Europarl Training Corpus

	Spanish ↔ English		French ↔ English		German ↔ English	
Sentences	1,411,589		1,428,799		1,418,115	
Words	40,067,498	41,042,070	44,692,992	40,067,498	39,516,645	37,431,872
Distinct words	154,971	108,116	129,166	107,733	320,180	104,269

News Commentary Training Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		Czech ↔ English	
Sentences	74,512		64,223		82,740		79,930	
Words	2,052,186	1,799,312	1,831,149	1,560,274	2,051,369	1,977,200	1,733,865	1,891,559
Distinct words	56,578	41,592	46,056	38,821	92,313	43,383	105,280	41,801

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

Hunglish Training Corpus

	Hungarian ↔ English	
Sentences	1,517,584	
Words	26,114,985	31,467,693
Distinct words	717,198	192,901

CzEng Training Corpus

	Czech ↔ English	
Sentences	1,096,940	
Words	15,336,783	17,909,979
Distinct words	339,683	129,176

Europarl Language Model Data

	English	Spanish	French	German
Sentence	1,658,841	1,607,419	1,676,435	1,713,715
Words	44,983,136	45,382,287	50,577,097	41,457,414
Distinct words	117,577	162,604	138,621	348,197

News Language Model Data

	English	Spanish	French	German	Czech	Hungarian
Sentence	21,232,163	1,626,538	6,722,485	10,193,376	5,116,211	4,209,121
Words	504,094,159	48,392,418	167,204,556	185,639,915	81,743,223	86,538,513
Distinct words	1,141,895	358,664	660,123	1,668,387	929,318	1,313,578

News Test Set

	English	Spanish	French	German	Czech	Hungarian	Italian
Sentences	2525						
Words	65,595	68,092	72,554	62,699	55,389	54,464	64,906
Distinct words	8,907	10,631	10,609	12,277	15,387	16,167	11,046

News System Combination Development Set

	English	Spanish	French	German	Czech	Hungarian	Italian
Sentences	502						
Words	11,843	12,499	12,988	11,235	9,997	9,628	11,833
Distinct words	2,940	3,176	3,202	3,471	4,121	4,133	3,318