

Einführung in die Computerlinguistik

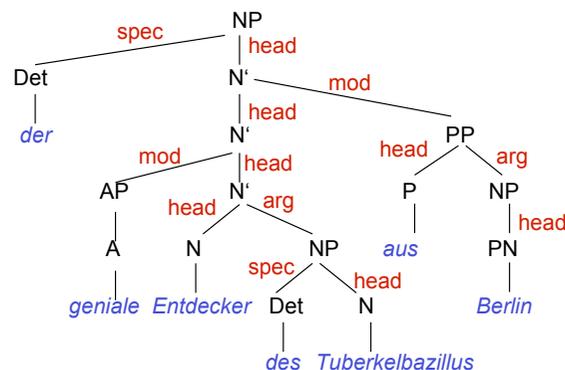
Syntax: Grammatische Funktionen / Grammatische Merkmale

WS 2008/2009

Manfred Pinkal

Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Grammatische Funktionen



Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Kategorie und Funktion

- **Syntaktische Kategorien** bezeichnen Klassen Ausdrücken mit ähnlicher innerer Struktur und ähnlichem distributionellem Verhalten.
- **Grammatische Funktionen** dagegen bezeichnen die Rolle, die eine Konstituente im größeren Ausdruck spielt. Eine NP kann, je nach Stellung im Satz unter anderem die Funktion von **Subjekt** oder (direktem oder indirektem) **Objekt** eines Satzes, (Genitiv-) **Attribut** einer anderen NP oder **Argument** einer Präpositionalphrase bilden. - Grammatische Funktionen sind relationale Konzepte!
- Unterschiedliche Kategorien können die gleiche Funktion ausüben: Subjekte können zum Beispiel Nominalphrasen oder Sätze sein: *Dass das Wetter schön ist, gefällt mir* - *Das schöne Wetter gefällt mir*.

Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Grammatische Funktionen

- **Köpfe** („heads“) sind die Kernbestandteile einer Konstituente. Die Merkmale des „lexikalischen Kopfes“ werden über die „Kopflinie“ nach oben zur Phrase weitergeleitet.
- **Argumente** („arg“) werden durch lexikalische Köpfe „subkategorisiert“ oder „regiert“: Ein lexikalischer Ausdruck (V, N, A, P) kann ein oder mehrere Argumente mit bestimmten grammatischen Eigenschaften verlangen, z.B. Subjekt, Objekt im Akkusativ, Präpositionales Objekt etc.
- **Modifikatoren** („mod“) sind freie, optionale Ergänzungen, die einen Ausdruck erweitern, ohne seine Kategorie zu verändern. Nominale Modifikatoren werden **Attribute** genannt (Genitivattribut, präpositionales Attribut), Satzmodifikatoren **Adjunkte** (auch „adverbiale Bestimmungen“).

Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Grammatische Merkmale

- Wie finden Sie **die** angehängten Bilder? Das **sind** Fotos, **die** im Rahmen des TALK-Projektes entstanden **sind**, uns gehören, und von BMW schon freigegeben **waren**. Außerdem vermitteln **sie** besser den Bezug zur Forschung.

Merkmalsabhängigkeiten

- Grammatische Merkmale von Ausdrücken in der syntaktischen Struktur hängen in systematischer Weise voneinander ab.
- Die grundlegenden Typen solcher Beziehungen sind
 - Kongruenz und
 - Rektion oder Subkategorisierung

Kongruenz

- Kongruenz ist die Übereinstimmung von zwei oder mehreren Ausdrücken in Genus, Numerus, Kasus, Person, ...
 - Nominalkongruenz innerhalb der NP zwischen Artikel, Nomen, Adjektiv, Relativpronomen: **die[pl] angehängten[pl] Bilder[pl]**
 - Subjekt-Verb-Kongruenz: **sie[pl] vermitteln[pl]**
 - Pronominalkongruenz zwischen einem „anaphorischen“ Pronomen und der NP, auf die er sich bezieht
Fotos[pl] ... **sie[pl]**

