LS 2: Einführung in die Computerlinguistik / Blockseminar SoSe 2008

Abschließende Hausaufgabe

Teil I: Theoretische Grundlagen

Aufgabe 1: PS-Regeln, I

Gegeben seien die folgenden Daten:

- 1) unter dem Bett
- 2) auf dem Tisch
- 3) auf dem Blechdach

Diese Syntagmen können bei entsprechendem Lexikon über die folgenden, schwach äquivalenten PS-Regeln beschrieben werden:

- (a) $PP \rightarrow P \text{ Det } N$ *oder*
- (b) $PP \rightarrow P \text{ NP}$
 $NP \rightarrow \text{Det } N$

Erklären Sie, wieso die Regeln unter (b) im Vergleich zu (a) genereller und also vorzuziehen sind. Ziehen Sie dafür weitere sprachliche Daten heran.

Aufgabe 2: PS-Regeln, II

Gegeben seien die folgenden Daten:

- 1) I know that he loves her
- 2) I wonder if he loves her

Hier haben wir es mit komplexen Sätzen zu tun, also Sätzen, die einen Haupt- und einen Nebensatz enthalten. Um diese Sätze, genauer gesagt die Verbalphrasen dieser Sätze, zu beschreiben, führen wir zunächst die folgenden lexikalischen Kategorien ein: Ves (Verb, das einen eingebetteten Satz als Komplement hat) und Conj (Konjunktion):

Ves $\rightarrow$ {wonder, know}

Conj $\rightarrow$ {that, if}

Um die Struktur der oa. Verbalphrasen zu beschreiben, könnten wir die folgenden, schwach äquivalenten PS-Regeln einsetzen ('eS' steht für 'eingebetteter Satz'):

- (a) $VP \rightarrow V \text{ Conj } S$ *oder*
- (b) $VP \rightarrow V \text{ eS}$
 $eS \rightarrow \text{Conj } S$

Erklären Sie, wieso die Regeln unter (b) im Vergleich zu (a) die Konstituentenstruktur der Beispielsätze besser abbildet. Unterziehen Sie dafür die Beispielsätze den 'Methoden zur Satzanalyse'

Aufgabe 3: Merkmale

Geben Sie für jede der nachstehenden Merkmalstrukturen eine Wortform aus dem Deutschen an, durch die sie realisiert sein kann:

- (a)
$$\left[\begin{array}{l} \text{Kat} \quad A \\ \text{Agr} \quad \left[\begin{array}{ll} \text{Num} & \text{sg} \\ \text{Pers} & 3 \\ \text{Kas} & \text{akk} \\ \text{Gen} & \text{fem} \end{array} \right] \end{array} \right]$$
- (b)
$$\left[\begin{array}{l} \text{Kat} \quad N \\ \text{Agr} \quad \left[\begin{array}{ll} \text{Num} & \text{pl} \\ \text{Pers} & 3 \\ \text{Kas} & \text{dat} \\ \text{Gen} & \text{mask} \end{array} \right] \end{array} \right]$$
- (c)
$$\left[\begin{array}{l} \text{Kat} \quad \text{Vt} \\ \text{Agr} \quad \left[\begin{array}{ll} \text{Num} & \text{sg} \\ \text{Pers} & 2 \end{array} \right] \\ \text{Temp} \quad \text{Präsens} \\ \text{Modus} \quad \text{Indikativ} \end{array} \right]$$
- (d)
$$\left[\begin{array}{l} \text{Kat} \quad \text{Vi} \\ \text{Agr} \quad \left[\begin{array}{ll} \text{Num} & \text{pl} \\ \text{Pers} & 3 \end{array} \right] \\ \text{Temp} \quad \text{Präteritum} \\ \text{Modus} \quad \text{Indikativ} \end{array} \right]$$

Aufgabe 4: Unifikation

Welche der folgenden Merkmalstrukturen lassen sich miteinander unifizieren und wie sieht das Ergebnis jeweils aus? Geben Sie bei nicht unifizierbaren Strukturen an, wieso diese nicht unifiziert werden können.

- (a)
$$\left[\begin{array}{l} A \quad \alpha \\ B \quad \left[\begin{array}{ll} C & \beta \\ D & \chi \\ E & \delta \end{array} \right] \\ F \quad \epsilon \end{array} \right]$$
- (b)
$$\left[\begin{array}{l} A \quad \alpha \\ B \quad \left[\begin{array}{ll} C & \phi \\ D & \chi \end{array} \right] \\ F \quad \epsilon \end{array} \right]$$
- (c)
$$\left[\begin{array}{l} A \quad \alpha \\ B \quad \left[\begin{array}{ll} D & \chi \end{array} \right] \\ G \quad \phi \end{array} \right]$$
- (d)
$$\left[\begin{array}{l} A \quad \alpha \\ B \quad \left[\begin{array}{ll} C & \phi \\ E & \delta \end{array} \right] \\ G \quad \gamma \end{array} \right]$$

Aufgabe 5: Merkmalsbasierte Grammatik

Gegeben sei die folgende Grammatik:

S	→ NP VP	PP	→ P NP
	<NP Agr> = <VP Agr>	VP	→ Vt NP
	<NP Case> = subj		<NP Case> = obj
NP	→ Det N (PP)		<VP Agr> = <Vt Agr>
	<Det Agr> = <N Agr>		
	<NP Agr> = <N Agr>		
	<NP Case> = <N Case>		

Welche der folgenden Ketten lassen sich bei entsprechendem Lexikon aus dieser Grammatik ableiten; welche nicht und warum nicht?

