

Einführung in die Computerlinguistik

Syntax: Merkmalsstrukturen

WS 2011/2012

Manfred Pinkal

Vorlesung "Einführung in die CL" 2011/2012 © 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" 2011/2012 © 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" 2011/2012 © M. Pinkal UdS Computerlinguistik

Status von Merkmalen

- Inhärente Merkmale (Genus beim Substantiv)
- Unabhängige Merkmale (Numerus beim Substantiv, Tempus beim Verb)
- Abhängige Merkmale

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" 2011/2012 © 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[pf] angehängten[pf] Bilder[pf]*
 - **Subjekt-Verb-Kongruenz**: *sie[pf] vermitteln[pf]*
 - **Pronominalkongruenz** zwischen einem „anaphorischen“ Pronomen und der NP, auf die er sich bezieht
Fotos[pf] ... sie[pf]

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Grammatische Merkmale in der CFG

- Beispielgrammatik 1: Ohne Merkmalsinformation

$S \rightarrow NP VP$ $VP \rightarrow VT NP$

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

$VI \rightarrow$ *schläft* | *arbeitet*

$VT \rightarrow$ *kennt* | *studiert*

$N \rightarrow$ *Student* | *Studentin* | *Studenten* | *Studentinnen* | *Fach*

$DET \rightarrow$ *der* | *die* | *das* | *den*

- Nicht-Berücksichtigung von Kongruenz und Rektion führt zu massiver Übergenerierung:
 - *die Studenten arbeitet*
 - *der Student kennt der Student*
- Wie können Merkmale und Merkmalsabhängigkeiten in der CFG modelliert werden?

Vorlesung "Einführung in die CL" 2011/2012 © 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" 2011/2012 © M. Pinkal UdS Computerlinguistik

Versuch: Verfeinerung der Kategorien

- Beispielgrammatik 2: Merkmalsinformation in Kategoriensymbole integriert

$S \rightarrow$ NPSgNom VPSg

$VPSg \rightarrow$ VISg

$VPSg \rightarrow$ VTSg NPAkk

$NPSgNom \rightarrow$ DETSgNomM NSgNomM

$NPPINom \rightarrow$ DETPINom NPINom

$DETSgNomM \rightarrow$ *der*

$NSgNomM \rightarrow$ *Student*

...

$S \rightarrow$ NPPINom VPPI

$VPPI \rightarrow$ VIPI

$VPPI \rightarrow$ VTPI NPAkk

$NPSgNom \rightarrow$ DETSgNomF NSgNomF

$DETSgNomF \rightarrow$ *die*

$NSgNomF \rightarrow$ *Studentin*.

...

- Nachteile:
 - Regularitäten können nicht ausgedrückt werden
 - Das Regelsystem wird aufgebläht (2 Numeri x 3 Genera x 4 Kasus x 3 Personen x ...)

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Merkmalsstrukturen und CFG

- Eine Merkmalsstruktur ist eine Menge von „Attribut-Wert-Paaren“: Die Merkmalsstruktur des NP-Knotens im Beispiel hat drei Merkmale, das erste Merkmal $[num\ pl]$ besteht aus dem Attribut „num“ und dem atomaren Wert „pl“.
- Die Symbole der CFG sind komplex: Paar aus Kategorien-/Terminalsymbol und Merkmalsstruktur.

Integration von Merkmalsstrukturen in die CFG, 1. Versuch:

- Erzeugungsregeln enthalten statt einfacher Symbole komplexe Symbole.
- Ableitung erfolgt wie bisher.
- Problem: Die Regelzahl wird nicht reduziert. Regularitäten sind besser erkennbar, werden aber nicht genutzt.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Explizite Kodierung von Merkmalen

