

Pipeline effectiveness in Sketch Engine - charts and tables

Matúš Kostka

November 2022

List of Tables

| | | |
|----|------------------------------------|---|
| 1 | FM measured pipelines | 5 |
| 2 | FM: Overall stats 10,000 tokens | 6 |
| 3 | FM: Overall stats 100,000 tokens | 6 |
| 4 | FM: Overall stats 1,000,000 tokens | 6 |
| 5 | SM: Overall stats 5,000 tokens | 6 |
| 6 | SM: Overall stats 10,000 tokens | 6 |
| 7 | SM: Overall stats 100,000 tokens | 6 |
| 8 | SM: Overall stats 250,000 tokens | 6 |
| 9 | SM: Overall stats 500,000 tokens | 7 |
| 10 | SM: Overall stats 750,000 tokens | 7 |
| 11 | SM: Overall stats 1,000,000 tokens | 7 |

List of Figures

| | | |
|----|--|----|
| 1 | FM: Execution time 10,000 tokens | 8 |
| 2 | FM: CPU usage 10,000 tokens | 9 |
| 3 | FM: RAM usage 10,000 tokens | 10 |
| 4 | FM: Execution time 100,000 tokens | 11 |
| 5 | FM: CPU usage 100,000 tokens | 12 |
| 6 | FM: RAM usage 100,000 tokens | 13 |
| 7 | FM: Execution time 1,000,000 tokens | 14 |
| 8 | FM: CPU usage 1,000,000 tokens | 15 |
| 9 | FM: RAM usage 1,000,000 tokens | 16 |
| 10 | FM: Initialization time | 17 |
| 11 | FM: Linear regression (bad) | 18 |
| 12 | SM: Execution time 5,000 tokens | 19 |
| 13 | SM: CPU usage 5,000 tokens | 20 |
| 14 | SM: RAM usage 5,000 tokens | 21 |
| 15 | SM: Execution time usage 10,000 tokens | 22 |
| 16 | SM: CPU usage 10,000 tokens | 23 |

| | | |
|----|--|----|
| 17 | SM: RAM usage 10,000 tokens | 24 |
| 18 | SM: Execution time 100,000 tokens | 25 |
| 19 | SM: CPU usage 100,000 tokens | 26 |
| 20 | SM: RAM usage 100,000 tokens | 27 |
| 21 | SM: Execution time 250,000 tokens | 28 |
| 22 | SM: CPU usage 250,000 tokens | 29 |
| 23 | SM: RAM usage 250,000 tokens | 30 |
| 24 | SM: Execution time 500,000 tokens | 31 |
| 25 | SM: CPU usage 500,000 tokens | 32 |
| 26 | SM: RAM usage 500,000 tokens | 33 |
| 27 | SM: Execution time 750,000 tokens | 34 |
| 28 | SM: CPU usage 750,000 tokens | 35 |
| 29 | SM: RAM usage 750,000 tokens | 36 |
| 30 | SM: Execution time 1,000,000 tokens | 37 |
| 31 | SM: CPU usage 1,000,000 tokens | 38 |
| 32 | SM: RAM usage 1,000,000 tokens | 39 |
| 33 | SM: Linear regression (better) | 40 |
| 34 | SM: Linear regression Bulgarian | 41 |
| 35 | SM: Linear regression Catalan | 41 |
| 36 | SM: Linear regression French V2 | 42 |
| 37 | SM: Linear regression French V3 | 42 |
| 38 | SM: Linear regression Italian V1 | 43 |
| 39 | SM: Linear regression Portuguese | 43 |
| 40 | SM: Linear regression Spanish V5 | 44 |
| 41 | SM: Linear regression Spanish V6 | 44 |
| 42 | SM: Linear regression Korean | 45 |
| 43 | SM: Linear regression Hungarian | 45 |
| 44 | SM: Linear regression Serbian hunpos | 46 |
| 45 | SM: Linear regression Swedish | 46 |
| 46 | SM: Linear regression Japanese | 47 |
| 47 | SM: Linear regression Croatian | 47 |
| 48 | SM: Linear regression Czech | 48 |
| 49 | SM: Linear regression Danish | 48 |
| 50 | SM: Linear regression Dutch | 49 |
| 51 | SM: Linear regression English | 49 |
| 52 | SM: Linear regression Estonian | 50 |
| 53 | SM: Linear regression Finnish | 50 |
| 54 | SM: Linear regression German | 51 |
| 55 | SM: Linear regression Greek tt2 | 51 |
| 56 | SM: Linear regression Greek | 52 |
| 57 | SM: Linear regression Hebrew tok1 | 52 |
| 58 | SM: Linear regression Hindi | 53 |
| 59 | SM: Linear regression Irish | 53 |
| 60 | SM: Linear regression Italian | 54 |
| 61 | SM: Linear regression Lao | 54 |
| 62 | SM: Linear regression Maldivian | 55 |

| | | |
|----|---|----|
| 63 | SM: Linear regression Polish | 55 |
| 64 | SM: Linear regression Romanian | 56 |
| 65 | SM: Linear regression Russian | 56 |
| 66 | SM: Linear regression Scottish | 57 |
| 67 | SM: Linear regression Serbian rft1 | 57 |
| 68 | SM: Linear regression Slovak | 58 |
| 69 | SM: Linear regression Slovene | 58 |
| 70 | SM: Linear regression Swahili | 59 |
| 71 | SM: Linear regression Tagalog | 59 |
| 72 | SM: Linear regression Tibetan | 60 |
| 73 | SM: Linear regression Ukrainian | 60 |
| 74 | SM: Linear regression Yoruba | 61 |
| 75 | SM: Linear regression Arabic stanford | 61 |
| 76 | SM: Linear regression Hebrew yap | 62 |

Introduction

This document contains charts and tables from pipeline measurements. There are 2 types of measurement:

1. Original one which measure: **1; 10,000; 100,000** and **1,000,000** tokens.
2. More complex with: **5,000; 10,000** and **50,000 - 1,000,000** tokens.

The second one contains a linear regression with low error rate, the linear regression show dependency of execution time (in minutes) on number of tokens.

First measurement is refered as FM (first measurement)

Second measurement is refered as SM (second measurement).

In attachments is **Statfile.big.ods** with detailed data and tables from whole measurement.

https://drive.google.com/file/d/16Rya-3dPsFP33_Wg0BcfTc4i65tmvqNw/view?usp=share_link

| No. | Language | Pipeline |
|-----|---------------------|---|
| 1. | Arabic | /opt/stanford_nlp_tools/stanford_arabic.sh |
| 2. | Bulgarian | /opt/skepipelines/bulgarian_tt2/pipeline.sh |
| 3. | Catalan | /opt/freeling_pipe/freeling_catalan_v2.py |
| 4. | Czech | /opt/skepipelines/czech_mj2/pipeline.sh |
| 5. | Danish | /opt/skepipelines/danish_tt2/pipeline.sh |
| 6. | German | /opt/skepipelines/german_rft3/pipeline.sh |
| 7. | Irish | /opt/skepipelines/irish_rft1/pipeline.sh |
| 8. | Maldivian | /opt/skepipelines/maldivian_tok1/pipeline.sh |
| 9. | Greek | /opt/skepipelines/greek_tt2/pipeline.sh |
| 10. | Greek | /opt/skepipelines/greek_tt3/pipeline.sh |
| 11. | English | /opt/skepipelines/english_tt31/pipeline.sh |
| 12. | Spanish | /opt/freeling_pipe/freeling_spanish_v6.sh |
| 13. | Spanish | /opt/freeling_pipe/freeling_spanish_v5.py |
| 14. | Estonian | /opt/skepipelines/estonian_estnltk3/pipeline.sh |
| 15. | Finnish | /opt/skepipelines/finnish_tt3/pipeline.sh |
| 16. | French | /opt/freeling_pipe/freeling_french_v3.sh |
| 17. | French | /opt/freeling_pipe/freeling_french_v2.py |
| 18. | Scottish | /opt/skepipelines/scottish_tok1/pipeline.sh |
| 19. | Hebrew | /opt/yap_pipe/yap_he_v1.sh |
| 20. | Hebrew | /opt/skepipelines/hebrew_tok1/pipeline.sh |
| 21. | Hindi | /opt/skepipelines/hindi_siva1/pipeline.sh |
| 22. | Croatian | /opt/skepipelines/croatian_rft1/pipeline.sh |
| 23. | Hungarian | /opt/hunpos_pipe/hunpos_hungarian_v2.sh |
| 24. | Icelandic | /opt/icenlp/pipeline/pipeline_v1.sh |
| 25. | Italian | /opt/freeling_pipe/freeling_italian_v1.sh |
| 26. | Italian | /opt/skepipelines/italian_tt2/pipeline.sh |
| 27. | Japan | /opt/mecab_pipe/mecab_unidic_comainu_jpn.sh |
| 28. | Korean | /opt/HanNanumTagger/runHanNanumTagger_v2.sh |
| 29. | Lao | /opt/skepipelines/lao_rft1/pipeline.sh |
| 30. | Marathi | /opt/skepipelines/universal_tok3/pipeline.sh |
| 31. | Dutch | /opt/skepipelines/dutch_tt3/pipeline.sh |
| 32. | Norwegian-Nynorsk | /opt/obtagger_pipe/oslo-bergen-tagger_nynorsk_v2.sh |
| 33. | Polish | /opt/skepipelines/polish_rft1/pipeline.sh |
| 34. | Portuguese | /opt/freeling_pipe/freeling_portuguese_v4.py |
| 35. | Romanian | /opt/skepipelines/romanian_tt1/pipeline.sh |
| 36. | Russian | /opt/skepipelines/russian_rft1/pipeline.sh |
| 37. | Slovak | /opt/skepipelines/slovak_rft1/pipeline.sh |
| 38. | Slovene | /opt/skepipelines/slovene_tt2/pipeline.sh |
| 39. | Serbian | /opt/skepipelines/serbian_rft1/pipeline.sh |
| 40. | Serbian | /opt/hunpos_pipe/hunpos_serbian.sh |
| 41. | Swahili | /opt/skepipelines/swahili_tt1/pipeline.sh |
| 42. | Thai | /opt/skepipelines/thai_sw1/pipeline.sh |
| 43. | Tibetan | /opt/skepipelines/tibetan_tt1/pipeline.sh |
| 44. | Tagalog | /opt/skepipelines/tagalog_stf1/pipeline.sh |
| 45. | Turkish | /opt/skepipelines/turkish_tm2/pipeline.sh |
| 46. | Ukrainian | /opt/skepipelines/ukrainian_rft1/pipeline.sh |
| 47. | Yoruba | /opt/skepipelines/yoruba_tok1/pipeline.sh |
| 48. | Chinese-traditional | /opt/corenlp_pipe/stf_chinese_v3.sh |
| 49. | Chinese-simplified | /opt/corenlp_pipe/stf_chinese_v3.sh |

Table 1: **FM measured pipelines**

Table 2: FM: Overall stats 10,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 2.47 | 762.7 | 78.27 | 54.46 |
| CPU usage (%) | 0 | 100 | 26 | 18 |
| RAM usage (GB) | 0.007 | 2.326 | 0.252 | 0.141 |

Table 3: FM: Overall stats 100,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 4.21 | 3300.14 | 271.02 | 108.76 |
| CPU usage (%) | 0 | 127 | 40 | 38 |
| RAM usage (GB) | 0.008 | 5.443 | 0.388 | 0.187 |

Table 4: FM: Overall stats 1,000,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 23.7 | 8109.08 | 1020.49 | 395.78 |
| CPU usage (%) | 0 | 222 | 75 | 77 |
| RAM usage (GB) | 0.008 | 5.629 | 0.733 | 0.209 |

Table 5: SM: Overall stats 5,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 0.09 | 379.19 | 28.87 | 7.98 |
| CPU usage (%) | 3 | 228 | 82 | 75 |
| RAM usage (GB) | 0.008 | 2.149 | 0.285 | 0.194 |

Table 6: SM: Overall stats 10,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 0.16 | 551.47 | 39.59 | 9.99 |
| CPU usage (%) | 2 | 214 | 85 | 77 |
| RAM usage (GB) | 0.008 | 2.141 | 0.289 | 0.195 |

Table 7: SM: Overall stats 100,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 1.28 | 3784.92 | 240.98 | 48.01 |
| CPU usage (%) | 0 | 244 | 89 | 83 |
| RAM usage (GB) | 0.008 | 3.717 | 0.411 | 0.227 |

Table 8: SM: Overall stats 250,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|---------|
| Execution time (sec) | 3.03 | 7346.25 | 452.56 | 102.325 |
| CPU usage (%) | 39 | 238 | 90 | 89 |
| RAM usage (GB) | 0.008 | 5.497 | 0.607 | 0.249 |

Table 9: SM: Overall stats 500,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 6.02 | 3136.27 | 413.54 | 165.21 |
| CPU usage (%) | 50 | 241 | 101 | 100 |
| RAM usage (GB) | 0.008 | 3.316 | 0.628 | 0.243 |

Table 10: SM: Overall stats 750,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 9.41 | 4630.42 | 489.01 | 179.07 |
| CPU usage (%) | 51 | 241 | 108 | 101 |
| RAM usage (GB) | 0.008 | 6.330 | 0.850 | 0.243 |

Table 11: SM: Overall stats 1,000,000 tokens

| | Min value | Max value | Average | Median |
|-----------------------------|-----------|-----------|---------|--------|
| Execution time (sec) | 11.77 | 2851.95 | 475.74 | 171.24 |
| CPU usage (%) | 57 | 224 | 111 | 105 |
| RAM usage (GB) | 0.007 | 5.600 | 0.866 | 0.243 |

