

# Übersicht

## 7 Unifikationsgrammatiken

### 7.1 Modellierung von Subkategorisierung und Agreement

### 7.2 Unifikationsbasierte Erweiterungen

7.2.1 Kasus und Agreement im Deutschen

7.2.2 Subkategorisierung in GPSG und HPSG

7.2.3 Auxiliare und Inversion

7.2.4 *long distance dependencies*

# 7 Unifikationsgrammatiken

# **7.1 Modellierung von Subkategorisierung und Agreement**

## Modellierung von Subkategorisierung

- **Anzahl und Art der verbalen Argumente** (Valenz) muss in formaler Modellierung berücksichtigt werden, um Überproduktion zu vermeiden
  - falscher Argumenttyp: *\*Er jagt, dass er kommt.*
  - falsche Anzahl: *\*Der Hund bellt den Vogel.*
- **Subkategorisierung = Unterteilung der Klasse der Verben nach Valenztypen**

- **Subkategorisierungsrahmen** = formale Repräsentation der syntaktischen Valenz eines Wortes
- *rennen* : [  $\_$   $NP_{NOM}$  ], *jagen* : [  $\_$   $NP_{NOM}$   $NP_{ACC}$  ]
- **Subkategorisierungsprinzip**: Ein Verb kann nur in einer **Umgebung** auftreten, die seinem **Subkategorisierungsrahmen** entspricht
- Subkategorisierung als **Beschränkung** (*Constraint*) der **syntaktischen Umgebung** (Kontext) von Verben, in der sie vorkommen können

## Modellierung von Subkategor. mit kontextsensitiven Regeln

- **Berücksichtigung Kontext** zur Modellierung von Subkategorisierung
- **kontextsensitive Regeln**: formale Grammatik kann **mehr als ein Nichtterminal auf der linken Seite** enthalten (Kontext)
- Ersetzung einzelner Nichtterminale (hier: V) nur in Kontext
- Problem: kontextsensitive Regeln **komplex in der Verarbeitung**
- **Status** von *V<sub>intrans</sub>* usw. als Subklasse von V wird nur **indirekt** über kontextsensitive Regel **sichtbar**

Auflistung 1: *Subkategorisierung mit kontextsensitiven Regeln*

- 1  $VP \rightarrow V(NP)(NP \mid PP)$
- 2
- 3  $V \rightarrow Vintrans$
- 4  $V NP \rightarrow Vtrans NP$
- 5  $V PP \rightarrow Vprepobj PP$
- 6  $V NP NP \rightarrow Vditrans NP NP$
- 7  $V NP PP \rightarrow Vplace NP PP$
- 8  $V S\text{-}BAR \rightarrow Vclause S\text{-}BAR$

## Modellierung von Subkategorisierung mit Merkmalsstrukturen

- **Subkategorisierung als Merkmal in Lexikoneinträgen der Verben**
- **Status als Subkategorie der Wortklasse Verb direkt modelliert:**  
V[SUBCAT=intrans]
- Verwendung **SUBCAT-Merkmal als Index** von PSG-Regeln zur Angabe, welche Argumente ein Verb verlangt



*Auflistung 2: PSG-Regeln für Subkategorisierungsmerkmal*

- 1  $VP \rightarrow V[\text{SUBCAT}=\text{intrans}]$
- 2  $VP \rightarrow V[\text{SUBCAT}=\text{trans}] \quad NP$
- 3  $VP \rightarrow V[\text{SUBCAT}=\text{prepobj}] \quad PP$
- 4  $VP \rightarrow V[\text{SUBCAT}=\text{ditrans}] \quad NP \quad NP$
- 5  $VP \rightarrow V[\text{SUBCAT}=\text{place}] \quad NP \quad PP$
- 6  $VP \rightarrow V[\text{SUBCAT}=\text{clause}] \quad S\text{-BAR}$
- 7
- 8  $\text{bellt} \rightarrow V[\text{SUBCAT}=\text{intrans}]$
- 9  $\text{jagt} \rightarrow V[\text{SUBCAT}=\text{trans}]$

## Subkategorisierung als direkte Valenzkodierung

- **Alternativ** können in einem SUBCAT-Merkmal auch **direkt die verlangten Komplementtypen** (als komplexes Merkmal) kodiert werden

$$\begin{array}{c}
 \text{bellt} \left[ \begin{array}{cc} \text{CAT} & V \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]
 \end{array}
 \quad
 \begin{array}{c}
 \text{jagt} \left[ \begin{array}{cc} \text{CAT} & V \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \left[ \begin{array}{cc} \text{CAT} & \text{NP} \\ \text{CASE} & \text{ACC} \end{array} \right] \end{array} \right]
 \end{array}$$

## Constraintregeln für Subkategorisierung

- Regel für intransitive VP:

$$\begin{bmatrix} \text{CAT} & VP \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \boxed{1} \\ \text{SUBCAT} & \text{NONE} \end{bmatrix}$$

