Stage	Collection Size	%
GIGAWORD (1995)	600,000	10%
Documents Retrieved	150,000	2.5%
Document Pairs Selected	10,000	0.25%
Sentence Pairs Extracted	270,000	0.1%
Fragment Pairs Extracted	90,000	0.01%