. Jekat

ZHW

E-mail: jes@zhwin.ch, Subject: MTZH

Übersicht

Themenbereiche:

1. Erstellung und Standardisierung von Ressourcen für die Maschinelle Übersetzung
- 2. Grammatikformalismen für die Maschinelle Übersetzung**
3. Maschinelles Dolmetschen
4. Computergestützte Übersetzung
5. Evaluation von Systemen zur Maschinellen Übersetzung

Thema 2: Grammatikformalismen für die Maschinelle Übersetzung

Übersicht Thema 2:

1. Einführung
2. Head Driven Phrase Structure Grammar
3. **Lexical Functional Grammar**

Thema 2: Grammatikformalismen für die Maschinelle Übersetzung

Wissensgebiete zu Thema 2:

- a) **Syntax und Morphologie**
- b) **Merkmalsstrukturen**
- c) **Kontrastive Linguistik**
- d) **MT (Grundlagen)**
- e) **Translationswissenschaft? (eher Probleme der Translation)**

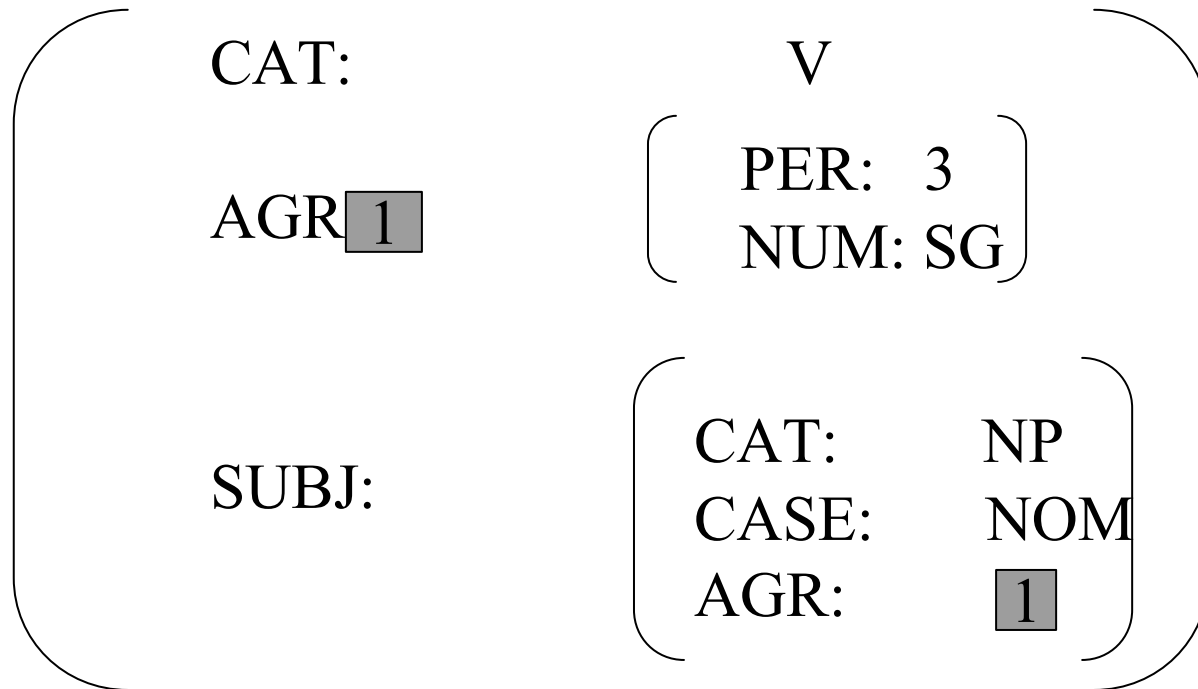
Offene Fragen

Wie geht es weiter mit
der Repräsentation
von Der Hund bellt
in HPSG – Head Feature
Principle?



Darstellung von Der Hund bellt

Kongruenz zwischen Verb (Prädikat) und Subjekt (vereinfacht)



Lexical Functional Grammar

als kontextfreie Grammatik hohe generative
Kapazität

kontextfrei?

