

Reading assignment 3

Due date: On CDF at 13h10, Wednesday 21 October 2015.
Late write-ups will not be accepted without a valid excuse.
This assignment is worth 5% of your final grade.

Read and write up this paper

Ratnaparkhi, Adwait. Statistical models for unsupervised prepositional phrase attachment. *Proceedings, 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics*, Montreal, 1998, 1079–1085.

Errata

Note the following typos in this paper, both in the second column of page 1082 (fourth page):

- On line 14, $p(\phi = true | n)$ should be $Pr(\phi = true | n)$.
- On the 4th line from the bottom, “ p , and $n2$ ” should be “and p ”.

Note also on the same page that the equation at the bottom of the first column continues at the top of the next column. The right-hand side is supposed to be the product of four probabilities.

Some terminology

General <ul style="list-style-type: none">• the PP attachment task• resource intensive• heuristic• portable	Machine learning <ul style="list-style-type: none">• classification task• supervised/unsupervised• bootstrapping• development/training/test [data] sets
Data-driven CL <ul style="list-style-type: none">• corpus-based• raw text• annotation• treebank/Penn Treebank• partial parser• chunker	Statistics and probability models <ul style="list-style-type: none">• formula for joint probability of dependent or independent events• random variable• argmax• back off• interpolation/smoothing