

Language Pair	# tokens		# types	
	Src	Tgt	Src	Tgt
Tr - En	2,7M	2,0M	171K	53K
Ar - En	3,9M	4,9M	220K	120K
Cs - En	2,0M	2,3M	118K	50K
De - En	4,0M	4,3M	144K	69K
It - En	3,5M	3,8M	95K	63K