Phrasenstrukturgrammatik, die nur solche
Regeln enthält, für deren Anwendung
keinerlei Kontextbedingungen gelten

im Unterschied zu Kontextgebunden (kontextsensitiv)

Phrasenstrukturregel der Form $X \rightarrow Y / A B$,
worin das Symbol X nur dann durch Y
ersetzt werden kann, wenn X in einer Kette
der Form $A X B$ vorkommt

Beispiel: $V \rightarrow V_{\text{trans}} / \text{NP}_{\text{Akk}}$

ersetze V durch ein transitives Verb, wenn
ihm eine NP im Akkusativ folgt

Lexical Functional Grammar

Sätze werden durch eine kontextfreie Grammatik beschrieben (C-Struktur, *constituent structure*), deren Kategorien mit einer Menge von Attribut-Wert-Paaren annotiert sind (F-Struktur, *functional structure*).

LFG

d.h. es gibt zwei Ebenen der syntaktischen
Repräsentation
(C-Struktur und F-Struktur)

Die F-Struktur integriert die Informationen aus der C-Struktur und dem Lexikon (Idee: jeder Satz hat neben einer Konstituentenstruktur auch eine funktionale Struktur, vgl. Schwarze 1996, an dieser Stelle herzlichen Dank an C. Schwarze für die Bereitstellung seines Arbeitspapiers).

C-Struktur

variiert zwischen Sprachen

kodiert lineare Ordnung, hierarchische
Strukturen und die syntaktischen
Kategorien der Konstituenten

Beispiel: Phrasenstrukturregeln für Englisch

$S \rightarrow NP \quad VP \quad VP \rightarrow V \quad NP$

Lexikon

enthält Informationen über die Bedeutung, die Argument-Struktur und die grammatischen Funktionen eines lexikalischen Eintrags

Beispiel: hit

hit, Verb

(↑ PRED) = {Bedeutung von hit} < SUBJ, OBJ >

Lexikon

Beispiel: hit

hit, Verb

(↑ PRED) = {Bedeutung von hit} < SUBJ, OBJ >

↑ Variable, verweist hier auf die lexikalische Einheit,
unter der der Eintrag *hit* gefunden wird

PRED ist ein zweiwertiges Prädikat (Angabe in
spitzen Klammern) und hat als Wert die
Bedeutung von *hit*

F-Struktur

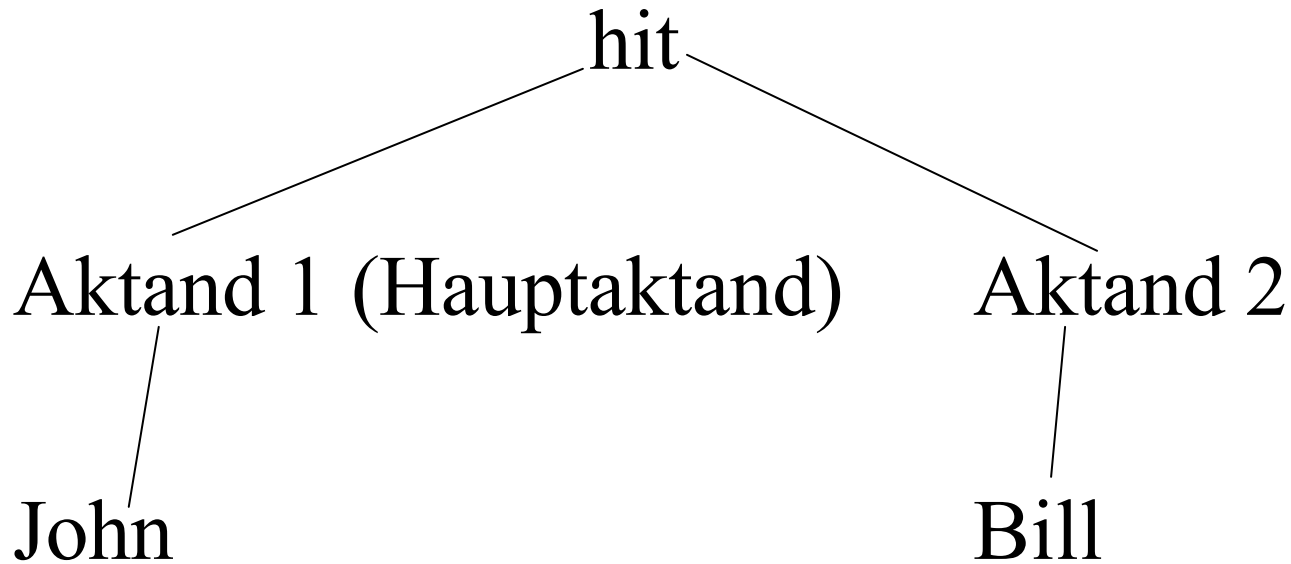
Integriert in die LFG sind Mechanismen der
Valenztheorie
und der
Theorie der Grammatischen Funktionen (vgl.
Schwarze 1996, Lektion 1).