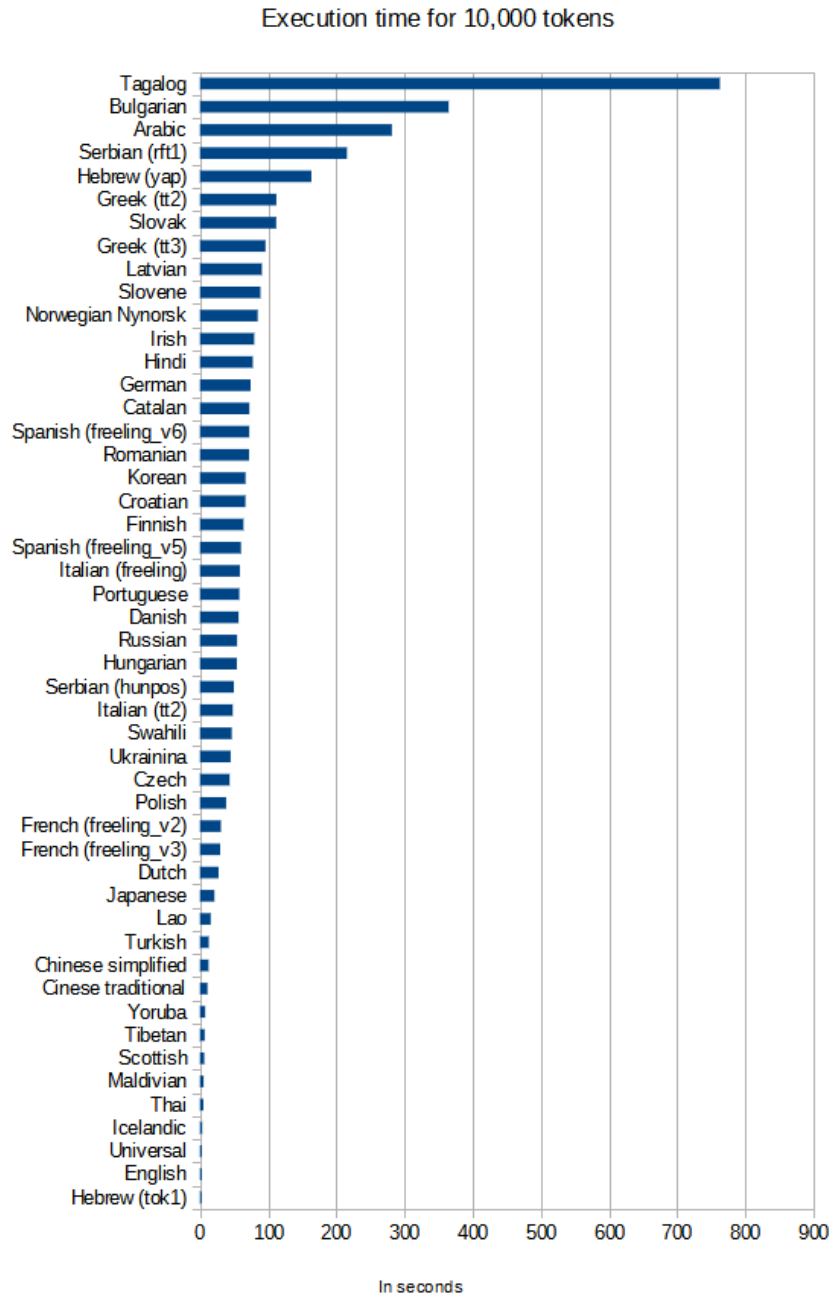


Figure 1: FM: Execution time 10,000 tokens

CPU usage for 10,000 tokens

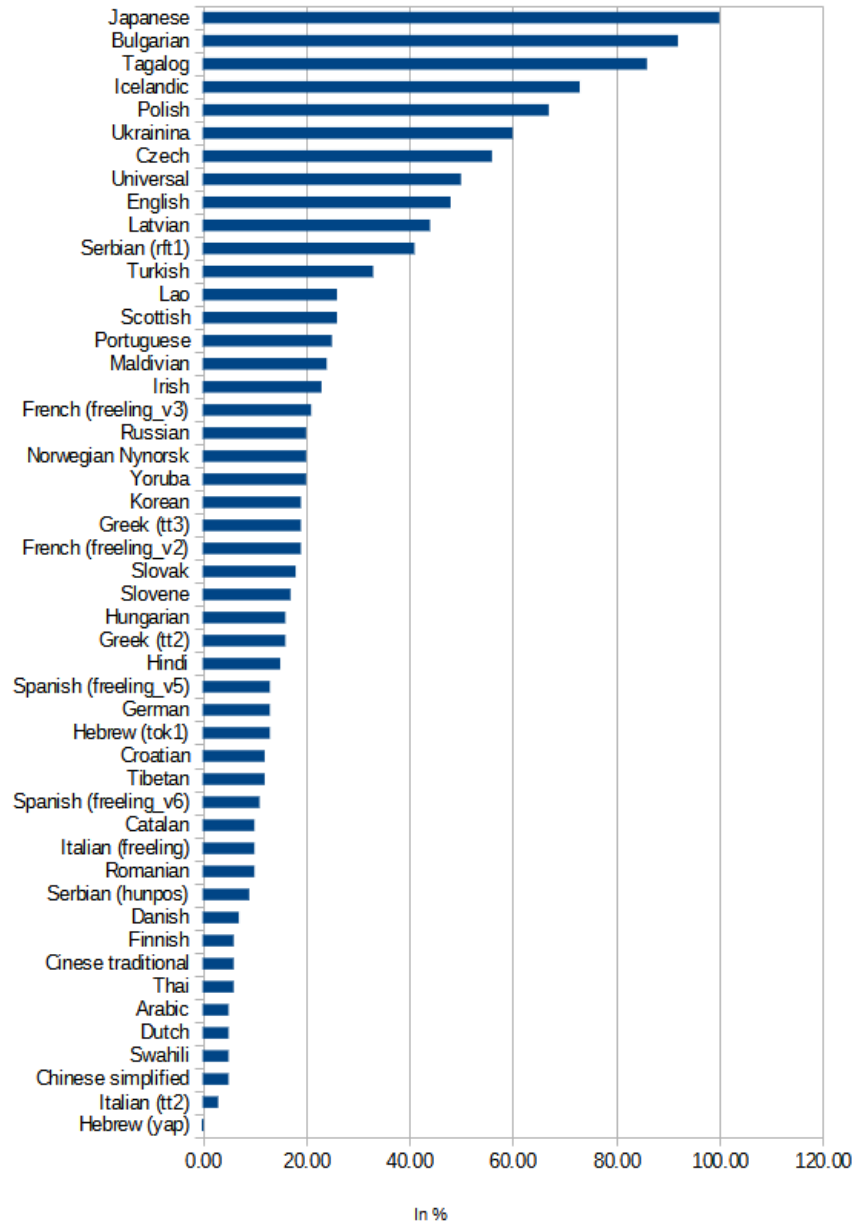


Figure 2: FM: CPU usage 10,000 tokens

RAM usage for 10,000 tokens

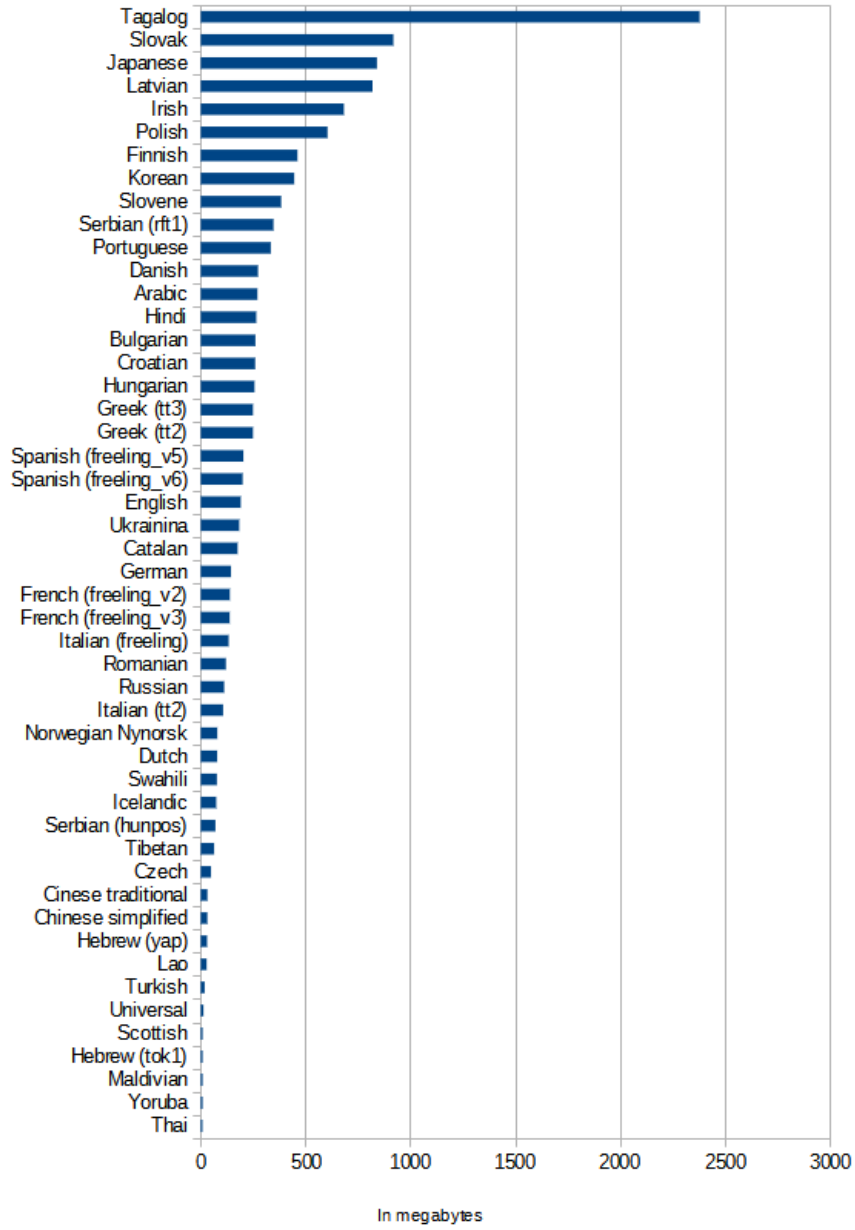


Figure 3: FM: RAM usage 10,000 tokens

Execution time for 100,000 tokens

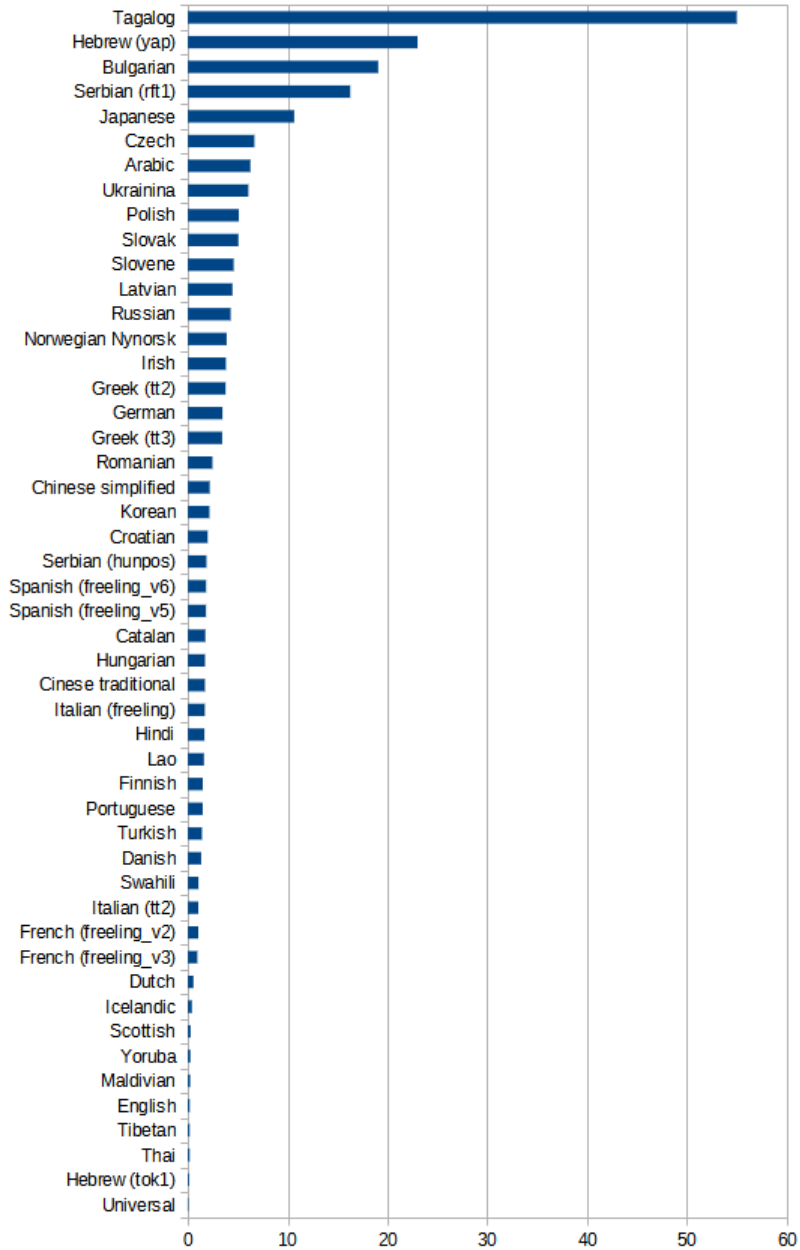


Figure 4: FM: Execution time 100,000 tokens

CPU usage for 100,000

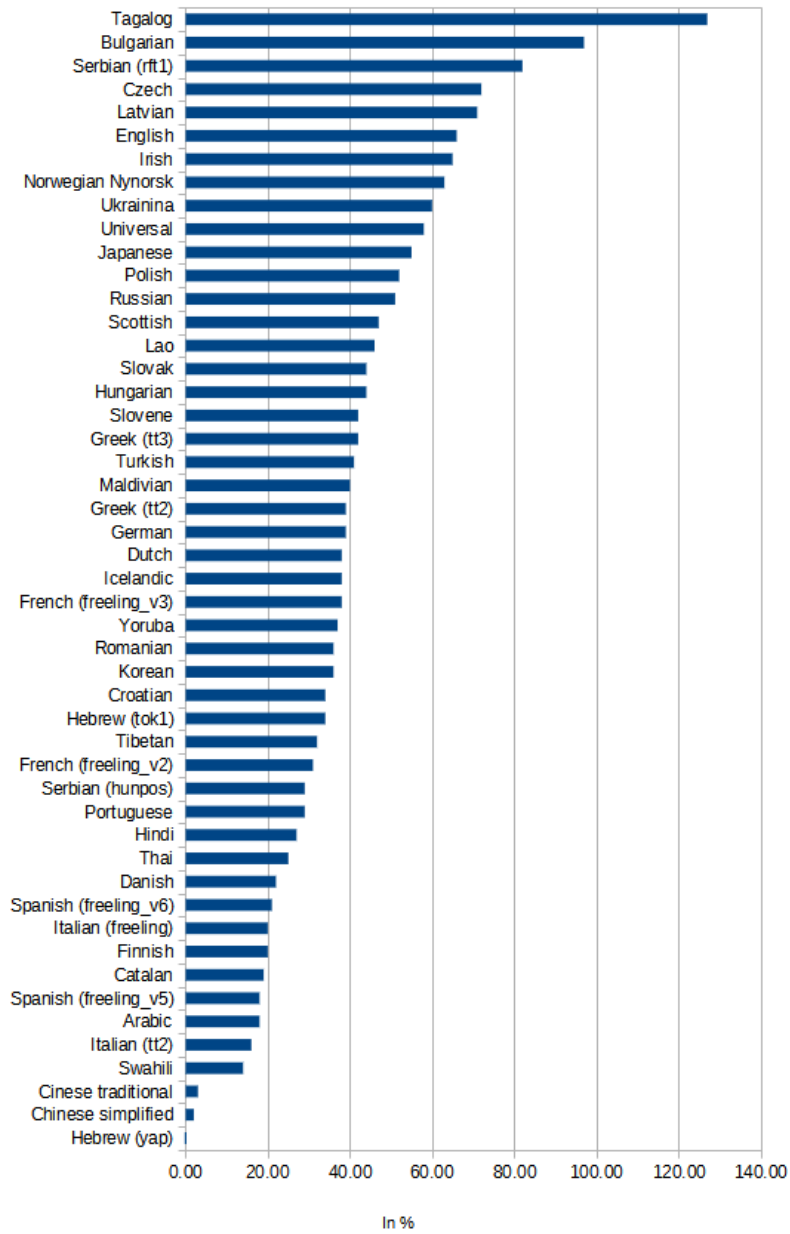


Figure 5: FM: CPU usage 100,000 tokens

RAM usage for 100,000 tokens

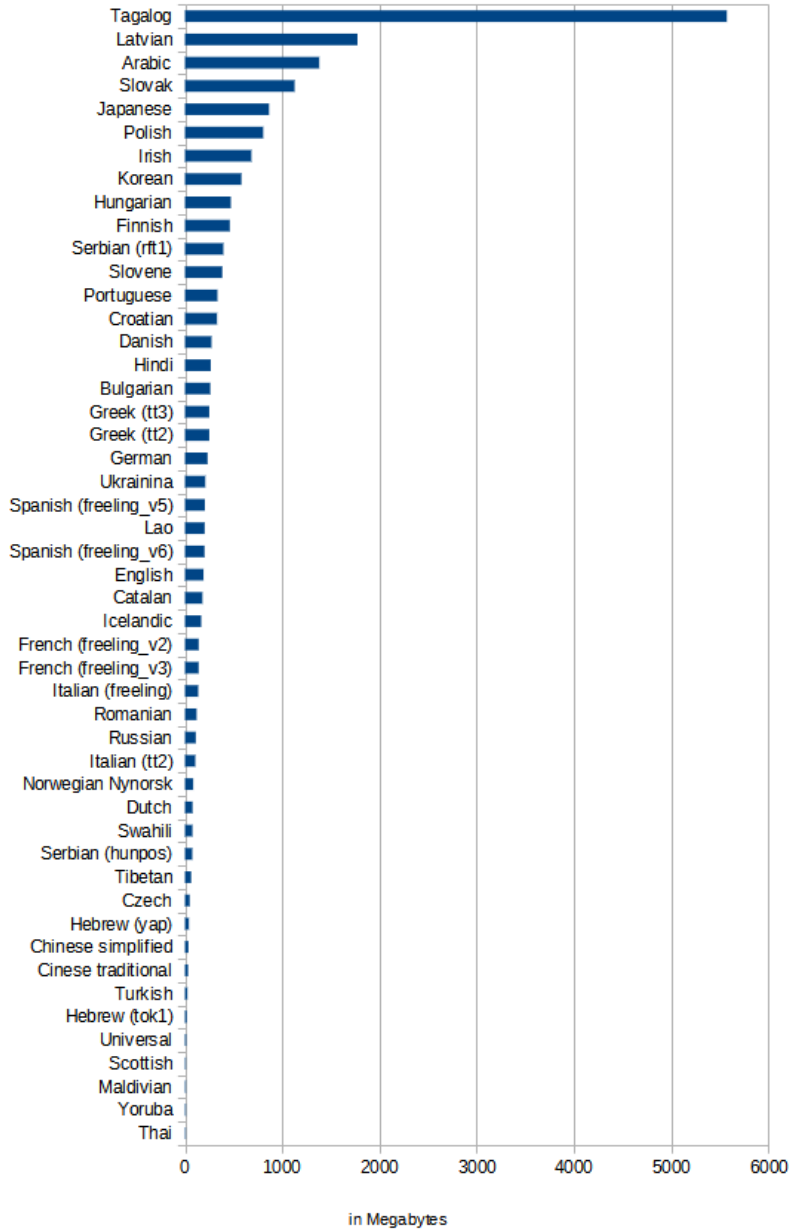


Figure 6: FM: RAM usage 100,000 tokens

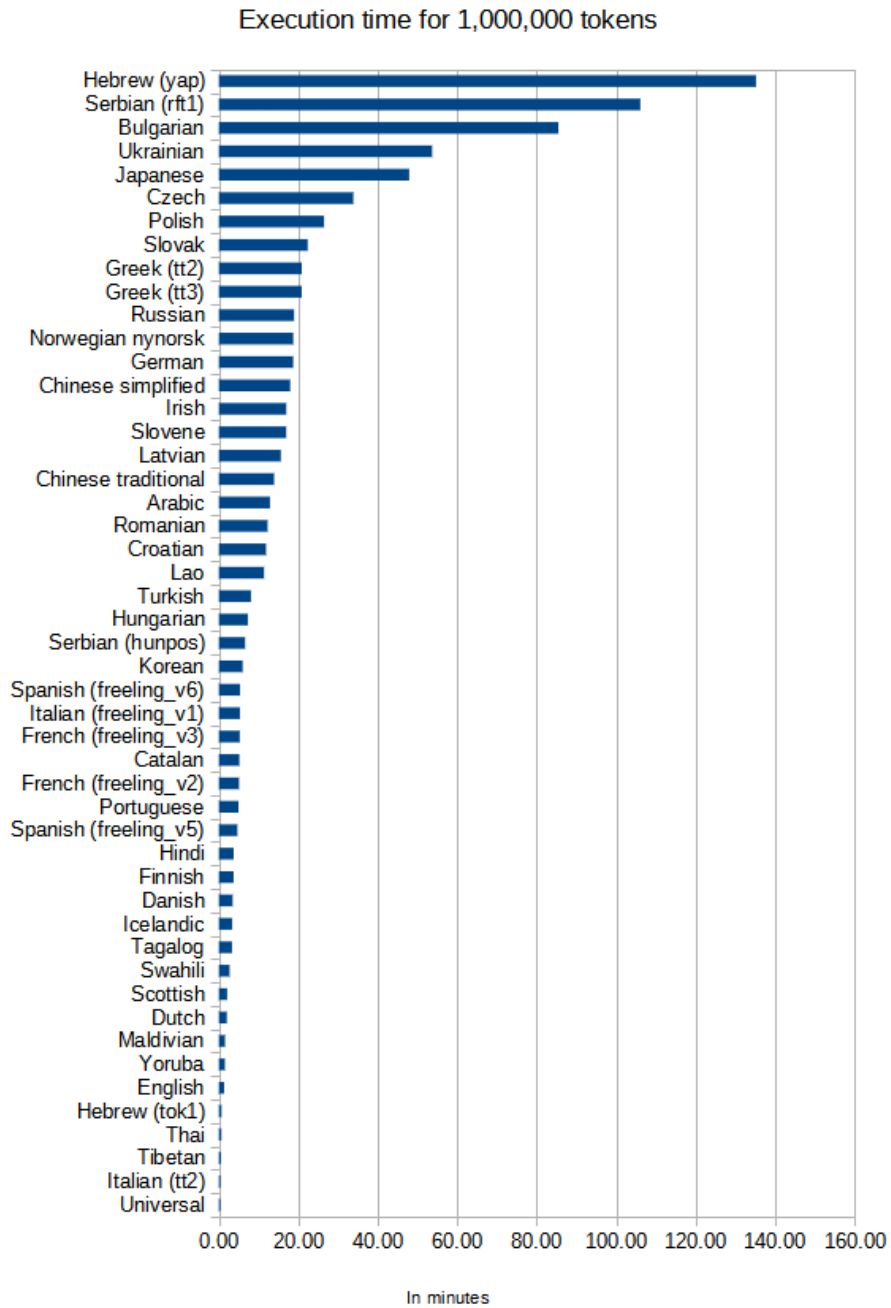


Figure 7: FM: Execution time 1,000,000 tokens

CPU usage for 1,000,000 tokens

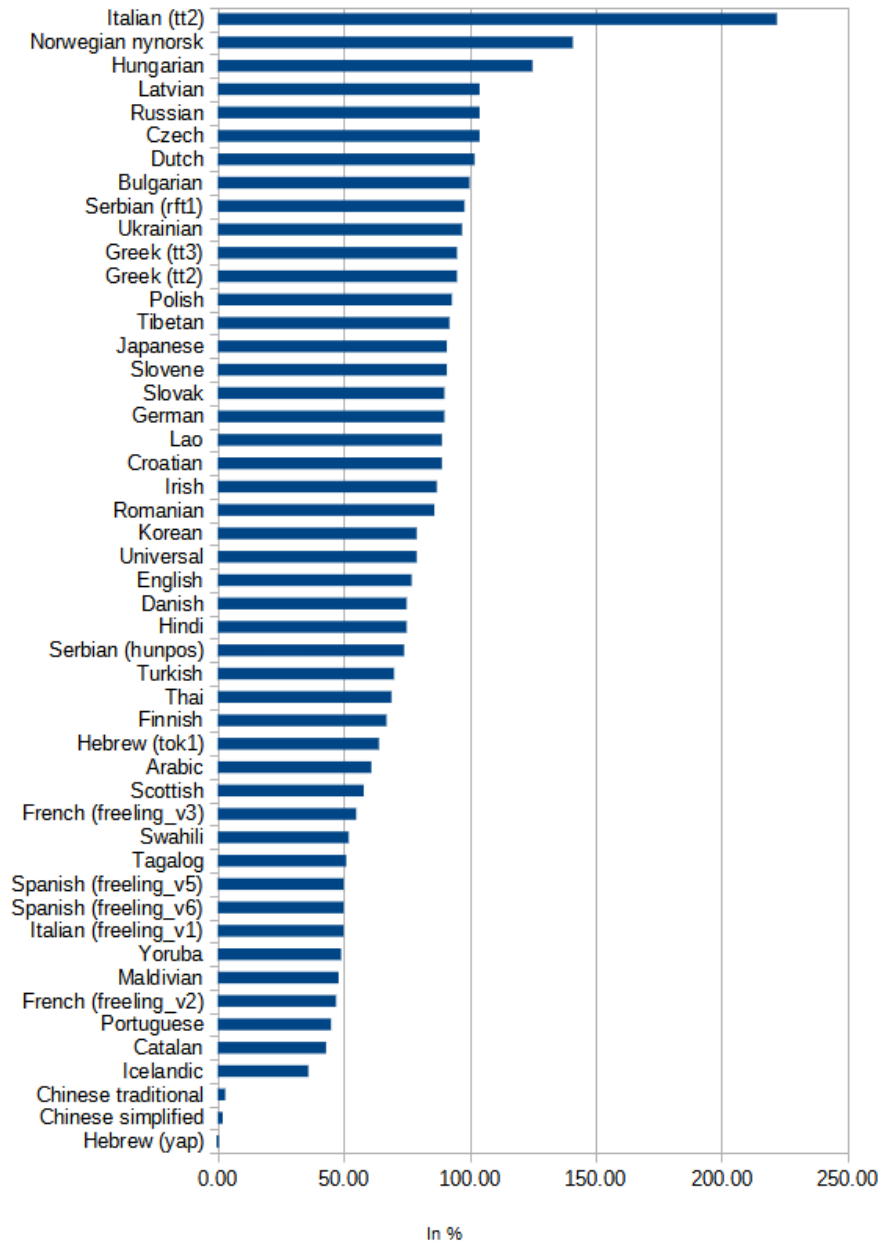


Figure 8: FM: CPU usage 1,000,000 tokens

Maximal RAM usage for 1 million tokens

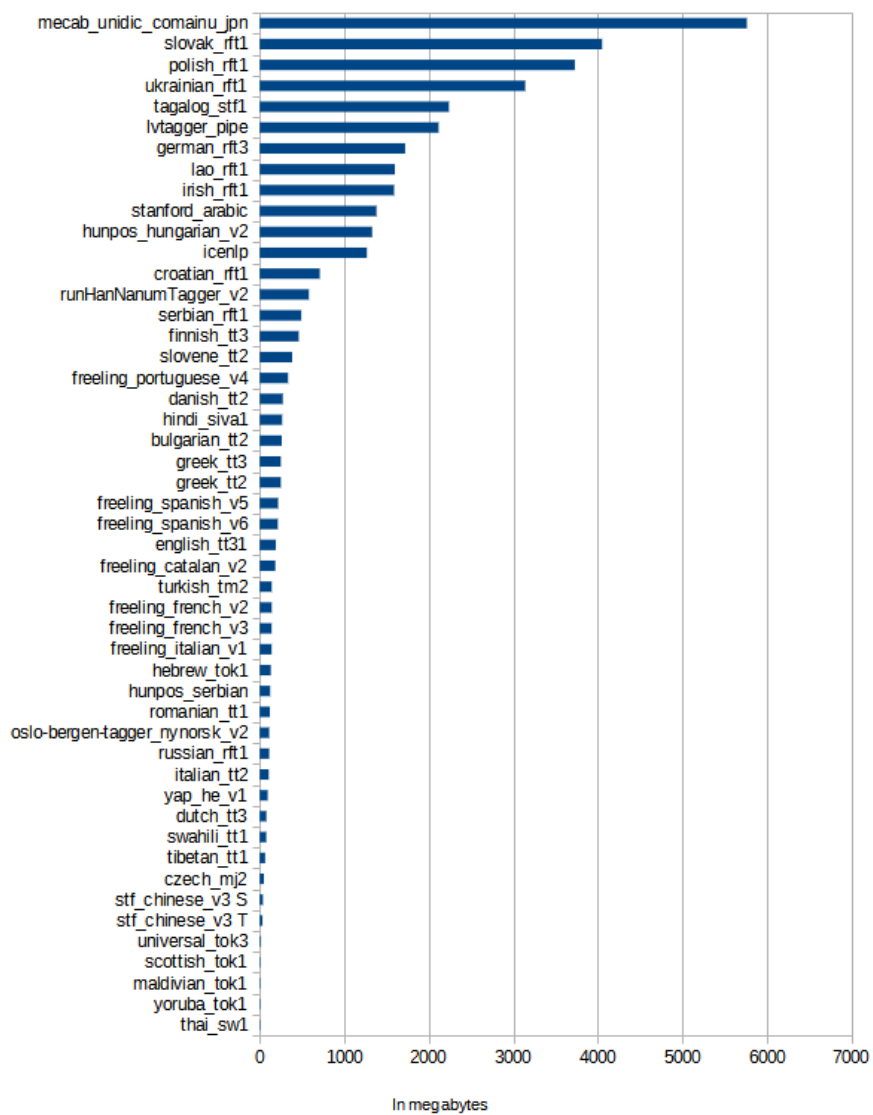


Figure 9: FM: RAM usage 1,000,000 tokens



Figure 10: FM: Initialization time

| Language | Slope | Intercept | Error |
|-----------------------|------------|-----------|---------|
| Tagalog | 0.00058 | 787.45 | 1757.59 |
| Bulgarian | 0.0048 | 313.65 | 330.8 |
| Japanese | 0.0027 | 110.093 | 210.69 |
| Czech | 0.0019 | 71.22 | 111.16 |
| Arabic | 0.00011 | 85.94 | 96.89 |
| Latvian | 0.00085 | 85.73 | 90.109 |
| Irish | 0.00095 | 65.26 | 66.37 |
| Norwegian Nynorsk | 0.001 | 64.79 | 63.51 |
| Greek (tt2) | 0.001187 | 67.84 | 60.17 |
