Corpus	Lang	train	dev	test	$|s|$	$|v|$
PT	Eng	40K	3K	4K	21.1	9999
SB	Eng	945K	10K	5.2K	10.39	47283
GW	Chi	531K	165K	260K	10.79	5123

$|s|$ denotes the average number of words in the sentences.

$|v|$ denotes the size of the vocabulary.

PT denotes Penn Treebank Corpus.

SB denotes Switchboard Corpus.

GW denotes Gigaword Corpus.