

## 6. Übungsblatt - Abgabe: 06.12.2010

Ausführliche Informationen zu den vorgestellten Parsingverfahren finden Sie auch im Lehrbuch von Jurafsky und Martin in Kapitel 13 (2. Auflage) bzw. Kapitel 10 (1. Auflage).

### Aufgabe 6.1

Gegeben folgende Grammatik:

$S \rightarrow NP VP$   
 $VP \rightarrow VP ADV$   
 $VP \rightarrow geht$   
 $NP \rightarrow Hans$   
 $ADV \rightarrow heute \mid schnell$

Betrachten Sie den Satz „Hans geht schnell“.

- Warum kann ein Top-down-Parser (mit Backtracking, wie in der Vorlesung vorgestellt) mit dieser Grammatik den Satz nicht parsen? Wie lässt sich dieses Problem durch Anpassen der Grammatik beheben?
- Zeigen Sie, wie der Bottom-up-Parser (=Shift-Reduce-Parser) diesen Satz verarbeiten würde. Geben Sie für jeden Schritt den Stack und die verwendete Grammatikregel an.
- Modifizieren Sie den Top-down-Parser so, dass er zuerst immer das offene Nicht-Terminal ganz rechts expandiert, also von rechts nach links vorgeht. Parsen Sie den Satz mit dem veränderten Algorithmus. Geben Sie auch hier für jeden Schritt die benutzte Regel an.

### Aufgabe 6.2

Betrachten Sie den Earley-Algorithmus, wie er in der Vorlesung besprochen wurde (Er ist nicht ganz identisch mit dem Algorithmus im Jurafsky & Martin).

- Benutzen Sie den Earley-Algorithmus, um beide Analysen für den folgenden Satz zu finden:

„Der Student sieht den Mann mit der Brille.“

Benutzen Sie folgende Grammatik:

S → NP VP  
VP → V NP  
VP → VP PP  
NP → Det N  
NP → NP PP  
PP → P NP  
V → sieht  
N → Mann  
N → Student  
N → Brille  
Det → der  
Det → den  
P → mit

Geben Sie die Chart als Tabelle an (Achtung, es sind an die 60 einzelne Charteinträge!).

Hinweis: Der selbe Eintrag braucht nicht mehrfach auf einer Chartposition abgelegt zu werden.

- (b) Geben Sie durch geeignete (alternative) Verlinkung der vollständigen Charteinträge die beiden Lesarten des Satzes an (vgl. die letzte Vorlesungsfolie).
- (c) Stellen Sie für jede Lesart die Teilchart, die nur die vollständigen Einträge enthält, graphisch dar.

---

Abgabe in Gruppen von bis zu drei Studierenden bis **06.12.2010** 18 Uhr entweder als Email im pdf-Format an **i2cl@coli.uni-sb.de** oder auf Papier im Briefkasten an der Tür von Raum 1.04 in C7.2.