- Lexikoneintrag (intransitives Verb = ohne Komplement):

$$bel\bar{l}t \begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & PRES \\ \text{NUM} & SG \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \text{NONE} \end{bmatrix}$$

- **Regel für transitive VP** (auch für Präpositionalobjekt usw.):

$$\begin{bmatrix} \text{CAT} & VP \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \boxed{1} \\ \text{SUBCAT} & \boxed{2} \end{bmatrix} \boxed{2}$$

- **Lexikoneintrag (transitives Verb mit Akkusativ-Komplement):**

$$jagt \begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & PRES \\ \text{NUM} & SG \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \begin{bmatrix} \text{CAT} & NP \\ \text{CASE} & ACC \end{bmatrix} \end{bmatrix}$$

- **Beispiel: Ablehnung intransitives Verb + NP**

→ **Unifikation schlägt fehl**, inkompatiblen Werte im SUBCAT-Merkmal (Constraintanweisung:  $\langle V \text{ SUBCAT} \rangle = \langle NP \rangle$ ):

$$\text{bellt} \left[ \begin{array}{cc} \text{CAT} & v \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right] \text{ den Hund} \left[ \begin{array}{cc} \text{CAT} & NP \\ \text{CASE} & \text{ACC} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{array} \right] \end{array} \right]$$

$$\text{NONE} \sqsubset \left[ \begin{array}{cc} \text{CAT} & NP \\ \text{CASE} & \text{ACC} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{array} \right] \end{array} \right] = \text{FAIL!}$$

## Constraintregel für verbales Agreement und Subjekt-Kasus

- Berücksichtigung von **Subjekt-Verb-Kongruenz und Kasus des Subjekts** zur Vermeidung von Überproduktion: *\*Der Hund bel-*  
*len; \*Den Hund bellt.*

$$[ \text{CAT} \quad s ] \rightarrow \left[ \begin{array}{cc} \text{CAT} & NP \\ \text{CASE} & NOM \\ \text{AGR} & \boxed{1} \end{array} \right] \left[ \begin{array}{cc} \text{CAT} & VP \\ \text{AGR} & \boxed{1} \end{array} \right]$$

- mit Regeln für **NP** und **intransitive VP** von oben:

$$\begin{bmatrix} \text{CAT} & NP \\ \text{CASE} & \boxed{2} \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & DET \\ \text{CASE} & \boxed{2} \\ \text{AGR} & \boxed{1} \end{bmatrix} \begin{bmatrix} \text{CAT} & N \\ \text{CASE} & \boxed{2} \\ \text{AGR} & \boxed{1} \end{bmatrix}$$

$$\begin{bmatrix} \text{CAT} & VP \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \boxed{1} \\ \text{SUBCAT} & NONE \end{bmatrix}$$

- **Akzeptierung:**

→ Unifikation gelingt (**keine inkompatiblen Strukturen**)

$$\begin{array}{c} \text{der} \end{array} \left[ \begin{array}{cc} \text{CAT} & \text{DET} \\ \text{CASE} & \text{NOM} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{array} \right] \end{array} \right] \quad \text{Hund} \left[ \begin{array}{cc} \text{CAT} & \text{N} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{array} \right] \end{array} \right] \quad \text{bellt} \left[ \begin{array}{cc} \text{CAT} & \text{V} \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]$$

$$\left[ \begin{array}{cc} \text{CAT} & \text{NP} \\ \text{CASE} & \text{NOM} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{array} \right] \end{array} \right], \quad \left[ \begin{array}{cc} \text{CAT} & \text{VP} \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]$$



- **Ablehnung:**

→ **Subjekt-Verb-Agreement-Constraint wird verletzt:**  $\langle \text{NP AGR NUM} \rangle = \langle \text{VP AGR NUM} \rangle$

→ (Subjekt-Kasus-Constraint erfüllt,  $\langle \text{NP CASE} \rangle$  unterspezifiziert)

$$\begin{array}{c}
 \text{die} \left[ \begin{array}{cc} \text{CAT} & \text{DET} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \end{array} \right] \end{array} \right]
 \end{array}
 \quad
 \text{Katze} \left[ \begin{array}{cc} \text{CAT} & \text{N} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \\ \text{PER} & 3 \end{array} \right] \end{array} \right]
 \quad
 \text{bellen} \left[ \begin{array}{cc} \text{CAT} & \text{V} \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{PL} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]$$
  

$$\left[ \begin{array}{cc} \text{CAT} & \text{NP} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \\ \text{PER} & 3 \end{array} \right] \end{array} \right], \quad
 \left[ \begin{array}{cc} \text{CAT} & \text{VP} \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{PL} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]$$

