

Einführung in die Computerlinguistik

Syntax: Merkmalsstrukturen

WS 2012/2013

Manfred Pinkal

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Eigenschaften der syntaktischen Struktur [1]

- *Er hat die Übungen gemacht.*
- *Der Student hat die Übungen gemacht.*
- *Der interessierte Student hat die Übungen gemacht.*
- *Der an computerlinguistischen Fragestellungen interessierte Student hat die Übungen gemacht.*
- *Der an computerlinguistischen Fragestellungen interessierte Student im ersten Semester hat die Übungen gemacht.*
- *Der an computerlinguistischen Fragestellungen interessierte Student im ersten Semester, der im Hauptfach Informatik studiert, hat die Übungen gemacht.*
- *Der an computerlinguistischen Fragestellungen interessierte Student im ersten Semester, der im Hauptfach, für das er sich nach langer Überlegung entschieden hat, Informatik studiert, hat die Übungen gemacht.*

Vorlesung "Einführung in die CL" 2012/13 © 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.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Einige grammatische Merkmale

- Bei Substantiven:
 - Numerus, Genus, Kasus
- Bei Adjektiven:
 - Numerus, Genus, Kasus, Steigerungsform
- Bei Pronomen:
 - Numerus, Genus, Kasus, Person
- Bei Verben:
 - Numerus, Person
 - Tempus
 - Modus (Indikativ, Konjunktiv, Imperativ)
 - Genus verbi (engl. „voice“: Aktiv, Passiv)

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Status von Merkmalen

- Inhärente Merkmale (Genus beim Substantiv)
- Unabhängige Merkmale (Numerus beim Substantiv, Tempus beim Verb, Steigerungsform beim Adjektiv)
- Abhängige Merkmale (Numerus und Genus beim Adjektiv, Kasus beim Substantiv)

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

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

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[p]* *angehängten[p]* *Bilder[p]*
 - **Subjekt-Verb-Kongruenz**: *sie[p]* *vermitteln[p]*
 - **Pronominalkongruenz** zwischen einem „anaphorischen“ Pronomen und der NP, auf die er sich bezieht
Fotos[p] ... *sie[p]*

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

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*
 - *seinen Freunden verpflichtet*

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Grammatische Merkmale in der CFG

- Die Standard-CFG hat zunächst keinen Mechanismus zur Modellierung grammatischer Merkmale und ihrer Abhängigkeiten.
- Nicht-Berücksichtigung von Kongruenz und Rektion führt zu massiver Übergenerierung:
- Beispielgrammatik 1:
 $S \rightarrow NP VP$ $VP \rightarrow VT NP$
 $VP \rightarrow VI$ $NP \rightarrow DET N$
 $VI \rightarrow schläft | arbeitet$
 $VT \rightarrow kennt | studiert$
 $N \rightarrow Student | Studentin | Studenten | Studentinnen | Fach$
 $DET \rightarrow der | die | das | den$
- *Die Studenten arbeitet – Der Student kennt der Student*

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Versuch: Verfeinerung der Kategorien

- Beispielgrammatik 2:

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

- Integration von Merkmalsinformation in Kategoriensymbole – Nachteile:
 - Das Regelsystem wird aufgebläht (2 Numeri x 3 Genera x 4 Kasus x 3 Personen x ...)
 - Regularitäten können nicht ausgedrückt werden

Merkmalsstrukturen

- Wir kodieren Merkmale explizit in Form von **Merkmalsstrukturen**.

Statt "NPSgNomM" schreiben wir $NP \begin{bmatrix} \text{num} & \text{sg} \\ \text{kas} & \text{nom} \\ \text{gen} & \text{m} \end{bmatrix}$

- Die Beschreibung einer Konstituente ist ein Paar, bestehend aus einem Kategoriensymbol und einer Merkmalsstruktur.
- Die Merkmalsstruktur besteht aus „Attribut-Wert-Paaren“: Die obige Merkmalsstruktur hat drei Merkmale, das erste Merkmal $[\text{num} \text{sg}]$ besteht aus dem Attribut „num“ und dem Wert „pl“.

