Europarl Training Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		German ↔ Spanish	
Sentences	1,258,778		1,288,074		1,266,520		1,237,537	
Words	36,424,186	35,060,653	38,784,144	36,046,219	33,404,503	35,259,758	32,652,649	35,780,165
Distinct words	149,159	96,746	119,437	97,571	301,006	96,802	298,040	148,206

News Commentary Training Corpus

	Spanish ↔ English		French ↔ English		German ↔ English		German ↔ Spanish	
Sentences	64,308		55,030		72,291		63,312	
Words	1,759,972	1,544,633	1,528,159	1,329,940	1,784,456	1,718,561	1,597,152	1,751,215
Distinct words	52,832	38,787	42,385	36,032	84,700	40,553	78,658	52,397

Hunglish Training Corpus

	Hungarian ↔ English	
Sentences	1,517,584	
Words	26,082,667	31,458,540
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	1,412,546	1,426,427	1,438,435	1,467,291
Words	34,501,453	36,147,902	35,680,827	32,069,151
Distinct words	100,826	155,579	124,149	314,990

Europarl test set

	English	Spanish	French	German
Sentences	2,000			
Words	60,185	61,790	64,378	56,624
Distinct words	6,050	7,814	7,361	8,844

News Commentary test set

	English	Czech
Sentences	2,028	
Words	45,520	39,384
Distinct words	7,163	12,570

News Test Set

	English	Spanish	French	German	Czech	Hungarian
Sentences	2,051					
Words	43,482	47,155	46,183	41,175	36,359	35,513
Distinct words	7,807	8,973	8,898	10,569	12,732	13,144