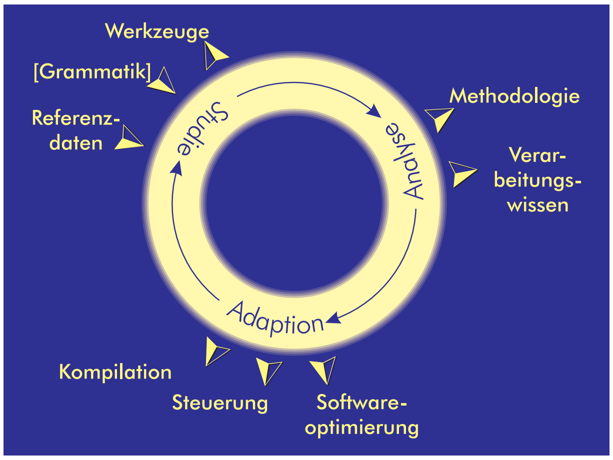
PERFORM: Performanzmodellierung für deklarative Grammatikmodelle



Im Projekt PERFORM (B4) werden Performanzmodelle für die effiziente und robuste Sprachverarbeitung mit HPSG-Grammatiken entwickelt, implementiert und evaluiert. Ausgehend von großen bestehenden Systemen und Grammatiken werden dabei ingenieuerwissenschaftliche (Steuerungs-, Kompilations- und Lernansätze) sowie kognitionswissenschaftliche Verfahren der Performanzmodellierung experimentell erprobt und empirisch genau analysiert. In der ersten Projektphase stand die (traditionell) computerlinguistische Perspektive deutlich im Vordergrund und ermöglichte die Erarbeitung von neuen, methodologisch und technologisch grundlegenden Verfahren für die analytisch genaue Arbeit mit den derzeit umfangreichsten HPSG-Grammatiken. In Abhängigkeit von Verarbeitungsmethoden und linguistischen Phänomenen wurde der Verbrauch der Ressourcen Arbeitsspeicher und Zeit in detaillierten Messungen bestimmt. Die Ergebnisse wurden ausgewertet und in Kombinationen von Verarbeitungsmethoden umgesetzt, die deutliche Performanzsteigerungen erbrachten. Zudem wurden Erkenntnisse über die Grenzen des rein ingenieurwissenschaftlichen Ansatzes zur Performanzmodellierung gewonnen.

Die Forsetzung des Projektes bezieht daher verstärkt empirische Erkenntnisse und Modellierungsansätze der Psycholinguistik in die Arbeit ein und zielt sowohl auf technologisch vorteilhaftere als auch auf kognitiv plausiblere Performanzmodelle, die im Zuge der geplanten Integration und Erweiterung (upscaling) aktueller Forschungsergebnisse synergetisch zusammenfließen sollen. Dabei werden überschneidungen, aber auch Divergenzen im Phänomenbereich, in den angewendeten Methoden und in den vorgeschlagenen Lösungen der aktuellen Computerlinguistik- und Psycholinguistikforschung identifiziert und wissenschaftlich erschlossen.

Der neue integrative Forschungsansatz zur Performanzmodellierung kombiniert die folgenden Elemente:

- fortgeschrittenste linguistische Grammatikmodelle (am konkreten Beispiel der HPSG);
- große deklarative Grammatiken (am Beispiel der ERGO- und VM-Grammatiken);
- mindestens ein großes parametrisierbares Verarbeitungssystem (am Beispiel von PAGE);
- ingenieurwissenschaftliche Methoden der Performanzverbesserung;
- psycholinguistische Methoden und Erkenntnisse zur Satzverarbeitung;
- systematisches Experimentieren mit diagnostischen Referenzdaten.

Die erwarteten Ergebnisse sind (i) weitere praktische Performanzverbesserungen in der Sprachverarbeitung mit HPSG, (ii) annotiertes Referenzmaterial zu problematischen Phänomenen aus der Linguistik, Psychologie und maschineller Sprachverarbeitung (ergänzt durch Daten zum Ressourcenverbrauch in Abhängigkeit von verschiedenen Verarbeitungsmethoden) und (iii) wichtige Erkenntnisse zur Eignung der HPSG und vergleichbarer Grammatikmodelle für die kognitionswissenschaftliche Modellierung von Satzverarbeitungsphänomenen.

Prof. Dr. Werner H. Tack
SFB 378 – der Sprecher
Universität des Saarlandes
Postfach 151150
D-66041 Saarbrücken

URL: http://www.coli.uni-sb.de/cl/projects/perform.html
Kontakt: M.A. Stephan Oepen
Telefon: +49-681-302-4176
Telefax: +49-681-302-4700
email: oe@coli.uni-sb.de