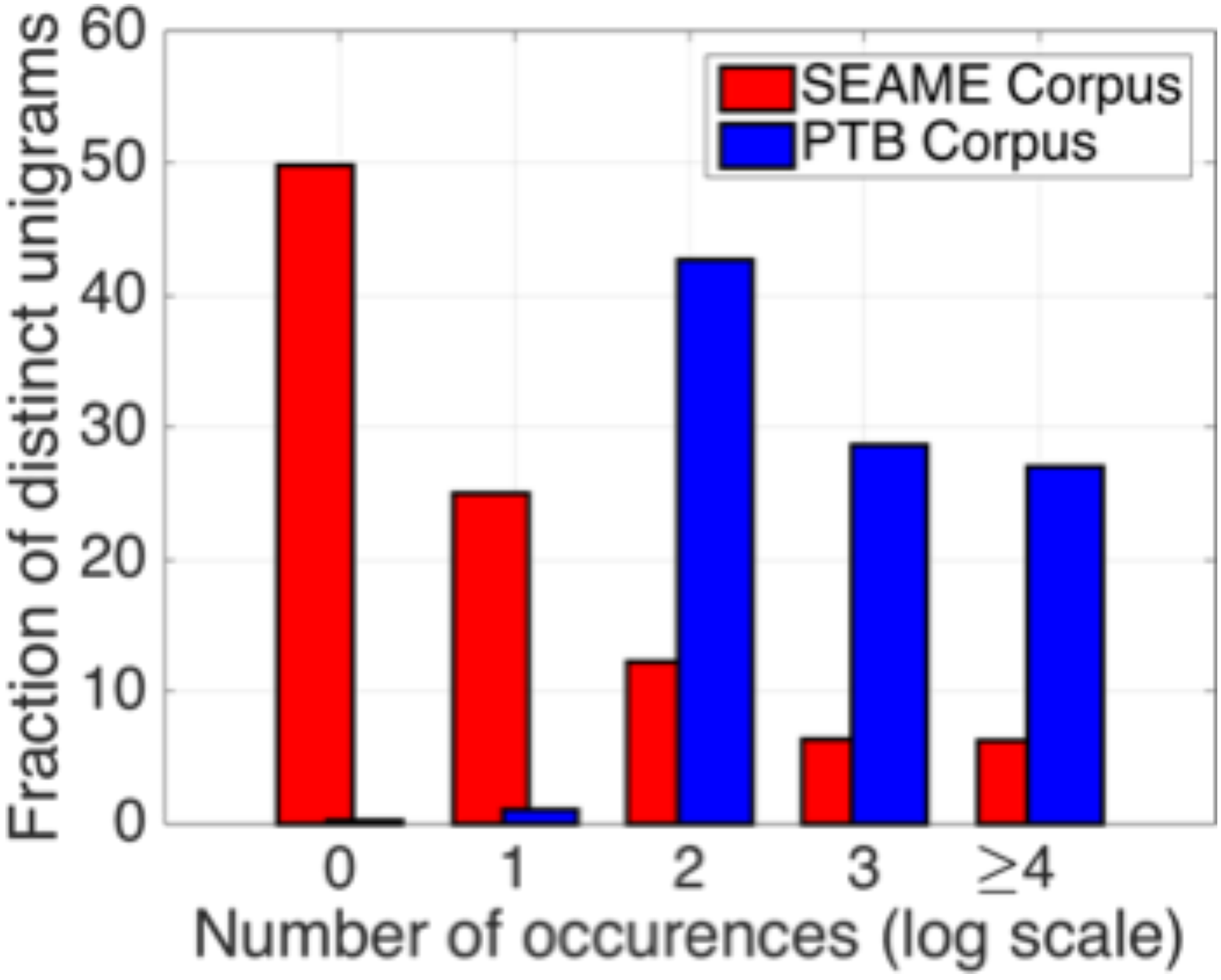
Generalisation Experiments

15

High frequency of low-occurring unigrams

Need for better generalisation


SEAME Corpus
PTB Corpus
Fraction of distinct unigrams
60
50
40
30
20
10
0
0
1
2
3
≥4
Number of occurences (log scale)

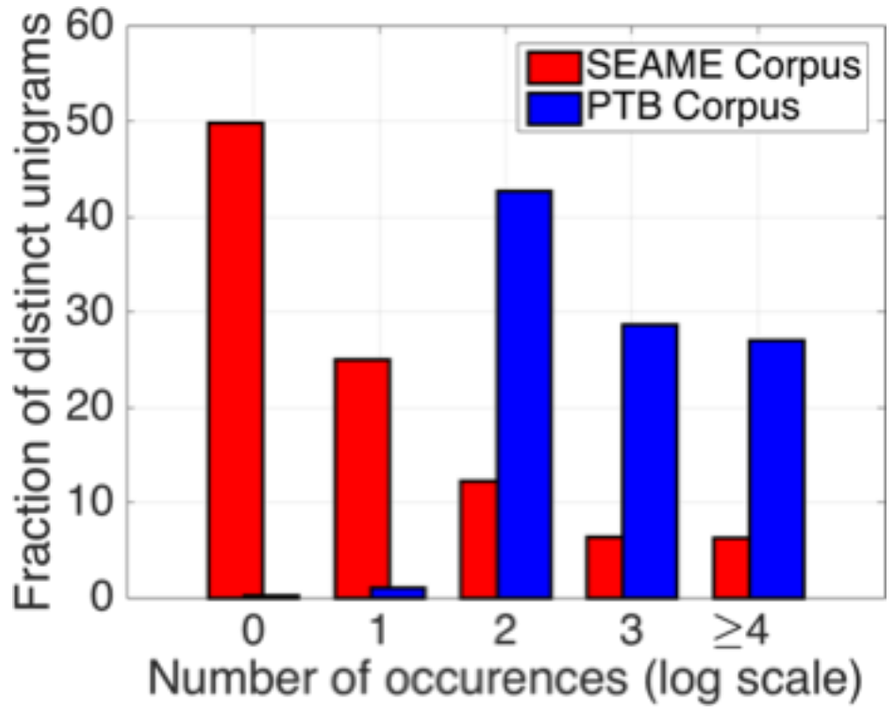
Training data	Dev		Test	
	mixed LM	DLM	mixed LM	DLM
Full	329.6725	**324.9268**	376.0968	**369.9355**
1/2	362.0966	**350.5860**	400.5831	**389.7618**
1/3	368.6205	**356.012**	408.562	**394.2131**

Gap increases!

Generalisation Experiments

High frequency of low-occurring unigrams

Need for better generalisation



Training data	Dev		Test	
	mixed LM	DLM	mixed LM	DLM
Full	329.6725	**324.9268**	376.0968	**369.9355**
1/2	362.0966	**350.5860**	400.5831	**389.7618**
1/3	368.6205	**356.012**	408.562	**394.2131**

Gap increases!

ASR Experiments

SAT : Speaker Adapted Triphone models
(Standard MFCC + delta + double delta features with MLLR transforms)

TDNN : Time delayed neural network