- **Ablehnung:**

→ **Subjekt-Kasus-Constraint** wird verletzt:  $\langle \text{NP CASE} \rangle = \text{NOM}$

$$\begin{array}{c} \text{den} \end{array} \left[ \begin{array}{cc} \text{CAT} & \text{DET} \\ \text{CASE} & \text{ACC} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{array} \right] \end{array} \right] \quad \text{Hund} \left[ \begin{array}{cc} \text{CAT} & \text{N} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{array} \right] \end{array} \right] \quad \text{bellt} \left[ \begin{array}{cc} \text{CAT} & \text{V} \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]$$

$$\left[ \begin{array}{cc} \text{CAT} & \text{NP} \\ \text{CASE} & \text{ACC} \\ \text{AGR} & \left[ \begin{array}{cc} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{array} \right] \end{array} \right], \quad \left[ \begin{array}{cc} \text{CAT} & \text{VP} \\ \text{AGR} & \left[ \begin{array}{cc} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{array} \right] \\ \text{SUBCAT} & \text{NONE} \end{array} \right]$$

# **7.2 Unifikationsbasierte Erweiterungen**

## 7.2.1 Kasus und Agreement im Deutschen

- **Grundlage: NLTK Kapitel 9.3.5:**

<http://www.nltk.org/book/ch09.html#code-germancfg>

- zu *feature structures* und *feature-based grammars* im NLTK siehe auch:

<http://www.nltk.org/howto/featgram.html>

<http://www.nltk.org/howto/featstruct.html>

- Beispielgrammatik für Berücksichtigung von **Kasusreaktion** und **verbalem Agreement** mit **merkmalsstrukturbasierter Grammatik** zur Vermeidung von Überproduktion:

*\*den Hund (CASE) sehen (AGR) dem Vogel (CASE)*

- einfache Lösung für **Subkategorisierung (Anzahl Argumente)** über **Kategorienerweiterung** (*\* der Hund kommt den Vogel*):

IV=intransitives Verb, TV=transitives Verb

- **Rektionsbeziehung über Merkmalconstraint, insbesondere Kasus der Objekt-NP eines TV:**

→ TV-Merkmal OBJCASE muss mit CASE-Merkmal von NP unifizierbar sein, als Pfadgleichung:  $\langle \text{TV} \text{ OBJCASE} \rangle = \langle \text{NP} \text{ CASE} \rangle$

## Kasus-Rektion als Merkmalconstraint

$$\begin{bmatrix} \text{CAT} & VP \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & IV \\ \text{AGR} & \boxed{1} \end{bmatrix}$$

$$\begin{bmatrix} \text{CAT} & VP \\ \text{AGR} & \boxed{1} \end{bmatrix} \rightarrow \begin{bmatrix} \text{CAT} & TV \\ \text{AGR} & \boxed{1} \\ \text{OBJCASE} & \boxed{2} \end{bmatrix} \quad \begin{bmatrix} \text{CAT} & NP \\ \text{CASE} & \boxed{2} \end{bmatrix}$$

$$\textit{kommt} \begin{bmatrix} \text{CAT} & IV \\ \text{AGR} & \begin{bmatrix} \text{NUM} & SG \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix} \quad \textit{folgt} \begin{bmatrix} \text{CAT} & TV \\ \text{AGR} & \begin{bmatrix} \text{NUM} & SG \\ \text{PER} & 3 \end{bmatrix} \\ \text{OBJCASE} & DAT \end{bmatrix}$$

Auflistung 3: *Merkmalstrukturbasierte Grammatik für das Deutsche*

```
## Natural Language Toolkit: german.fcfg
```

```
% start S
```

```
#####
```

```
# Grammar Productions
```

```
S → NP[CASE=nom, AGR=?a] VP[AGR=?a]
```

```
NP[CASE=?c, AGR=?a] → PRO[CASE=?c, AGR=?a]
```

```
NP[CASE=?c, AGR=?a] → Det[CASE=?c, AGR=?a] N[CASE=?c, AGR=?a]
```

```
VP[AGR=?a] → IV[AGR=?a]
```

```
VP[AGR=?a] → TV[OBJCASE=?c, AGR=?a] NP[CASE=?c]
```

#####

### *Lexical Productions (Auswahl):*

# *Singular determiners masc Sg*

Det[CASE=nom, AGR=[GND=masc,PER=3,NUM=sg]] → 'der'

Det[CASE=dat, AGR=[GND=masc,PER=3,NUM=sg]] → 'dem'

Det[CASE=acc, AGR=[GND=masc,PER=3,NUM=sg]] → 'den'

# *Nouns*

N[AGR=[GND=masc,PER=3,NUM=sg]] → 'Hund'

N[CASE=nom, AGR=[GND=masc,PER=3,NUM=pl]] → 'Hunde'

N[CASE=dat, AGR=[GND=masc,PER=3,NUM=pl]] → 'Hunden'

N[CASE=acc, AGR=[GND=masc,PER=3,NUM=pl]] → 'Hunde'

# *Pronouns*

PRO[CASE=nom, AGR=[PER=1,NUM=sg]] → 'ich'

PRO[CASE=acc, AGR=[PER=1,NUM=sg]] → 'mich'

PRO[CASE=dat, AGR=[PER=1,NUM=sg]] → 'mir'

PRO[CASE=nom, AGR=[PER=3,NUM=sg]] → 'er' | 'sie' | 'es'