| German | 0.00107 | 53.03 | 50.66 |
| Greek (tt3) | 0.001197 | 55.61 | 49.17 |
| Hindi | 0.00017 | 52.53 | 43.65 |
| Spanish (freeling_v6) | 0.00027 | 50.38 | 42.48 |
| Spanish (freeling_v5) | 0.00023 | 47.31 | 42.06 |
| Catalan | 0.00026 | 49.25 | 41.4 |
| Finnish | 0.00018 | 43.53 | 38.78 |
| Serbian (rft1) | 0.00035 | 39.46 | 37.82 |
| Italian (freeling) | 0.00027 | 42.75 | 37.44 |
| Danish | 0.00017 | 38.22 | 33.96 |
| Italian (tt2) | -8.56 | 36.25 | 33.15 |
| Croatian | 0.00067 | 36.37 | 32.3 |
| Portuguese | 0.00025 | 40.41 | 31.15 |
| Korean | 0.0000041 | 24.42 | 26.69 |
| French (freeling_v2) | 0.00028 | 20.78 | 16.92 |
| Ukrainian | 0.00024 | -8.87 | 16.62 |
| Lao | 0.00067 | 12.23 | 14.87 |
| French (freeling_v3) | 0.00029 | 18.07 | 14.33 |
| Slovene | 0.00016 | 11.27 | 14.1 |
| Chinese | 0.0001 | 8.29 | 12.64 |
| Romanian | 0.00014 | -5.37 | 10.3 |
| Slovak | 0.00016 | 11.55 | 8.78 |
| Universal | 3.35 | 3.35 | 3.77 |
| Serbian (hunpos) | 0.000029 | 2.35 | 2.81 |
| Russian | 0.000082 | 1.2 | 2.71 |
| Maldivian | 8.96 | 3.17 | 2.47 |
| Polish | 0.00022 | 5.95 | 2.33 |
| Hebrew | 3.76 | 1.88 | 1.86 |
| Icelandic | 0.00019 | 1.67 | 1.84 |
| English | 7.8 | 1.59 | 1.3 |
| Dutch | 0.00001001 | 0.93 | 1.24 |
| Scottish | 0.000015 | 1.11 | 0.57 |
| Swahili | 0.0000117 | 0.95 | 0.41 |
| Hebrew (yap) | 0.0000014 | 0.4 | 0.35 |
| Yoruba | 0.000016 | 0.801 | 0.32 |
| Turkish | 0.00007 | 0.57 | 0.27 |
| Thai | 1.67 | 0.21 | 0.201 |
| Tibetan | 4.78 | 0.07 | 0.19 |