Subkategorisierung/ Rektion

- Von Rektion oder Subkategorisierung spricht man, wenn ein lexikalischer Kopf Argumente mit bestimmten grammatischen Eigenschaften verlangt.
- Subkategorisierung/ Rektion von Verben
 - Sie **vermitteln** den Bezug **[NP im Akkusativ]**
 - Die Bilder **gefallen** dem Betrachter **[NP im Dativ]**
 - Sie **erinnern** uns **[NP im Akkusativ]** an den Urlaub **[PP mit Akkusativ]**
- Präpositionen
 - **um** das Haus
 - **bei** dem Haus
 - **wegen** des Hauses
- Adjektive
 - **an** computerlinguistischen Fragestellungen interessiert
 - **einem** Baum ähnlich

Grammatische Merkmale in der CFG

- Beispielgrammatik:

$S \rightarrow NP VP$ $VP \rightarrow VT NP$
 $VP \rightarrow VI$ $NP \rightarrow DET N$

$VI \rightarrow schl\ddot{a}ft \mid arbeitet$
 $VT \rightarrow kennt \mid studiert$
 $N \rightarrow Student \mid Studentin \mid Studenten \mid Studentinnen \mid Fach$
 $DET \rightarrow der \mid die \mid das \mid den$

- Massive Übergenerierung ohne Berücksichtigung von Kongruenz und Rektion:

- *die Studenten arbeitet*
- *der Studentin arbeiten*
- *der Student kennt der Student*

- Wie können Merkmale in der CFG berücksichtigt werden?

Explizite Kodierung von Merkmalen

$S \rightarrow NP \begin{bmatrix} num & sg \\ kas & nom \end{bmatrix} VP[num \ sg]$ $S \rightarrow NP \begin{bmatrix} num & pl \\ kas & nom \end{bmatrix} VP[num \ pl]$

$VP[num \ sg] \rightarrow VI[num \ sg]$ $VP[num \ pl] \rightarrow VI[num \ pl]$

$NP \begin{bmatrix} num & sg \\ kas & nom \\ gen & m \end{bmatrix} \rightarrow Det \begin{bmatrix} kas & nom \\ gen & m \end{bmatrix} N \begin{bmatrix} num & sg \\ kas & nom \\ gen & m \end{bmatrix}$

$NP \begin{bmatrix} num & pl \\ kas & nom \\ gen & m \end{bmatrix} \rightarrow Det \begin{bmatrix} kas & nom \\ gen & m \end{bmatrix} N \begin{bmatrix} num & pl \\ kas & nom \\ gen & m \end{bmatrix}$

Versuch: Verfeinerung der Kategorien

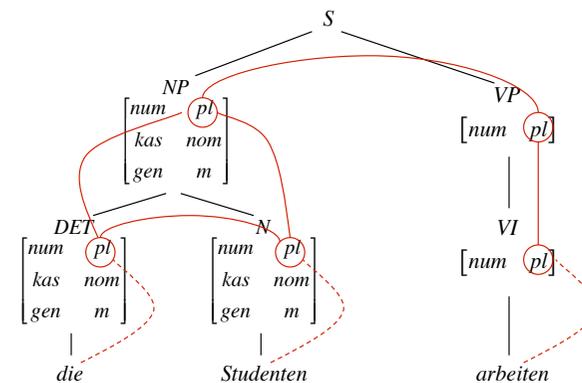
- Beispielgrammatik:

$S \rightarrow NPSgNom VPSg$	$S \rightarrow NPPINom VPPI$
$VPSg \rightarrow VISg$	$VPPI \rightarrow VIPI$
$VPSg \rightarrow VTSg NPAkk$	$VPPI \rightarrow VTPI NPAkk$
$NPSgNom \rightarrow DETSgNomM NSgNomM$	$NPSgNom \rightarrow DETSgNomF NSgNomF$
$NPPINom \rightarrow DETPINom NPINom$	
$DETSgNomM \rightarrow der$	$DETSgNomF \rightarrow die$
$NSgNomM \rightarrow Student$	$NSgNomF \rightarrow Studentin \quad etc.$

- Nachteile:

- Information aus dem Lexikon (Kategorie und inhärente Merkmale) und aus der morphologischen Analyse (Variable Merkmale) müssen zu neuen Kategorien fusioniert werden.
- Regularitäten können nicht ausgedrückt werden
- Das Regelsystem wird aufgebläht (2 Numeri x 3 Genera x 4 Kasus x 3 Personen x ...)