# *Verbs*

IV[AGR=[NUM=sg,PER=3]] → 'kommt'

TV[OBJCASE=acc, AGR=[NUM=sg,PER=3]] → 'sieht' | 'mag'

TV[OBJCASE=dat, AGR=[NUM=sg,PER=2]] → 'folgst' | 'hilfst'



Auflistung 4: *Parsing Beispielsatz*

```
1 from nltk import load_parser
2
3 parser =
    load_parser('./grammars/book_grammars/german.fcfg',
    trace=0)
4
5 trees = parser.parse("der Hund folgt
    mich".split())
6 i = -1
7 for i, tree in enumerate(trees):
8     print(tree)
9
10 if i < 0:
11     print("Satz nicht akzeptiert!")
12 #Satz nicht akzeptiert!
```

```
13 trees = parser.parse("der Hund folgt  
    mir".split())  
14 i = -1  
15 for i, tree in enumerate(trees):  
16     print(tree)  
17 #(S []  
18 #   (NP [AGR=[GND='masc', NUM='sg', PER=3],  
        CASE='nom']  
19 #     (Det [AGR=[GND='masc', NUM='sg', PER=3],  
          CASE='nom'] der)  
20 #     (N [AGR=[GND='masc', NUM='sg', PER=3]]  
        Hund))  
21 #   (VP [AGR=[NUM='sg', PER=3]]  
22 #     (TV [AGR=[NUM='sg', PER=3], OBJCASE='dat']  
        folgt)  
23 #     (NP [AGR=[NUM='sg', PER=1], CASE='dat']
```

24 # (PRO [AGR=[NUM='sg', PER=1], CASE='dat']  
 mir)))))

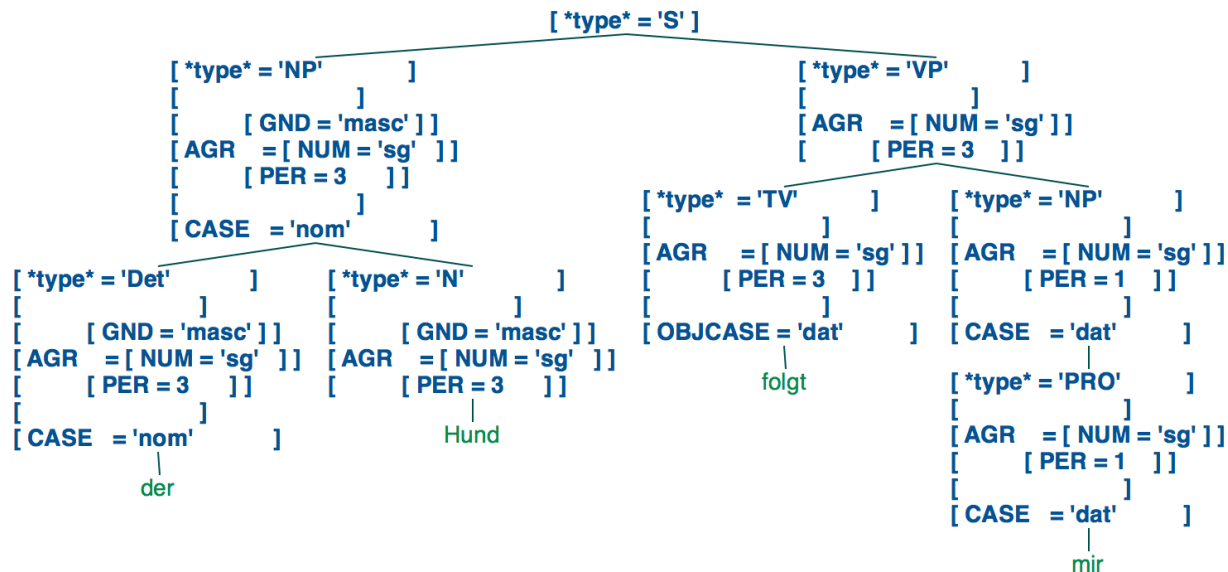


Abbildung 1: Syntaxbaum zur Ableitung

## **7.2.2 Subkategorisierung in GPSG und HPSG**

## Modellierung syntaktischer Phänomene des Englischen

- **Grundlage: NLTK Kapitel 9.3.1-4:**

<http://www.nltk.org/book/ch09.html>

- **Merkmalstrukturbasierte Modellierung** von syntaktischen Phänomenen wie **Subkategorisierung** und **Wortstellung**

## Subkategorisierung als Index

- Ansatz der **GPSG** (*Generalized Phrase Structure Grammar*)
- SUBCAT-Wert als **Index der VP-Produktionsregeln**
- **atomare Werte:** intrans, trans, clause
- auch Subkategorisierung nach Komplementsätzen
- **Grammatik besteht im Kern aus PSG-Regeln**, die um Merkmalsbeschränkungen erweitert sind