Grammatik mit Merkmalsstrukturen

- Beispielgrammatik 3:

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

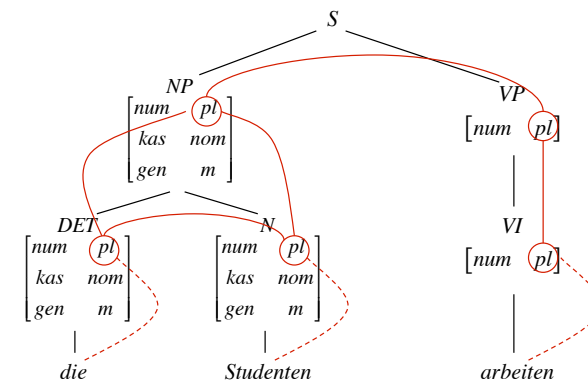
$VP[\text{num} \text{sg}] \rightarrow VI[\text{num} \text{sg}]$ $VP[\text{num} \text{pl}] \rightarrow VI[\text{num} \text{pl}]$

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

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

... ..

Direkte Ableitung von Merkmalsstrukturen



Merkmalsconstraints

- Problem: Bei der direkten Erzeugung von Ableitungsbäumen mit Merkmalsstrukturen wird das Regelsystem nicht kleiner: Regularitäten sind besser sichtbar, werden aber nicht genutzt.
- Sinnvoller ist es, Merkmale zur Formulierung von Bedingungen oder "Constraints" zu verwenden, die grammatisch zulässige Verteilungen von Merkmalsstrukturen in Ableitungsbäumen spezifizieren.
- Grammatikregeln werden erweitert: Jede Regel besteht aus
 - einer Ersetzungsregel (wie gehabt), und
 - einer Menge von Constraints, die zulässige Merkmalskonfigurationen spezifizieren.
- Beispiel:

$$S \rightarrow NP VP$$

Numerus der NP muss gleich Numerus der VP sein
Kasus der NP ist Nominativ

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Formulierung von Merkmalsconstraints

- Wie können wir einen Constraint über Merkmalsstrukturen wie "Numerus der NP muss gleich Numerus der VP sein" formal repräsentieren?
- Was sind Merkmalsstrukturen eigentlich?
- Antwort 1: (Partielle) Funktionen, die Attribute in Werte abbilden

$$\begin{bmatrix} num & sg \\ kas & nom \\ gen & m \end{bmatrix} \text{ ist die Funktion } f \text{ mit: } \begin{array}{l} f(num) = sg \\ f(kas) = nom \\ f(gen) = m \end{array}$$

- Beispielregel mit Constraints: $S \rightarrow NP VP$

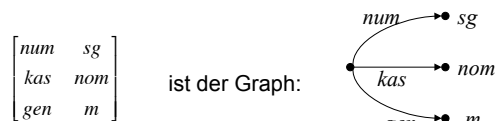
$$f_{NP}(num) = f_{VP}(num)$$

$$f_{NP}(kas) = nom$$

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Formulierung von Merkmalsconstraints

- Was sind Merkmalsstrukturen eigentlich?
- Antwort 2: **Gerichtete azyklische Graphen** mit
 - einem ausgezeichneten Wurzelknoten
 - Attributen als Kantenbeschriftungen
 - Merkmalswerten an den (End-)Knoten.



- Wir bezeichnen ein spezifisches Merkmal in einer Merkmalsstruktur durch einen vom Wurzelknoten ausgehenden Pfad, beginnend mit der Konstituente, mit der die Struktur assoziiert ist:

