Statistische Methoden in der Computerlinguistik

General Information

Organizer:	Bryan Jurish (`moocow@ling.uni-potsdam.de`)
Course Web Site:	http://www.ling.uni-potsdam.de/~moocow/class/statcl
Office:	II.24.182
Telephone:	(0331) 977 2726
Office Hours:	by appointment

Course Requirements

Proseminar:	in-class presentation,
	or small programming project,
	or written paper
Hauptseminar:	in-class presentation and written paper,
	or extended programming project

- Topics for in-class presentations, papers, and programming projects must be arranged with me.

- Students presenting a topic in class must meet with me to discuss the respective topic at least one week before the presentation is scheduled

The Goat Problem

Consider the following situation: you are a contestant in a TV game show. The host presents you with three closed doors. Behind one of these doors is the Grand Prize; behind each of the other two doors is a goat. The host explains your task to you: "You must first choose (but not open) one of the three doors *(first choice)*. After this, I will open a different door, behind which I guarantee will be a goat. You may then decide which of the remaining two unopened doors you wish to open *(second choice)* – whatever lies behind this door is the prize (or goat) you get."

Are you more likely to get the Grand Prize if you stick with your first choice?

Course Syllabus

Week	Date	Topic(s)	Reading
1	14.04.	Administrivia	-
2	21.04.	Probability Theory	M&S, Ch. 1
3	28.04.	Probability Theory	M&S, Ch. 2.1
4	05.05.	*no class*	
5	12.05.	Probability Theory	M&S, Ch. 2.1
6	19.05.	Information Theory	M&S, Ch. 2.2
7	26.05.	Corpus Linguistics	M&S, Ch. 4, Krenn and Samuelsson (1997, Appendix B)
8	02.06.	Collocations	M&S, Ch. 5
9	09.06.	N-Gram Models	M&S, Ch. 6
10	16.06.	N-Gram Models	M&S, Ch. 6
11	23.06.	Hidden Markov Models	M&S, Ch. 9, Charniak et al. (1993),
12	30.06.	Statistical Context-Free Grammars	M&S, Ch. 11, Charniak (1993)
13	07.07.	Statistical Parsing	M&S, Ch. 11
14	14.07.	Decision Trees	M&S, pp. 579–589, Mitchell (1997, Ch. 3)

References

E. Charniak. *Statistical Language Learning*. MIT Press, Cambridge, MA, 1993.

E. Charniak, C. Hendrickson, N. Jacobson, and M. Perkowitz. Equations for part-of-speech tagging. In *Proceedings of the Eleventh National Conference on Artificial Intelligence, Menlo Park*. AAAI Press/MIT Press, 1993. URL `http://www.cs.brown.edu/people/ec`.

B. Krenn and C. Samuelsson. *The Linguist's Guide to Statistics*. URL `http://www.coli.uni-sb.de/~christer/stat_cl.ps`. Unpublished manuscript, 1997.

C. D. Manning and H. Shütze. *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA, 1999.

T. M. Mitchell. *Machine Learning*. McGraw-Hill, Boston, MA, 1997.