Valenztheorie

Eigenschaft eines Lexems, seine syntaktischen Umgebungen vorzustrukturieren, indem es anderen Konstituenten im Satz Bedingungen bezüglich ihrer grammatischen Eigenschaften auferlegt (wird Tesnière 1959 zugeschrieben)

So fordert z.B. das Verb *hit* ein Subjekt und ein Akkusativ-Objekt (vgl. Folie 12) bzw. es regiert zwei Aktanden

Valenztheorie



Theorie der grammatischen Funktionen

aus der traditionellen Grammatik: Satzteilen werden funktionale Kategorien wie Subjekt, Prädikat, Objekt zugeordnet

diese Kategorien werden in der LFG zur Rekonstruktion des Valenzbegriffs verwendet (Schwarze 1996, Lektion 1)

Subkategorisierung über das Attribut PRED im lexikalischen Eintrag, es wird nur die grammatische Funktion aufgeführt, nicht die syntaktische Kategorie

F-Struktur

enthält alle notwendigen Informationen für die semantische Interpretation einer Äusserung, wird als universal angesehen (variiert nicht zwischen Sprachen)

Wenn die lexikalische Einheit (LE), die den Endknoten belegt, in die syntaktische Struktur eingefügt wird, wird die Information, die in LE enthalten ist inkl. der relevanten Gleichsetzungen in die F-Struktur integriert.

F-Struktur

John hit Bill

SUBJ [PRED `John`]
PRED hit < SUBJ, OBJ >
OBJ [PRED `Bill`]

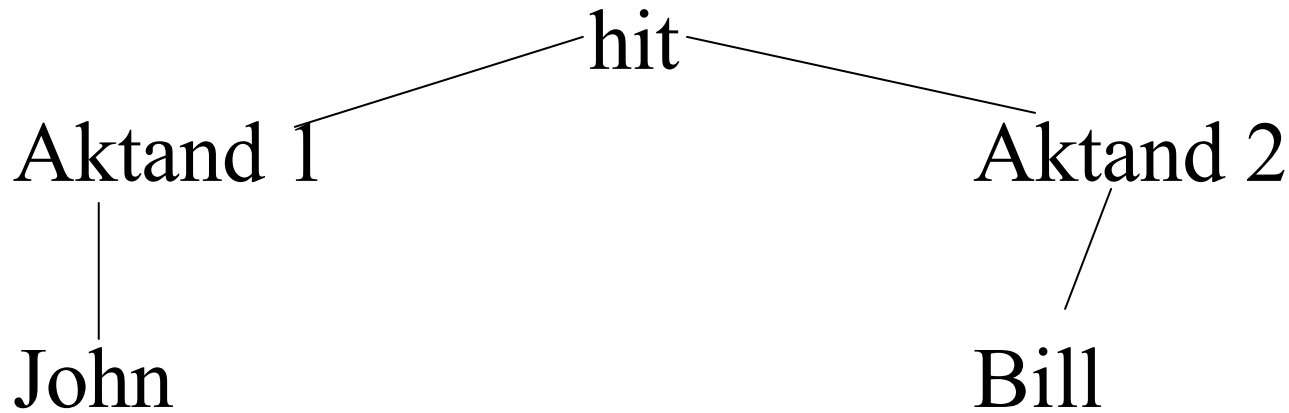
F-Struktur

die F-Struktur unterliegt verschiedenen Wohlgeformtheitsbedingungen:

