

**Wiebke Petersen****Proseminar****Fr 11-13****Raum 23.21.U1.85****Beginn 17.10.03****Teilgebiet:** Mathematische Linguistik / Computerlinguistik**Teilnehmerkreis:** Grundkenntnisse der Theorie formaler Sprachen sind wünschenswert aber nicht notwendig, da die benötigten Grundlagen in den ersten Sitzungen vermittelt bzw. wiederholt werden.**Inhalt/Seminarplan:** Die Computerlinguistik entwickelt Grammatikformalismen, die es ermöglichen, Modelle (Grammatiken) von Einzelsprachen zu formulieren. Mit Hilfe dieser Modelle ist es dann möglich, natürlichsprachliche Sätze zu analysieren oder zu generieren. Die algorithmische Komplexität der Sprachverarbeitung hängt von den formalen Eigenschaften des verwendeten Sprachmodells ab. Es ergibt sich die Frage: „Welche strukturelle Komplexität haben natürliche Sprachen?“ Eine Antwort darauf determiniert die gerade für Computerlinguisten bedeutende Frage nach dem geringsten nötigen Aufwand, um Sprache zu verarbeiten.

Die von Noam Chomsky entwickelte Hierarchie formaler Sprachen, erweist sich in der Informatik als probates Instrument zur Klassifikation von Programmiersprachen. Um die von Chomsky bereits 1957 aufgeworfene Frage, wo in dieser Hierarchie die natürlichen Sprachen anzusiedeln sind, entflammte jedoch ein heftiger Streit. Im Laufe des Seminars werden wir uns die zum Verständnis der Frage nötigen Grundlagen erarbeiten und den Streit an ausgewählten Veröffentlichungen nachvollziehen. Dabei wird sich herausstellen, dass gerade das Schweizerdeutsche Phänomene aufweist, die mächtigere Grammatiken erfordern, als zur Beschreibung des Englischen nötig scheinen. Das Schweizerdeutsche ist offenbar besonders komplex!

Die bisherigen Grammatiken für natürliche Sprachen sind aufgrund zu großer oder kleiner generativer Kapazität nur eingeschränkt in der Lage, natürliche Sprachen zu beschreiben. Unnötig große generative Kapazität eines Formalismus kann dazu führen, dass es keine effizienten Parsing-Algorithmen gibt, bzw. dass wichtige Fragen wie das Wortproblem, das Leerheitsproblem oder das Endlichkeitsproblem nicht mehr entscheidbar sind. Daher sind in den letzten Jahren diverse sogenannte schwach kontextsensitive Formalismen entwickelt worden, deren Mächtigkeit speziell auf die Analyse von natürlichen Sprachen zugeschnitten ist. Ein Beispiel dafür sind "Tree-Adjoining Grammars", die zum Ende des Seminars vorgestellt werden sollen.

**Folgeveranstaltungen:** keine**Arbeitsform:** Im ersten Teil des Seminars, wird es Übungsaufgaben geben, die von mir korrigiert und in der Folgesitzung besprochen werden. Darüber hinaus werden im zweiten Teil des Seminars Kurzreferate ausgegeben.**Leistungsnachweis (Zuordnung des Scheins zur Prüfungsordnung, Art des Scheinerwerbs):**

Beteiligungsnachweis: regelmäßige Teilnahme und Bearbeitung der Hausaufgaben

Leistungsnachweis: Hausarbeit

**Empfohlene Literatur:**B. H. Partee et al.: *Mathematical Methods in Linguistics*, Part E. Dordrecht et al.: Kluwer Acad. Publ., 1990.W. J. Savitch, E. Bach, W. Marsh und G. Safran-Naveh (Hrsg.): *The Formal Complexity of Natural Language*, D. Reidel Publishing Company, 1987.daraus im Besonderen: - G. K. Pullum und G. Gazdar: *Natural Languages and Context-Free Languages*.  
- S. M. Shieber: *Evidence Against the Context-Freeness of Natural Language*.A. K. Joshi und Y. Shabes: *Tree-Adjoining Grammars*. In G. Rosenberg und A. Salomaa (Hrsg.): *Handbook of Formal Languages*, 69-123, 1997.**Sprechzeiten:** Freitag 13-14 Uhr, Raum 23.21.04.45, Tel. 81-15295