Figure 11: FM: Linear regression (bad)

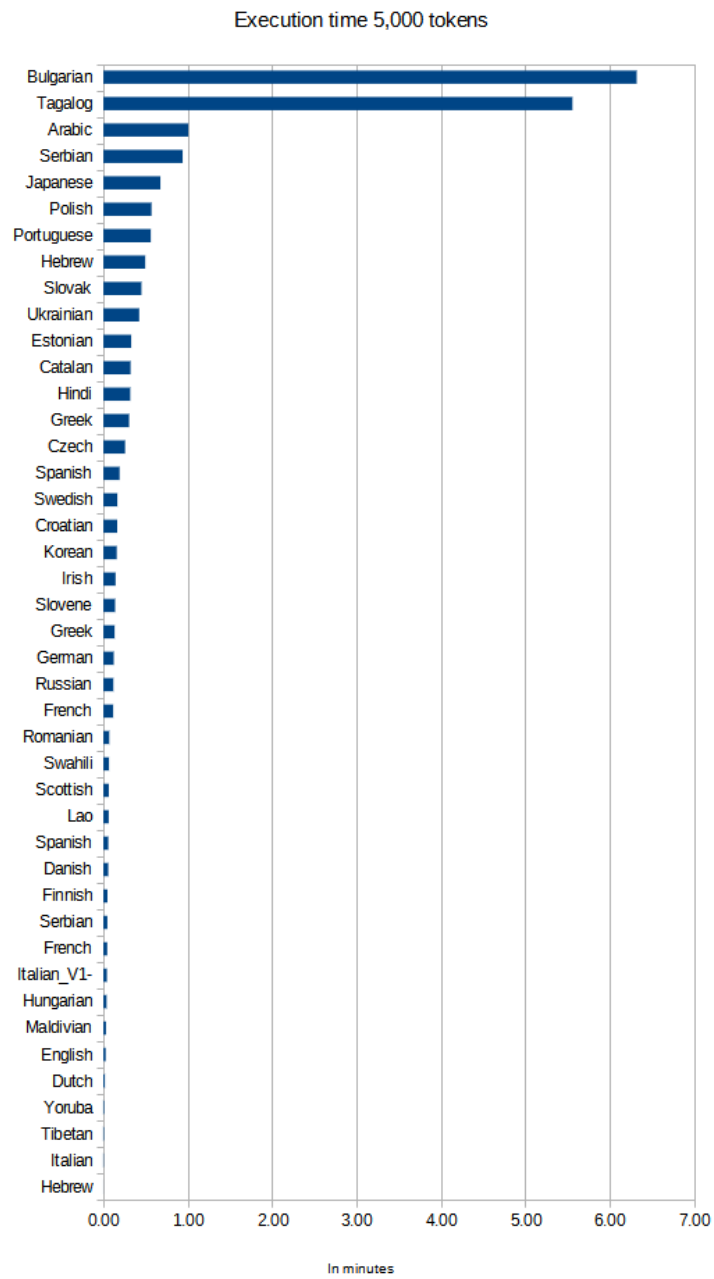


Figure 12: SM: Execution time 5,000 tokens

CPU usage 5,000 tokens

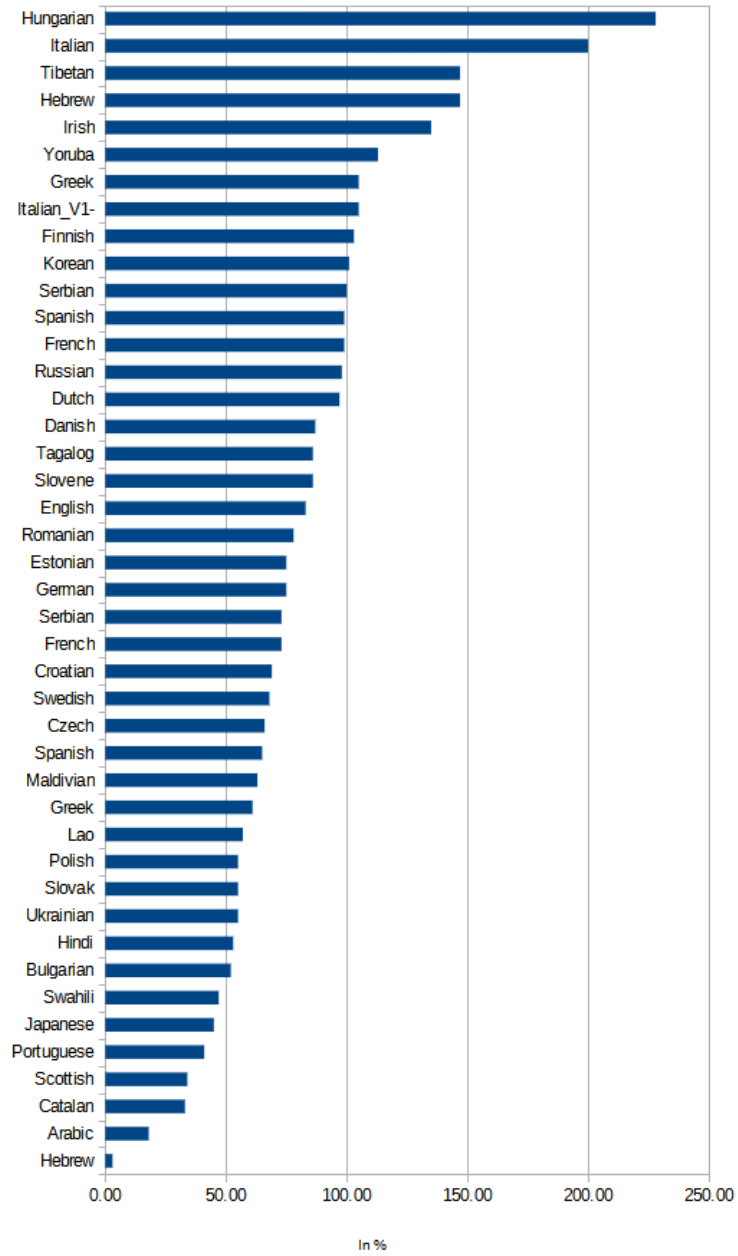


Figure 13: SM: CPU usage 5,000 tokens

RAM usage 5,000 tokens

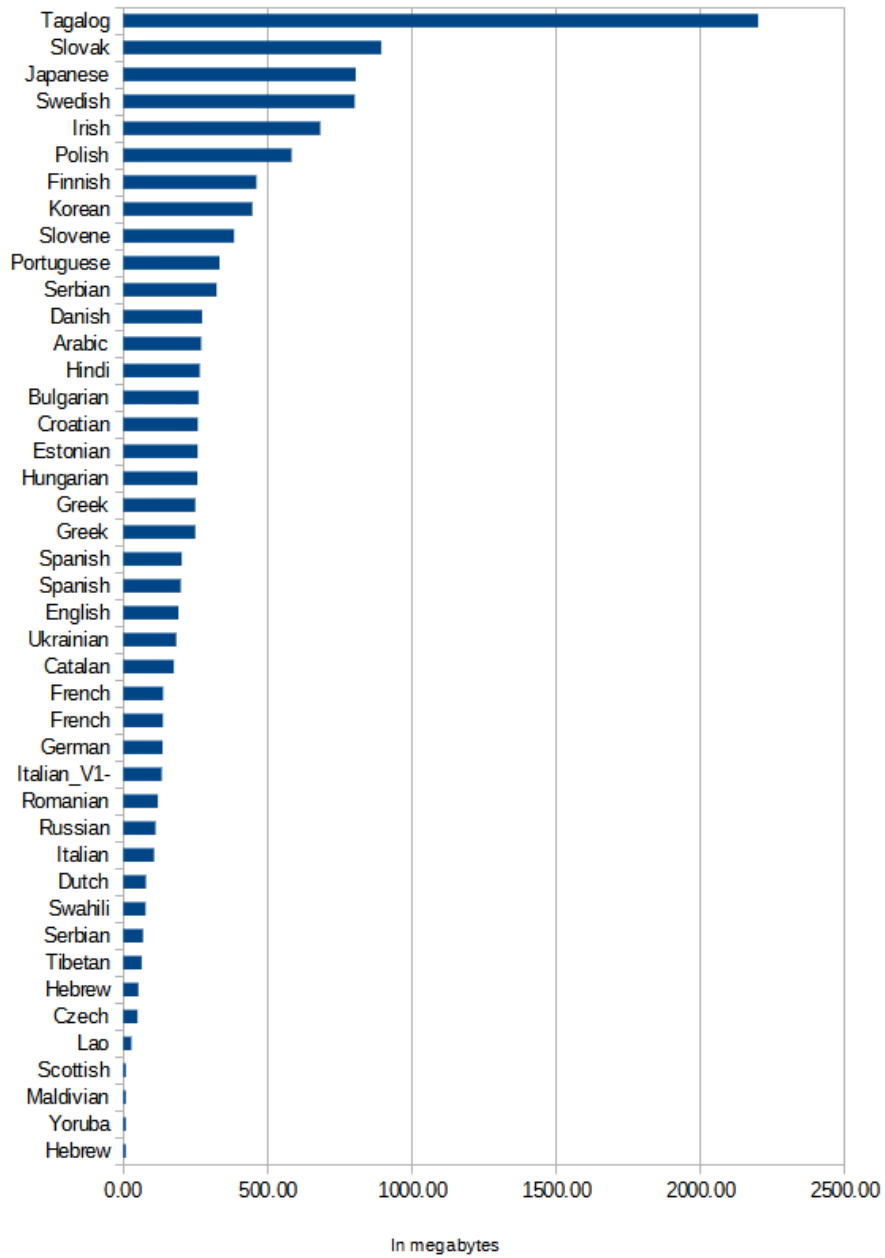


Figure 14: SM: RAM usage 5,000 tokens

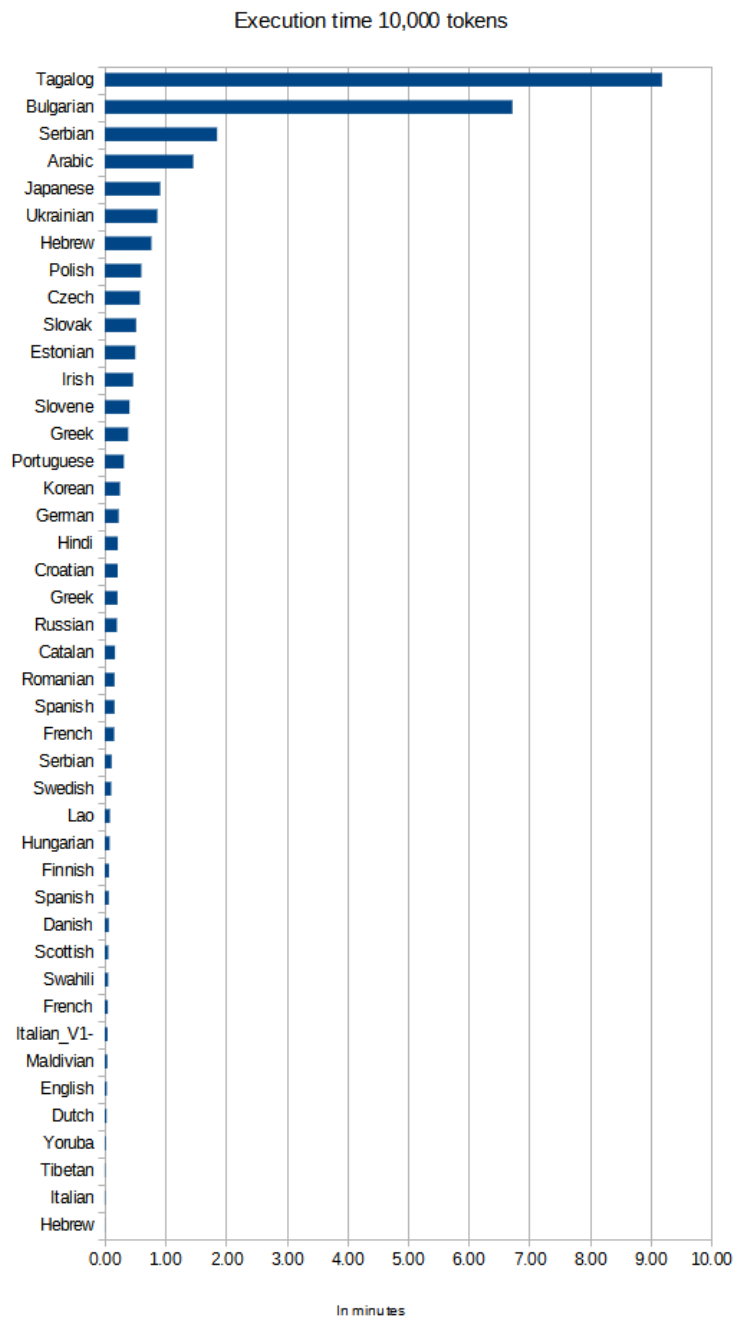


Figure 15: SM: Execution time usage 10,000 tokens

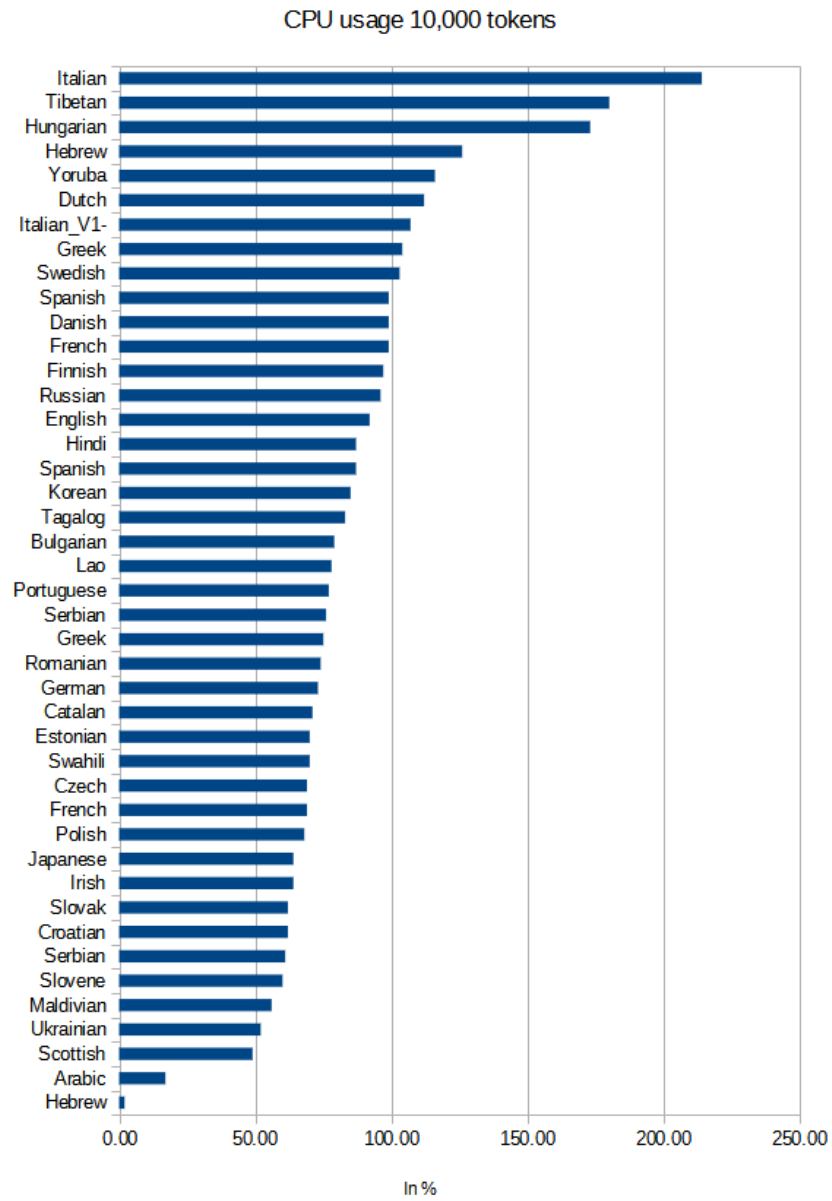


Figure 16: SM: CPU usage 10,000 tokens

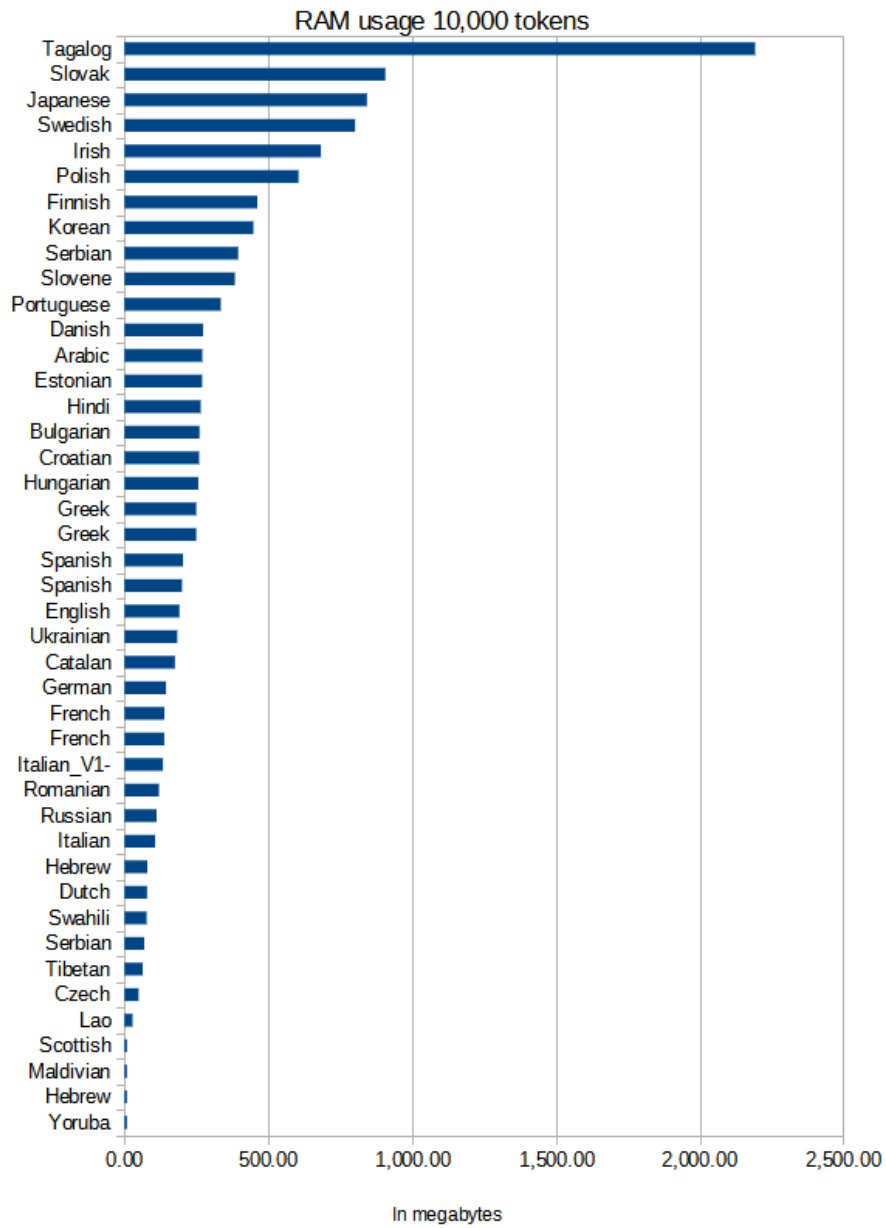


Figure 17: SM: RAM usage 10,000 tokens

Execution time 100,000 tokens

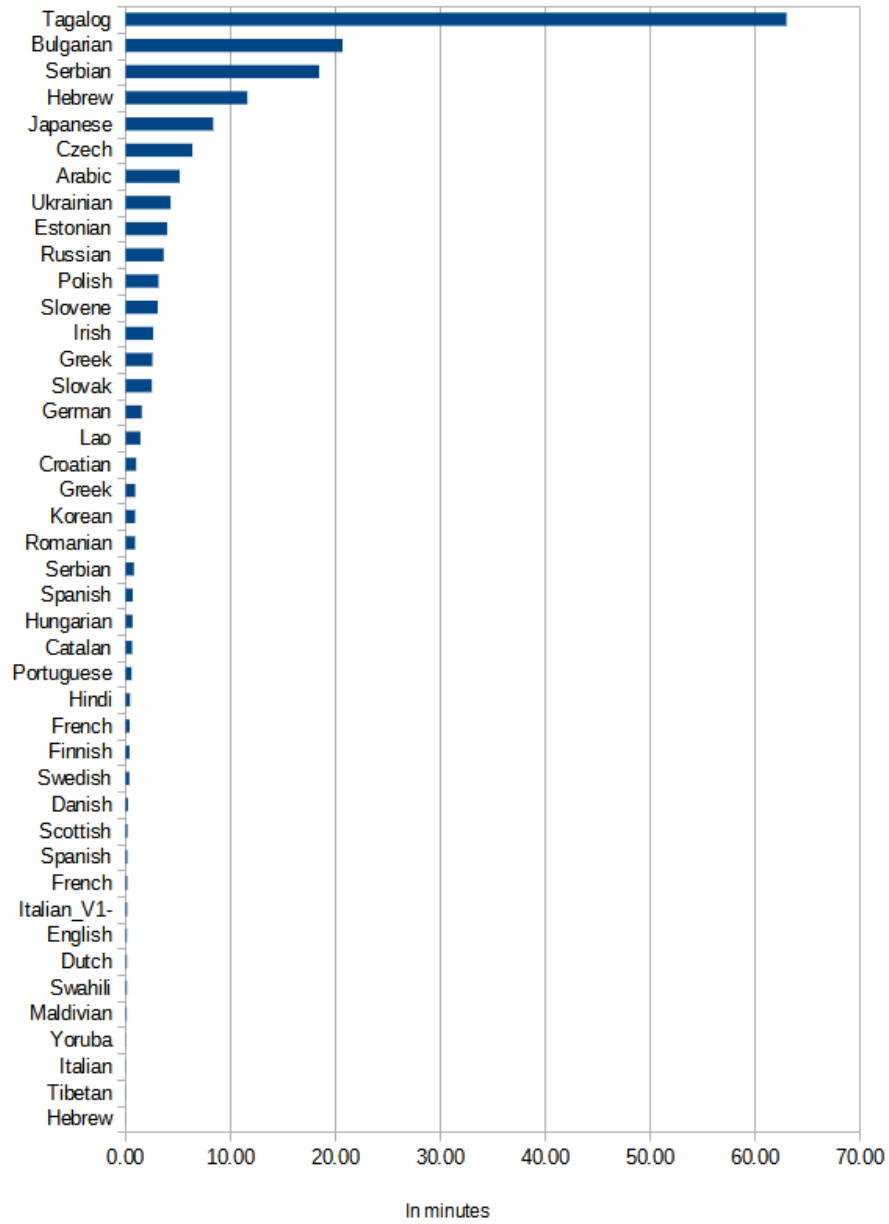


Figure 18: SM: Execution time 100,000 tokens

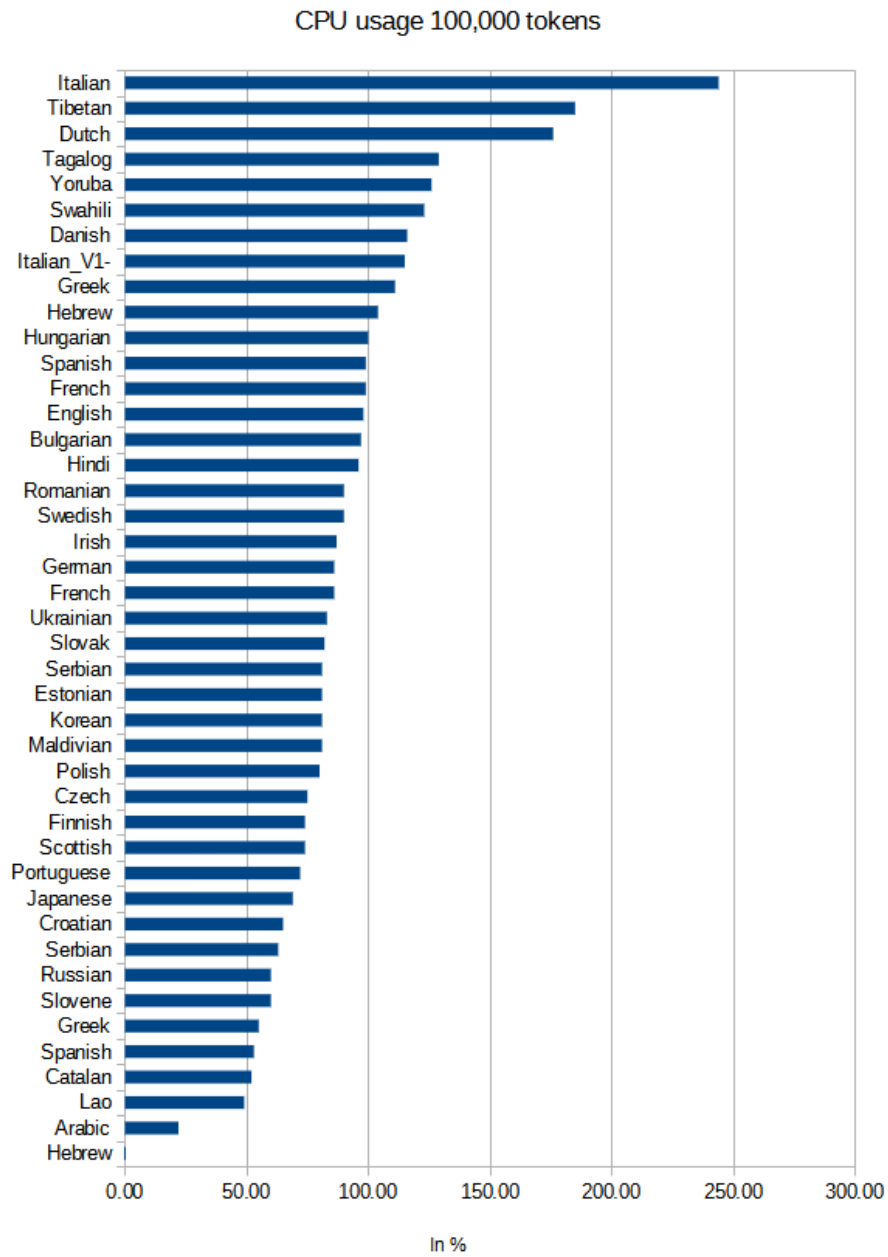


Figure 19: SM: CPU usage 100,000 tokens

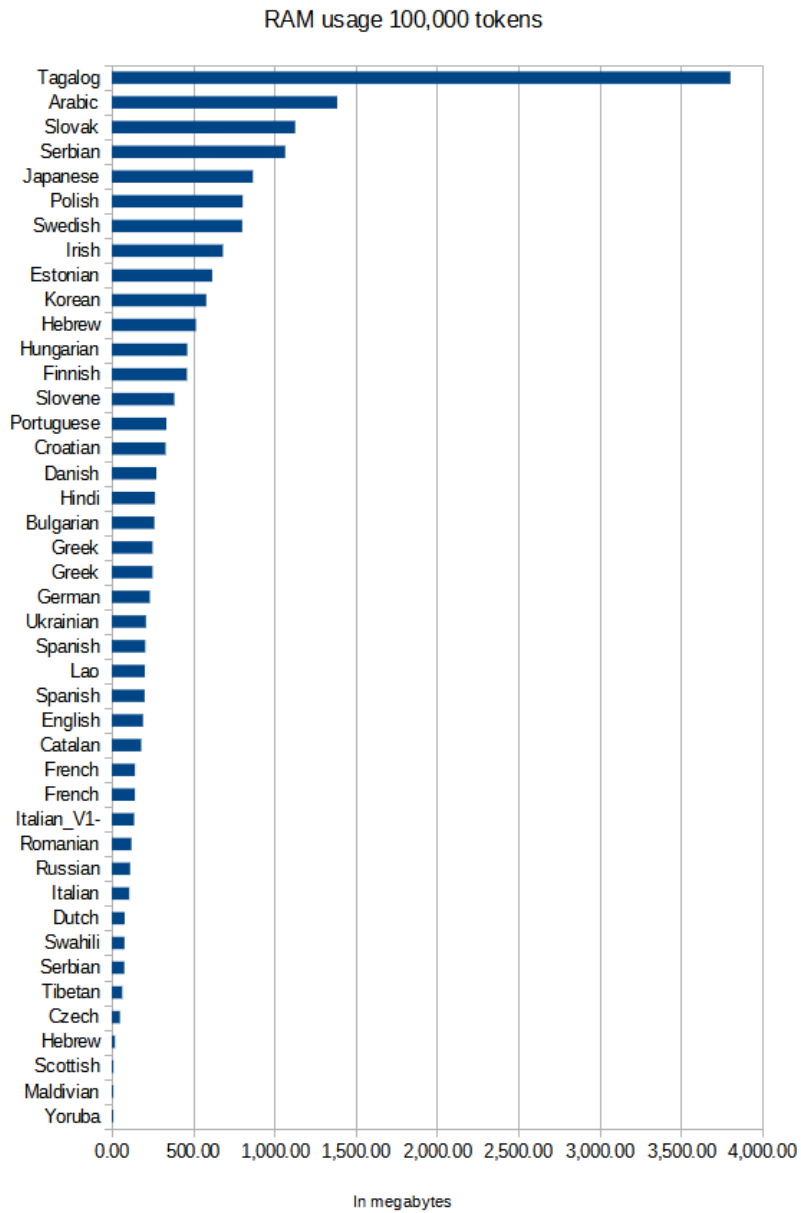


Figure 20: SM: RAM usage 100,000 tokens

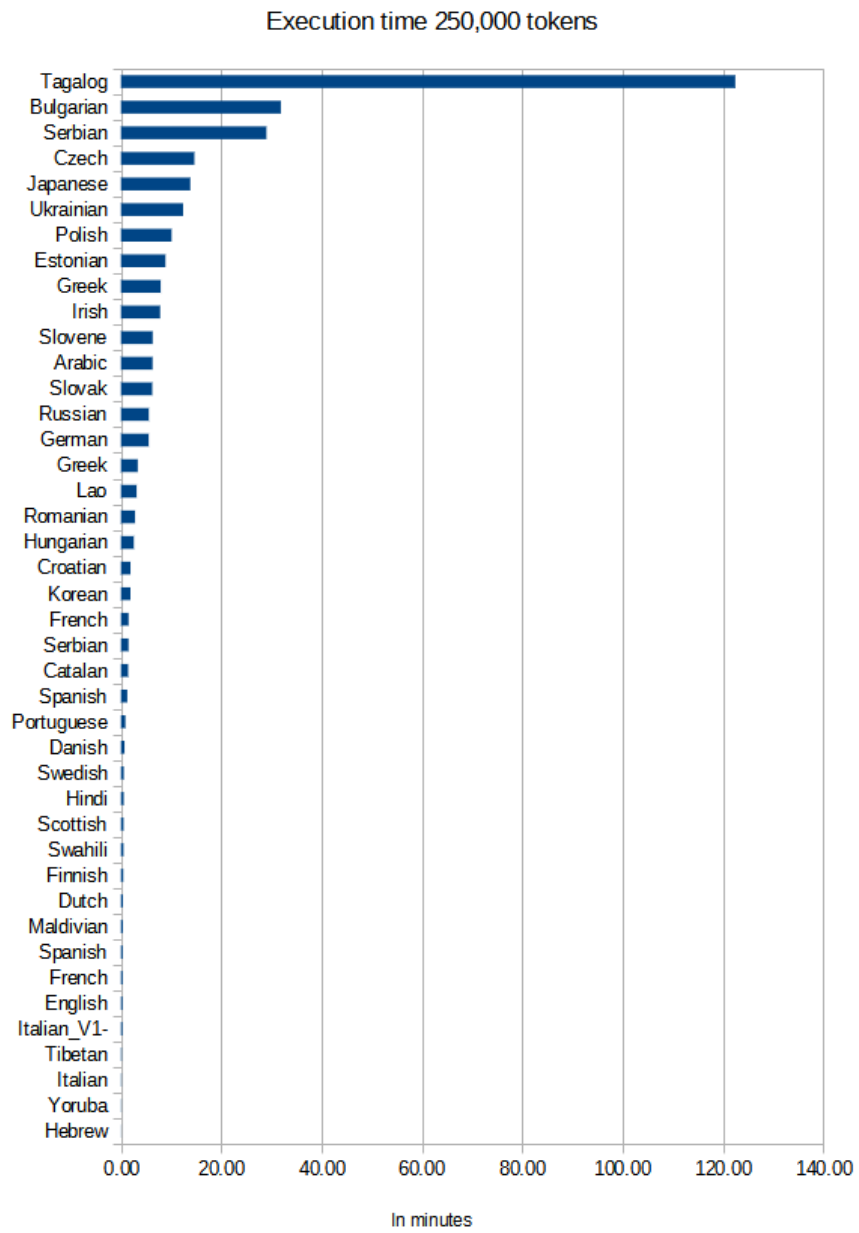


Figure 21: SM: Execution time 250,000 tokens

CPU usage 250,000 tokens

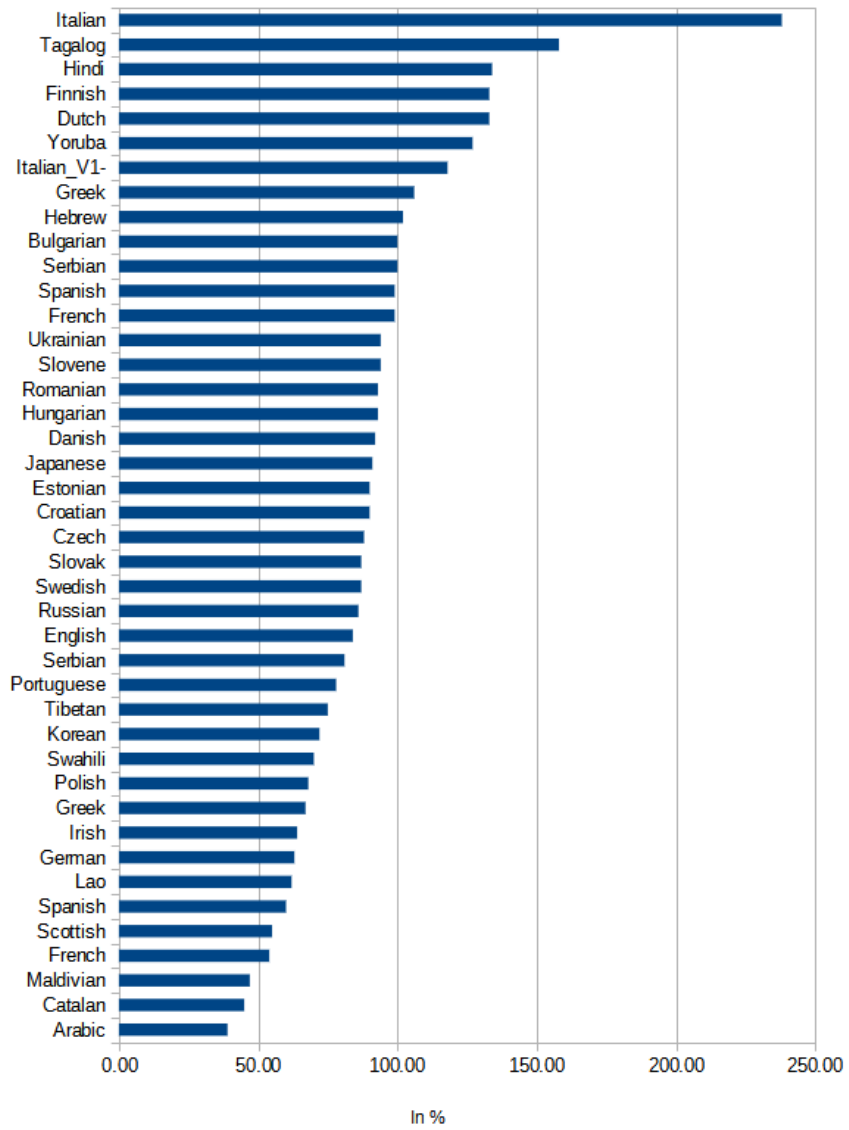


Figure 22: SM: CPU usage 250,000 tokens

RAM usage 250,000 tokens

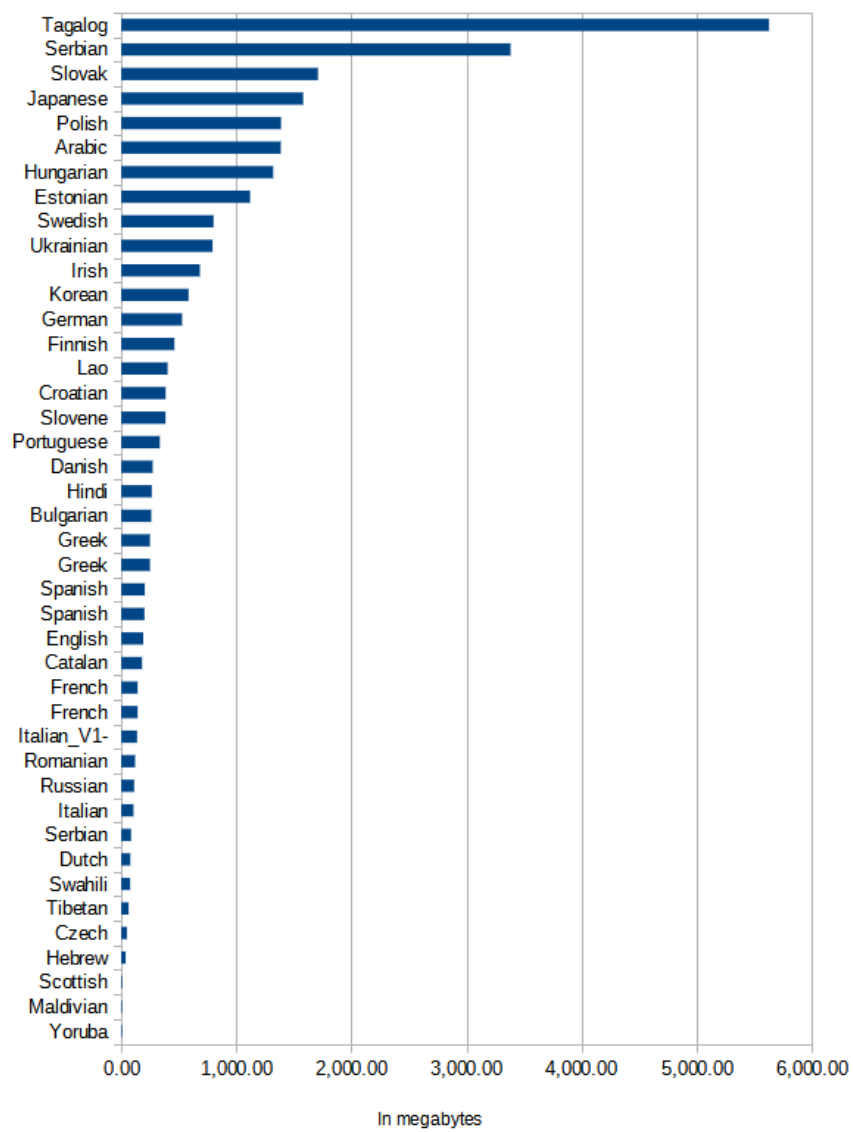


Figure 23: SM: RAM usage 250,000 tokens

Execution time 500,000 tokens

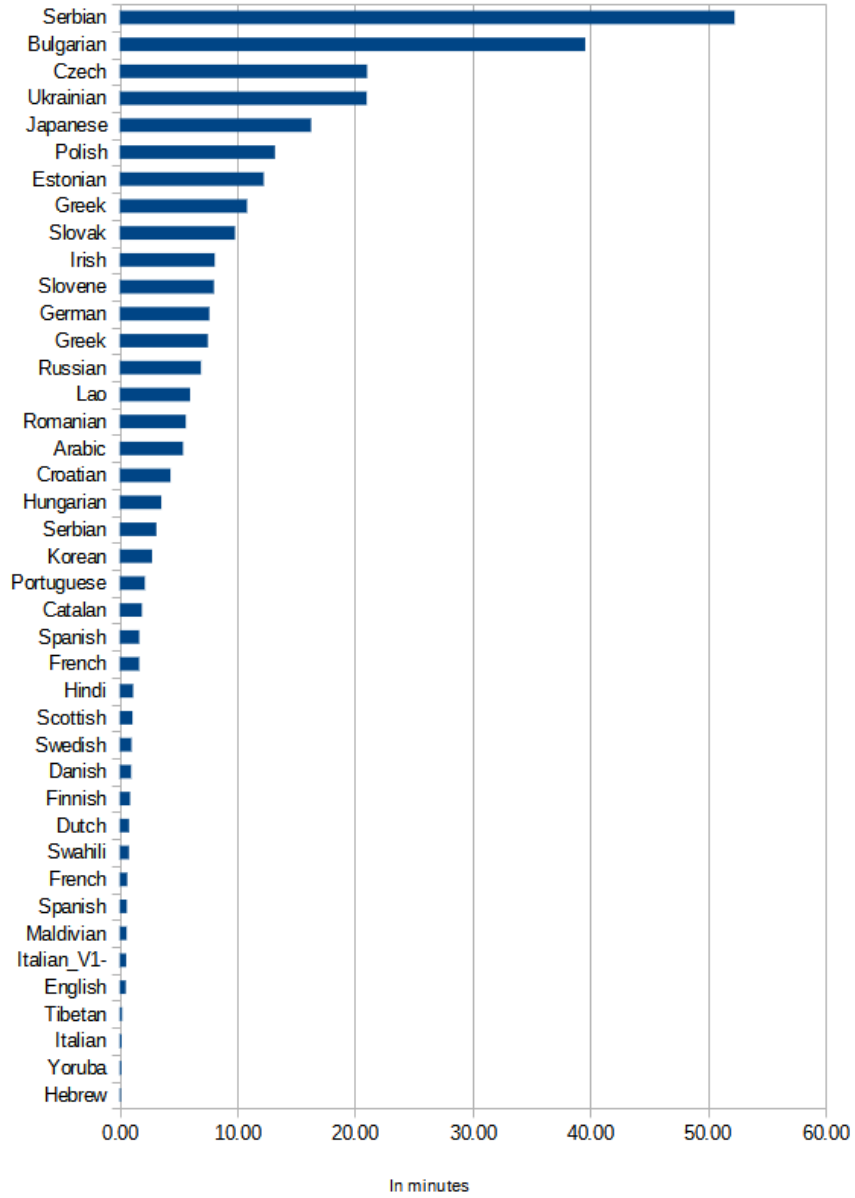


Figure 24: SM: Execution time 500,000 tokens

CPU usage 500,000 tokens

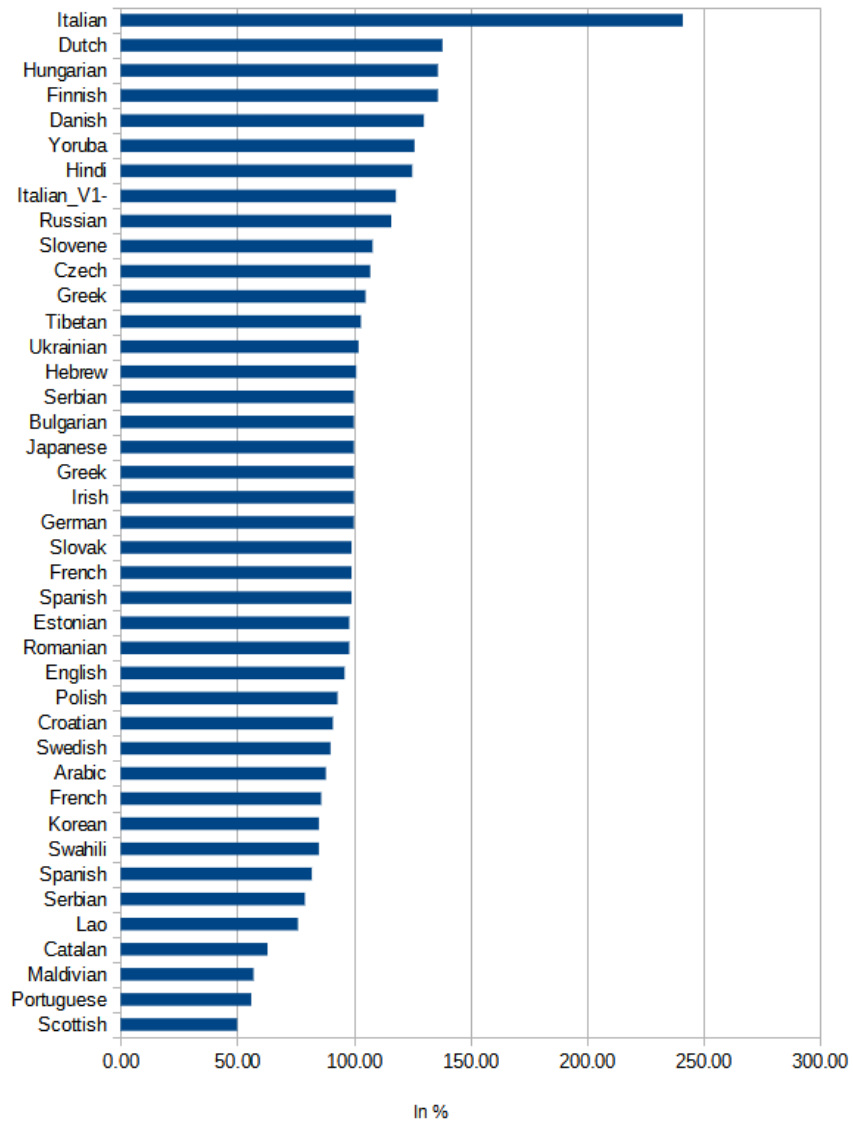


Figure 25: SM: CPU usage 500,000 tokens

RAM usage 500,000 tokens

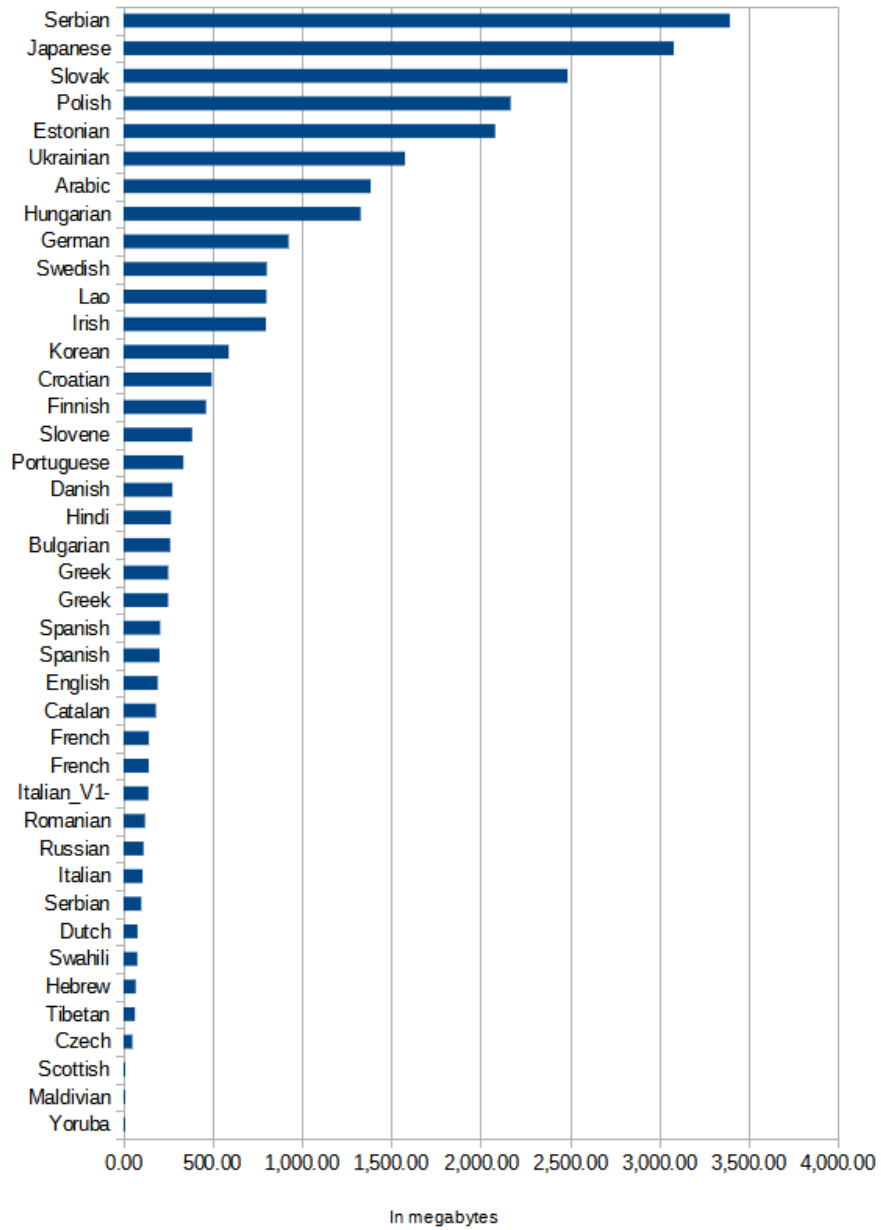


Figure 26: SM: RAM usage 500,000 tokens

Execution time 750,000 tokens

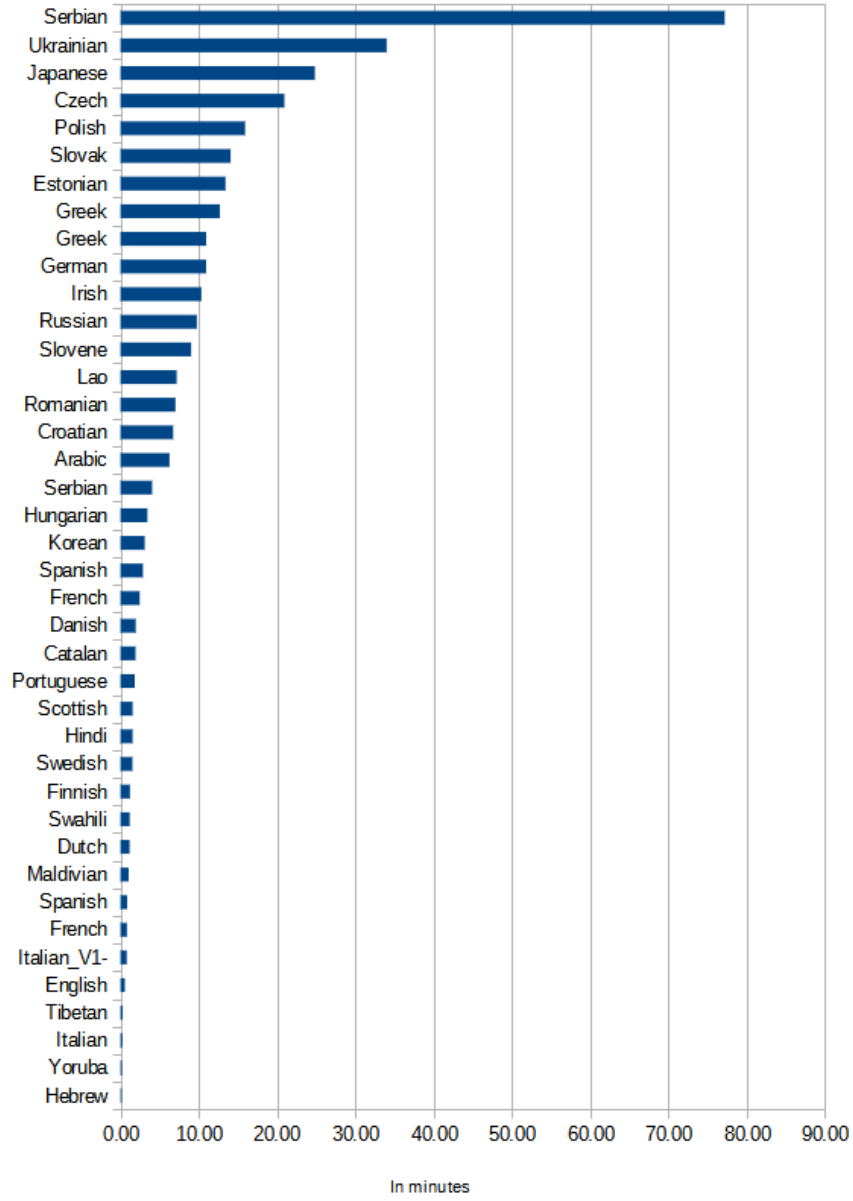


Figure 27: SM: Execution time 750,000 tokens

CPU usage 750,000 tokens

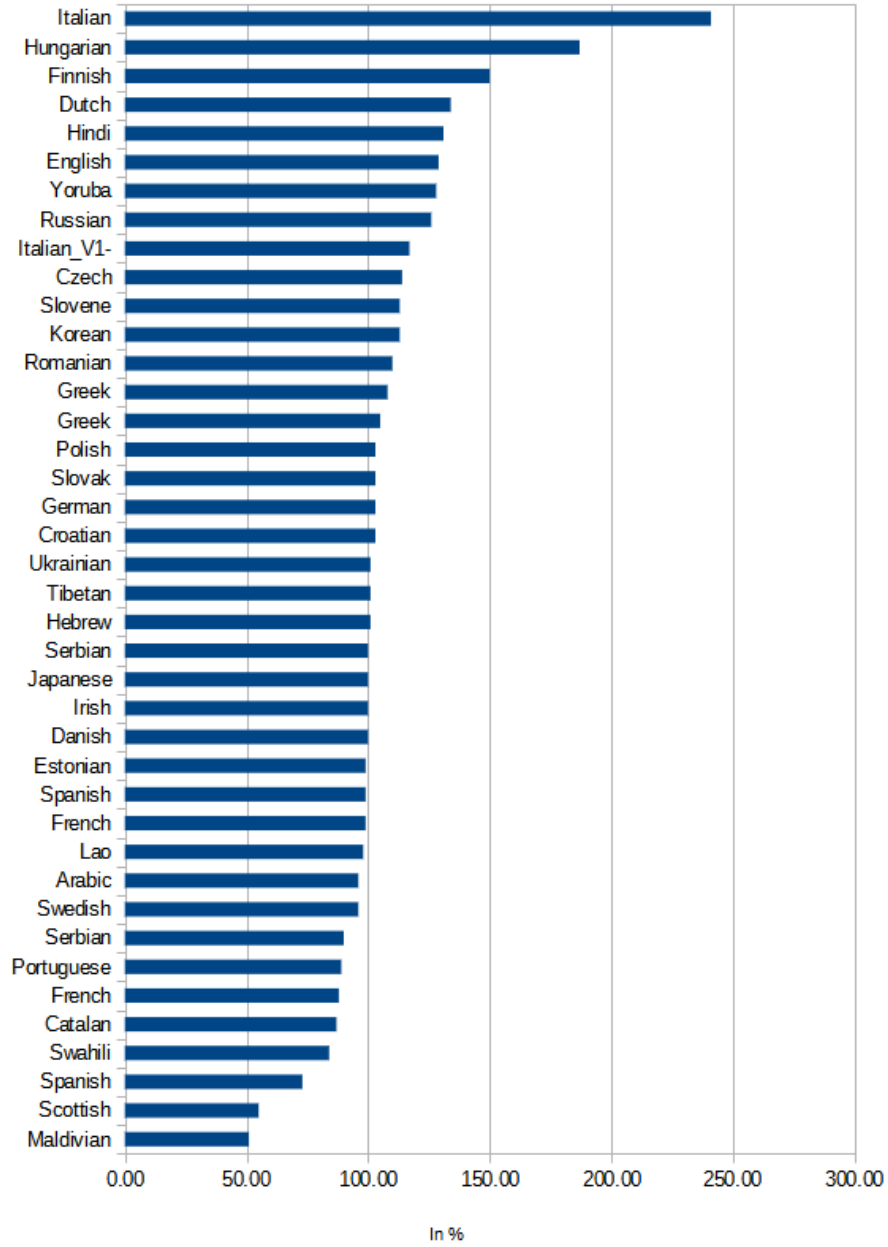


Figure 28: SM: CPU usage 750,000 tokens

RAM usage 750,000 tokens

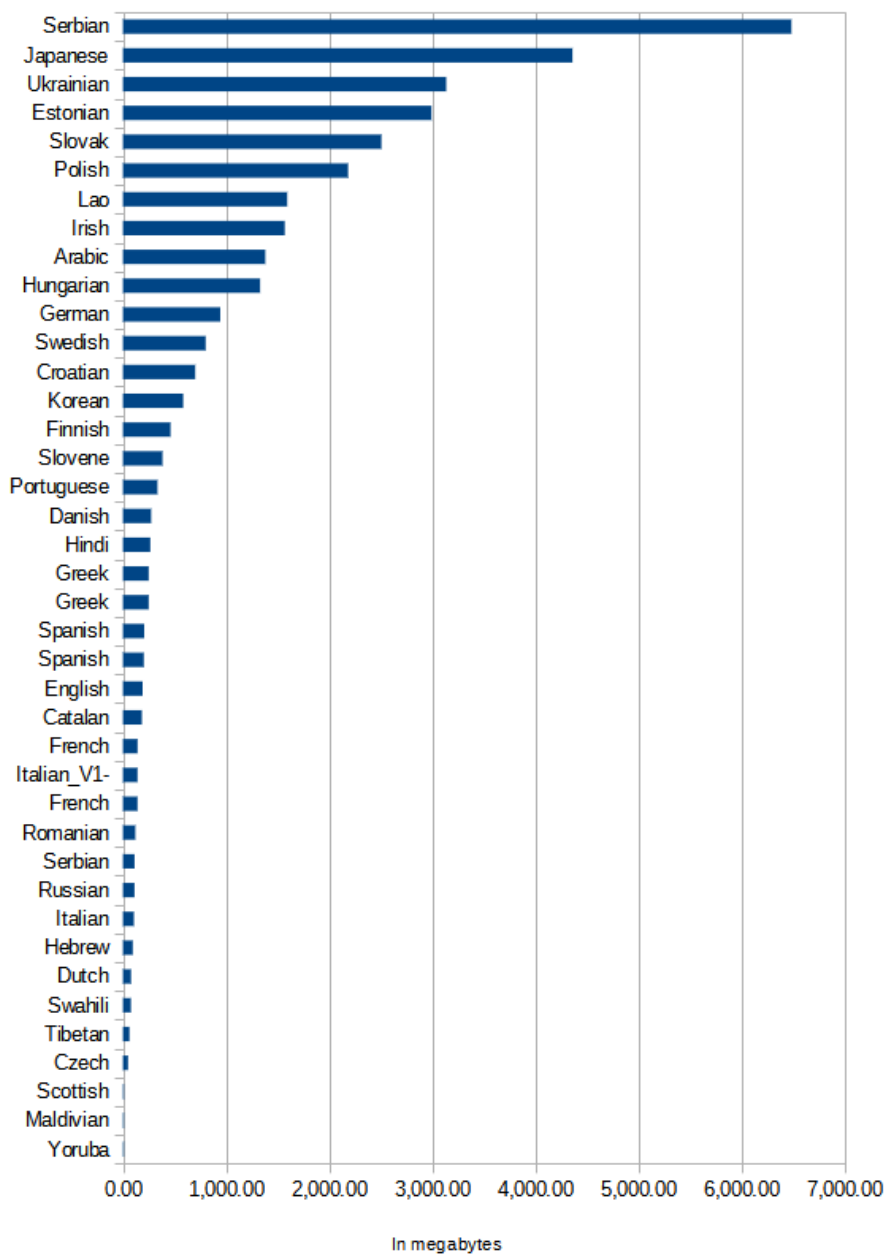


Figure 29: SM: RAM usage 750,000 tokens

Execution time 1,000,000 tokens

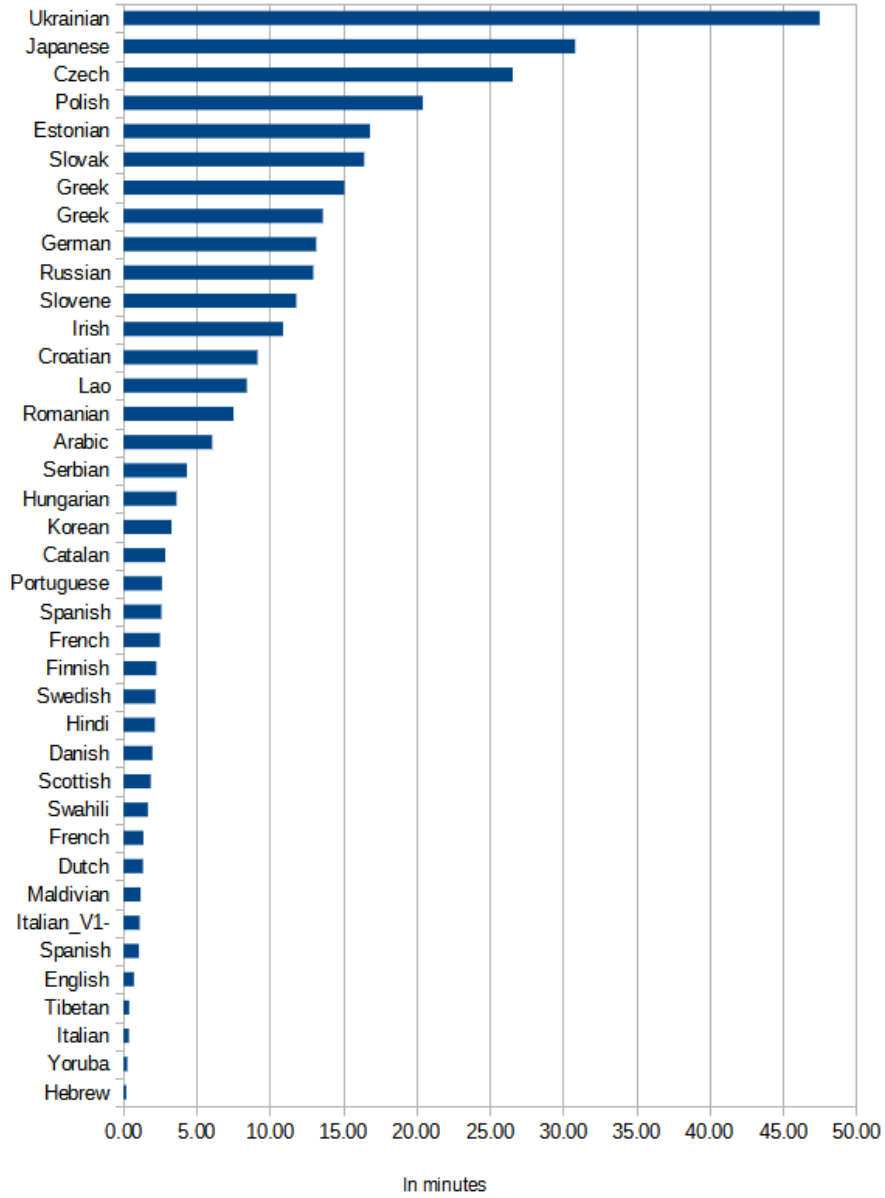


Figure 30: SM: Execution time 1,000,000 tokens

CPU usage 1,000,000 tokens

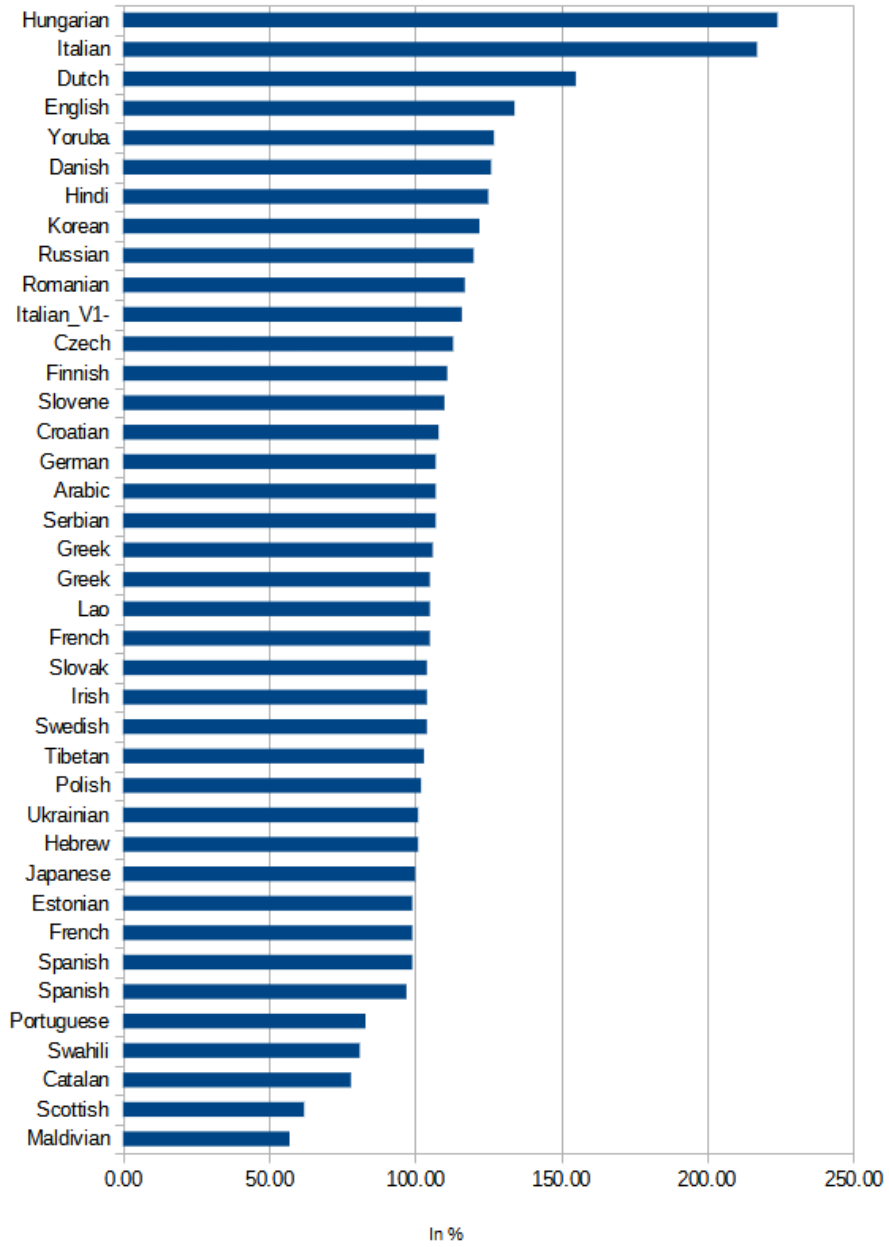


Figure 31: SM: CPU usage 1,000,000 tokens

RAM usage 1,000,000 tokens