1. The girl in the red dress watched the boy in the suit.
2. The girl watched the boy.
3. The man with the telescope watched the woman.
4. The woman watched the man with the telescope
5. *Her watched the boy
6. *These girl watched the boy

Teil II: Anwendung

Aufgabe 1: Grundlagen

Implementieren Sie die im Handout 'Übergenerierung, Subkategorisierung und Merkmale' auf den Seiten 8 – 10 diskutierte Grammatik unter [psg1.grm, psg1.lex]. Diese Grammatik beschreibt die folgenden Daten:

1) This cat loves her	5) *These dogs loves her	9) These very ugly dogs love this cat
2) These dogs love this cat	6) *This dogs love her	10) These dogs love this beautiful cat.
3) *Her loves this cat	7) I love this cat	11) Dogs love this cat
4) *These dogs love she	8) *I loves this cat	12) *Dog loves this cat

Anmerkung zur Regelkomponente

Alle für die Regekomponente relevanten Informationen finden Sie im o.a. Handout. Denken Sie daran, dass das Kasusmerkmal des Pronomens auch 'hochgereicht' werden muss (in der relevanten Regel: <NP Case> = <Pron Case>).

Anmerkung zum Lexikon

Wir haben im Handout 'Übergenerierung, Subkategorisierung und Merkmale' gesehen, nach welchem Muster die Lexikoneinträge aufgebaut sind, hier noch einmal einige Einträge:

loves	$\begin{bmatrix} \text{Kat} & \text{V} \\ \text{Agr} & \begin{bmatrix} \text{Num} & \text{sg} \\ \text{Pers} & 3 \end{bmatrix} \end{bmatrix}$	dogs	$\begin{bmatrix} \text{Kat} & \text{N} \\ \text{Agr} & \begin{bmatrix} \text{Num} & \text{pl} \\ \text{Pers} & 3 \end{bmatrix} \end{bmatrix}$	I	$\begin{bmatrix} \text{Kat} & \text{Pron} \\ \text{Agr} & \begin{bmatrix} \text{Num} & \text{sg} \\ \text{Pers} & 1 \end{bmatrix} \\ \text{Kas} & \text{subj} \end{bmatrix}$	ugly	[Kat A]
-------	---	------	---	---	--	------	---------

Die Frage ist: wie werden diese Informationen in PATR umgesetzt? Auf diese Frage gibt es mehrere Antworten, dh. es gibt verschiedene Möglichkeiten, wie genau die Merkmalstrukturen notiert werden müssen, um dann von der Regelkomponente entsprechend interpretiert werden zu können. Allen gemein ist, dass die einer Wortform zugeordneten Merkmalen durch das Tag *lf* ('f' für *Feature*) eingeleitet werden. Hinter dieses Tag werden dann in Spitzklammern die relevanten Merkmale gesetzt. Die oa. Wortformen können wir in PATR wie folgt erfassen:

\w loves	\w dogs	\w I	\w ugly
\c Vt	\c N	\c Pron	\c A
\f <Agr Num> = sg	\f <Agr Num> = pl	\f <Agr Num> = sg	
<Agr Pers> = 3	<Agr Pers> = 3	<Agr Pers> = 1	
		<Case> = subj	

Aufgabe 2: Fortgeschritten

Speichern Sie Grammatik 1 unter psg2.grm und psg2.lex ab und erweitern Sie diese Grammatik so, dass auch die folgenden Daten beschrieben werden:

1) I know that this cat loves her	3) *I know if this cat loves her
2) I wonder if this cat loves her	4) *I wonder that this cat loves her

Überlegen Sie sich, wie die Struktur von Sätzen, die selber einen Satz enthalten, aussehen könnte. Sehen Sie dazu Aufgabe 2 aus Teil 1. Achten Sie dann darauf, dass die Nebensatzleitende Konjunktion (*that* bzw. *if*) mit dem Verb des Matrixsatzes (*wonder* bzw. *know*) übereinstimmt: *know* funktioniert mit *that*, aber nicht mit *if*, bei *wonder* ist es genau andersherum. Führen Sie, um diese Beobachtung zu modellieren, entsprechende Merkmale ein.