

Corpus	Source Words	Target Words
Training		
Canadian Hansard	161.7 m	144.5 m
Tune-1 / Tune-2 / Test		
Medical	53k / 43k / 35k	46k / 38k / 30k
Science	92k / 120k / 120k	75k / 101k / 101k
Language Modeling and Comparable Corpus Selection		
Medical	-	5.9 m
Science	-	3.6 m