

Miriam Butt und Maribel Romero  
 G220/5109, G212/2728  
 {miriam.butt|maribel.romero}@uni-konstanz.de

**Advanced Computational Semantics**  
 Sommersemester 2008, Hauptseminar

Lösung zu Übung 7, 12.06.2008

**Aufgabe 2)**

a) Beispielsätze:

Many children laugh.

Pilots fly planes.

The man gives the monkey a telescope.

Does every child laugh?

A sad child does not laugh.

b) In Boxer werden Allquantoren durch eine Implikatur ausgedrückt, analog zu der Schreibweise, die wir auch im Kurs kennen gelernt haben. Was vielleicht sonst noch auffällt, ist dieser Abschnitt in der Repräsentation:

```
%%% Every child laughs .
...
pos(1001, 'DT'),
pos(1002, 'NN'),
pos(1003, 'VBZ'),
pos(1004, '.'')
...
```

Hier wird Part-of-Speech-Tagging vorgenommen, um der Syntax die Arbeit zu erleichtern. Die Tags sind Abkürzungen, die die genaue Wortart angeben (DT für Determinierer, NN für zählbares Nomen, VBZ für Verb in 3. Person Singular, usw.). Die Tags sind von der Penn Treebank übernommen worden, der bekanntesten Baumbank.<sup>1</sup>

---

<sup>1</sup>Die Tags aus dem Penn Treebank Projekt kann man zum Beispiel hier abrufen: [http://www.ling.upenn.edu/courses/Fall\\_2007/ling001/penn\\_treebank\\_pos.html](http://www.ling.upenn.edu/courses/Fall_2007/ling001/penn_treebank_pos.html)

c) Bei einer solchen Eingabe versteht `candc` den Punkt als zum letzten Wort dazugehörig und versucht, das so gebildete Wort (z.B. `runs.`) zu interpretieren, also mit dem Lexikon abzugleichen. Daher kommt bei der Eingabe nichts sinnvolles heraus. Da bei `candc` kein intelligenter Tokenizer (Programm zur Identifizierung von Wortgrenzen und Satzzeichen) eingebaut ist, kann es den Punkt nicht vom letzten Wort abtrennen; es schaut lediglich nach den Leerzeichen.