Beispielgrammatik 3: Direkte Erzeugung von Merkmalsstrukturen

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

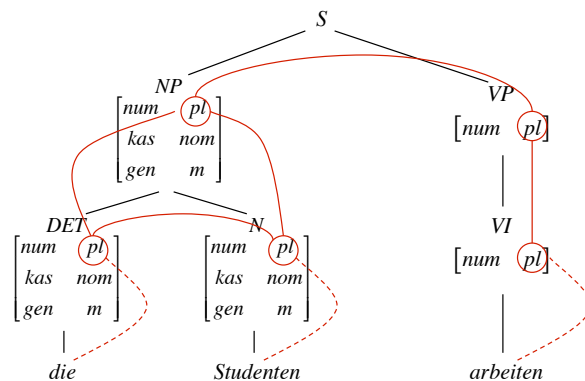
$$VP[num\ sg] \rightarrow VI[num\ sg] \quad VP[num\ pl] \rightarrow VI[num\ pl]$$

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

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

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Explizite Kodierung von Merkmalen



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Merkmalsconstraints

- Merkmalsstrukturen erlauben die Formulierung von Bedingungen/ Einschränkungen/ Constraints, z.B.
Beispiel: „Numerus von NP und Numerus von VP sind identisch“

- Die Regeln der Grammatik sind zweiteilig:
 - Ersetzungsregel, wie üblich über Kategorien und lexikalische Ausdrücke formuliert
 - eine Menge von Constraints über Merkmalsstrukturen

• Beispiel:

$$S \rightarrow NP\ VP$$

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

Notation:

$$S \rightarrow NP\ VP$$

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

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

CFG mit Merkmalsconstraints, Beispiel

Beispielgrammatik 4: Ersetzungsregeln + einfache Merkmalsconstraints

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

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

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

$NP \rightarrow DET N$
 $\langle DET \text{ num} \rangle = \langle N \text{ num} \rangle$
 $\langle DET \text{ gen} \rangle = \langle N \text{ gen} \rangle$
 $\langle DET \text{ kas} \rangle = \langle N \text{ kas} \rangle$
 $\langle NP \text{ num} \rangle = \langle N \text{ num} \rangle$
 $\langle NP \text{ gen} \rangle = \langle N \text{ gen} \rangle$
 $\langle NP \text{ kas} \rangle = \langle N \text{ kas} \rangle$

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

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

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

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

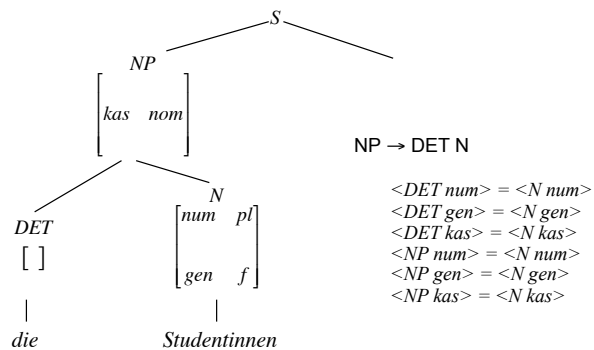
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Anwendung von Merkmalsconstraints, Variante 1

- Schritt 1: Generiere mit den Ersetzungsregeln Ableitungsbäume mit beliebig gewählten Merkmalsstrukturen für nicht-terminale Kategorien.
- Schritt 2: Teste, ob (1) die Merkmale an den präterminalen Knoten zu den Lexikoneinträgen passen, und ob (2) alle durch Regeln eingeführte Constraints erfüllt sind.
 - Der Constraint $\langle NP \text{ num} \rangle = \langle VP \text{ num} \rangle$ ist erfüllt, wenn $\langle NP \text{ num} \rangle = \text{sg}$ und $\langle VP \text{ num} \rangle = \text{sg}$ oder wenn $\langle NP \text{ num} \rangle = \text{pl}$ und $\langle VP \text{ num} \rangle = \text{pl}$
- Kommentar: Regeln werden kompakt repräsentiert, das Verfahren ist jedoch extrem ineffizient.

