Europarl Training Corpus

	Spanish ↔ English		French ↔ English		German ↔ English	
Sentences	1,650,152		1,683,156		1,540,549	
Words	47,694,560	46,078,122	50,964,362	47,145,288	40,756,801	43,037,967
Distinct words	173,033	95,305	123,639	95,846	316,365	92,464

News Commentary Training Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		Czech ↔ English	
Sentences	98,598		84,624		100,269		94,742	
Words	2,724,141	2,432,064	2,405,082	2,101,921	2,505,583	2,443,183	2,050,545	2,290,066
Distinct words	69,410	46,918	53,763	43,906	101,529	47,034	125,678	45,306

United Nations Training Corpus

	Spanish ↔ English		French ↔ English	
Sentences	6,222,450		7,230,217	
Words	213,877,170	190,978,737	243,465,100	216,052,412
Distinct words	441,517	361,734	402,491	412,815

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

CzEng Training Corpus

	Czech ↔ English	
Sentences	7,227,409	
Words	72,993,427	84,856,749
Distinct words	1,088,642	522,770

Europarl Language Model Data

	English	Spanish	French	German
Sentence	1,843,035	1,822,021	1,855,589	1,772,039
Words	50,132,615	51,223,902	54,273,514	43,781,217
Distinct words	99,206	178,934	127,689	328,628

News Language Model Data

	English	Spanish	French	German	Czech
Sentence	48,653,884	3,857,414	15,670,745	17,474,133	13,042,040
Words	1,148,480,525	106,716,219	382,563,246	321,165,206	205,614,201
Distinct words	1,451,719	548,169	998,595	1,855,993	1,715,376

News Test Set

	English	Spanish	French	German	Czech
Sentences	2489				
Words	62,988	65,654	68,107	62,390	53,171
Distinct words	9,457	11,409	10,775	12,718	15,825