$$S \rightarrow NP VP$$

$$\langle NP \ num \rangle = \langle VP \ num \rangle$$

$$\langle NP \ kas \rangle = nom$$

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

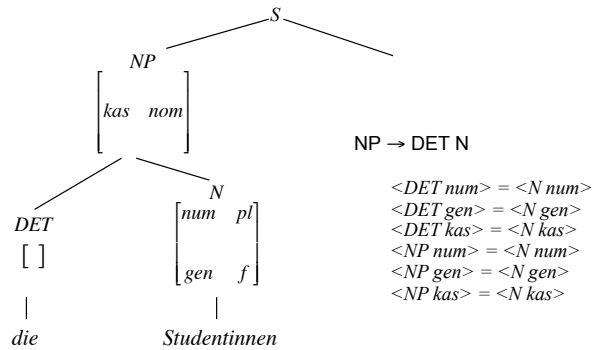
CFG mit Merkmalsconstraints, Beispiel

Beispielgrammatik 4: Ersetzungsregeln mit Merkmalsconstraints

$S \rightarrow NP VP$ $\langle NP \ num \rangle = \langle VP \ num \rangle$ $\langle NP \ kas \rangle = nom$ $VP \rightarrow VI$ $\langle VP \ num \rangle = \langle VI \ num \rangle$ $VP \rightarrow VT NP$ $\langle VP \ num \rangle = \langle VT \ num \rangle$ $\langle NP \ kas \rangle = akk$ $NP \rightarrow DET N$ $\langle DET \ num \rangle = \langle N \ num \rangle$ $\langle DET \ gen \rangle = \langle N \ gen \rangle$ $\langle DET \ kas \rangle = \langle N \ kas \rangle$ $\langle NP \ num \rangle = \langle N \ num \rangle$ $\langle NP \ gen \rangle = \langle N \ gen \rangle$ $\langle NP \ kas \rangle = \langle N \ kas \rangle$	$VI \rightarrow arbeitet$ $\langle VI \ num \rangle = sg$ $VI \rightarrow arbeiten$ $\langle VI \ num \rangle = pl$ $N \rightarrow Student$ $\langle N \ num \rangle = sg$ $\langle N \ gen \rangle = m$ $\langle N \ kas \rangle = nom$ $DET \rightarrow der$ $\langle DET \ num \rangle = sg$ $\langle DET \ gen \rangle = m$ $\langle DET \ kas \rangle = nom$
---	--

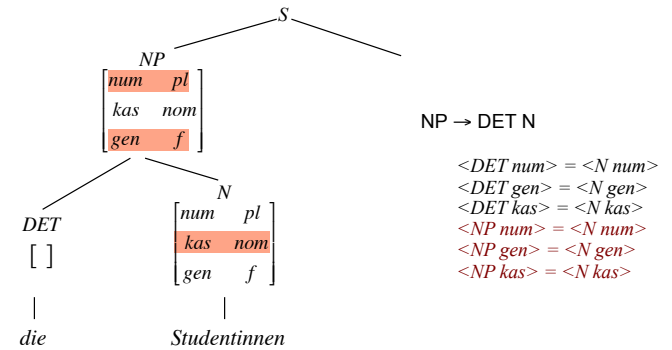
Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Die Anwendung von Merkmalsconstraints



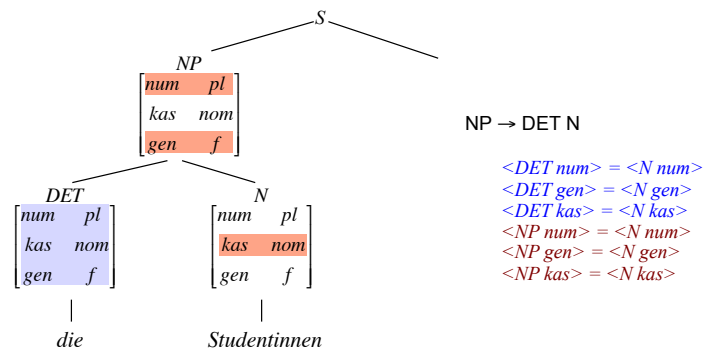
Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Die Anwendung von Merkmalsconstraints



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Die Anwendung von Merkmalsconstraints



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Die Anwendung von Merkmalsconstraints

