***Assignment week 1 (Johan, Laurence, Jarik)***

*Exercise 1:*

1D: (a) the most likely 20 collocations using PMI

[('1', '/'), ('31', 'Lyon'), ('ARTHUR', 'CONAN'), ('Alicia', 'Whittington'), ('Amateur', 'Mendicant'), ('Amoy', 'River'), ('Anderson', '",'), ('Apache', 'Indians'), ('Arabian', 'Nights'), ('Arnsworth', 'Castle'), ('Atkinson', 'brothers'), ('BERYL', 'CORONET'), ('BLUE', 'CARBUNCLE'), ('BOSCOMBE', 'VALLEY'), ('Bengal', 'Artillery'), ('Bermuda', 'Dockyard'), ('Blue', 'Carbuncle'), ('Botany', 'variable'), ('CONAN', 'DOYLE'), ('COPPER', 'BEECHES')]

(b) the most likely 20 collocations using χ2[('1', '/'), ('31', 'Lyon'), ('ARTHUR', 'CONAN'), ('Alicia', 'Whittington'), ('Amateur', 'Mendicant'), ('Amoy', 'River'), ('Anderson', '",'), ('Apache', 'Indians'), ('Arabian', 'Nights'), ('Arnsworth', 'Castle'), ('Atkinson', 'brothers'), ('BERYL', 'CORONET'), ('BLUE', 'CARBUNCLE'), ('BOSCOMBE', 'VALLEY'), ('Bengal', 'Artillery'), ('Bermuda', 'Dockyard'), ('Beryl', 'Coronet'), ('Blue', 'Carbuncle'), ('Botany', 'variable'), ('Briony', 'Lodge')]

(c) just looking at them, there are surely differences in the rankings. Can you provide an explanation?

The first 16 items are equal, the last 4 differ from each other. The square-method doesn’t work well with bigrams not occuring very often, so maybe that’s the case why at the lower end of the top 20 differences between the PMI and square-method are being shown.