Europarl Parallel Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		Czech ↔ English	
Sentences	1,965,734		2,007,723		1,920,209		646,605	
Words	56,895,229	54,420,026	60,125,563	55,642,101	50,486,398	53,008,851	14,946,399	17,376,433
Distinct words	176,258	117,481	140,915	118,404	381,583	115,966	172,461	63,039

News Commentary Parallel Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		Czech ↔ English		Russian ↔ English	
Sentences	174,441		157,168		178,221		140,324		150,217	
Words	5,116,388	4,520,796	4,928,135	4,066,721	4,597,904	4,541,058	3,206,423	3,507,249	3,841,950	4,008,949
Distinct words	84,273	61,693	69,028	58,295	142,461	61,761	138,991	54,270	145,997	57,991

Common Crawl Parallel Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		Czech ↔ English		Russian ↔ English	
Sentences	1,845,286		3,244,152		2,399,123		161,838		878,386	
Words	49,561,060	46,861,758	91,328,790	81,096,306	54,575,405	58,870,638	3,529,783	3,927,378	21,018,793	21,535,122
Distinct words	710,755	640,778	889,291	859,017	1,640,835	823,480	210,170	128,212	764,203	432,062

United Nations Parallel Corpus

	Spanish ↔ English		French ↔ English	
Sentences	11,196,913		12,886,831	
Words	318,788,686	365,127,098	411,916,781	360,341,450
Distinct words	593,567	581,339	565,553	666,077

10^9 Word Parallel Corpus

	French ↔ English	
Sentences	22,520,400	
Words	811,203,407	668,412,817
Distinct words	2,738,882	2,861,836

Yandex 1M Parallel Corpus

	Russian ↔ English	
Sentences	1,000,000	
Words	24,121,459	26,107,293
Distinct words	701,809	387,646

CzEng Parallel Corpus

	Czech ↔ English	
Sentences	14,833,358	
Words	200,658,857	228,040,794
Distinct words	1,389,803	920,824

Wiki Headlines Parallel Corpus

	Russian ↔ English	
Sentences	514,859	
Words	1,191,474	1,230,644
Distinct words	282,989	251,328

Europarl Language Model Data

	English	Spanish	French	German	Czech
Sentence	2,218,201	2,123,835	2,190,579	2,176,537	668,595
Words	59,848,044	60,476,282	63,439,791	53,534,167	14,946,399
Distinct words	123,059	181,837	145,496	394,781	172,461

News Language Model Data

	English	Spanish	French	German	Czech	Russian
Sentence	68,521,621	13,384,314	21,195,476	54,619,789	27,540,749	19,912,911
Words	1,613,778,461	386,014,234	524,541,570	983,818,841	456,271,247	351,595,790
Distinct words	3,392,137	1,163,825	1,590,187	6,814,953	2,655,813	2,195,112

News Test Set

	English	Spanish	French	German	Czech	Russian
Sentences	3000					
Words	64,810	73,659	73,659	63,412	57,050	58,327
Distinct words	8,935	10,601	11,441	12,189	15,324	15,736