

Corpus (Abbreviation):	Date range	# Words	# Texts	Referen
Twitter (TW)	9/9/2008 to 3/3/2010	$9.07 \times 10^9$	$8.21 \times 10^8$ tweets	
Google Books Project, English (GB)	1520 to 2008	$3.61 \times 10^{11}$	$3.29 \times 10^6$ books	
The New York Times (NYT)	1/1/1987 to 6/30/2007	$1.02 \times 10^9$	$1.8 \times 10^6$ articles	
Music lyrics (ML)	1960 to 2007	$5.86 \times 10^7$	$2.95 \times 10^5$ songs	