CS674, Spring 2002
Lecture Handout, 4/24/02

Today's papers:

- Pasi Tapanainen and Atro Voutilainen, 1994. Tagging accurately - Don't guess if you know. *Proceedings of the Fourth ACL Conference on Applied Natural Language Processing.*
- Jean-Pierre Chanod and Pasi Tapanainen, 1995. Tagging French – comparing a statistical and a constraint-based method. *Proceedings of the EACL.*
- Christer Samuelsson and Atro Voutilainen, 1997. Comparing a linguistic and a stochastic tagger. *Proceedings of the 35th ACL/8th EACL* .

1. From Figure 3 of TV94. Errors are words for which the correct tag has been ruled out. The test corpus contained 26,711 words.

	Ambiguous words	Readings/word	Error rate (%)
Morphological analysis	37.6 %	1.77	0.09 %
(*) Morph. + combined system	0.0 %	1.00	1.46 %
Morph. + ENGCG	3.6 %	1.04	0.63 %
Morph. + XT	0.7 %	1.01	6.38 %

2. Table 1 from SV97 (generated from LaTeXsource provided by xxx.arXiv.cornell.edu). The two columns for the statistical tagger are for the Viterbi path version (δ) and the thresholded state-probability version (γ). Parentheses indicate interpolation. The held-out test set contained 55,000 words.

Ambiguity	Error rate (%)		
(Tags/word)	Statistical Tagger		EngCG
	(δ)	(γ)	
1.000	4.72	4.68	
1.012		4.20	
1.025		3.75	
1.026		(3.72)	0.43
1.035		(3.48)	0.29
1.038		3.40	
1.048		(3.20)	0.15
1.051		3.14	
1.059		(2.99)	0.12
1.065		2.87	
1.070		(2.80)	0.10
1.078		2.69	
1.093		2.55	