Explizite Kodierung von Merkmalen



Kontextfreie Grammatik mit Merkmalsstrukturen

- Konstituenten werden mit Paaren aus Kategoriensymbolen und Merkmalsstrukturen ausgezeichnet.
- Eine Merkmalsstruktur ist eine Menge von Merkmal-Wert-Paaren (auch „Attribut-Wert-Paaren“): Die Merkmalsstruktur des NP-Knotens im Beispiel hat drei Merkmale, das erste besteht aus dem Attribut „num“ und dem atomaren Wert „sg“.
- Die explizite Kodierung von Merkmalen erlaubt die Formulierung von Bedingungen / Constraints, z.B. „Numerus von NP und Numerus von VP sind identisch“, oder „Subjekts-NP hat Kasus Nominativ“.
- Regeln der Grammatik sind zweiteilig: Sie bestehen aus einer Ersetzungsregel (wie üblich über Kategorien und lexikalische Ausdrücke formuliert) und einer Menge von Constraints über Merkmalsstrukturen.

CFG mit Merkmalsconstraints, Beispiel

<p>S → NP VP <i>Numerus der NP = Numerus der VP</i> <i>Kasus der NP = nom</i></p> <p>VP → VI <i>Numerus der VP = Numerus von VI</i></p> <p>VP → VT NP <i>Numerus der VP = Numerus von VT</i> <i>Kasus der NP = akk</i></p> <p>NP → DET N <i>Numerus von DET = Numerus von N</i> <i>Genus von DET = Genus von N</i> <i>Kasus von DET = Kasus von N</i> <i>Numerus der NP = Numerus von N</i> <i>Genus der NP = Genus von N</i> <i>Kasus der NP = Kasus von N</i></p>	<p>VI → <i>arbeitet</i> <i>Numerus von VI = sg</i></p> <p>VI → <i>arbeiten</i> <i>Numerus von VI = pl</i></p> <p>N → <i>Student</i> <i>Numerus von N = sg</i> <i>Genus von N = m</i> <i>Kasus von N = nom</i></p> <p>DET → <i>der</i> <i>Numerus von DET = sg</i> <i>Genus von DET = m</i> <i>Kasus von DET = nom</i></p>
--	---

Merkmalstrukturen: Erste Erweiterung

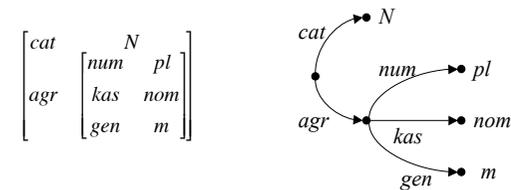
Die Notation von Constraints lässt sich stark vereinfachen, wenn gleichzeitig auf Mengen von Merkmalen Bezug nehmen kann. Wir erlauben komplexe Merkmalsstrukturen, in denen Attribute nicht nur atomare Werte, sondern auch Merkmalsstrukturen als Werte haben können. Beispiel:

$$\left[\begin{array}{c} agr \\ \end{array} \left[\begin{array}{cc} num & pl \\ kas & nom \\ gen & m \end{array} \right] \right]$$

„agr“ für englisch „agreement“ (Kongruenz) nimmt als Wert eine Merkmalsstruktur, die die Kongruenzmerkmale spezifiziert. Wir können

Statt der Aufzählung einzelner Kongruenzmerkmale in der NP-Regel können wir formulieren
 „Kongruenzmerkmale von DET = Kongruenzmerkmale von N“
 „Kongruenzmerkmale von NP = Kongruenzmerkmale von N“

Was sind Merkmalsstrukturen eigentlich?

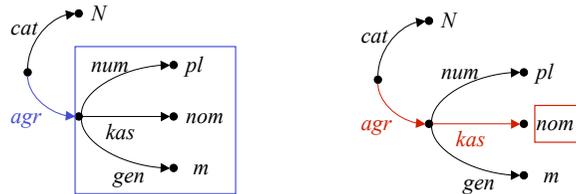


Merkmalsstrukturen sind formal als gerichtete azyklische Graphen mit Kanteninschriften darstellbar.