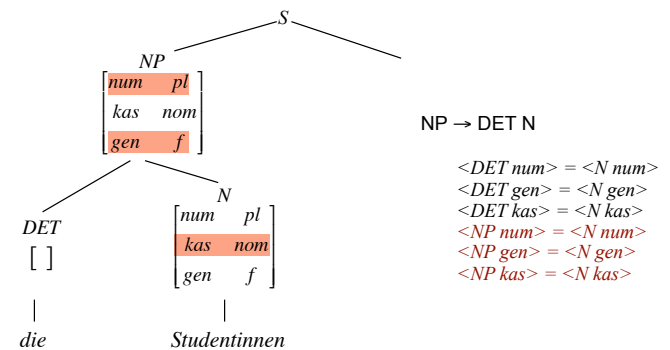
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Anwendung von Merkmalsconstraints, Variante 2



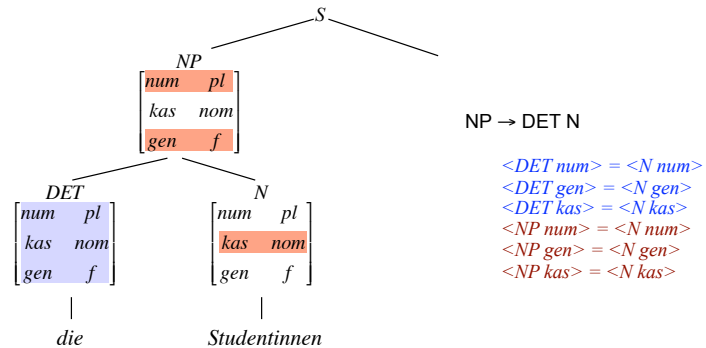
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Die Anwendung von Merkmalsconstraints



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UoS Computerlinguistik

Die Anwendung von Merkmalsconstraints



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Anwendung von Merkmalsconstraints, Variante 2

- Schritt 1: Erzeuge mit den Ersetzungsregeln der Grammatik einen kompletten Ableitungsbaum; die Constraints der verwendeten Regeln werden aufgesammelt.
- Schritt 2: Annotiere (prä-)terminale Symbole mit der im Lexikon vorhandenen sicheren Merkmalsinformation.
- Schritt 3: Wende Constraints an, um (1) die Konsistenz der vorhandenen Merkmalsinformation zu testen, und (2) um nicht vorhandene Merkmalsinformation aufzufüllen.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Anwendung von Merkmalsconstraints, Variante 2

- Wir arbeiten mit partiell definierten Merkmalsstrukturen.
- 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 $\langle DET\ num \rangle$ offen, setze $\langle DET\ num \rangle := \langle N\ num \rangle$ und akzeptiere
 Wenn $\langle N\ num \rangle$ offen, setze $\langle N\ num \rangle := \langle DET\ num \rangle$ und akzeptiere
 Wenn $\langle DET\ num \rangle \neq \langle N\ num \rangle$: Fehlschlag!

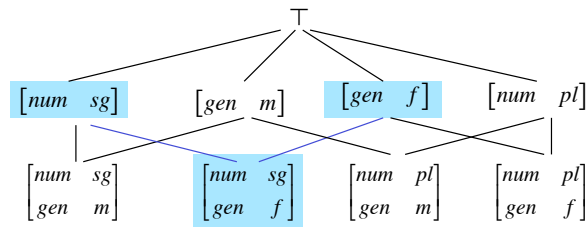
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Unifikation

- Wir schreiben $A \sqcup B$ für „A unifiziert mit B“, und notieren undefinierte Merkmale mit \perp , Fehlschlag mit \perp .
- Unifikation ist definiert durch:
 $a \sqcup a = a \sqcup \perp = \perp \sqcup a = a$
 $a \sqcup b = \perp$, falls $a \neq b$
- Der Effekt des Constraints $\langle DET\ num \rangle = \langle N\ num \rangle$ lässt sich auf diese Weise einfach beschreiben:
 $\langle DET\ num \rangle := \langle DET\ num \rangle \sqcup \langle N\ num \rangle$
 $\langle N\ num \rangle := \langle DET\ num \rangle \sqcup \langle N\ num \rangle$

