

Syntax natürlicher Sprachen

Vorlesung 7: Unifikationsgrammatiken

A. Wisiorek

Centrum für Informations- und Sprachverarbeitung,
Ludwig-Maximilians-Universität München

06.12.2022

Themen der heutigen Vorlesung

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - **Beispiel 1: Subkategorisierung in GPSG und HPSG**
 - **Beispiel 2: Kasus und Agreement im Deutschen**
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

1. Feature-Modellierungen für Subkategorisierung

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

Subkategorisierung

- **Unterteilung der Klasse der Verben nach Valenztypen:**
- **Anzahl und Art** verbaler Argumente (Komplemente)
- **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.*

Subkategorisierungsrahmen

- formale Repräsentation der syntaktischen Valenz eines Wortes

Beispiele Subkategorisierungsrahmen

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

Subkategorisierung mit kontextsensitiven Regeln

- **Berücksichtigung Kontext** zur Modellierung von Subkategorisierung unter Erhalt der Klasse V
 - **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 $V_{intrans}$ usw. als Subklasse von V wird nur **indirekt** über kontextsensitive Regel **sichtbar**

Kontextsensitive Regeln

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

- z.B. darf V mit rechtem Kontext NP nur zu Vtrans abgeleitet werden.

- **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
- verwendet u.a. in Feature-Grammar der GPSG (*Generalized Phrase Structure Grammar*)

CFG-Regeln mit SUBCAT-Feature

```
1 VP → V[SUBCAT=intrans]
2 VP → V[SUBCAT=trans] NP
3 VP → V[SUBCAT=prepobj] PP
4 VP → V[SUBCAT=ditrans] NP NP
5 VP → V[SUBCAT=place] NP PP
6 VP → V[SUBCAT=clause] S-BAR
7
8 V[SUBCAT=intrans] → bellt
9 V[SUBCAT=trans] → jagt
```

Subkategorisierung als direkte Valenzkodierung

- **Alternativ** können in einem **SUBCAT-Merkmal** auch **direkt die verlangten Komplementtypen** (als komplexes Merkmal) kodiert werden
- Grundidee der Categorical Grammar
- verwendet u.a. in Feature-Grammar der HPSG (*Head-driven Phrase Structure Grammar*)

bellt	CAT	V	
	AGR	TEMP	PRES
		NUM	SG
		PER	3
	SUBCAT	NONE	

jagt	CAT	V	
	AGR	TEMP	PRES
		NUM	SG
		PER	3
	SUBCAT	CAT	NP
		CASE	ACC

Constraintregeln für Subkategorisierung

- Regel für intransitive VP:

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

- Lexikoneintrag (intransitives Verb = ohne Komplement):

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

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

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

- **Constraintanweisung:** z.B. $\langle V \text{ SUBCAT} \rangle = \langle NP \rangle$ usw.
- **Lexikoneinträge (transitives Verb mit Akkusativ-Komplement):**

$$\begin{array}{l} \text{jagt} \end{array} \begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & PRES \\ \text{NUM} & SG \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \begin{bmatrix} \text{CAT} & NP \\ \text{CASE} & ACC \end{bmatrix} \end{bmatrix}, \text{ den Hund} \begin{bmatrix} \text{CAT} & NP \\ \text{CASE} & ACC \\ \text{AGR} & \begin{bmatrix} \text{NUM} & SG \\ \text{GEN} & MASK \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix}$$

- **Beispiel: Ablehnung intransitives Verb + NP (transitive VP-Regel)**
 → **Unifikation schlägt fehl**, inkompatiblen Werte im SUBCAT-Merkmal
 (Constraintanweisung: $\langle V \text{ SUBCAT} \rangle = \langle NP \rangle$):

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

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

2. Feature-Modellierung Agreement

- 1 Feature-Modellierungen für Subkategorisierung
- 2 **Feature-Modellierung Agreement**
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

Constraintregel für verbales Agreement und Subjekt-Kasus

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

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

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

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

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

- **Akzeptanz:**

→ *Unifikation gelingt (keine inkompatiblen Strukturen)*

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

- **Ablehnung:**

→ **Subjekt-Verb-Agreement-Constraint wird verletzt:** $\langle NP \ AGR \ NUM \rangle$

= $\langle VP \ AGR \ NUM \rangle$

→ (Subjekt-Kasus-Constraint erfüllt, $\langle NP \ CASE \rangle$ underspezifiziert)

die	$\begin{bmatrix} \text{CAT} & \text{DET} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \end{bmatrix} \end{bmatrix}$	Katze	$\begin{bmatrix} \text{CAT} & \text{N} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix}$	bellen	$\begin{bmatrix} \text{CAT} & \text{V} \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{PL} \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \text{NONE} \end{bmatrix}$
	$\begin{bmatrix} \text{CAT} & \text{NP} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{FEM} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix},$		$\begin{bmatrix} \text{CAT} & \text{VP} \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{PL} \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \text{NONE} \end{bmatrix}$		