Merkmalsstrukturen, die wir bisher betrachten, sind (ungeordnete) Bäume mit einer Wurzel. Die Kanteninschriften sind Attribut-Label. Die Blätter sind mit atomaren Werten gelabelt.

Merkmalspfade sind Folgen von Kantenlabeln. Wir schreiben sie in der Form <agr> bzw. <agr kas> und können

Merkmalspfade, Pfadgleichungen



Merkmalspfade sind Folgen von Kantenlabeln. Wir schreiben sie in der Form $\langle \text{agr} \rangle$ bzw. $\langle \text{agr kas} \rangle$ und beziehen uns mit ihnen eindeutig auf einen Knoten im Graph und den mit dem Knoten verbundenen Wert.

CFG mit Merkmalsconstraints, Beispiel

$S \rightarrow NP VP$
 $\langle NP \text{ agr num} \rangle = \langle VP \text{ agr num} \rangle$
 $\langle NP \text{ agr kas} \rangle = \text{nom}$

$VP \rightarrow VI$
 $\langle VP \text{ agr num} \rangle = \langle VI \text{ agr num} \rangle$

$VP \rightarrow VT NP$
 $\langle VP \text{ agr num} \rangle = \langle VT \text{ agr num} \rangle$
 $\langle NP \text{ agr kas} \rangle = \text{akk}$

$NP \rightarrow DET N$
 $\langle DET \text{ agr} \rangle = \langle N \text{ agr} \rangle$
 $\langle NP \text{ agr} \rangle = \langle N \text{ agr} \rangle$

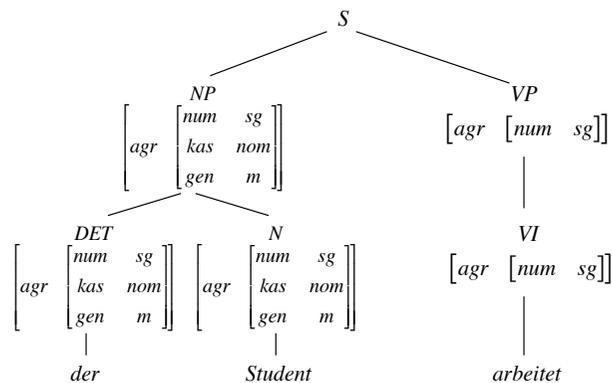
$VI \rightarrow \text{arbeitet}$
 $\langle VI \text{ agr num} \rangle = \text{sg}$

$VI \rightarrow \text{arbeiten}$
 $\langle VI \text{ agr num} \rangle = \text{pl}$

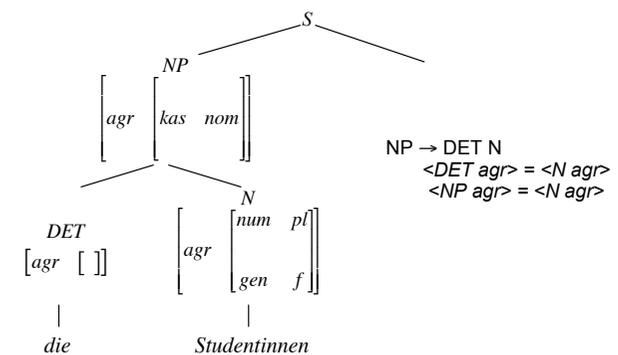
$N \rightarrow \text{Student}$
 $\langle N \text{ agr num} \rangle = \text{sg}$
 $\langle N \text{ agr gen} \rangle = \text{m}$
 $\langle N \text{ agr kas} \rangle = \text{nom}$

$DET \rightarrow \text{der}$
 $\langle DET \text{ agr num} \rangle = \text{sg}$
 $\langle DET \text{ agr gen} \rangle = \text{m}$
 $\langle DET \text{ agr kas} \rangle = \text{nom}$

