

set	lines	Urdu		English	
		tokens	types	tokens	types
training	202k	1.7M	56k	1.7M	51k
dev	981	21k	4k	19k	4k
devtest	883	22k	4k	19-20k	4k
test	1792	42k	6k	38-41k	5k