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

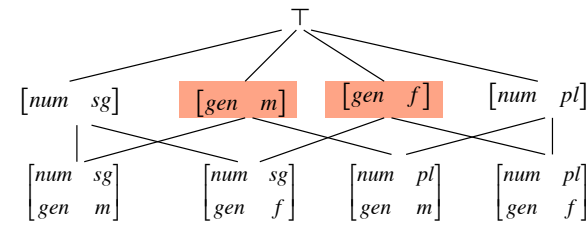
Unifikation



$$\begin{bmatrix} \text{num} & \text{sg} \end{bmatrix} \sqcup \begin{bmatrix} \text{gen} & \text{f} \end{bmatrix} = \begin{bmatrix} \text{num} & \text{sg} \\ \text{gen} & \text{f} \end{bmatrix}$$

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

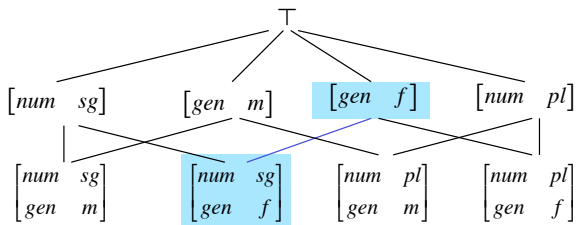
Unifikation



$$\begin{bmatrix} \text{gen} & \text{m} \end{bmatrix} \sqcup \begin{bmatrix} \text{gen} & \text{f} \end{bmatrix} = \perp$$

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Unifikation



$$\begin{bmatrix} \text{num} & \text{sg} \\ \text{gen} & \text{f} \end{bmatrix} \sqcup \begin{bmatrix} \text{gen} & \text{f} \end{bmatrix} = \begin{bmatrix} \text{num} & \text{sg} \\ \text{gen} & \text{f} \end{bmatrix}$$

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Die CFG mit Merkmalsconstraints

Beispielgrammatik 4: Ersetzungsregeln + einfache Merkmalsconstraints

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

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

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

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

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

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

$NP \rightarrow DET N$
 $\langle DET \text{ num} \rangle = \langle N \text{ num} \rangle$
 $\langle DET \text{ gen} \rangle = \langle N \text{ gen} \rangle$
 $\langle DET \text{ kas} \rangle = \langle N \text{ kas} \rangle$
 $\langle NP \text{ num} \rangle = \langle N \text{ num} \rangle$
 $\langle NP \text{ gen} \rangle = \langle N \text{ gen} \rangle$
 $\langle NP \text{ kas} \rangle = \langle N \text{ kas} \rangle$

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

Vorlesung "Einführung in die CL" 2011/2012 © 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. Beispiel:

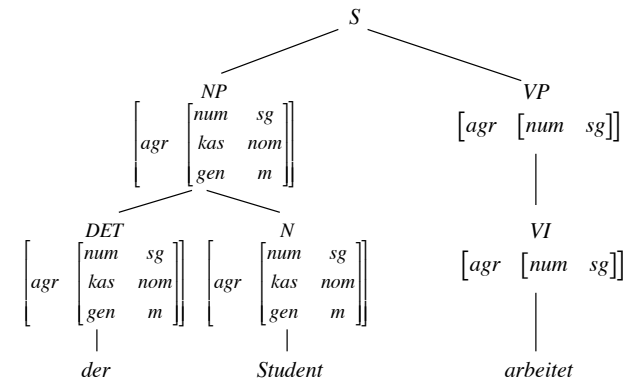
$$\left[\begin{array}{c} agr \\ \left[\begin{array}{cc} num & pl \\ kas & nom \\ gen & m \end{array} \right] \end{array} \right]$$

„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 formulieren

Kongruenzmerkmale von DET = Kongruenzmerkmale von N
Kongruenzmerkmale von NP = Kongruenzmerkmale von N