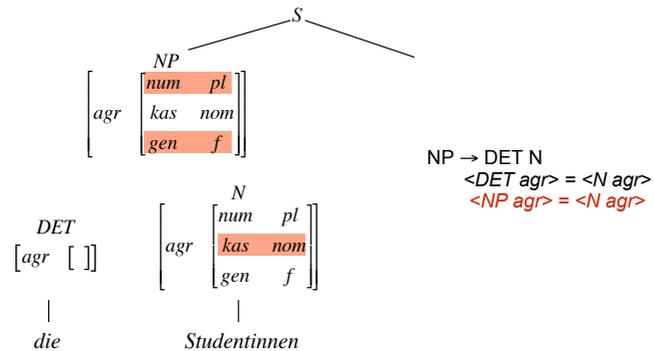
Die Anwendung von Merkmals-Constraints



Die Anwendung von Merkmals-Constraints

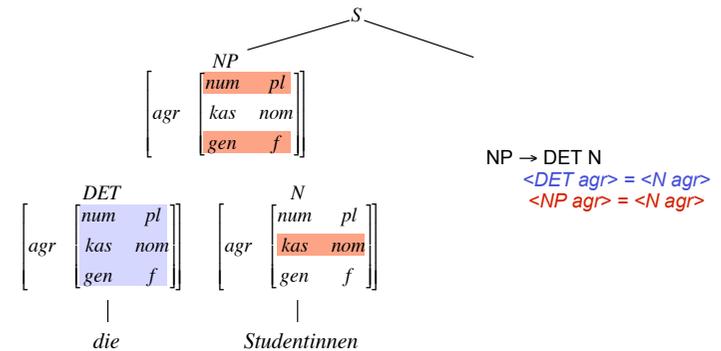


Die Anwendung von Merkmals-Constraints



Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Die Anwendung von Merkmalsconstraints



Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Unifikation

- Bei der Anwendung von Pfadgleichungen wird nicht die Identität aktueller Werte abgeprüft, sondern
 - die Kompatibilität der vorhandenen Merkmalsinformation getestet, indem
 - die Information auf beiden Seiten der Gleichung zusammengeführt.
- Dies geschieht, indem die entsprechenden Merkmalsinformationen **unifiziert** werden. Wir schreiben $A \sqcup B$.
- Resultat der Unifikation ist die (allgemeinste) Merkmalsstruktur, die die Information aus beiden Merkmalsstrukturen umfasst, wenn es eine solche Struktur gibt. Ansonsten schlägt sie fehl.
- Wenn Merkmalsstruktur B die Information aus Merkmalsstruktur A enthält, d.h., wenn A allgemeiner ist als B, sprechen wir von Subsumption: A subsumiert B, oder $A \sqsubseteq B$.
- Das Resultat der Unifikation von A und B: $A \sqcup B$ ist die allgemeinste Struktur C, sodass $A \sqsubseteq C$ und $B \sqsubseteq C$.

Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Unifikation

- Wie berechnet sich das Resultat der Unifikation aus A und B?
- Fall 1: Beide Seiten sind atomar:
 - Unifikation ergibt $A (=B)$, falls $A=B$
 - Unifikation schlägt fehl, wenn $A \neq B$
- Fall 2: Eine Seite ist atomar, die andere komplexe Merkmalsstruktur
 - Unifikation schlägt fehl
- Fall 3: Beide Seiten sind komplexe Merkmalstrukturen:
 - Unifikation ergibt eine Merkmalstruktur, deren Attribute die Vereinigung der Attribute von A und B sind.
 - Der Wert für jedes Attribut ist
 - der Wert des Attributes in A (B), wenn das Attribut nur in A (B) instantiiert ist
 - die Unifikation der Werte, die das Attribut in A und B hat, wenn es in beiden vorkommt.

Vorlesung "Einführung in die CL" 2008/2009 © M. Pinkal UdS Computerlinguistik

Unifikation, Beispiele

$sg \cup sg = sg$

$sg \cup pl = fail$

$$\begin{bmatrix} cat \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & f \end{smallmatrix} \right] \end{bmatrix} \cup \begin{bmatrix} agr \\ \quad [kas \quad akk] \end{bmatrix} = \begin{bmatrix} cat \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & f \end{smallmatrix} \right] \\ \quad \left[\begin{smallmatrix} kas & \\ & akk \end{smallmatrix} \right] \end{bmatrix}$$

$$\begin{bmatrix} cat \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & f \end{smallmatrix} \right] \end{bmatrix} \cup \begin{bmatrix} agr \\ \quad [gen \quad m] \end{bmatrix} = fail$$

