

Data	Corpus	Sentence	<u>Tokens</u>	
			En	Ja
Train	BTEC	464K	3.60M	4.97M
	KFTT	377K	7.77M	8.04M
Dev	BTEC	510	3.8K	5.3K
	KFTT	1160	24.3K	26.8K
Test	BTEC	508	3.8K	5.5K
	KFTT	1169	26.0K	28.4K