

Corpus	Language	Tokens	Types	Letters
DUC'04	English	294 236	17 780	1 834 167
MEDICINA CLÍNICA	Spanish	125 024	9 657	793 937
PISTES	French	380 992	18 887	2 590 623