*Auflistung 5: Unifikationsgrammatik mit SUBCAT als Index*

VP[TENSE=?t, NUM=?n] → V[SUBCAT=intrans, TENSE=?t, NUM=?n]  
VP[TENSE=?t, NUM=?n] → V[SUBCAT=trans, TENSE=?t, NUM=?n] NP  
VP[TENSE=?t, NUM=?n] → V[SUBCAT=clause, TENSE=?t, NUM=?n] SBar

V[SUBCAT=intrans, TENSE=pres, NUM=sg] → 'disappears' | 'walks'  
V[SUBCAT=trans, TENSE=pres, NUM=sg] → 'sees' | 'likes'  
V[SUBCAT=clause, TENSE=pres, NUM=sg] → 'says' | 'claims'

V[SUBCAT=intrans, TENSE=pres, NUM=pl] → 'disappear' | 'walk'  
V[SUBCAT=trans, TENSE=pres, NUM=pl] → 'see' | 'like'  
V[SUBCAT=clause, TENSE=pres, NUM=pl] → 'say' | 'claim'

V[SUBCAT=intrans, TENSE=past, NUM=?n] → 'disappeared' | 'walked'  
V[SUBCAT=trans, TENSE=past, NUM=?n] → 'saw' | 'liked'  
V[SUBCAT=clause, TENSE=past, NUM=?n] → 'said' | 'claimed'

SBar → Comp S  
Comp → 'that'

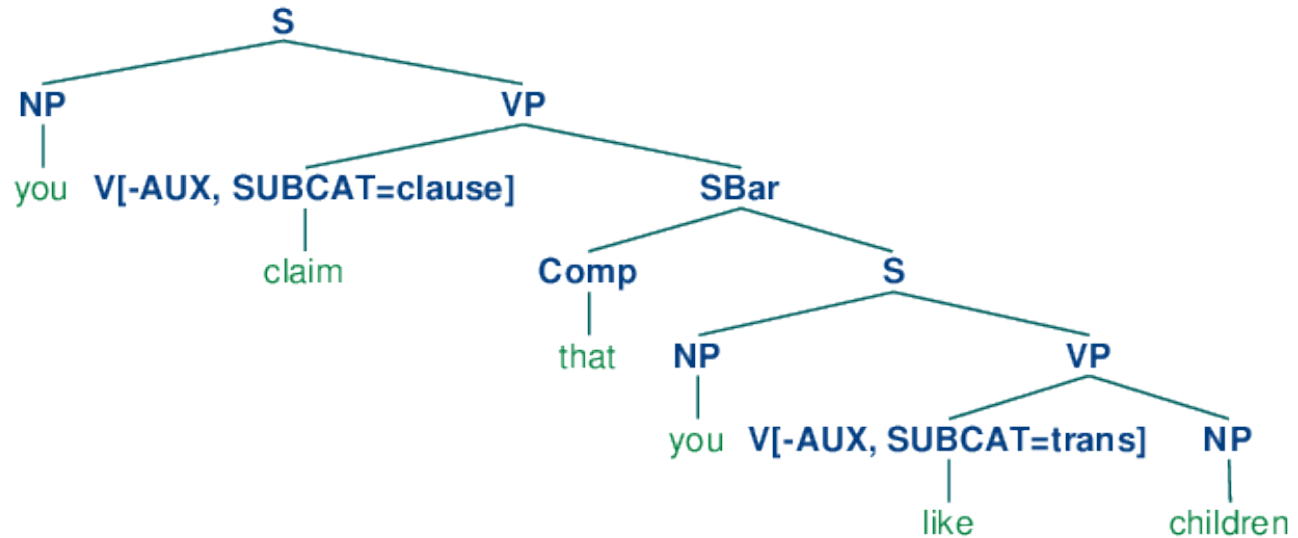


Abbildung 2: Subkategorisierung nach Komplementsatz  
([http://www.nltk.org/book/tree\\_images/ch09-tree-10.png](http://www.nltk.org/book/tree_images/ch09-tree-10.png))



## Subkategorisierung als direkte Valenzkodierung

- Ansatz der **HPSG** (*Head-driven Phrase Structure Grammar*)
- Wert des SUBCAT-Merkmals ist eine **Liste der Argumente**, in deren Umgebung das Verb auftreten kann
- **kein PSG-Regelkern** mehr notwendig
- **Modellierung** syntaktischer Kategorien **durch komplexe Merkmalsstrukturen unterschiedlicher Spezifität**:
  - **Strukturinformation in Kategorien** statt in Regeln
  - **Idee der *Categorical Grammar***

- **Argument-Liste im SUBCAT-Merkmal:**

$$\left[ \begin{array}{cc} \text{CAT} & V \\ \text{AGR} & \boxed{1} \\ \text{SUBCAT} & < \left[ \begin{array}{cc} \text{CAT} & N \\ \text{CASE} & \text{NOM} \\ \text{AGR} & \boxed{1} \end{array} \right], \left[ \begin{array}{cc} \text{CAT} & N \\ \text{CASE} & \text{ACC} \end{array} \right] > \end{array} \right]$$