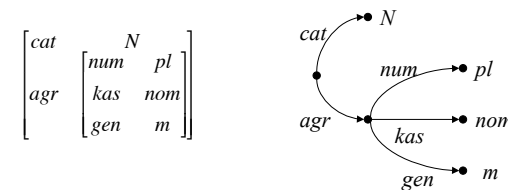
Beispiel



Berechnung von $A \sqcup B$

- Wie berechnet sich das Resultat der Unifikation aus A und B?
- Fall 1: Beide Seiten sind atomar:
 - $a \sqcup a = a$
 - $a \sqcup b = \perp$, wenn $a \neq b$
- Fall 2: Eine Seite ist atomar, die andere komplexe Merkmalsstruktur
 - $a \sqcup F = \perp$
- Fall 3: Beide Seiten sind komplexe Merkmalstrukturen:
 - $F \sqcup G = H$, wobei die Attribute von H Attribute die Vereinigung der Attribute von F und G sind.
 - Der Wert für jedes Attribut f in H ist
 - der Wert von f in F (G), wenn f nur in F (G) instantiiert ist
 - $A \sqcup B$, wenn A und B die Werte von f in F bzw. G sind.

Was sind Merkmalsstrukturen eigentlich?

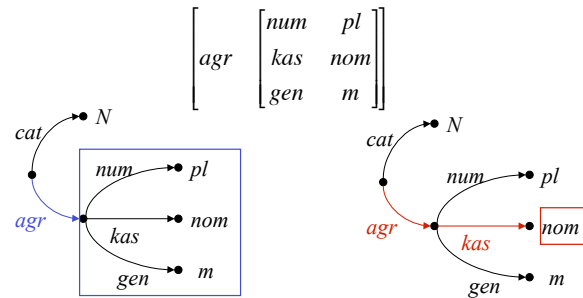


Eine Sicht auf Merkmalsstrukturen: (Partielle) Funktionen

Eine andere Sicht: Merkmalstrukturen sind als gerichtete (azyklische) Graphen mit Kanteninschriften darstellbar (in dem einfachen Fall, den wir hier betrachten: Bäume)

Die Kanteninschriften sind Attribut-Label. Die Blätter sind mit atomaren Werten beschriftet.

Merkmalspfade



Folgen wie $\langle N \text{ agr} \rangle$ bzw. $\langle N \text{ agr kas} \rangle$, die in mit Grammatikregeln assoziierten Merkmalsconstraints vorkommen, spezifizieren eindeutig eine Position in den beteiligten Merkmalsstrukturen/ Merkmalsgraphen: Kategoriensymbol + „Merkmalspfad“

Sie bezeichnen den atomaren Wert oder den komplexen Graphen, der sich unter dieser Position befindet.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Beispielgrammatik 5: Komplexe Merkmale

$S \rightarrow NP VP$ $\langle NP \text{ AGR NUM} \rangle = \langle VP \text{ AGR NUM} \rangle$ $\langle NP \text{ AGR KAS} \rangle = \text{nom}$	wählt: $\begin{bmatrix} CAT & VT \\ AGR & [NUM \text{ sg}] \end{bmatrix}$
$VP \rightarrow VI$ $\langle VP \text{ AGR NUM} \rangle = \langle VI \text{ AGR NUM} \rangle$	arbeitet: $\begin{bmatrix} CAT & VI \\ AGR & [NUM \text{ sg}] \end{bmatrix}$
$VP \rightarrow VT NP$ $\langle VP \text{ AGR NUM} \rangle = \langle VT \text{ AGR NUM} \rangle$ $\langle NP \text{ AGR KAS} \rangle = \text{akk}$	Studentin: $\begin{bmatrix} CAT & N \\ AGR & [NUM \text{ sg}] \\ & [GEN \text{ f}] \end{bmatrix}$
$NP \rightarrow DET N$ $\langle DET \text{ AGR} \rangle = \langle N \text{ AGR} \rangle$ $\langle NP \text{ AGR} \rangle = \langle N \text{ AGR} \rangle$	Fach: $\begin{bmatrix} CAT & N \\ AGR & [NUM \text{ sg}] \\ & [GEN \text{ n}] \end{bmatrix}$
	die: $\begin{bmatrix} CAT & DET \\ AGR & [GEN \text{ f}] \end{bmatrix}$
	das: $\begin{bmatrix} CAT & DET \\ AGR & [NUM \text{ sg}] \\ & [GEN \text{ n}] \end{bmatrix}$

