

	IMDB		BBC		Penn TreeBank	
	# Sentences	# Words	# Sentences	# Words	# Sentences	# Words
Training	930,139	21M	37,207	890K	42,068	888K
Validation	152,987	3M	1,998	49K	3,370	70K
Test	151,987	3M	2,199	53K	3,761	79K