- **Alternativ für jeden Argumenttyp ein Merkmal:**

$$\left[ \begin{array}{cc} \text{CAT} & V \\ \text{AGR} & \boxed{1} \\ \text{SBJ} & \left[ \begin{array}{cc} \text{CAT} & N \\ \text{CASE} & \text{NOM} \\ \text{AGR} & \boxed{1} \end{array} \right] \\ \text{OBJ} & \left[ \begin{array}{cc} \text{CAT} & N \\ \text{CASE} & \text{ACC} \end{array} \right] \end{array} \right]$$

```
1  #V put = verbale Kategorie, die 3 Argumente  
   benötigt:  
2  V[SUBCAT=<NP, NP, PP>]  
3  
4  #VP = verbale Kategorie, die 1 Argument  
   benötigt (Subjekt-NP):  
5  V[SUBCAT=<NP>]  
6  
7  #SATZ = verbale Kategorie, die kein Argument  
   benötigt:  
8  V[SUBCAT=<>]
```

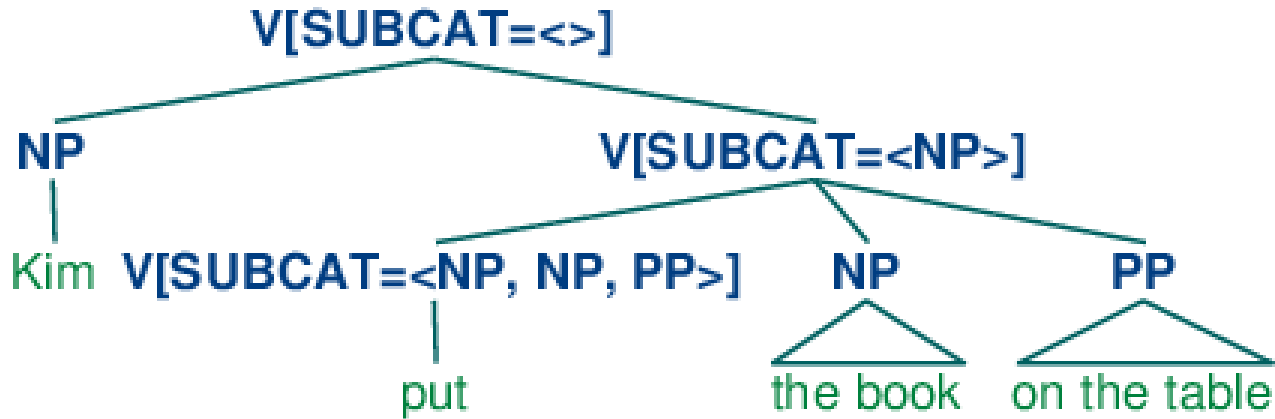


Abbildung 3: Subkategorisierung mit direkter Kodierung der Valenz  
([http://www.nltk.org/book/tree\\_images/ch09-tree-11.png](http://www.nltk.org/book/tree_images/ch09-tree-11.png))

## 7.2.3 Auxiliare und Inversion

- Beim **Entscheidungsfragesatz** vertauscht sich im Englischen die **Stellung von finitem Hilfsverb und Subjekt (= Inversion)**
- Modellierung durch **Zusatzregel mit invertierter Wortstellung** für Fragesatz mit (boolschem) *Inversionsmerkmal*
- **Auxiliar als Merkmal**

1 |  $S [+INV] \rightarrow V [+AUX] \ NP \ VP$

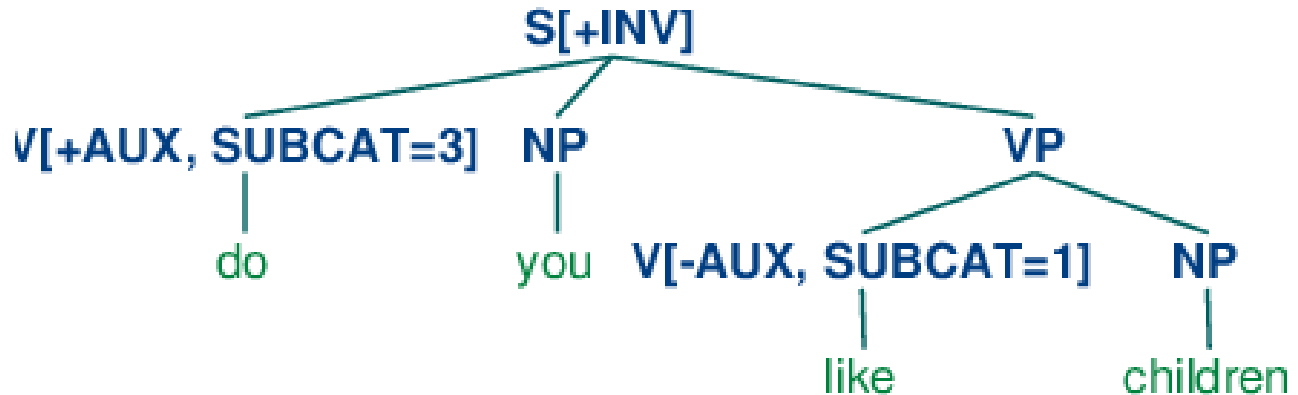


Abbildung 4: Auxiliare und Inversion  
([http://www.nltk.org/book/tree\\_images/ch09-tree-15.png](http://www.nltk.org/book/tree_images/ch09-tree-15.png))

