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


Errata

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

- On line 14, \( p(\phi = \text{true} \mid n) \) should be \( Pr(\phi = \text{true} \mid n) \).
- 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

<table>
<thead>
<tr>
<th>General</th>
<th>Machine learning</th>
</tr>
</thead>
</table>
| • the PP attachment task  
• resource intensive  
• heuristic  
• portable | • classification task  
• supervised/unsupervised  
• bootstrapping  
• development/training/test [data] sets |

<table>
<thead>
<tr>
<th>Data-driven CL</th>
<th>Statistics and probability models</th>
</tr>
</thead>
</table>
| • corpus-based  
• raw text  
• annotation  
• treebank/Penn Treebank  
• partial parser  
• chunker | • formula for joint probability of dependent or independent events  
• random variable  
• argmax  
• back off  
• interpolation/smoothing |

Copyright © 2015 Frank Rudzicz, Suzanne Stevenson, and Graeme Hirst. All rights reserved.