Language	Sentence Pairs	Foreign Words	English Words	Corpora
Arabic	9,542,054	205,508,319	204,862,233	GALE
Bulgarian	406,934	9,306,037	9,886,401	Europarl-v7
Chinese	11,097,351	229,364,807	244,690,254	GALE
Czech	596,189	12,285,430	14,277,300	Europarl-v7
Dutch	1,997,775	49,533,217	50,661,711	Europarl-v7
Estonian	651,746	11,214,489	15,685,939	Europarl-v7
Finnish	1,924,942	32,330,289	47,526,505	Europarl-v7
French	52,004,519	932,475,412	821,546,279	Europarl-v7, 10^9 word parallel corpus, JRC, OpenSubtitles, UN
German	1,720,573	39,301,114	41,212,173	Europarl-v7
Greek	1,235,976	32,031,068	31,939,677	Europarl-v7
Hungarian	624,934	12,422,462	15,096,547	Europarl-v7
Italian	1,909,115	48,011,261	49,732,033	Europarl-v7
Latvian	637,599	11,957,078	15,412,186	Europarl-v7
Lithuanian	635,146	11,394,858	15,342,163	Europarl-v7
Polish	632,565	12,815,795	15,269,016	Europarl-v7
Portugese	1,960,407	49,961,396	49,283,373	Europarl-v7
Romanian	399,375	9,628,356	9,710,439	Europarl-v7
Russian	2,376,138	40,765,979	43,273,593	CommonCrawl, Yandex 1M corpus, News Commentary
Slovak	640,715	15,442,442	12,942,700	Europarl-v7
Slovenian	623,490	12,525,860	15,021,689	Europarl-v7
Swedish	1,862,234	45,767,032	41,602,279	Europarl-v7