- **SQ (Penn-Treebank):** *"Inverted yes/no question, or main clause of a wh-question, following the wh-phrase in SBARQ."*

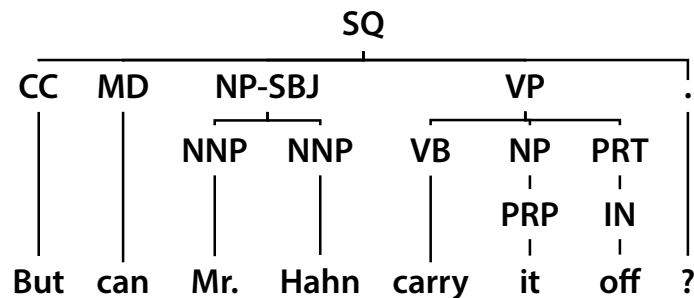


Abbildung 5: Entscheidungsfragesatz (SQ)

- **SINV (Penn-Treebank):** *"Inverted declarative sentence, i.e. one in which the subject follows the tensed verb or modal."*  
→ Beispiel: *Rarely see you Kim.*

## 7.2.4 *long distance dependencies*

- ***Wh-Extraction***: beim **Ergänzungsfragesatz** nach dem Objekt wird die **Objekt-NP aus der VP herausbewegt** und satzinitial gestellt: *Who do you like \_?*  
→ ***long distance dependency***
- An Ausgangspunkt im Syntaxbaum bleibt Leerstelle (***trace***) zurück
- in GPSG: Modellierung dieser ***diskontinuierlichen Strukturen*** durch ***Slash-Kategorien***



- Einführung einer **Satzkategorie mit NP-Lücke**: S/NP  
→ **Slash-Kategorie**: Satzkonstituente fehlt NP-Subkonstituente
- **Zusatzregel für Ergänzungsfragesätze** mit vorangestelltem Fragepronomen (*filler*):  $S \rightarrow NP \ S/NP$  ('*gap-introduction*')
- Slash-Kategorie kann als **Merkmal mit fehlender Kategorie als Wert** modelliert werden:  $S \ [SLASH=NP]$   
→ NLTK: Parser interpretiert S/NP entsprechend
- über **Variable** wird die **gap-Information** heruntergereicht bis  
 $NP/NP: S/?x \rightarrow AUX \ NP \ VP/?x; VP/x? \rightarrow V \ NP/?x$
- Realisierung der Lücke als **leeren String** über  $NP/NP \rightarrow \epsilon$

```
1  nltk.data.show_cfg('grammars/book_grammars/feat1.fcfg')
2  % start S
3  # #####
4  # Grammar Productions
5  # #####
6  S[-INV] → NP VP
7  S[-INV]/?x → NP VP/?x
8  S[-INV] -> NP S/NP
9  S[-INV] → Adv[+NEG] S[+INV]
10 S[+INV] → V[+AUX] NP VP
11 S[+INV]/?x -> V[+AUX] NP VP/?x
12 SBar → Comp S[-INV]
13 SBar/?x → Comp S[-INV]/?x
14 VP → V[SUBCAT=intrans, -AUX]
15 VP → V[SUBCAT=trans, -AUX] NP
16 VP/?x -> V[SUBCAT=trans, -AUX] NP/?x
```

```
17 VP → V[SUBCAT=clause, -AUX] SBar
18 VP/?x → V[SUBCAT=clause, -AUX] SBar/?x
19 VP → V[+AUX] VP
20 VP/?x → V[+AUX] VP/?x
21 # #####
22 # Lexical Productions
23 # #####
24 V[SUBCAT=intrans, -AUX] → 'walk' | 'sing'
25 V[SUBCAT=trans, -AUX] → 'see' | 'like'
26 V[SUBCAT=clause, -AUX] → 'say' | 'claim'
27 V[+AUX] → 'do' | 'can'
28 NP[-WH] → 'you' | 'cats'
29 NP[+WH] → 'who'
30 Adv[+NEG] → 'rarely' | 'never'
31 NP/NP ->
32 Comp → 'that'
```