Vorlesung "Einführung in die CL" 2011/2012 © 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" 2011/2012 © 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" 2011/2012 © 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: $\begin{bmatrix} CAT & VT \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \end{bmatrix}$
arbeitet: $\begin{bmatrix} CAT & VI \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \end{bmatrix}$
Studentin: $\begin{bmatrix} CAT & N \\ HEAD & \begin{bmatrix} AGR & \begin{bmatrix} NUM sg \\ GEN f \end{bmatrix} \end{bmatrix} \end{bmatrix}$
Fach: $\begin{bmatrix} CAT & N \\ HEAD & \begin{bmatrix} AGR & \begin{bmatrix} NUM sg \\ GEN n \end{bmatrix} \end{bmatrix} \end{bmatrix}$
die: $\begin{bmatrix} CAT & DET \\ AGR & [GEN f] \end{bmatrix}$
das: $\begin{bmatrix} CAT & DET \\ AGR & \begin{bmatrix} NUM sg \\ GEN n \end{bmatrix} \end{bmatrix}$

Vorlesung "Einführung in die CL" 2011/2012 © 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" 2011/2012 © 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: $\begin{bmatrix} CAT & V \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \\ SUBCAT & [HEAD [AGR [KAS akk]]] \end{bmatrix}$
gefällt: $\begin{bmatrix} CAT & V \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \\ SUBCAT & [HEAD [AGR [KAS dat]]] \end{bmatrix}$

- 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.
- Die übliche Modellierung der Subkategorisierung in Grammatik-Formalisten ist noch allgemeiner und eleganter: beliebig viele Ergänzungen von beliebiger Kategorie.

Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Beispielgrammatik 7: SUBCAT- Merkmale

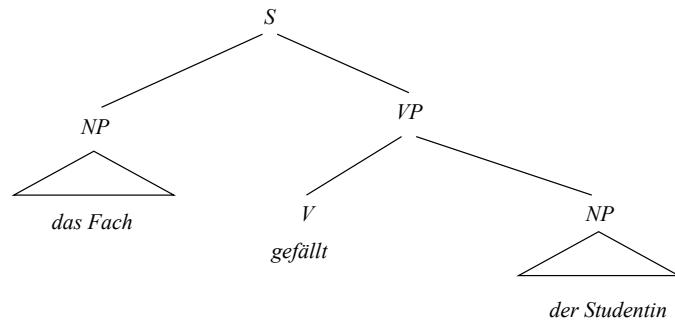
$S \rightarrow NP VP$
 $\langle S HEAD \rangle = \langle VP HEAD \rangle$
 $\langle S HEAD AGR NUM \rangle = \langle NP HEAD AGR NUM \rangle$
 $\langle NP HEAD AGR KAS \rangle = nom$
 $VP \rightarrow V$
 $\langle VP HEAD \rangle = \langle V HEAD \rangle$
 $\langle V SUBCAT \rangle = empty$
 $VP \rightarrow V NP$
 $\langle VP HEAD \rangle = \langle V HEAD \rangle$
 $\langle V SUBCAT \rangle = \langle NP \rangle$
 $NP \rightarrow DET N$
 $\langle NP HEAD \rangle = \langle N HEAD \rangle$
 $\langle NP HEAD AGR \rangle = \langle DET AGR \rangle$

wählt: $\begin{bmatrix} CAT & V \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \\ SUBCAT & [HEAD [AGR [KAS akk]]] \end{bmatrix}$
gefällt: $\begin{bmatrix} CAT & V \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \\ SUBCAT & [HEAD [AGR [KAS dat]]] \end{bmatrix}$
arbeitet: $\begin{bmatrix} CAT & V \\ HEAD & \begin{bmatrix} VFORM & finit \\ AGR & [NUM sg] \end{bmatrix} \\ SUBCAT & empty \end{bmatrix}$
Studentin: $\begin{bmatrix} CAT & N \\ HEAD & \begin{bmatrix} AGR & \begin{bmatrix} NUM sg \\ GEN f \end{bmatrix} \end{bmatrix} \end{bmatrix}$
Fach: $\begin{bmatrix} CAT & N \\ HEAD & \begin{bmatrix} AGR & \begin{bmatrix} NUM sg \\ GEN n \end{bmatrix} \end{bmatrix} \end{bmatrix}$
die: $\begin{bmatrix} CAT & DET \\ AGR & [GEN f] \end{bmatrix}$
der: $\begin{bmatrix} CAT & DET \\ AGR & [] \end{bmatrix}$