- **Ablehnung:**

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

den	$\begin{bmatrix} \text{CAT} & \text{DET} \\ \text{CASE} & \text{ACC} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \end{bmatrix} \end{bmatrix}$	Hund	$\begin{bmatrix} \text{CAT} & N \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix}$	bellt	$\begin{bmatrix} \text{CAT} & V \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \text{NONE} \end{bmatrix}$
	$\begin{bmatrix} \text{CAT} & NP \\ \text{CASE} & \text{ACC} \\ \text{AGR} & \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEN} & \text{MASK} \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix},$		$\begin{bmatrix} \text{CAT} & VP \\ \text{AGR} & \begin{bmatrix} \text{TEMP} & \text{PRES} \\ \text{NUM} & \text{SG} \\ \text{PER} & 3 \end{bmatrix} \\ \text{SUBCAT} & \text{NONE} \end{bmatrix}$		

3. Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK**
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

3.1. Beispiel 1: Subkategorisierung in GPSG und HPSG

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - **Beispiel 1: Subkategorisierung in GPSG und HPSG**
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

NLTK-Kapitel zur Feature-Modellierung syntaktischer Phänomene des Englischen

- Grundlage = **NLTK 9.3.1**:
<https://www.nltk.org/book/ch09.html#subcategorization>
- 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>

- 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

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'
```

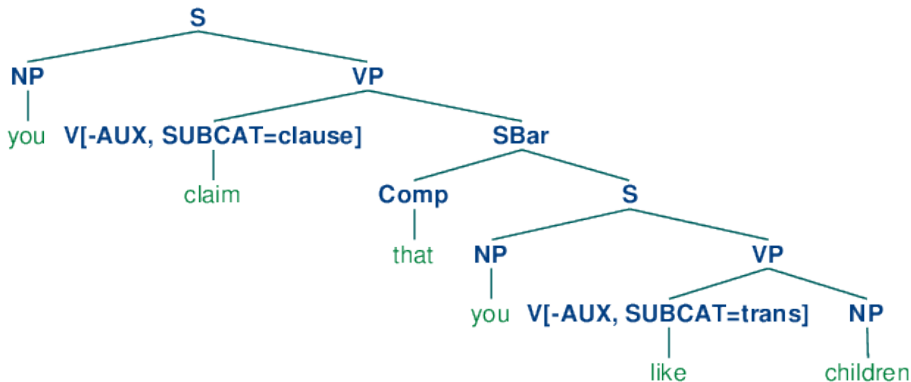



Abbildung: Subkategorisierung nach Komplementsatz
(http://www.nltk.org/book/tree_images/ch09-tree-10.png)

- 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:**

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

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

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

```
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=<>]
```

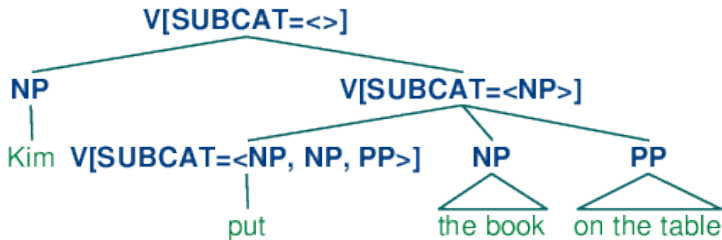


Abbildung: Subkategorisierung mit direkter Kodierung der Valenz
(http://www.nltk.org/book/tree_images/ch09-tree-11.png)

- VP als verbale Kategorie, die ein Subjekt-Argument benötigt
- Satz als verbale Kategorie, die keine weiteren Argumente fordert

3.2. Beispiel 2: Kasus und Agreement im Deutschen

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - **Beispiel 2: Kasus und Agreement im Deutschen**
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

- Grundlage = **NLTK Kapitel 9.3.5:**

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

- 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: <TV OBJCASE>=<NP CASE>*

Kasus-Rektion als Merkmalconstraint

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

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

$$\text{kommt} \begin{bmatrix} \text{CAT} & IV \\ \text{AGR} & \begin{bmatrix} \text{NUM} & SG \\ \text{PER} & 3 \end{bmatrix} \end{bmatrix}$$

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

Merkmalsstrukturbasierte Grammatik für das Deutsche (german.fcfg)

```
## 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'