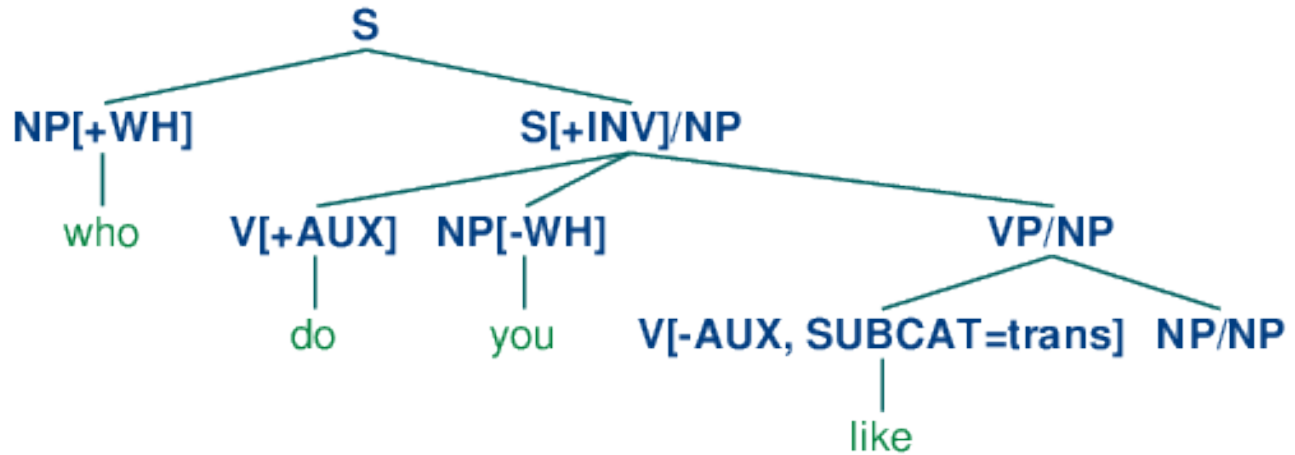


Abbildung 6: Modellierung Wh-Extraction mit Slash-Merkmal  
([http://www.nltk.org/book/tree\\_images/ch09-tree-16.png](http://www.nltk.org/book/tree_images/ch09-tree-16.png))

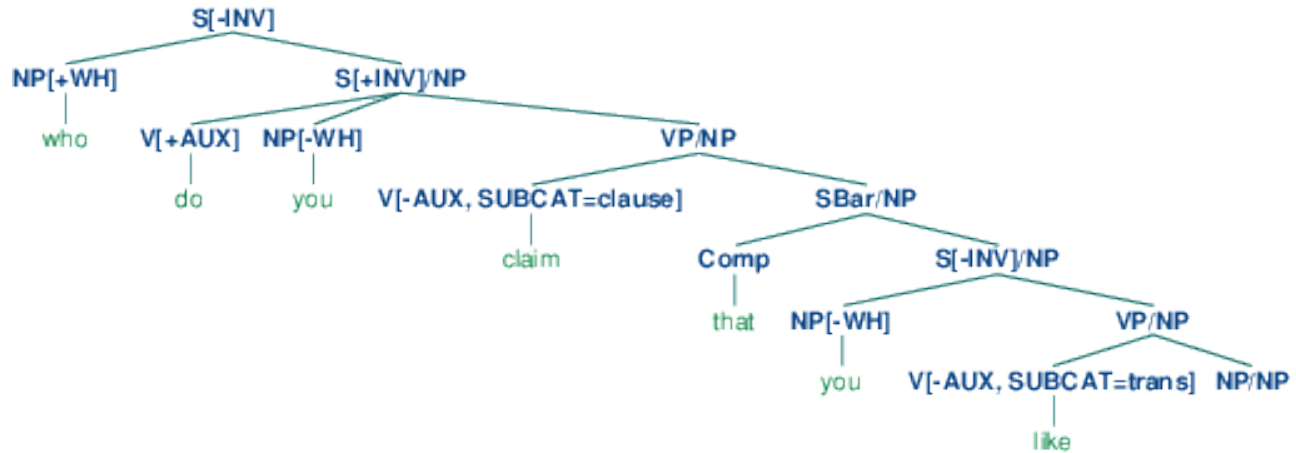


Abbildung 7: Syntaxbaum Wh-Extraction mit Slash-Merkmal  
 ([http://www.nltk.org/book/tree\\_images/ch09-tree-17.png](http://www.nltk.org/book/tree_images/ch09-tree-17.png))

- **SBARQ (Penn-Treebank):** *"Direct question introduced by a wh-word or a wh-phrase. Indirect questions and relative clauses should be bracketed as SBAR, not SBARQ"*

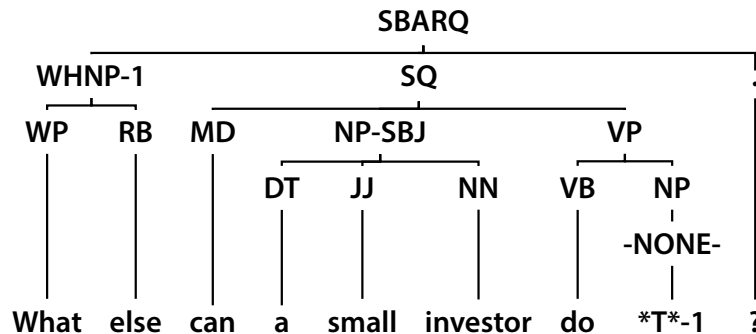


Abbildung 8: Penn-Treebank: Beispiel für long distance dependency durch Wh-Extraction; Beachte auch: Inversion in SQ