1. Kohärenz (Coherence)
2. Vollständigkeit (Completeness)
3. Konsistenz (Consistency)

Kohärenz in der F-Struktur

In jeder Unterstruktur der F-Struktur werden alle regierbaren grammatischen Funktionen auch vom Prädikat der Unterstruktur regiert, vgl. Folie 16,



Vollständigkeit in der F-Struktur

eine F-Struktur ist genau dann lokal vollständig, wenn jede ihrer Unterstrukturen alle regierbaren grammatischen Funktionen enthält, die das Prädikat der Unterstruktur regiert (d.h. die Valenz des Verbs muss erfüllt sein)

z.B. * Speaks ist als Satz unvollständig, da kein Wert für SUBJ vorhanden ist

Konsistenz in der F-Struktur

auch *functional uniqueness*

in einer F-Struktur darf jedes Attribut nur einen Wert besitzen

also ist eine F-Struktur inkonsistent, die die

Matrix GEND + FEM

 GEND - FEM

enthält.

Übung (nach Schwarze 1996)

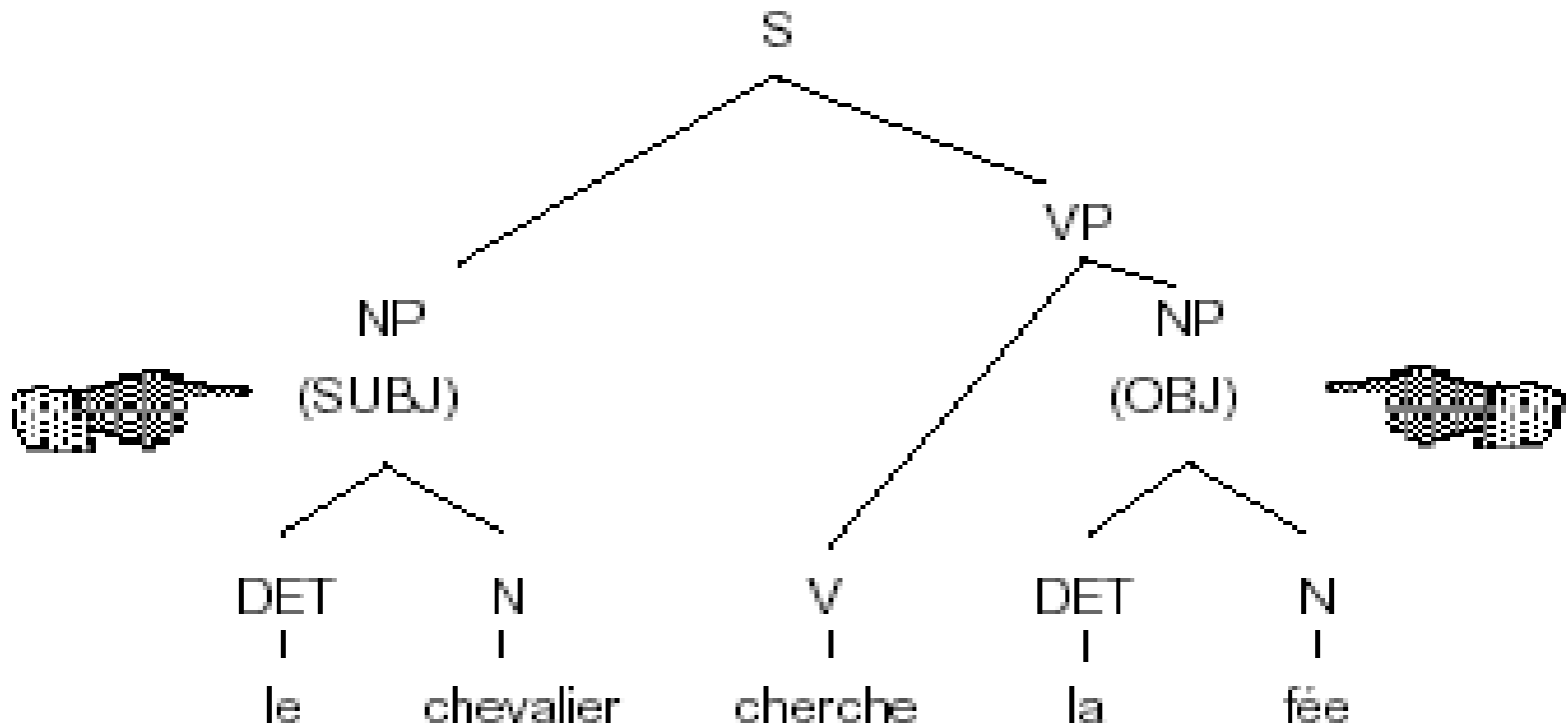
(1) Le chevalier
cherche la fée

N.B. Diese Grammatik
erzeugt auch andere Sätze
als (1), darunter auch
solche, die die
Kongruenzregeln
verletzen.

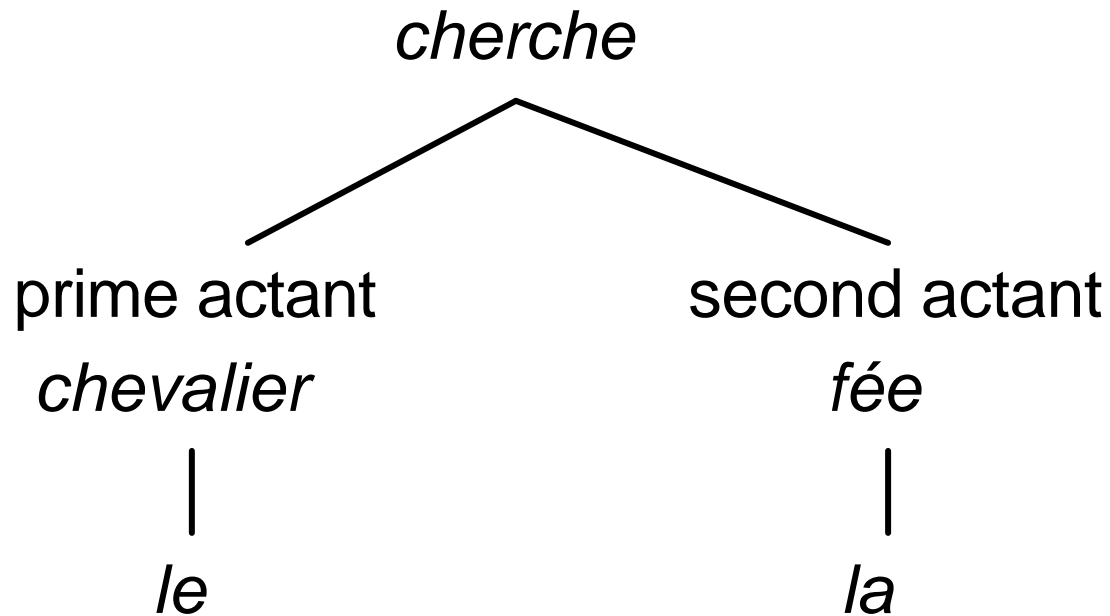
Phrasenstruktur

S	→	NP VP
VP	→	V NP
NP	→	DET N
V	→	cherche
N	→	chevalier
N	→	fée
DET	→	le
DET	→	la

Funktionale Information in der Phrasenstruktur (Schwarze 1996)



Valenzstruktur (Schwarze 1996)



Unifikation

fée hat für das Attribut Genus den Wert *Femininum* und für das Attribut Numerus den Wert *Singular*

Bei dem ungrammatischen Ausdruck **les belle fées* hingegen unifiziert der Wert des Attributs Numerus nicht: Er ist *Singular* bei *belle* und *Plural* bei *les* und *fées*.