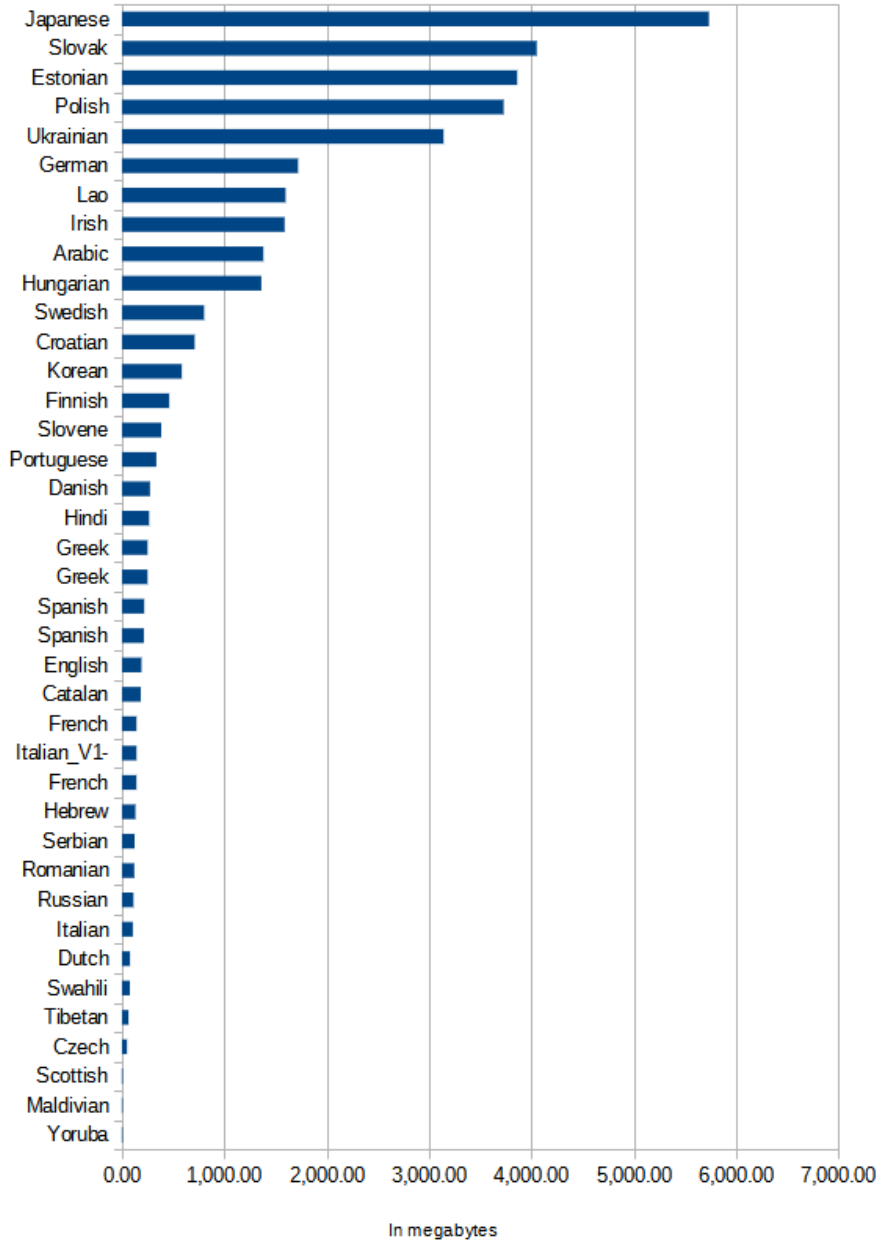


Figure 32: SM: RAM usage 1,000,000 tokens

| Language | Pipeline | Slope | Interception | Standard error |
|------------|---|-------------|--------------|----------------|
| Bulgarian | /mnt/toad7_disk1/kostka/pipelines/bulgarian_tt2/pipeline.sh | 0.000059580 | 12.187397746 | 3.955823 |
| Catalan | /opt/freeling_pipe/freeling_catalan_v2.py | 0.000002383 | 0.413853870 | 0.307906 |
| French | /opt/freeling_pipe/freeling_french_v2.py | 0.000001144 | 0.037698395 | 0.050148 |
| French | /opt/freeling_pipe/freeling_french_v3.sh | 0.000002507 | 0.397923658 | 0.272726 |
| Italian | /opt/freeling_pipe/freeling_italian_v1.sh | 0.000001039 | 0.039905863 | 0.015197 |
| Portuguese | /opt/freeling_pipe/freeling_portuguese_v4.py | 0.000002178 | 0.528692641 | 0.279007 |
| Spanish | /opt/freeling_pipe/freeling_spanish_v5.py | 0.000001064 | 0.063660510 | 0.019560 |
| Spanish | /opt/freeling_pipe/freeling_spanish_v6.sh | 0.000002479 | 0.425727571 | 0.266124 |
| Korean | /opt/HanNanumTagger/runHanNanumTagger_v2.sh | 0.000002979 | 0.799925837 | 0.360117 |
| Hungarian | /opt/hunpos_pipe/hunpos_hungarian_v2.sh | 0.000003952 | 0.820771538 | 0.753740 |
| Serbian | /opt/hunpos_pipe/hunpos_serbian.sh | 0.000004414 | 0.492097096 | 0.409840 |
| Swedish | /opt/hunpos_pipe/hunpos_swedish_v4.sh | 0.000001857 | 0.124014875 | 0.156347 |
| Japanese | /opt/mecab_pipe/mecab_unidic_comainu_jpn_lempos.sh | 0.000027175 | 3.707441805 | 1.672457 |
| Croatian | /opt/skepipelines/croatian_rft1/pipeline.sh | 0.000008731 | 0.105013428 | 0.406663 |
| Czech | /opt/skepipelines/czech_mj2/pipeline.sh | 0.000022130 | 6.355693608 | 3.442488 |
| Danish | /opt/skepipelines/danish_tt2/pipeline.sh | 0.000001998 | 0.061193564 | 0.143533 |
| Dutch | /opt/skepipelines/dutch_tt3/pipeline.sh | 0.000001421 | 0.030182895 | 0.084195 |
| English | /opt/skepipelines/english_tt31/pipeline.sh | 0.000000662 | 0.074969636 | 0.045265 |
| Estonian | /opt/skepipelines/estonian_estnltk3/pipeline.sh | 0.000013993 | 3.662536472 | 2.019407 |
| Finnish | /opt/skepipelines/finnish_tt3/pipeline.sh | 0.000001938 | 0.069874492 | 0.254756 |