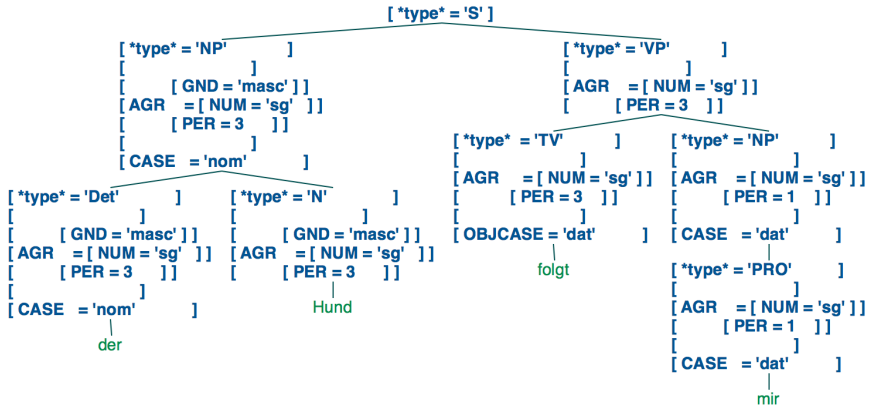


Abbildung: Syntaxbaum zu Ableitung der german.fcfg

4. Feature-Modellierung für Auxiliare und Inversion

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

Invertierte Wortstellung im Englischen

Entscheidungsfragesatz

- siehe auch **NLTK 9.3.3**: <https://www.nltk.org/book/ch09.html#auxiliary-verbs-and-inversion>

Inversion

- Beim **Entscheidungsfragesatz** vertauscht sich im Englischen die Stellung von finitem Hilfsverb (AUX) und Subjekt-NP

Feature-Modellierung

- **Zusatzregel mit invertierter Wortstellung** für Fragesatz:
 - (boolsches) *Inversionsmerkmal*: [+/-INV]
 - (boolsches) *Auxiliarmerkmal*: [+/-AUX]

Inverted-Clause-Regeln

S[+INV] -> V[+AUX] NP VP[-AUX]
VP[-AUX] -> V[-AUX] NP

Beispiel-Parsebaum Inversion

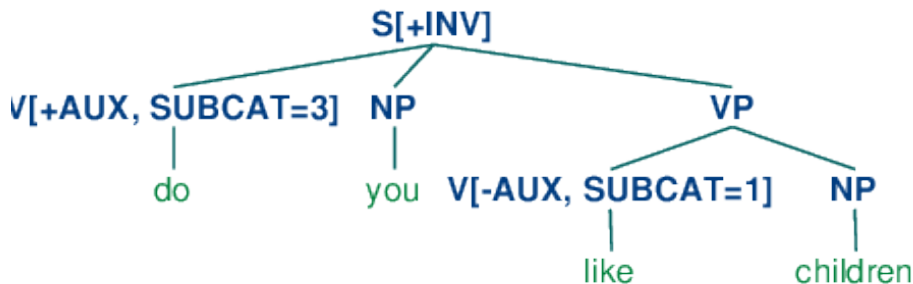
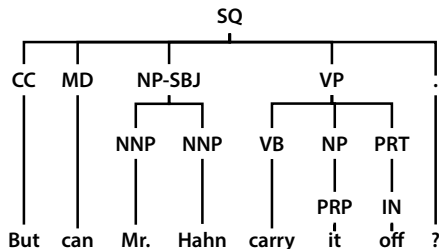


Abbildung: Auxiliare und Inversion

(http://www.nltk.org/book/tree_images/ch09-tree-15.png)

Inversion in der Penn-Treebank

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



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

5. Gap-Feature für *Long Distance Dependencies*

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*

Wh-Movement im Englischen (auch Wh-Extraction)

- **NLTK 9.3.4:** <https://www.nltk.org/book/ch09.html#unbounded-dependency-constructions>

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** oder **Unbounded Dependency**
- An Ausgangspunkt im Syntaxbaum bleibt Leerstelle (*trace*) zurück

Feature-Modellierung

- in GPSG: Modellierung durch *Slash-Kategorien*
 - VP/NP = 'VP ohne NP'; entspricht VP mit extrahierter Objekt-NP
 - NP/NP kann mit entsprechender Regel als leerer String realisiert werden (bzw. als trace-Element)
- Slash-Kategorien (Leerstellen) können als Feature in einer FCFG umgesetzt werden

Modellierung Movement im Ergänzungsfragesatz mit Gap-Feature (feat1.fcfg)

