

	French	German	Spanish	Russian	Japanese	Chinese
Segments	10.2M	7.50M	5.67M	3.58M	1.70M	1.42M
Source Tokens	128M	79.9M	71.5M	34.7M	9.91M	8.14M
Target Tokens	118M	87.5M	67.6M	36.7M	19.1M	14.8M
	Arabic	Bulgarian	Czech	Korean	Tamil	Urdu
Segments	1.21M	909K	848K	756K	116K	52.1K
Source Tokens	13.1M	8.48M	7.42M	6.56M	1.01M	734K
Target Tokens	13.5M	8.61M	8.20M	7.58M	996K	685K
	Bengali	Farsi	Telugu	Somali	Kannada	Pashto
Segments	59.9K	44.2K	50.6K	52.6K	34.5K	28.0K
Source Tokens	573K	477K	336K	318K	305K	208K
Target Tokens	537K	459K	358K	325K	297K	218K