| German | /opt/skepipelines/german_rft3/pipeline.sh | 0.000012523 | 1.166339907 | 0.811342 |
| Greek | /opt/skepipelines/greek_tt2/pipeline.sh | 0.000014582 | -0.072279046 | 0.546484 |
| Greek | /opt/skepipelines/greek_tt3/pipeline.sh | 0.000014403 | 2.298646122 | 1.318167 |
| Hebrew | /opt/skepipelines/hebrew_tok1/pipeline.sh | 0.000000201 | 0.002573504 | 0.005076 |
| Hindi | /opt/skepipelines/hindi_siva1/pipeline.sh | 0.000001869 | 0.262587912 | 0.180194 |
| Irish | /opt/skepipelines/irish_rft1/pipeline.sh | 0.000010108 | 2.511076479 | 1.458310 |
| Italian | /opt/skepipelines/italian_tt2/pipeline.sh | 0.000000354 | 0.003748209 | 0.010077 |
| Lao | /opt/skepipelines/lao_rft1/pipeline.sh | 0.000008450 | 0.705932111 | 0.594833 |
| Maldivian | /opt/skepipelines/maldivian_tok1/pipeline.sh | 0.000001177 | 0.015102960 | 0.086811 |
| Polish | /opt/skepipelines/polish_rft1/pipeline.sh | 0.000018693 | 2.566120501 | 1.504030 |
| Romanian | /opt/skepipelines/romanian_tt1/pipeline.sh | 0.000007574 | 0.980703440 | 0.882332 |
| Russian | /opt/skepipelines/russian_rft1/pipeline.sh | 0.000011465 | 1.407620802 | 0.709361 |
| Scottish | /opt/skepipelines/scottish_tok1/pipeline.sh | 0.000001749 | 0.068630762 | 0.161388 |
| Serbian | /opt/skepipelines/serbian_rft1/pipeline.sh | 0.000009365 | 4.936876503 | 2.520044 |
| Slovak | /opt/skepipelines/slovak_rft1/pipeline.sh | 0.000016095 | 2.065423544 | 1.556371 |
| Slovene | /opt/skepipelines/slovene_tt2/pipeline.sh | 0.000009785 | 2.731909492 | 1.443692 |
| Swahili | /opt/skepipelines/swahili_tt1/pipeline.sh | 0.000001412 | 0.053728272 | 0.134696 |
| Tagalog | /opt/skepipelines/tagalog_stf1/pipeline.sh | 0.000446177 | 8.540654090 | 7.866597 |
| Tibetan | /opt/skepipelines/tibetan_tt1/pipeline.sh | 0.000000415 | -0.002901591 | 0.026728 |
| Ukrainian | /opt/skepipelines/ukrainian_rft1/pipeline.sh | 0.000045308 | 0.394277086 | 1.330862 |
| Yoruba | /opt/skepipelines/yoruba_tok1/pipeline.sh | 0.000000258 | 0.012347669 | 0.007360 |
| Arabic | /opt/stanford_nlp_tools/stanford_arabic.sh | 0.000003267 | 3.792886145 | 1.248445 |
| Hebrew | /opt/yap_pipe/yap_he_v1.sh | 0.000161489 | -1.741799084 | 2.586092 |

Figure 33: SM: Linear regression (better)