- Im folgenden Anwendungsbeispiel nehmen wir außerdem an, dass zu jeder Regel mit linker Seite X ein Constraint existiert, der dem CAT-Attribut der Regel als Wert die Kategorie X zuweist. Kann hier ignoriert werden.

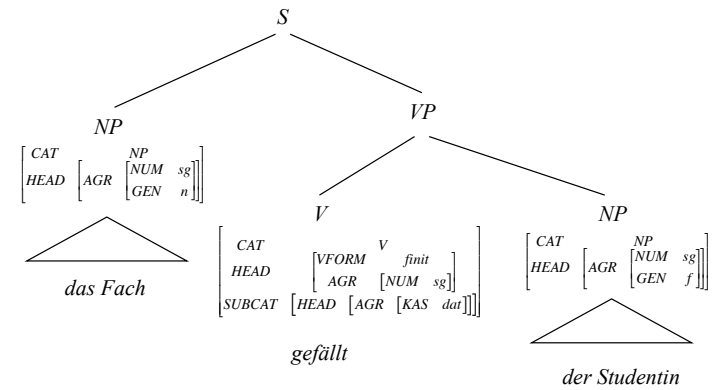
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



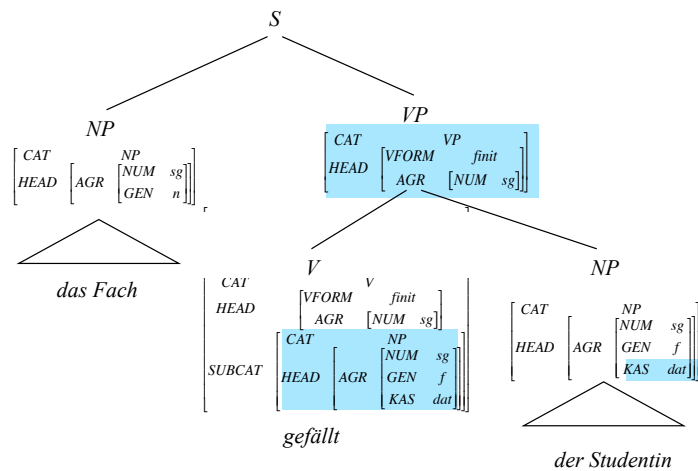
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



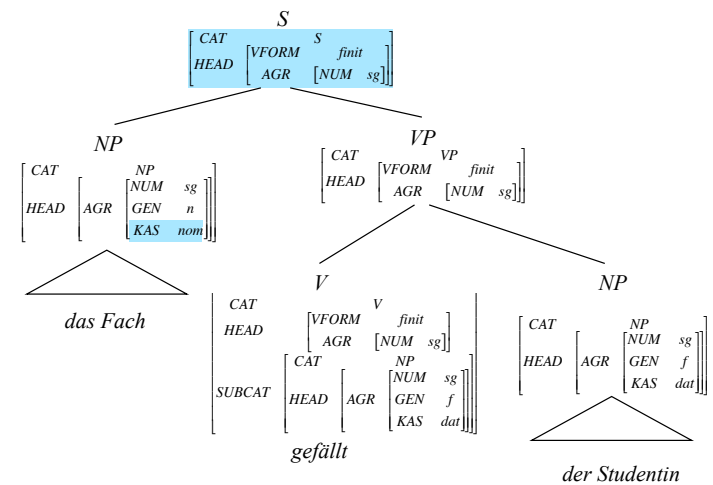
Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik

Eine Analyse mit Grammatik 7



Vorlesung "Einführung in die CL" 2011/2012 © M. Pinkal UdS Computerlinguistik