- 1 Einführung einer **Satzkategorie mit NP-Lücke**: S/NP
→ **Slash-Kategorie**: Satzkonstituente fehlt NP-Subkonstituente
- 2 **Zusatzregel für Ergänzungsfragesätze** mit vorangestelltem Fragepronomen (*filler*): $S \rightarrow NP \ S/NP$ ('gap-introduction')
- 3 Slash-Kategorie kann als **Merkmal mit fehlender Kategorie als Wert** modelliert werden: $S [SLASH=NP]$
→ *NLTK*: Parser interpretiert S/NP entsprechend
- 4 über **Variable** wird die **gap-Information heruntergereicht** bis NP/NP :
 $S/?x \rightarrow AUX \ NP \ VP/?x; VP/x? \rightarrow V \ NP/?x$
- 5 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 # Grammar Productions:
4 S[-INV] -> NP VP
5 S[-INV]/?x -> NP VP/?x
6 S[-INV] -> NP S/NP
7 S[-INV] -> Adv[+NEG] S[+INV]
8 S[+INV] -> V[+AUX] NP VP
9 S[+INV]/?x -> V[+AUX] NP VP/?x
10 SBar -> Comp S[-INV]
11 SBar/?x -> Comp S[-INV]/?x
12 VP -> V[SUBCAT=intrans, -AUX]
13 VP -> V[SUBCAT=trans, -AUX] NP
14 VP/?x -> V[SUBCAT=trans, -AUX] NP/?x
15 VP -> V[SUBCAT=clause, -AUX] SBar
16 VP/?x -> V[SUBCAT=clause, -AUX] SBar/?x
17 VP -> V[+AUX] VP
18 VP/?x -> V[+AUX] VP/?x
19
20

```

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

Modellierung Wh-Extraction mit Slash-Merkmal

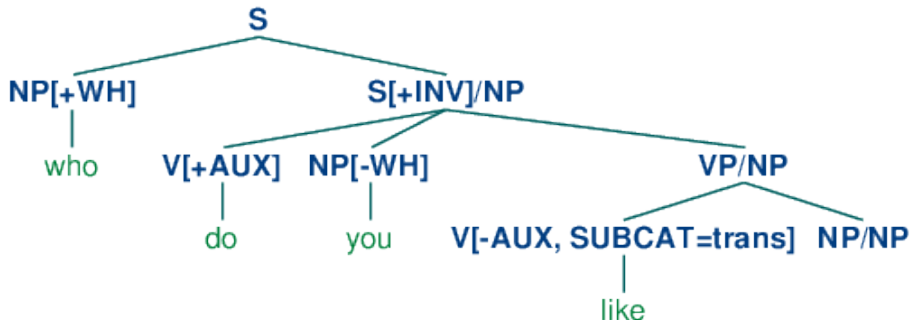


Abbildung: (http://www.nltk.org/book/tree_images/ch09-tree-16.png)

Extrahiertes Element kann beliebig tief rekursiv eingebettet sein (*Unbounded Dependency*)

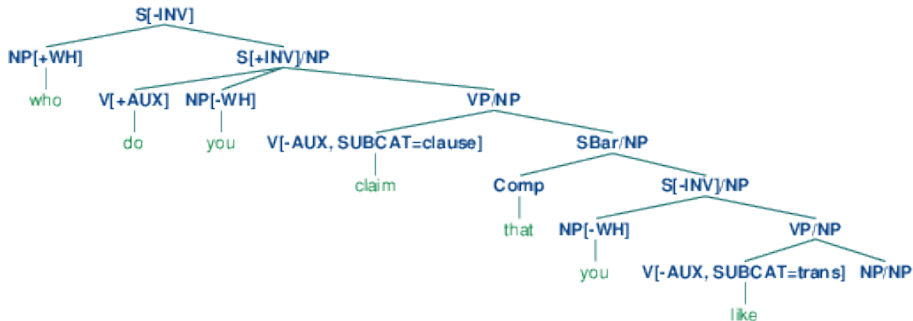


Abbildung: (http://www.nltk.org/book/tree_images/ch09-tree-17.png)

Wh-Extraction in Penn-Treebank

- Modellierung über entsprechende CFG-Kategorien mit *trace* (*T*)
- **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*"

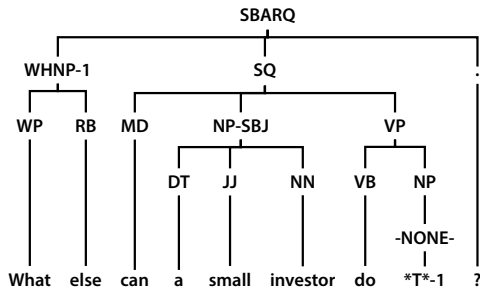


Abbildung: Penn-Treebank: Beispiel für *long distance dependency* durch *Wh-Extraction*; Beachte auch: Inversion in *SQ*

Rückblick auf heutige Themen

- 1 Feature-Modellierungen für Subkategorisierung
- 2 Feature-Modellierung Agreement
- 3 Beispiel-Modellierung mit Unifikationsgrammatiken im NLTK
 - Beispiel 1: Subkategorisierung in GPSG und HPSG
 - Beispiel 2: Kasus und Agreement im Deutschen
- 4 Feature-Modellierung für Auxiliare und Inversion
- 5 Gap-Feature für *Long Distance Dependencies*