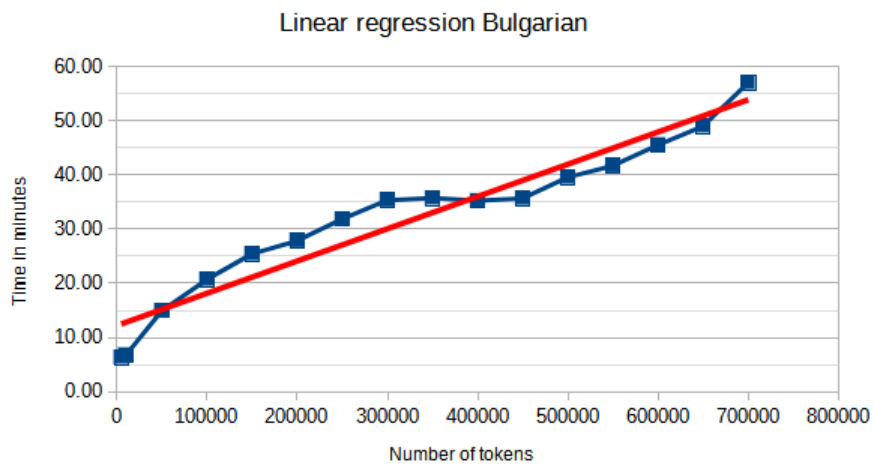


Figure 34: SM: Linear regression Bulgarian

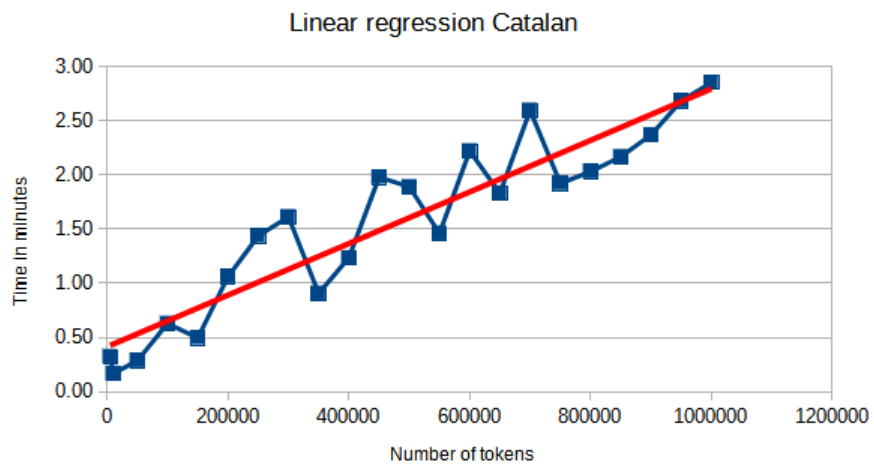


Figure 35: SM: Linear regression Catalan

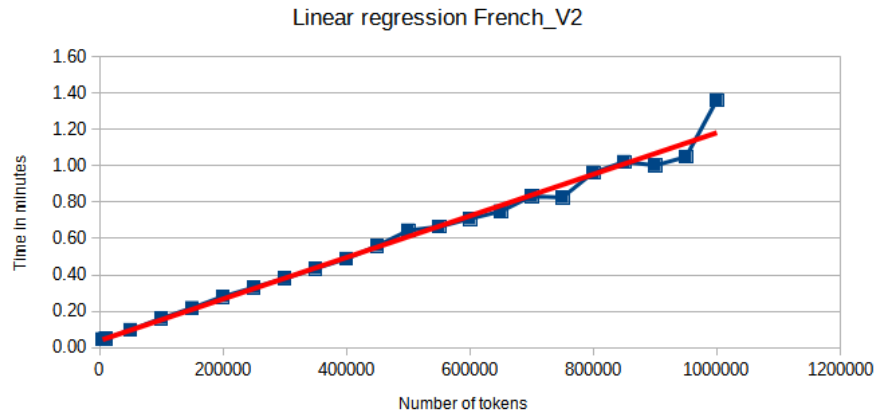


Figure 36: SM: Linear regression French V2

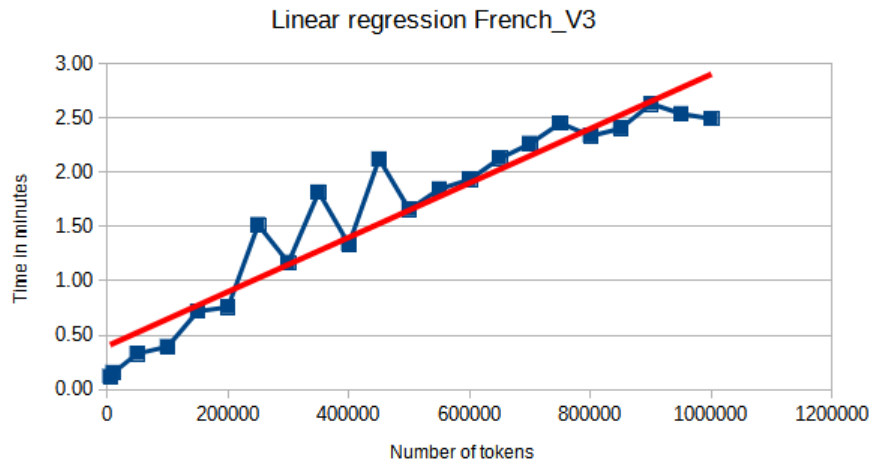


Figure 37: SM: Linear regression French V3

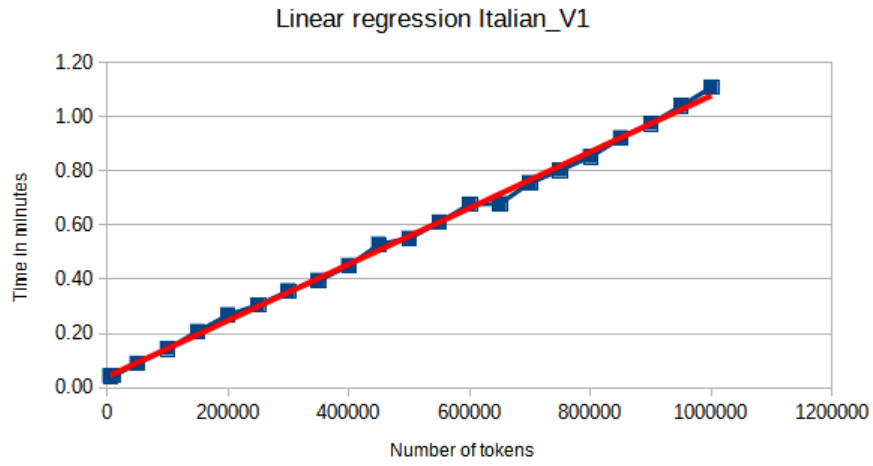


Figure 38: SM: Linear regression Italian V1

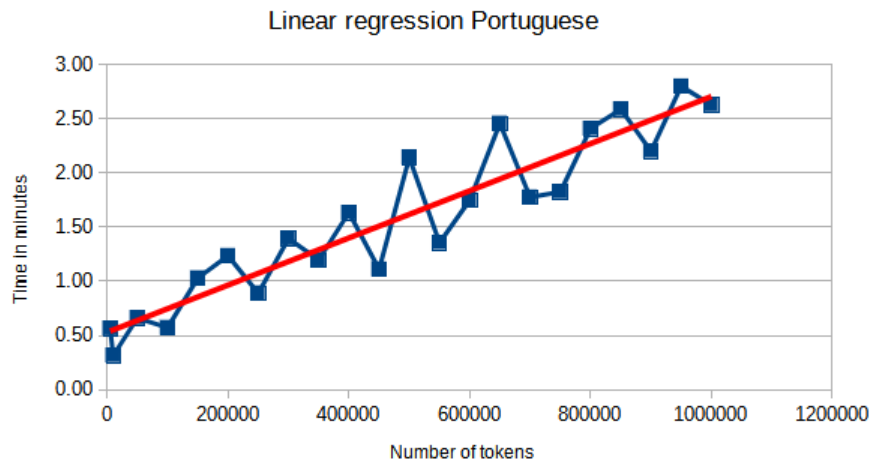


Figure 39: SM: Linear regression Portuguese

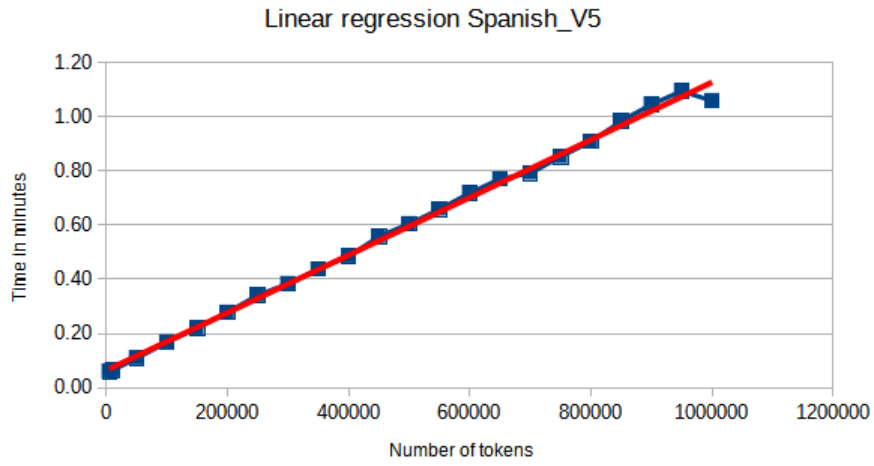


Figure 40: SM: Linear regression Spanish V5

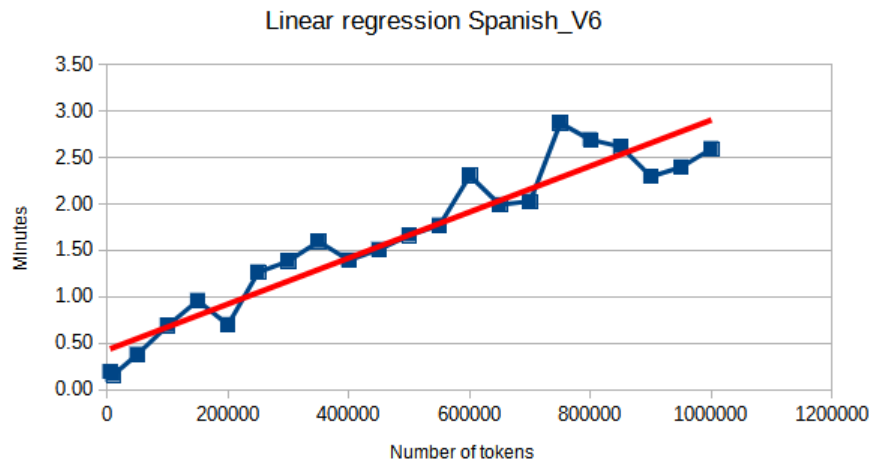


Figure 41: SM: Linear regression Spanish V6

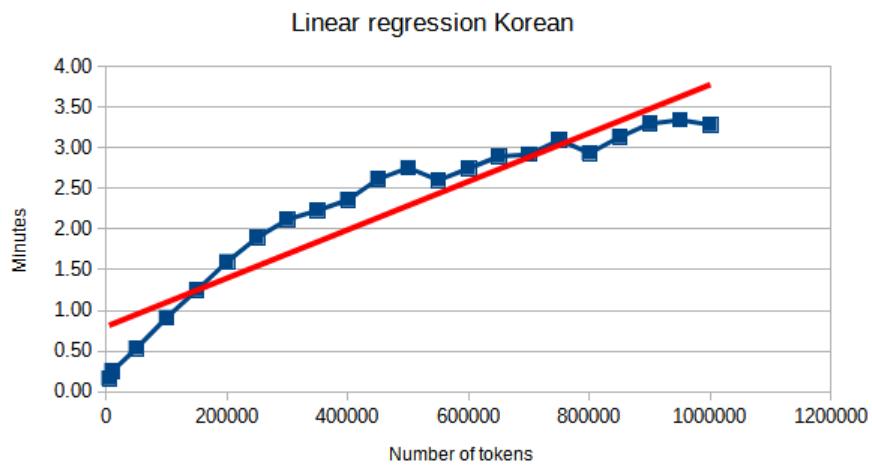


Figure 42: SM: Linear regression Korean

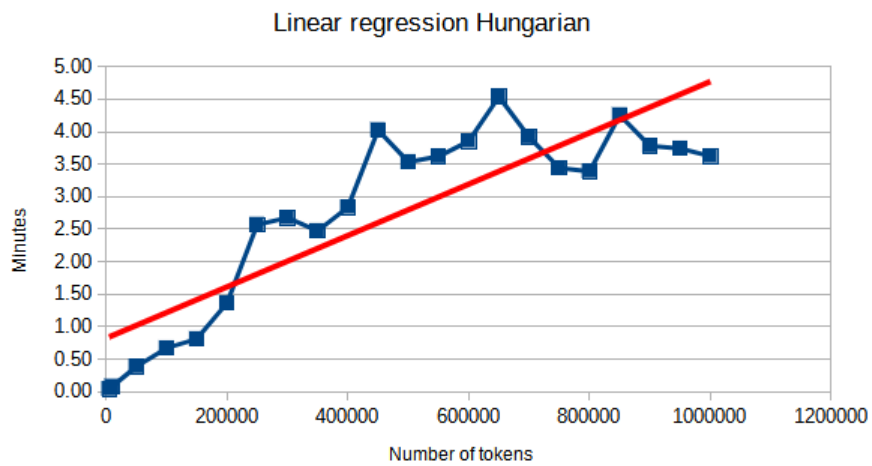


Figure 43: SM: Linear regression Hungarian

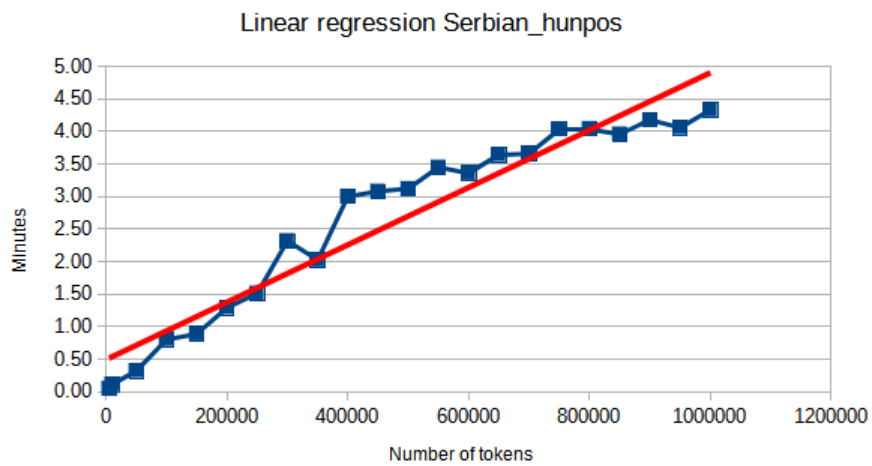


Figure 44: SM: Linear regression Serbian hunpos

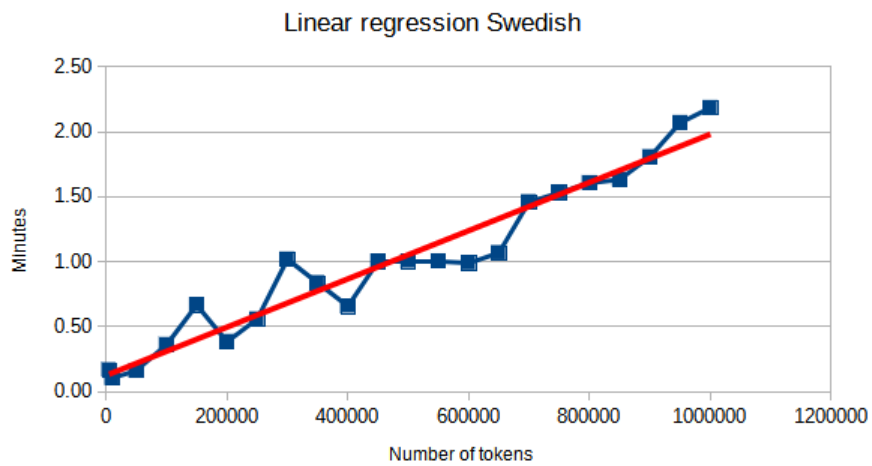


Figure 45: SM: Linear regression Swedish

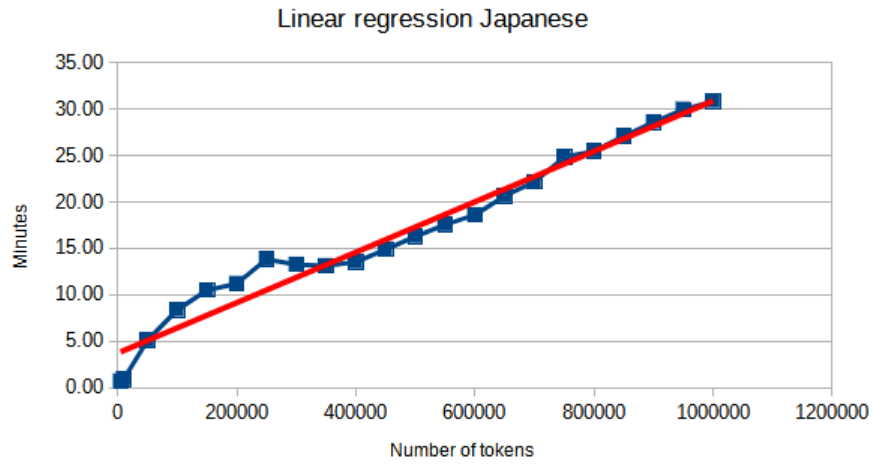


Figure 46: SM: Linear regression Japanese

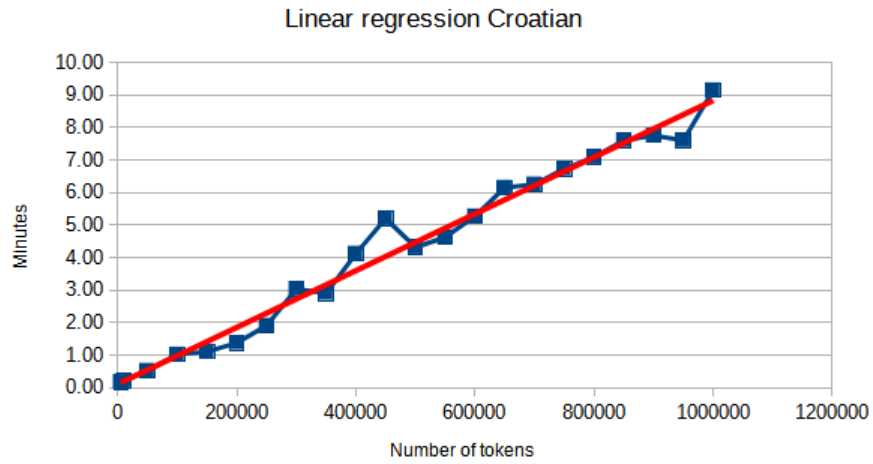


Figure 47: SM: Linear regression Croatian

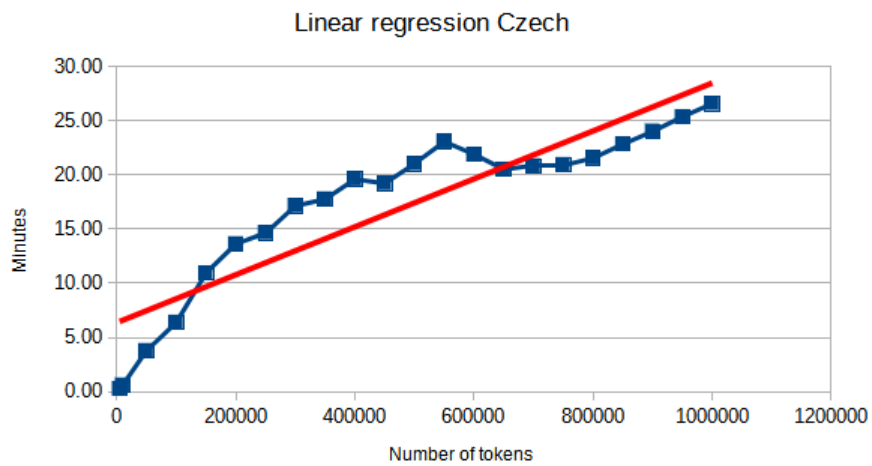


Figure 48: SM: Linear regression Czech

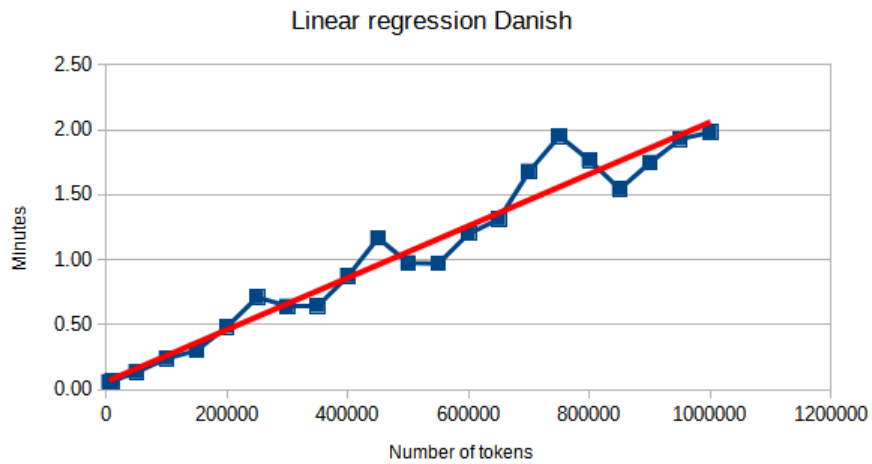


Figure 49: SM: Linear regression Danish

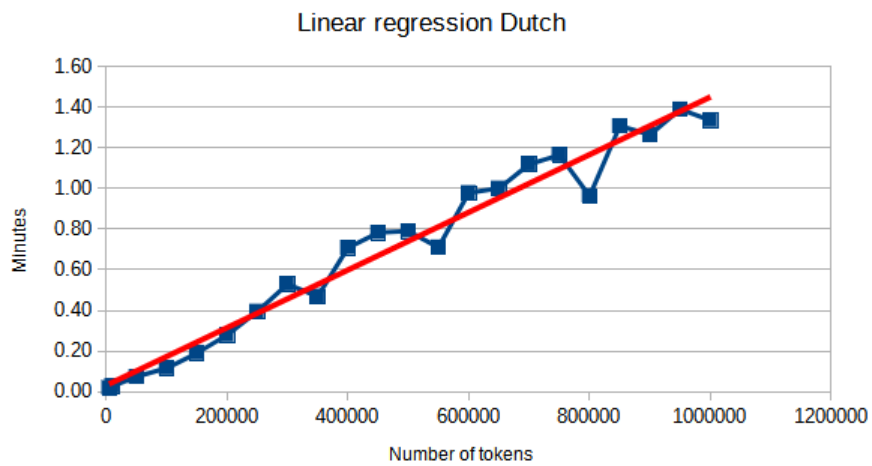


Figure 50: SM: Linear regression Dutch

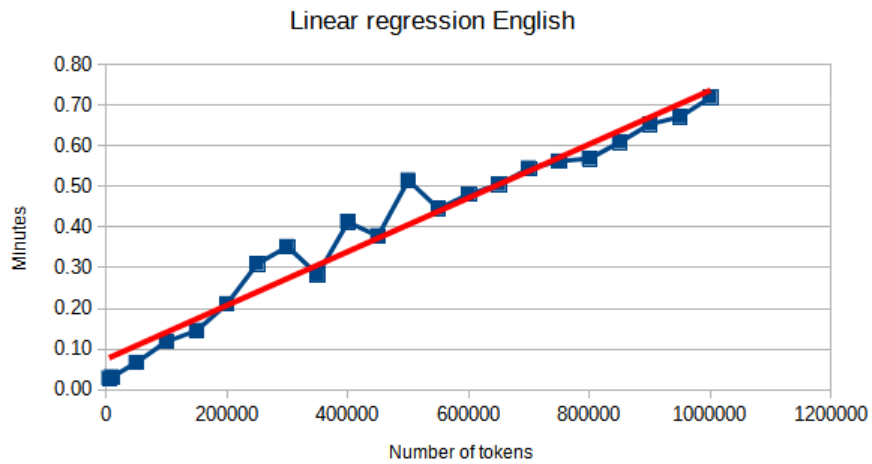


Figure 51: SM: Linear regression English

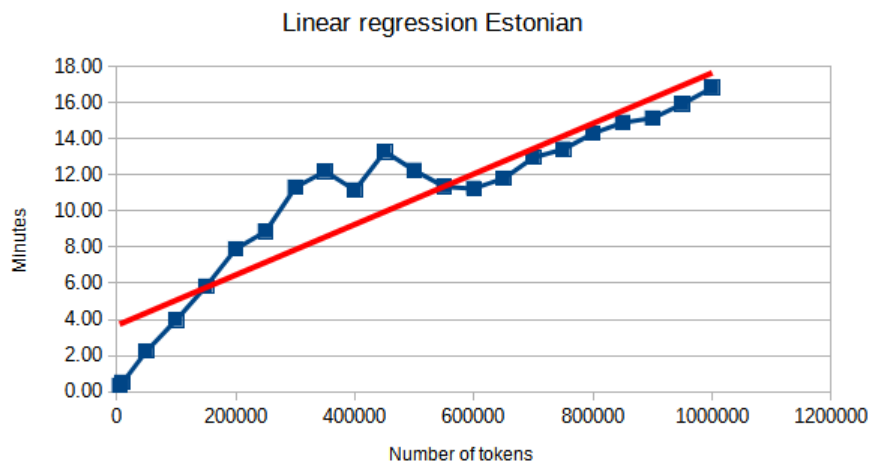


Figure 52: SM: Linear regression Estonian

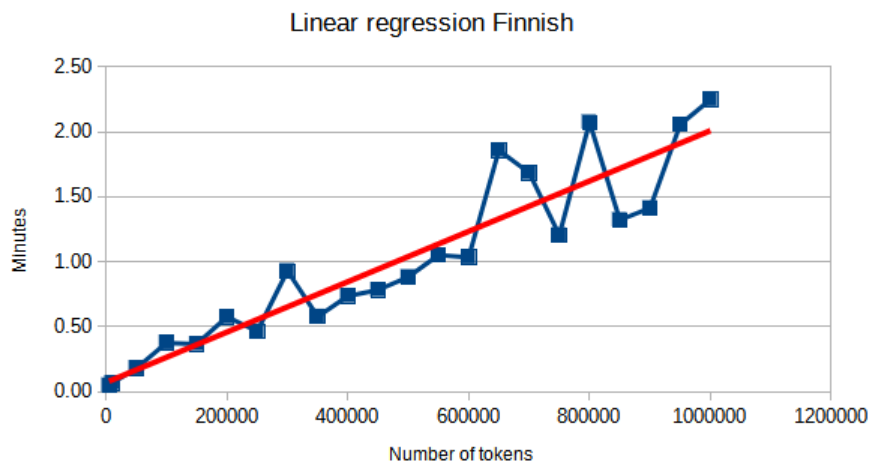