- Wir verwenden die Ersetzungsregeln, um einen Parse-Baum zu erzeugen.
- Wir wenden Constraints an, um die Konsistenz der vorhandenen Merkmalsinformation zu testen, und um Lücken in der Merkmalsinformation aufzufüllen.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Die Anwendung von Merkmalsconstraints

- Merkmalsgleichungen werden neu interpretiert: Sie sind nicht (nur) Identitätstests, sondern Anweisungen zur Ausführung einer Operation.
- Beispiel: $\langle DET\ num \rangle = \langle N\ num \rangle$ ist zu lesen als:
 - Wenn $\langle DET\ num \rangle$ und $\langle N\ num \rangle$ identisch: Akzeptiere!
 - Wenn $\langle DET\ num \rangle$ undefiniert, setze $\langle DET\ num \rangle := \langle N\ num \rangle$ und akzeptiere!
 - Wenn $\langle N\ num \rangle$ undefiniert, setze $\langle N\ num \rangle := \langle DET\ num \rangle$ und akzeptiere!
 - Wenn $\langle DET\ num \rangle \neq \langle N\ num \rangle$: Fehlschlag!
- Die Operation, die diesen Effekt hat, nennen wir **Unifikation**.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

CFG mit Merkmalsconstraints, Beispiel

Beispielgrammatik 4: Ersetzungsregeln mit Merkmalsconstraints

<p>$S \rightarrow NP\ VP$ $\langle NP\ num \rangle = \langle VP\ num \rangle$ $\langle NP\ kas \rangle = nom$</p> <p>$VP \rightarrow VI$ $\langle VP\ num \rangle = \langle VI\ num \rangle$</p> <p>$VP \rightarrow VT\ NP$ $\langle VP\ num \rangle = \langle VT\ num \rangle$ $\langle NP\ kas \rangle = akk$</p> <p>$NP \rightarrow DET\ N$ $\langle DET\ num \rangle = \langle N\ num \rangle$ $\langle DET\ gen \rangle = \langle N\ gen \rangle$ $\langle DET\ kas \rangle = \langle N\ kas \rangle$ $\langle NP\ num \rangle = \langle N\ num \rangle$ $\langle NP\ gen \rangle = \langle N\ gen \rangle$ $\langle NP\ kas \rangle = \langle N\ kas \rangle$ $\langle NP\ def \rangle = \langle DET\ def \rangle$</p>	<p>$VI \rightarrow arbeitet$ $\langle VI\ num \rangle = sg$</p> <p>$VI \rightarrow arbeiten$ $\langle VI\ num \rangle = pl$</p> <p>$N \rightarrow Student$ $\langle N\ num \rangle = sg$ $\langle N\ gen \rangle = m$ $\langle N\ kas \rangle = nom$</p> <p>$DET \rightarrow der$ $\langle DET\ num \rangle = sg$ $\langle DET\ gen \rangle = m$ $\langle DET\ kas \rangle = nom$ $\langle DET\ def \rangle = +$</p>
--	--

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Unifikation

- Wir schreiben $A \sqcup B$ für „die Unifikation von A und B“.
- Unifikation ist folgendermaßen definiert:
 Wenn $A=B=a$, oder $A=a$ und B undefiniert, oder umgekehrt, werden A und B auf den Wert a gesetzt.
 In diesen Fällen sagen wir, dass A und B "zu a unifizieren".
 Wenn A und B definiert und $A \neq B$, schlägt die Unifikation fehl.
- Der Constraints $\langle DET\ num \rangle = \langle N\ num \rangle$ ist zu lesen als eine Anweisung zur Ausführung der Unifikation.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Merkmalstrukturen: Erste Erweiterung

- Constraints lassen sich eleganter formulieren, wenn wir gleichzeitig auf Mengen von Merkmalen Bezug nehmen können.
- Wir erlauben komplexe Merkmalsstrukturen, in denen Attribute nicht nur atomare Werte, sondern auch Merkmalsstrukturen als Werte haben können. Als Beispiel eine Merkmalsstruktur für den bestimmten Artikel *der*:

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

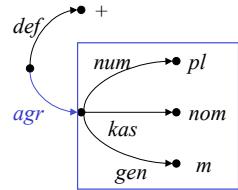
- „agr“ für englisch „agreement“ (Kongruenz) nimmt als Wert eine Merkmalsstruktur, die die Kongruenzmerkmale spezifiziert.
- Statt der Aufzählung einzelner Kongruenzmerkmale in der NP-Regel können wir nun Kongruenz generell verlangen:

$$\langle DET\ AGR \rangle = \langle N\ AGR \rangle$$

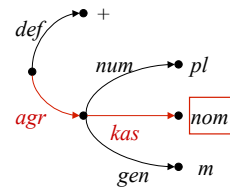
Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Merkmale als Graphen

<DET agr>:



<DET agr kas>:



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Berechnung von $A \sqcup B$

- Resultat der Unifikation ist die (allgemeinste) Merkmalsstruktur, die die Information aus beiden Merkmalsstrukturen umfasst, wenn es eine solche Struktur gibt. Ansonsten schlägt sie fehl.
- Wie berechnet sich das Resultat der Unifikation aus A und B?
- Fall 1: A und B sind atomar:
 - $A \sqcup B = a$, wenn $A=B=a$, oder eine Seite atomar und die andere undefiniert
 - $A \sqcup B$ schlägt fehl, wenn $a \neq b$
- Fall 2: Beide Seiten sind komplexe Merkmalstrukturen:
 - $F \sqcup G = H$, wobei die Attribute von H die Vereinigung der Attribute von F und G ist, und der Wert für jedes Attribut f in H
 - der Wert von f in F (bzw. G), wenn f nur in F (bzw. G) instantiiert ist
 - $A \sqcup B$, wenn A und B die Werte von f in F bzw. G sind.
 - $F \sqcup G$ schlägt fehl, wenn die Unifikation für ein Attribut fehlschlägt.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Beispielgrammatik 5: Komplexe Merkmale

$S \rightarrow NP VP$
 <NP AGR NUM> = <VP AGR NUM>
 <NP AGR KAS> = nom

$VP \rightarrow VI$
 <VP AGR NUM> = <VI AGR NUM>

$VP \rightarrow VT NP$
 <VP AGR NUM> = <VT AGR NUM>
 <NP AGR KAS> = akk

$NP \rightarrow DET N$
 <DET AGR> = <N AGR>
 <NP AGR> = <N AGR>
 <NP def> = <DET def>

wählt: $\begin{bmatrix} CAT & VT \\ AGR & [NUM \ 3g] \end{bmatrix}$

arbeitet: $\begin{bmatrix} CAT & VI \\ AGR & [NUM \ 3g] \end{bmatrix}$

Studentin: $\begin{bmatrix} CAT & N \\ AGR & [NUM \ 3g] \\ & [GEN \ f] \end{bmatrix}$

Fach: $\begin{bmatrix} CAT & N \\ AGR & [NUM \ 3g] \\ & [GEN \ n] \end{bmatrix}$

die: $\begin{bmatrix} CAT & DET \\ AGR & [GEN \ f] \\ DEF & + \end{bmatrix}$

das: $\begin{bmatrix} CAT & DET \\ AGR & [NUM \ 3g] \\ & [GEN \ n] \\ DEF & + \end{bmatrix}$

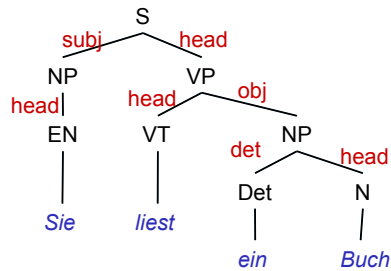
Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Anmerkungen zu Grammatik 5

- In der Grammatik 5 auf der folgenden Folie sind mehrere Neuerungen enthalten, die das Schreiben der Grammatik vereinfachen
 - Komplexe Merkmalsstrukturen mit dem Kongruenz-/Agreement-Merkmal „agr“.
 - Lexikoneinträge, bestehend aus einem Wort (Terminalsymbol) und einer Merkmalsstruktur.
 - Es gibt keine expliziten Erzeugungsregeln, die präterminale Symbole/lexikalische Kategorien in Terminalsymbole/Wörter überführen. In einer Ableitung kann für eine lexikalische Kategorie X jedes Wort aus dem Lexikon eingesetzt werden, dessen Merkmalsstruktur für das Attribut cat den Wert X besitzt.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Ein Beispiel