Kopfmerkmale und Subkategorisierung

$$\begin{array}{l} \text{wählt} \begin{bmatrix} cat \\ head \\ \quad \left[\begin{smallmatrix} vform & V & finit \\ agr & \left[\begin{smallmatrix} num & sg \\ per & 3 \end{smallmatrix} \right] \end{smallmatrix} \right] \\ subcat \\ \quad \left[\begin{smallmatrix} cat & np \\ kas & akk \end{smallmatrix} \right] \end{bmatrix} \\ \\ \text{das} \begin{bmatrix} cat \\ def \\ agr \\ \quad \left[\begin{smallmatrix} num & sg \\ gen & n \end{smallmatrix} \right] \end{bmatrix} \quad \text{Fach} \begin{bmatrix} cat \\ head \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & n \end{smallmatrix} \right] \end{bmatrix} \\ \\ \text{eine} \begin{bmatrix} cat \\ def \\ agr \\ \quad \left[\begin{smallmatrix} num & sg \\ gen & f \end{smallmatrix} \right] \end{bmatrix} \quad \text{Studentin} \begin{bmatrix} cat \\ head \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & f \end{smallmatrix} \right] \end{bmatrix} \end{array}$$

$S \rightarrow NP VP$
 $\langle S head \rangle = \langle VP head \rangle$
 $\langle NP head agr \rangle = \langle VP head agr \rangle$

 $VP \rightarrow V NP$
 $\langle VP head \rangle = \langle V head \rangle$
 $\langle V subcat \rangle = NP$
 $\langle V subcat kas \rangle = \langle NP head agr kas \rangle$

 $NP \rightarrow DET N$
 $\langle NP cat \rangle = NP$
 $\langle NP per \rangle = 3$
 $\langle NP head \rangle = \langle N head \rangle$
 $\langle DET agr \rangle = \langle NP head agr \rangle$
 $\langle DET def \rangle = \langle NP head def \rangle$

Semantik - Erster Ansatz

$$\begin{array}{l} \text{wählt} \begin{bmatrix} cat \\ head \\ \quad \left[\begin{smallmatrix} vform & V & finit \\ agr & \left[\begin{smallmatrix} num & sg \\ per & 3 \end{smallmatrix} \right] \\ sem & [pred \quad wählen] \end{smallmatrix} \right] \\ subcat \\ \quad \left[\begin{smallmatrix} cat & np \\ kas & akk \end{smallmatrix} \right] \end{bmatrix} \\ \\ \text{das} \begin{bmatrix} cat \\ def \\ agr \\ \quad \left[\begin{smallmatrix} num & sg \\ gen & n \end{smallmatrix} \right] \end{bmatrix} \quad \text{Fach} \begin{bmatrix} cat \\ head \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & n \end{smallmatrix} \right] \\ sem \\ \quad [fach'] \end{bmatrix} \\ \\ \text{eine} \begin{bmatrix} cat \\ def \\ agr \\ \quad \left[\begin{smallmatrix} num & sg \\ gen & f \end{smallmatrix} \right] \end{bmatrix} \quad \text{Studentin} \begin{bmatrix} cat \\ head \\ agr \\ \quad \left[\begin{smallmatrix} num & N \\ gen & f \end{smallmatrix} \right] \\ sem \\ \quad [studentin'] \end{bmatrix} \end{array}$$

$S \rightarrow NP VP$
 $\langle S head \rangle = \langle VP head \rangle$
 $\langle NP head agr \rangle = \langle VP head agr \rangle$
 $\langle VP head sem arg1 \rangle = \langle NP head sem \rangle$

 $VP \rightarrow V NP$
 $\langle VP head \rangle = \langle V head \rangle$
 $\langle V subcat \rangle = NP$
 $\langle V subcat kas \rangle = \langle NP head agr kas \rangle$
 $\langle V head sem arg2 \rangle = \langle NP head sem \rangle$

 $NP \rightarrow DET N$
 $\langle NP cat \rangle = NP$
 $\langle NP per \rangle = 3$
 $\langle NP head \rangle = \langle N head \rangle$
 $\langle DET agr \rangle = \langle NP head agr \rangle$
 $\langle DET def \rangle = \langle NP head def \rangle$