Monolingual Data vs. Lexicon Induction Performance

Accuracy in top-10, word frequency 200 (y-axis, 0 to 35)

Millions of Web Crawled Word Tokens (x-axis, 1e-01 to 1e+03)

Romanian, Uzbek, Albanian, Somali, Polish, Indonesian, Bosnian, Azeri, Serbian, Turkish, Bulgarian, Slovak, Welsh, Ukrainian, Latvian, Farsi, Nepali, Bengali, Hindi, Tamil