

## Assignment 2: Answer Sheet

Name

ID Number

### Part 1.b

What is the number of tokens in the corpus?  $N =$

### Part 1.d

What are the counts of the following words after lemmatization?

“that”:

“the”:

“negligent”:

“london”:

“.”:

### Part 1.e

How many times does the bigram (“fashionable”, “intelligence”) occur in windows of size 1?

How many times does the bigram (“high”, “chancery”) occur in windows of size 3?

### Part 1.f

Is (“mr.”, “skimpole”) among collocation candidates for window size 1? (Yes/No)

Is (“spontaneous”, “combustion”) among collocation candidates for window size 3? (Yes/No)

**Part 2.a**

Rank	Word	Score	$c(w_1w_2)$	$c(w_1)$	$c(w_2)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Table 1: Top collocation candidates, student's t-test (window size 1)

Rank	Word	Score	$c(w_1w_2)$	$c(w_1)$	$c(w_2)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Table 2: Top collocation candidates, chi-squared test (window size 1)

Rank	Word	Score	$c(w_1w_2)$	$c(w_1)$	$c(w_2)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Table 3: Top collocation candidates, likelihood ratio test (window size 1)

Rank	Word	Score	$c(w_1w_2)$	$c(w_1)$	$c(w_2)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Table 4: Top collocation candidates, student's t-test (window size 3)

Rank	Word	Score	$c(w_1w_2)$	$c(w_1)$	$c(w_2)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Table 5: Top collocation candidates, chi-squared test (window size 3)

Rank	Word	Score	$c(w_1w_2)$	$c(w_1)$	$c(w_2)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Table 6: Top collocation candidates, likelihood ratio test (window size 3)

### Part 3.a

Is (“cursitor”, “street”) a bigram according to the three tests with window size 1?

Test	Score	Threshold	Collocation? (Yes/No)
t-Test			
Chi-square Test			
Likelihood Ratio Test			

Table 7: Tests results for (“cursitor”, “street”)

Is (“good”, “one”) a bigram according to the three tests with window size 1?

Test	Score	Threshold	Collocation? (Yes/No)
t-Test			
Chi-square Test			
Likelihood Ratio Test			

Table 8: Tests results for (“good”, “one”)