Languages 101-200 Source Tokens **English Tokens** Sentences 10^{7} 10^{6} 10^{5} 10^{4}