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Kopfmerkmale

- Bestimmte Merkmale vererben sich „entlang der Kopflinie“.
- Insbesondere teilt die Phrase diese Merkmale mit ihrem lexikalischen Kopf,
 - zum Beispiel die Kongruenzmerkmale (*AGR*),
 - aber auch andere Merkmale, zum Beispiel die „Verbform“ (*VFORM*): Infinitivkonstruktionen enthalten Infinitive, finite Verbphrasen finite (flektierte) Verben, etc.
- Um die Regularität in der Grammatik ausdrücken, führen wir als zusätzliches Attribut *HEAD* ein (unter dem die „Kopfmerkmale“ aufgeführt werden).

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Beispielgrammatik 6: Kopfmerkmale

$S \rightarrow NP VP$
 $\langle S HEAD \rangle = \langle VP HEAD \rangle$
 $\langle S HEAD AGR NUM \rangle =$
 $\langle NP HEAD AGR NUM \rangle$
 $\langle NP HEAD AGR KAS \rangle = nom$
 $VP \rightarrow VI$
 $\langle VP HEAD \rangle = \langle VI HEAD \rangle$
 $VP \rightarrow VT NP$
 $\langle VP HEAD \rangle = \langle VT HEAD \rangle$
 $\langle NP HEAD AGR KAS \rangle = akk$
 $NP \rightarrow DET N$
 $\langle NP HEAD \rangle = \langle N HEAD \rangle$
 $\langle DET AGR \rangle = \langle N AGR \rangle$

wählt: $\begin{bmatrix} CAT \\ HEAD \\ \left[\begin{array}{l} VFORM \\ AGR \end{array} \right] \left[\begin{array}{l} VT \\ NUM \\ finit \\ sg \end{array} \right] \end{bmatrix}$
arbeitet: $\begin{bmatrix} CAT \\ HEAD \\ \left[\begin{array}{l} VFORM \\ AGR \end{array} \right] \left[\begin{array}{l} VI \\ NUM \\ finit \\ sg \end{array} \right] \end{bmatrix}$
Studentin: $\begin{bmatrix} CAT \\ HEAD \\ \left[\begin{array}{l} AGR \\ NUM \\ GEN \end{array} \right] \left[\begin{array}{l} N \\ sg \\ f \end{array} \right] \end{bmatrix}$
Fach: $\begin{bmatrix} CAT \\ HEAD \\ \left[\begin{array}{l} AGR \\ NUM \\ GEN \end{array} \right] \left[\begin{array}{l} N \\ sg \\ n \end{array} \right] \end{bmatrix}$
die: $\begin{bmatrix} CAT \\ AGR \\ DEF \\ + \end{bmatrix} \begin{bmatrix} DET \\ GEN \\ f \end{bmatrix}$
das: $\begin{bmatrix} CAT \\ AGR \\ DEF \\ + \end{bmatrix} \begin{bmatrix} DET \\ NUM \\ GEN \end{bmatrix} \left[\begin{array}{l} sg \\ n \end{array} \right]$

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Subkategorisierung

- Transitive und intransitive Verben gehören in unserer Beispielgrammatik zu verschiedenen Kategorien. Konsequenterweise müssten auch für ditransitive Verben (*geben*), Verben mit Dativobjekt (*gefallen*), Verben mit Präpositionalem Objekt (*warten auf*), Verben mit *zu*-Infinitiv (*versuchen*) unterschiedliche Kategorien vorgesehen werden. Ebenso für Präsuppositionen, Adjektive, Substantive, die unterschiedliche Argumente nehmen.
- Eleganter ist die Lösung, die Subkategorisierungseigenschaften durch ein Merkmal auszudrücken (traditionell „*SUBCAT*“).