Unifikation

la belle fée Kongruenz innerhalb einer
Konstituente (NP)

aber Subjekt und Prädikat sind verschiedene
Konstituenten (oberste NP und oberste VP des
Satzes), daher wird in LFG für jede Verbform
angegeben, welche Kongruenzmerkmale das
Subjekt haben muss

Kongruenz über Konstituenten

(Schwarze 1996)

die Kongruenz ist eine Forderung, die eine Form des Verbs an
das Subjekt stellt

rit, v

SUBJ NUM = SG

SUBJ PERS = 3

rient, v

SUBJ NUM = PL

SUBJ PERS = 3

rions, v

SUBJ NUM = PL

SUBJ PERS = 1

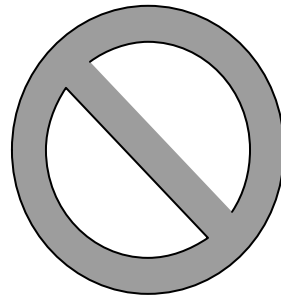
Kongruenz über Konstituenten

le chevalier

GEND = MAS

NUM = SING

PERS = 3



rions

NUM = PL

PERS = 1

F-Struktur

Die F-Struktur von *le chevalier cherche la fée* besteht aus den folgenden Fakten:

Der Satz besteht aus einem Subjekt, einem Verb und einem (direkten) Objekt.

Das Subjekt hat die lexikalische Bedeutung 'chevalier', es ist definit, und seine Kongruenzmerkmale sind Maskulinum, Singular und dritte Person.

Das Verb hat die lexikalische Bedeutung 'chercher'; seine Valenz legt fest, dass ein Subjekt und ein Objekt vorhanden sein müssen. Seine Kongruenzmerkmale sind Singular, dritte Person.

Das Objekt hat die lexikalische Bedeutung 'fée'; es ist definit, und seine Kongruenzmerkmale sind Femininum, Singular und dritte Person.

Erweiterung der Phrasenstrukturregeln (Schwarze 1996)

