

Language	Types	Tokens
Bengali	7,932	72,614
Tamil	14,264	70,258
Turkish	10,069	67,362
Zulu	13,628	58,027