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Subkategorisierung

$VP \rightarrow V NP$

$\langle VP HEAD \rangle = \langle V HEAD \rangle$

$\langle V SUBCAT \rangle = \langle NP \rangle$

$\langle NP HEAD AGR KAS \rangle = akk$

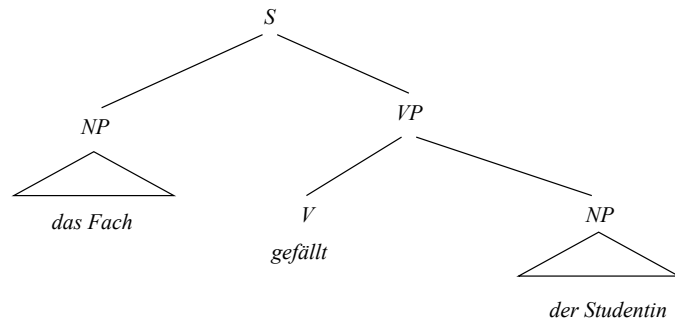
wählt: $\begin{bmatrix} CAT \\ HEAD \\ SUBCAT \end{bmatrix} \begin{bmatrix} V \\ VFORM \\ HEAD \end{bmatrix} \begin{bmatrix} finit \\ [NUM sg] \\ [AGR [KAS akk]] \end{bmatrix}$

gefällt: $\begin{bmatrix} CAT \\ HEAD \\ SUBCAT \end{bmatrix} \begin{bmatrix} V \\ VFORM \\ HEAD \end{bmatrix} \begin{bmatrix} finit \\ [NUM sg] \\ [AGR [KAS dat]] \end{bmatrix}$

- Die obige Regel beschreibt alle „V NP“-Konstruktionen.
- Wir identifizieren per Pfadgleichung die komplette Merkmalsstruktur der Objekt-NP mit dem SUBCAT-Merkmal des Verbs. Die spezifische Kasusforderung einzelner Verben steht im Lexikon.

Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Beispielgrammatik 7: SUBCAT-Merkmale

$S \rightarrow NP VP$
 $\langle S HEAD \rangle = \langle VP HEAD \rangle$
 $\langle S HEAD AGR NUM \rangle = \langle NP HEAD AGR NUM \rangle$
 $\langle NP HEAD AGR KAS \rangle = nom$

$VP \rightarrow V$
 $\langle VP HEAD \rangle = \langle VI HEAD \rangle$
 $\langle V SUBCAT \rangle = empty$

$VP \rightarrow V NP$
 $\langle VP HEAD \rangle = \langle V HEAD \rangle$
 $\langle V SUBCAT \rangle = \langle NP \rangle$

$NP \rightarrow DET N$
 $\langle NP HEAD \rangle = \langle N HEAD \rangle$
 $\langle NP HEAD AGR \rangle = \langle DET AGR \rangle$

das: $\begin{bmatrix} CAT \\ AGR \end{bmatrix} \begin{bmatrix} DET \\ [NUM sg] \\ [GEN n] \end{bmatrix}$ die: $\begin{bmatrix} CAT \\ AGR \end{bmatrix} \begin{bmatrix} DET \\ [GEN f] \end{bmatrix}$ der: $\begin{bmatrix} CAT \\ AGR \end{bmatrix} \begin{bmatrix} DET \\ [] \end{bmatrix}$

wählt: $\begin{bmatrix} CAT \\ HEAD \\ SUBCAT \end{bmatrix} \begin{bmatrix} V \\ VFORM \\ HEAD \end{bmatrix} \begin{bmatrix} finit \\ [AGR [NUM sg]] \\ [AGR [KAS akk]] \end{bmatrix}$

gefällt: $\begin{bmatrix} CAT \\ HEAD \\ SUBCAT \end{bmatrix} \begin{bmatrix} V \\ VFORM \\ HEAD \end{bmatrix} \begin{bmatrix} finit \\ [AGR [NUM sg]] \\ [AGR [KAS dat]] \end{bmatrix}$