S	∅	NP	VP
		(SUBJ)	
VP	∅	V	NP
			(OBJ)
NP	∅	DET	N

Funktionale Information im Lexikon

Hinzufügen von Prädikaten und lexikalischen
Formen

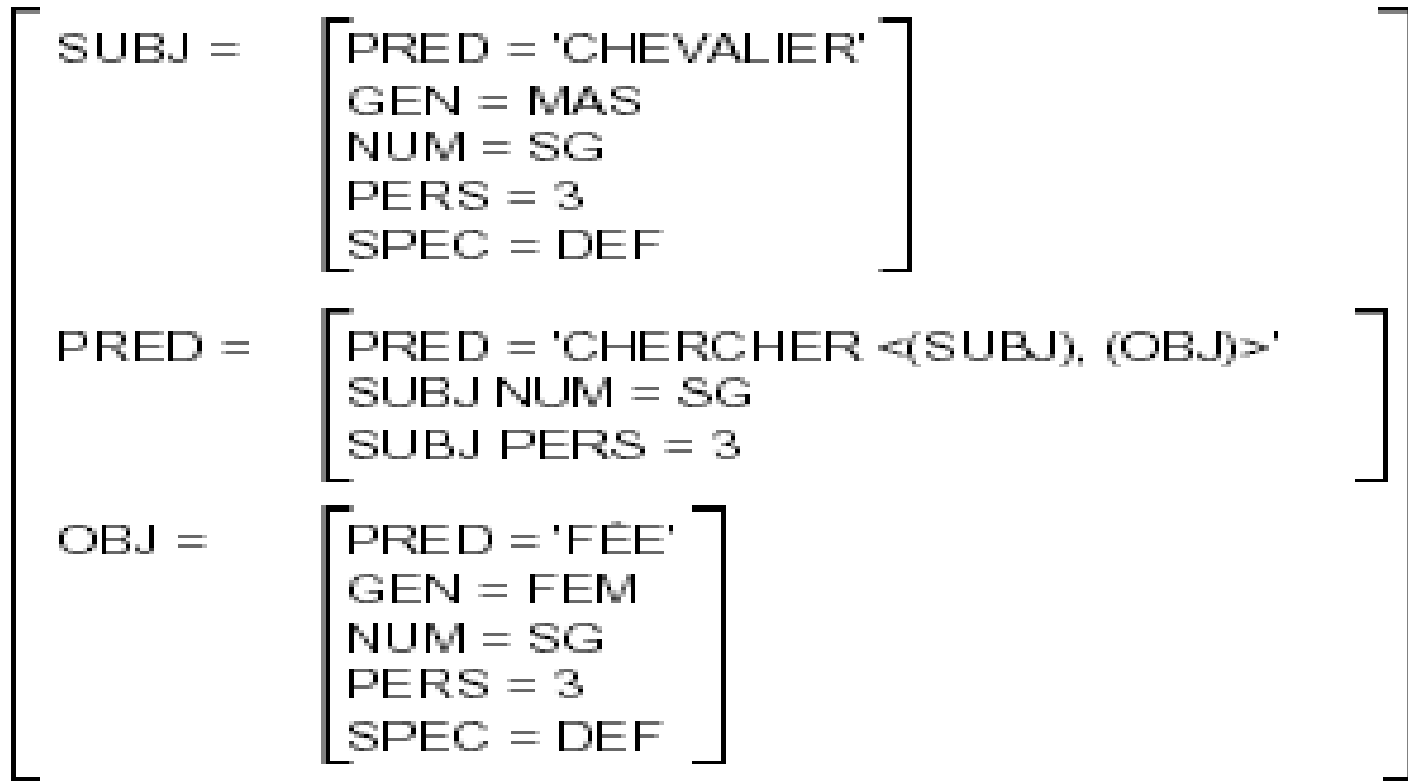
fée, N

PRED = 'FEE'

GEN = FEM

NUM = SG

F-Struktur (Schwarze 1996)



Der Grammatikschreiber

„Beim Grammatikschreiben muß jeweils entschieden werden, wo eine Information zu kodieren ist. Alles, was Eigenschaft einzelner Wörter ist, wird im Lexikon ausgedrückt, der Rest in den annotierten k-Regeln [hier C-Regeln, Anm. S.J.]. Außerdem muß man jeweils überlegen, welche f-Strukturen [hier F-Strukturen, Anm. S.J.] man erhalten will. Ein wichtiges Kriterium hierfür ist auch, daß die Semantik, deren Eingabe die s-Strukturen sind, alle Informationen bekommt, die sie braucht.“ (Schwarze, 1996, Lektion 9)

Kategorien der Konstituentenstruktur (C-Struktur)

a) Kategorien der Konstituentenstruktur

b) Kategorien der funktionalen Struktur.

Die Kategorien der Konstituentenstruktur
wiederum zerfallen in Hauptkategorien und
lexikalische Kategorien:

Hauptkategorien: S, NP, VP, AP, PP

Lexikalische Kategorien: N, V, A, P, DET

Kategorien der funktionale Struktur (F-Struktur)

Bei den funktionale Kategorien unterscheiden wir
zwischen den grammatischen Funktionen und
anderen funktionalen Merkmalen:

Grammatische Funktionen: SUBJ, OBJ, VCOMP, ...

Funktionale Merkmale: PRED, PERS, NUM, GEN,
INF, MODE, ...

