

Wesentliche Unterschiede zwischen dem Skript *Einführung in die Computerlinguistik* von Karsten Wenger und den Kursinhalten *Computerlinguistik II* Dozent Stefan Langer

Im Folgenden werden kurz die wesentlichen Unterschiede zwischen dem im Kurs Computerlinguistik II (CL II) entwickelten Prologprogramm und dem Skript von Karsten Wenger (*Einführung in die Computerlinguistik*) dargestellt. Hierbei geht es tatsächlich nur um die wirklich fundamentalen Differenzen, kleinere Unterschiede werden nicht erwähnt.

Verschränkung Syntaxmodul / Semantikmodul

Im Skript werden die Implementation der Syntax und der Semantik getrennt. Im Programm erfolgt zuerst die Syntaxanalyse, die einen Syntaxbaum erzeugt. Dieser Syntaxbaum ist Eingabe für das Modul, das die semantische Formel erzeugt. D.h. Syntaxregeln und Semantikregeln sind systematisch getrennt.

Im Kurs wird ein Programm vorgestellt, bei dem die Regeln zur Bildung der semantischen Formel direkt in die Syntaxregeln eingearbeitet sind, in Form eines zusätzlichen Arguments der syntaktischen Konstituenten. Syntaxanalyse und Aufbau der semantischen Formel laufen gleichzeitig ab.

Warum? Die Implementation im Skript zeigt klarer die Modularität des Systemaufbaus und ist dadurch etwas übersichtlicher. Das Programm im Kurs ist dafür reversibel - es ermöglicht auch die Generierung einer Aussage - d.h. eines Satzes - aus einer semantischen Formel. Damit kann das System nicht nur natürlichsprachliche Fragen verarbeiten, sondern gibt auch natürlichsprachliche Antworten.

Erzeugung des Syntaxbaums

Die endgültige Grammatik im Kurs generiert im Gegensatz zum Programm im Skript keinen Syntaxbaum.

Auch im Kurs wurde in einer Sitzung eine Grammatik vorgestellt, die für eine vorgenommene Analyse einen Syntaxbaum erzeugt. Allerdings ist diese Eigenschaft im endgültigen Programm unnötig und daher nicht vorhanden.

Warum? Dies liegt daran, dass Semantik und Syntax verschränkt sind (s.o.) - ein Syntaxbaum zur Eingabe in das Semantikmodul ist unnötig.

Sie sollten dennoch wissen, wie man einen Syntaxbaum in einer Prologgrammatik produziert.

Verwendung von Operatoren

Im Skript wird zur Darstellung der Lambda-Abstraktion der Operator " λ " verwendet, etwa in der Form " $X^\lambda A$ ". Im Kurs wird stattdessen ein Prolog-Term " $\text{lambda}(X,A)$ " gebildet. Dies ist nur ein Unterschied in der Form, nicht im Inhalt.

Warum? Da die meisten Studenten erst im Kurs Prolog lernen, reduziere ich die Vermittlung dieser Programmiersprache auf das absolut Notwendige. Definition von Operatoren durch den Benutzer gehört nicht unbedingt dazu.