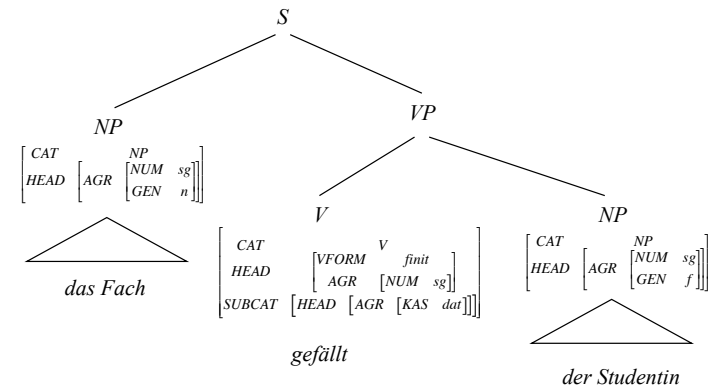
arbeitet: $\begin{bmatrix} CAT \\ HEAD \\ SUBCAT \end{bmatrix} \begin{bmatrix} V \\ VFORM \\ HEAD \end{bmatrix} \begin{bmatrix} finit \\ [AGR [NUM sg]] \\ empty \end{bmatrix}$

Studentin: $\begin{bmatrix} CAT \\ HEAD \end{bmatrix} \begin{bmatrix} N \\ [NUM sg] \\ [GEN f] \end{bmatrix}$

Fach: $\begin{bmatrix} CAT \\ HEAD \end{bmatrix} \begin{bmatrix} N \\ [NUM sg] \\ [GEN n] \end{bmatrix}$

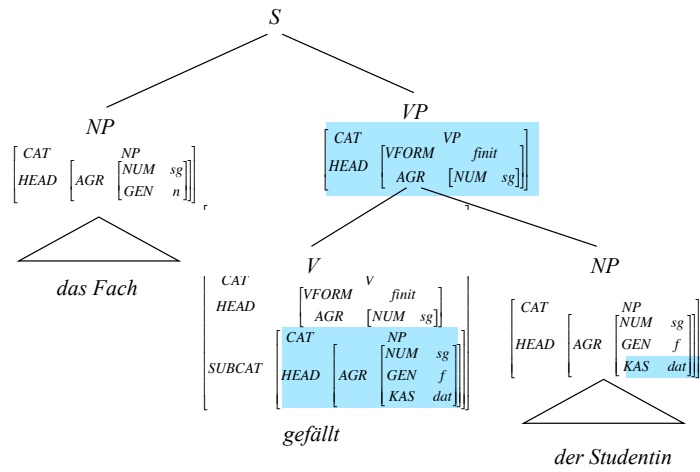
Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



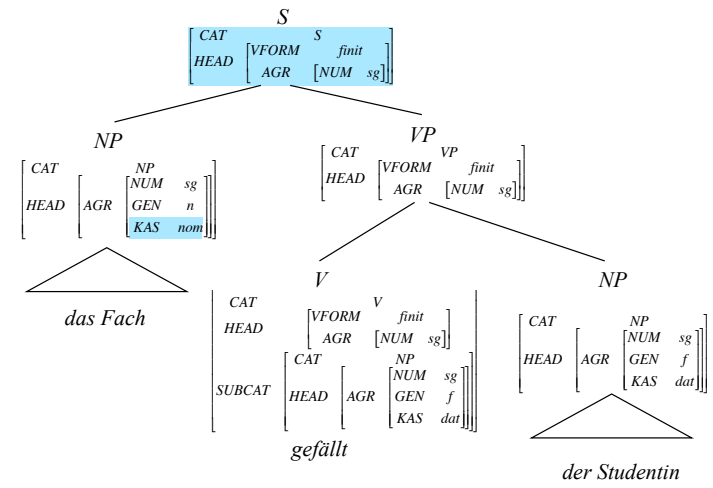
Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



Vorlesung "Einführung in die CL" 2012/13 © M. Pinkal UdS Computerlinguistik