

Dataset	Data Source	Vocabulary		Token		Sentence
		Source	Target	Source	Target	
EN-FR	europarl-v7	93,393	139,934	50,586,497	52,900,470	1,835,733
EN-DE	europarl-v7	93,033	323,367	48,625,466	46,484,368	1,763,744
EN-CS	europarl-v7	51,833	164,770	16,150,983	13,785,699	594,158
EN-FI	europarl-v7	91568	630184	48,584,379	34,819,783	1,779,397
EN-MG	CMU	57,668	76,469	1,592,66	2,023,336	80,306
EN-UR	NIST MT08	43,524	44,566	1,055,030	1,180,031	161,173