Aufbau der LFG

Jedes der funktionalen Merkmale besteht aus einem Attribut und einer Menge möglicher Werte:

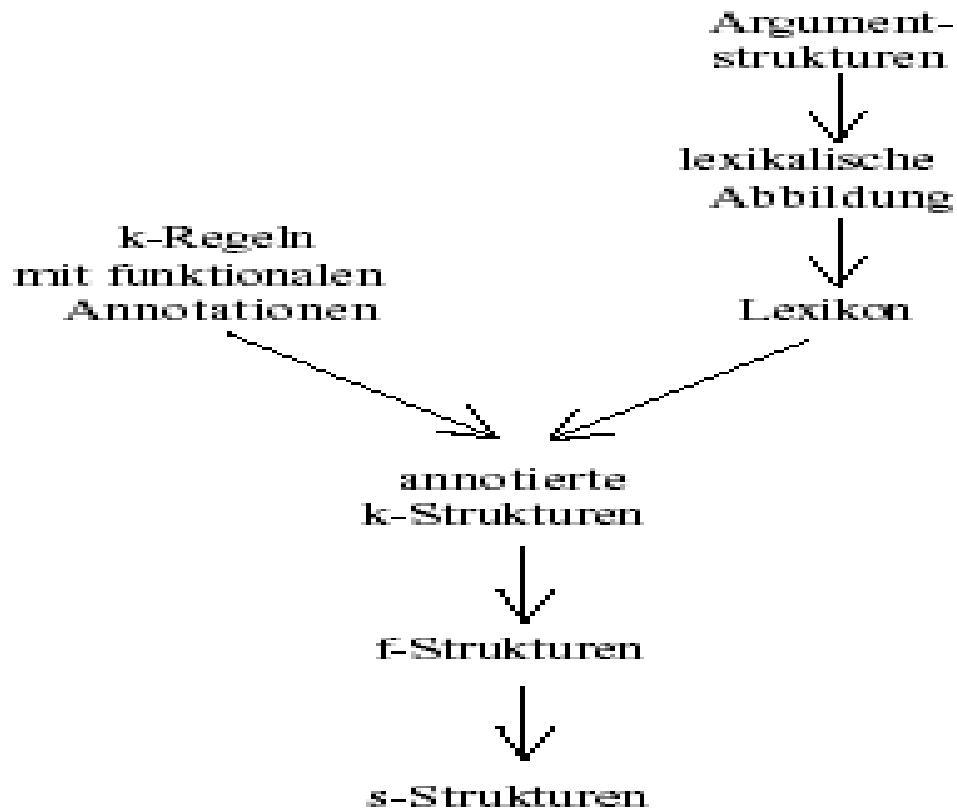
Attribut	mögliche Werte
PRED	{...}
PERS	{1,2,3}
NUM	{sg, pl}
GEND	{mas, fem}
INF	{part, infinitif}
MODE	{indicatif, subjonctif}

Aufbau der LFG

(Schwarze 1996, Lektion 9)

Die funktionalen Kategorien sind in den C-Regeln (als Annotationen) und in den Lexikoneinträgen kodiert und werden dann durch die schrittweise Unifikation zu F-Beschreibungen aufgebaut; diese werden schließlich, wenn sie den Anforderungen von Vollständigkeit und Kohärenz genügen, als F-Strukturen akzeptiert.

Aufbau der LFG (Schwarze 1996)



Quellen

- Bresnan, J., ed, (1982) *The Mental Representation of Grammatical Relations*, Cambridge, MA:MIT Press.
- Bussmann, Hadumod (2002) *Lexikon der Sprachwissenschaft*. Stuttgart:Kröner.
- Klabunde, Ralf et al., ed, (2004) *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg:Spektrum.
- Falk, Yehuda N. (2001) *Lexical-Functional Grammar: An introduction to Parallel Constraints-Based Syntax*. Stanford (Cal.):CSLI Publications.
- Neidle, Carol (o.J.) *Lexical Functional Grammar*, Ms.:Boston University.
- Schwarze, Christoph (1996) *Lexikalisch-funktionale Grammatik. Eine Einführung in 10 Lektionen, mit französischen Beispielen*. Fachbereich Sprachwissenschaft der Universität Konstanz, Arbeitspapier Nr. 76.
- Tesnière, L. (1959) *Eléments de syntaxe structurale*, Paris:Klincksieck.