6.4 RAT

---

1. Consider the two word senses for the word “newspaper”:

   - **newspaper[1]**: a company that publishes written news.
   - **newspaper[2]**: a single physical item published by the company.

   What term best describes the relations between these two senses?

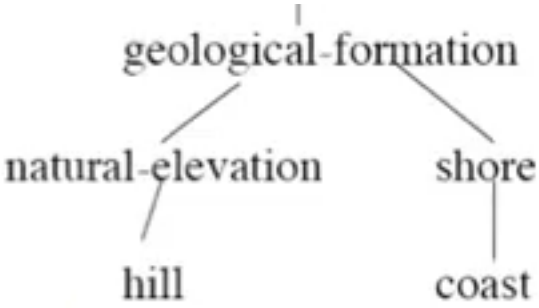
   a) homonymy

   b) polysemy

   c) metonymy

   d) hyponymy

2. Consider the information content of the two concepts “geological formation" and "hill". What is the relative information content of these two concepts?



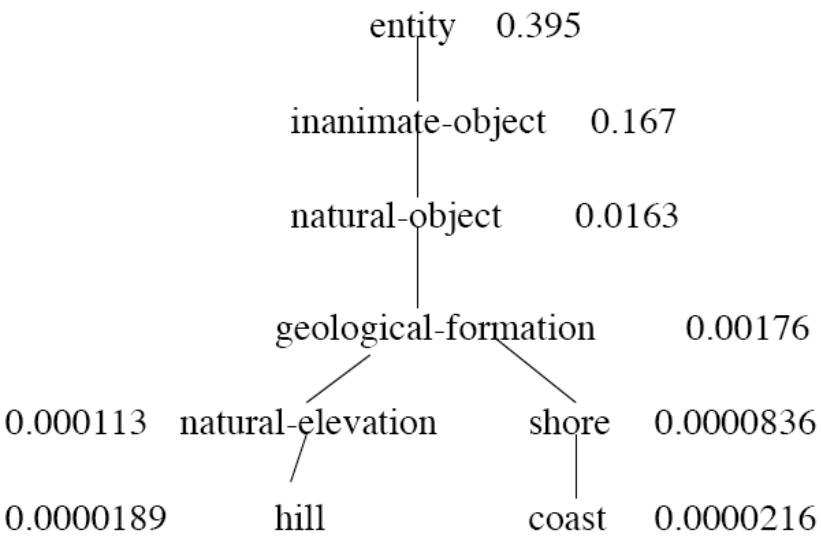
   a) IC(geological-formation) ≈ IC(hill)

   b) IC(geological-formation) < IC(hill)

   c) IC(geological-formation) > IC(hill)

   d) Not enough information

3. Given these probabilities:



P	$\log_2 P$
0.0000189	-15.69
0.0000216	-15.50
0.0000836	-13.55
0.000113	-13.11
0.00176	-9.15
0.0163	-5.94
0.100	-3.32
0.125	-3.00
0.167	-2.58
0.25	-2.00
0.4	-1.32
1.0	0.00
1.2	0.26
1.21	0.28
1.22	0.29
1.23	0.30
1.24	0.31
1.25	0.32

What is the Lin similarity between *hill* and *shore*?
Remember:

$$sim_{Lin}(c_1, c_2) = \frac{2 \log P(LCS(c_1, c_2))}{\log P(c_1) + \log P(c_2)}$$

4. Given the following word-context-matrix,

		Count(w,context)			
	computer	data	pinch	result	sugar
apricot	0	0	1	0	1
pineapple	0	0	1	0	1
digital	2	1	0	2	0
information	1	6	0	4	0

apply **add-one smoothing** and compute the PPMI(*information, result)*
remember:

$$pmi_{ij} = \log_2 \frac{p_{ij}}{p_{i*} p_{*j}}$$