Figure 53: SM: Linear regression Finnish

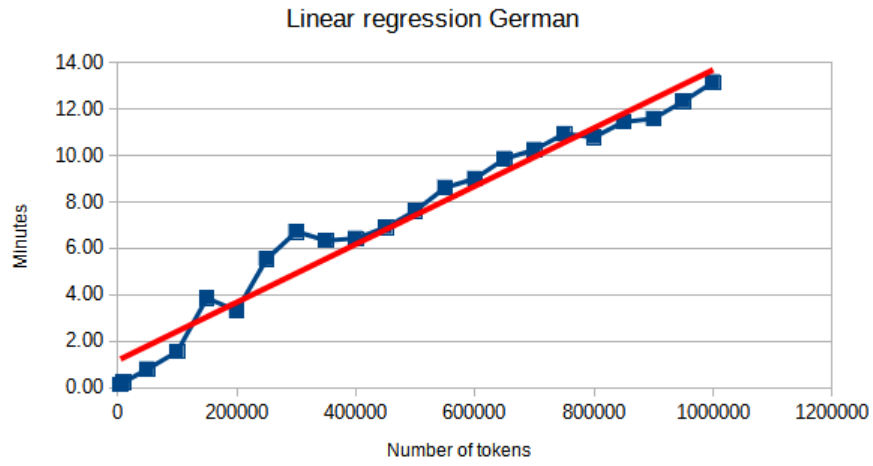


Figure 54: SM: Linear regression German

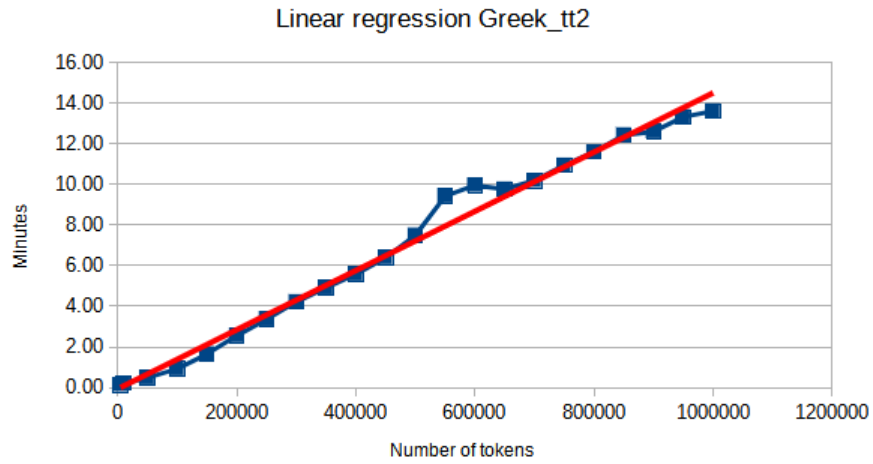


Figure 55: SM: Linear regression Greek tt2

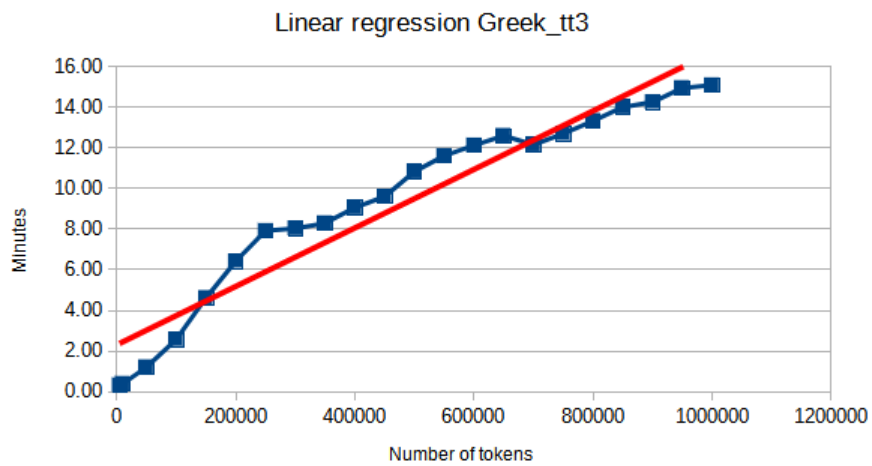


Figure 56: SM: Linear regression Greek

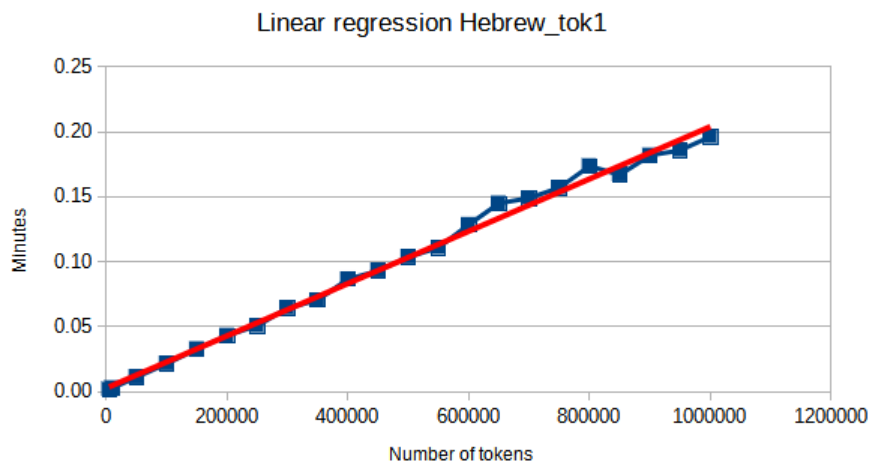


Figure 57: SM: Linear regression Hebrew tok1

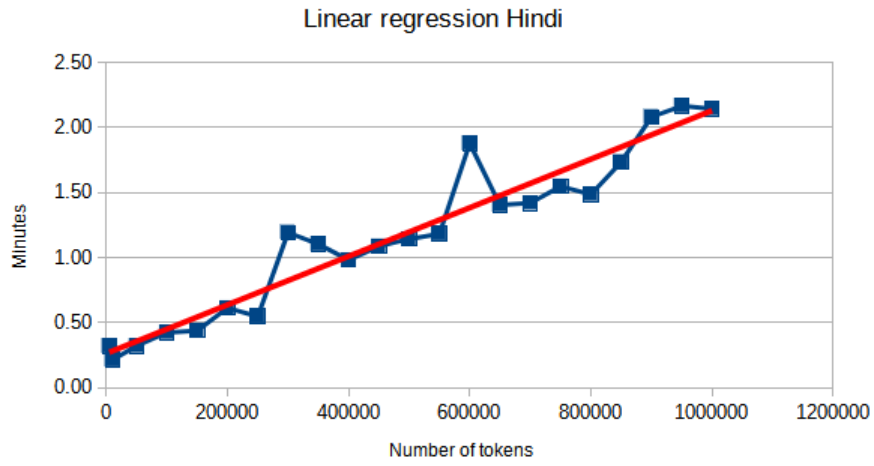


Figure 58: SM: Linear regression Hindi

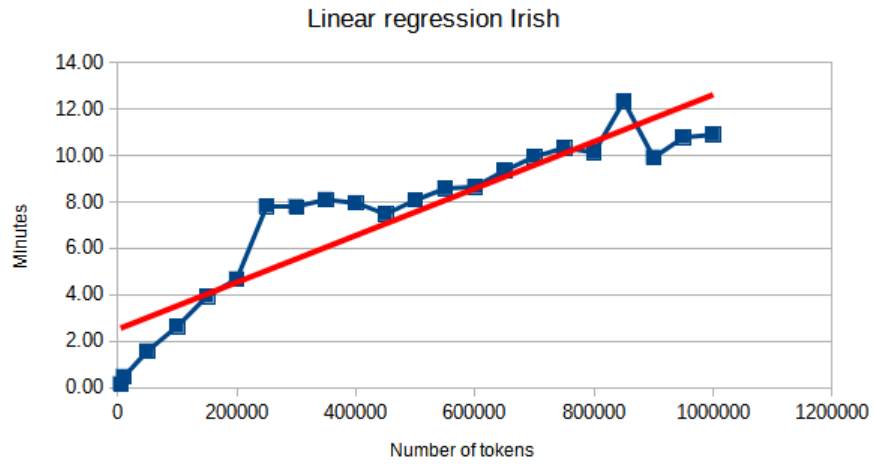


Figure 59: SM: Linear regression Irish

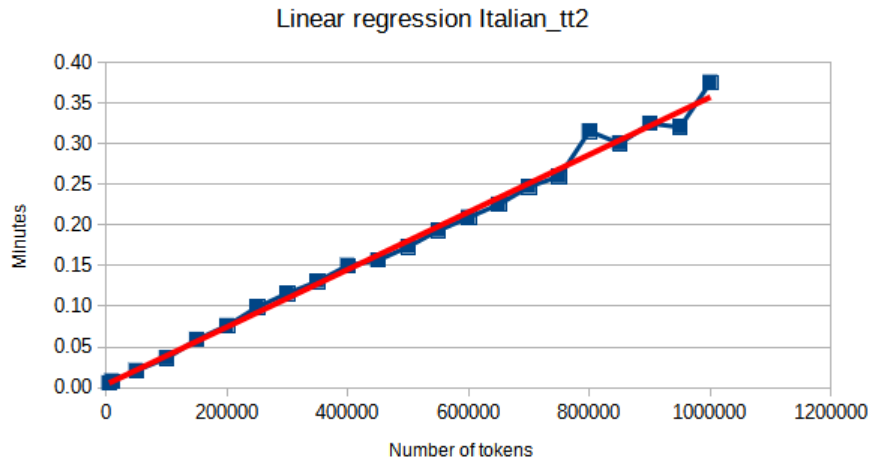


Figure 60: SM: Linear regression Italian

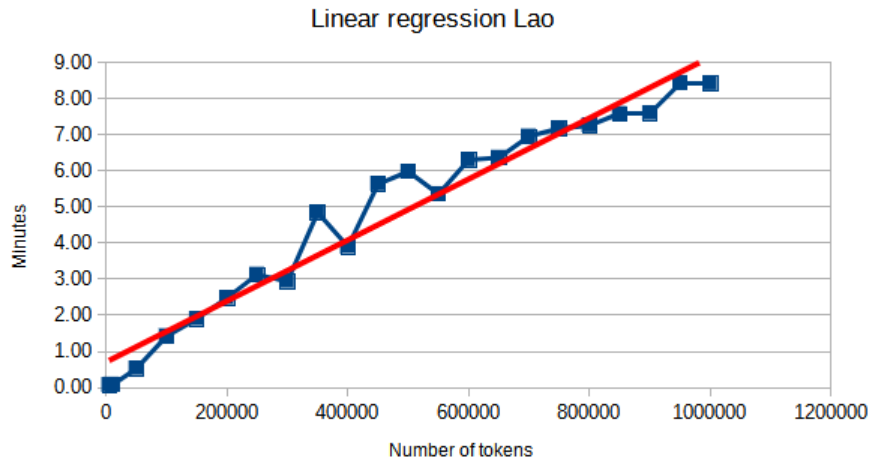


Figure 61: SM: Linear regression Lao

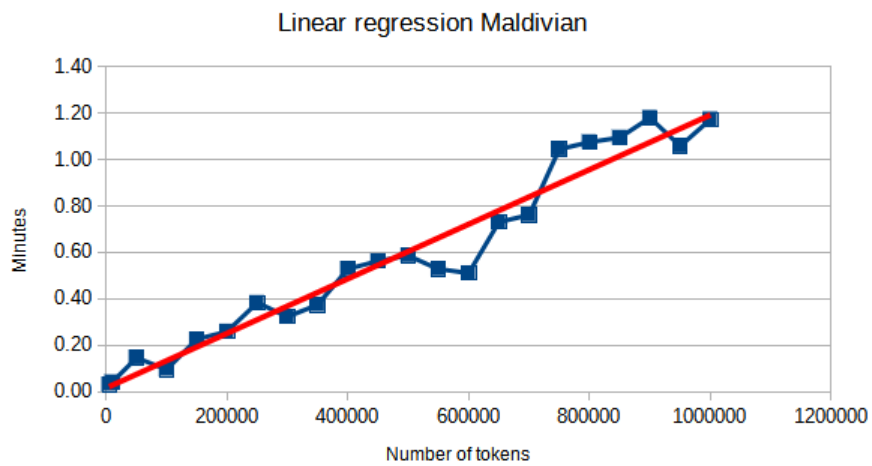


Figure 62: SM: Linear regression Maldivian

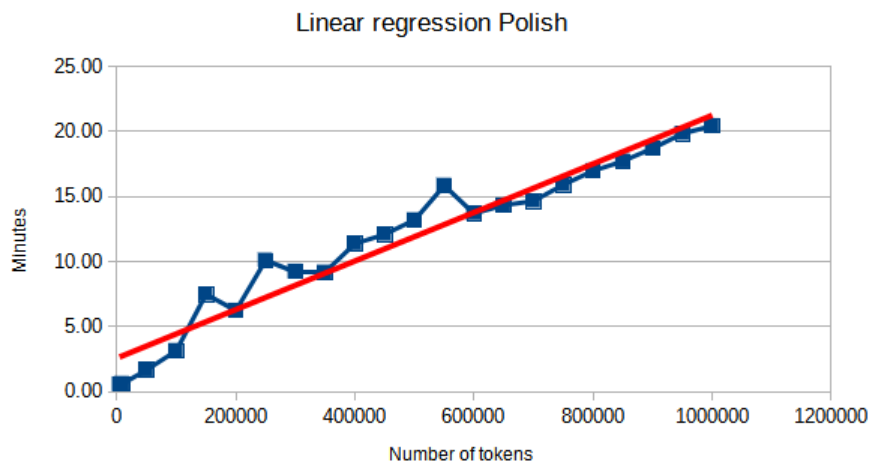


Figure 63: SM: Linear regression Polish

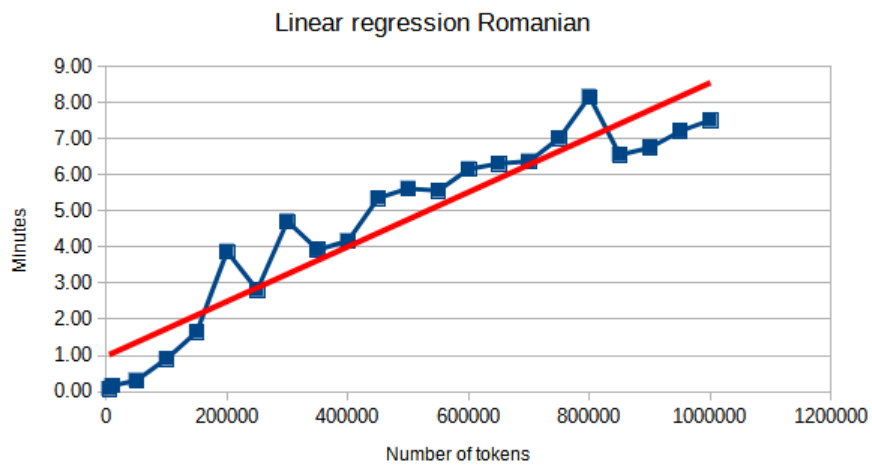


Figure 64: SM: Linear regression Romanian

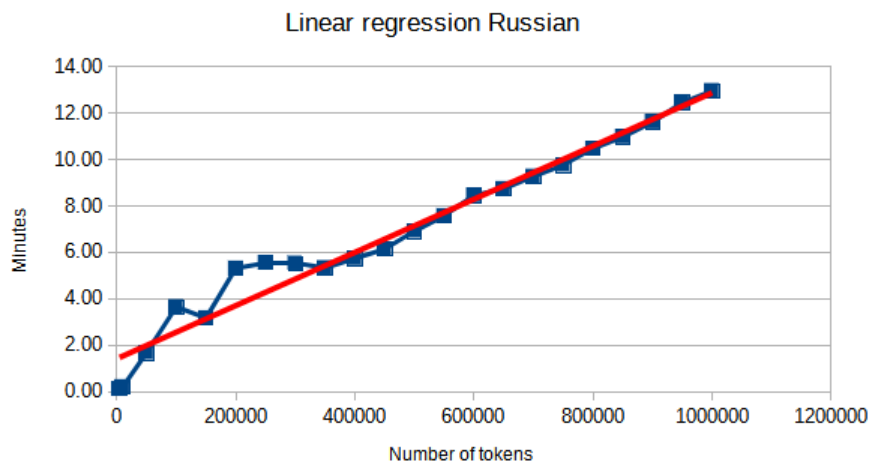


Figure 65: SM: Linear regression Russian

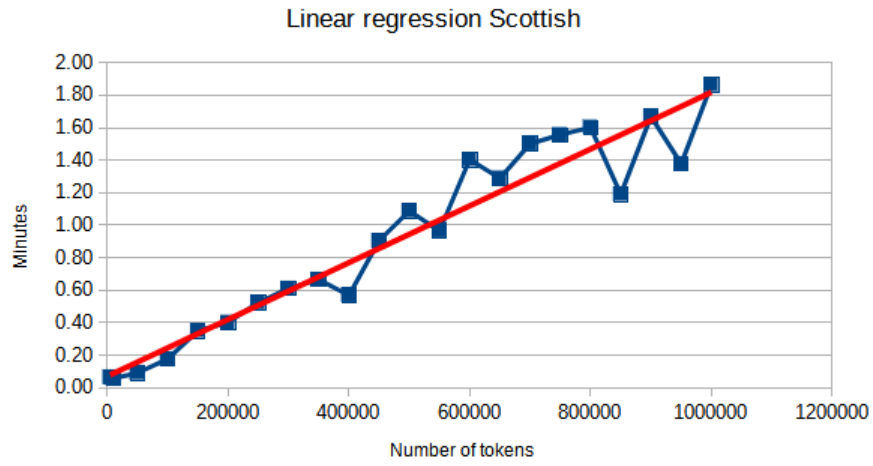


Figure 66: SM: Linear regression Scottish

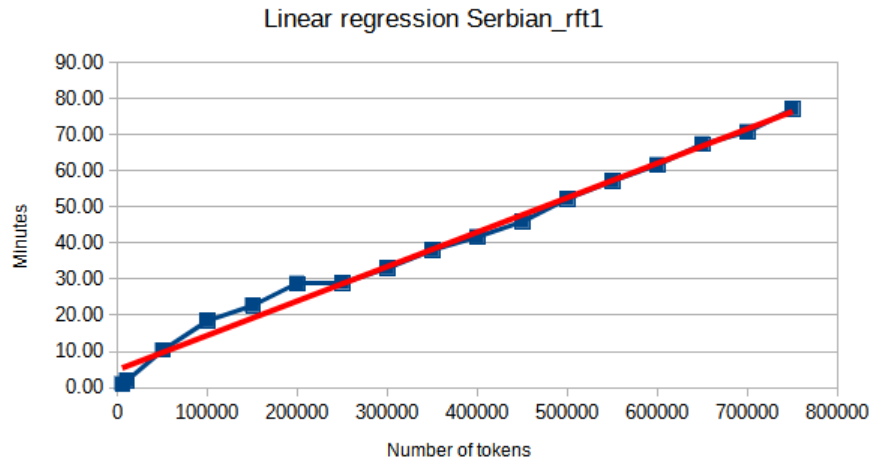


Figure 67: SM: Linear regression Serbian rft1

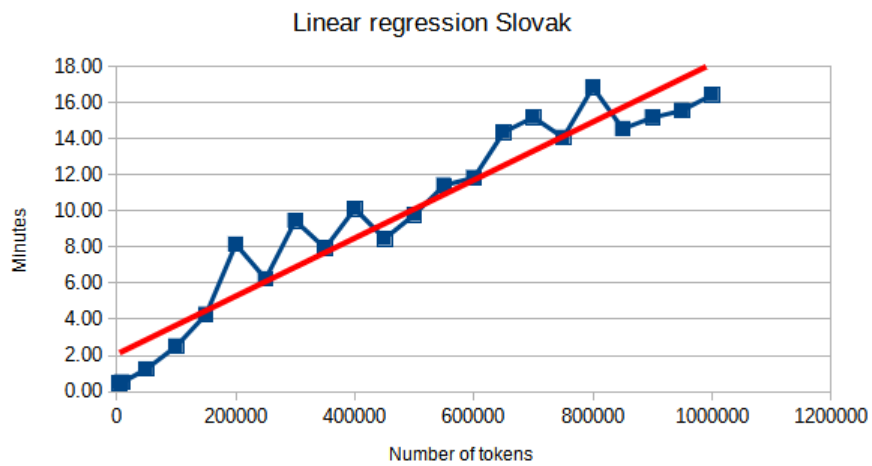


Figure 68: SM: Linear regression Slovak

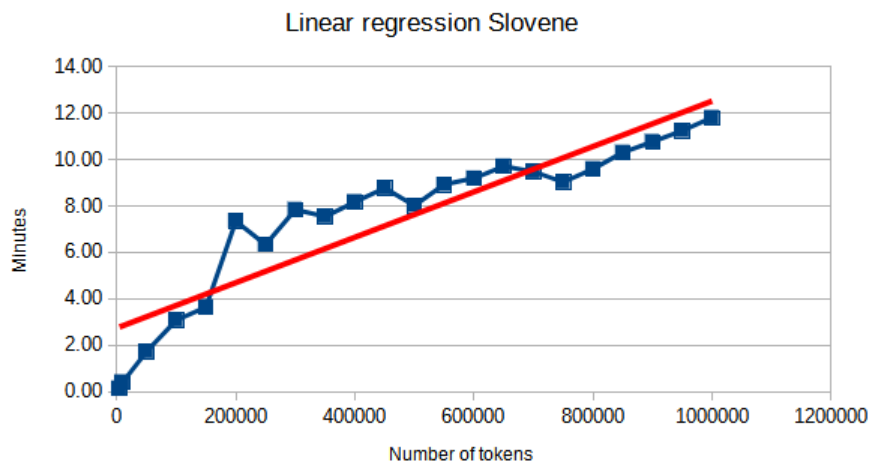


Figure 69: SM: Linear regression Slovene

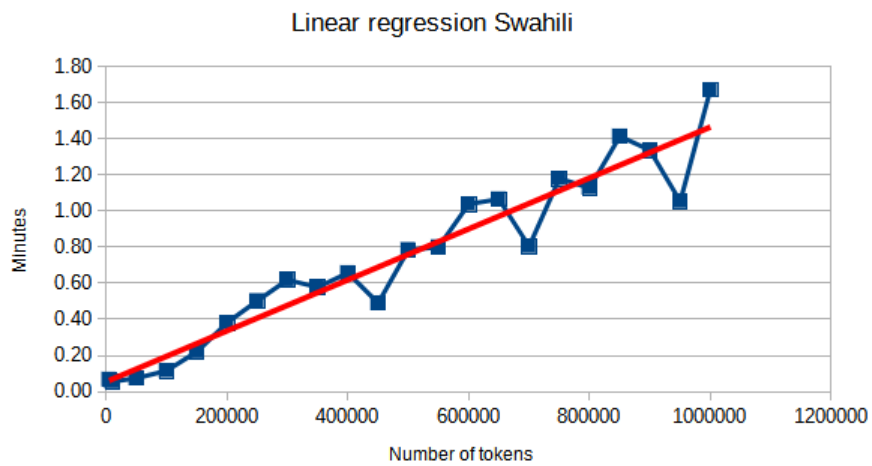


Figure 70: SM: Linear regression Swahili

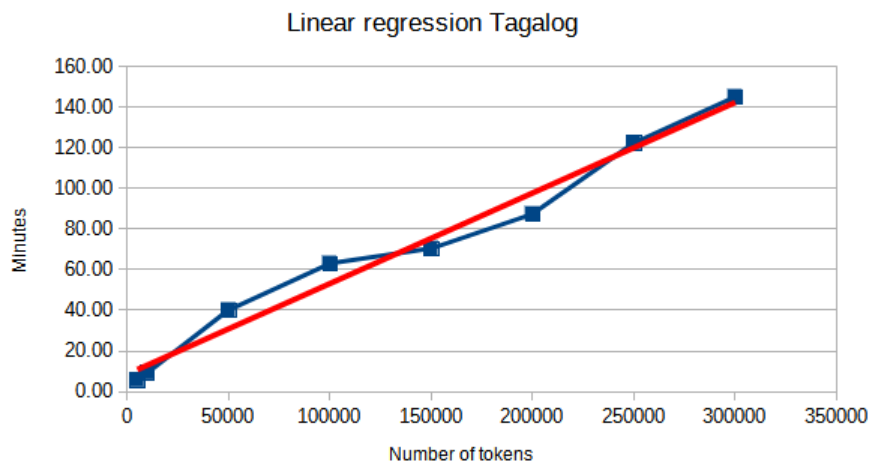


Figure 71: SM: Linear regression Tagalog

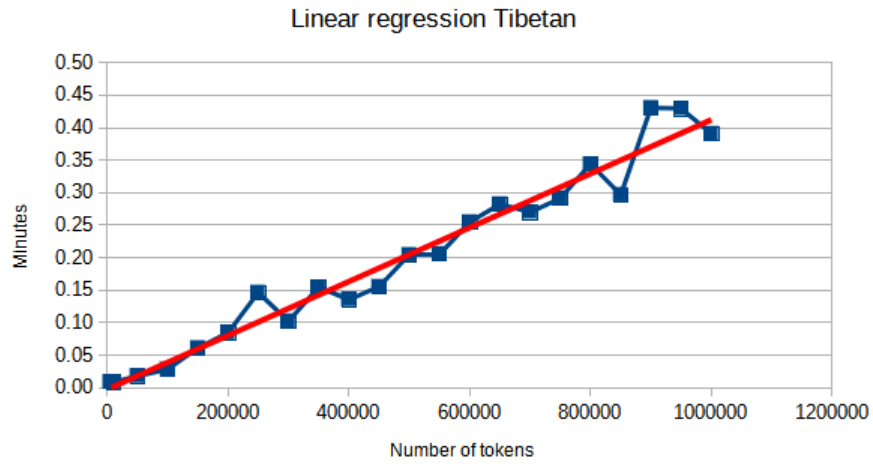


Figure 72: SM: Linear regression Tibetan

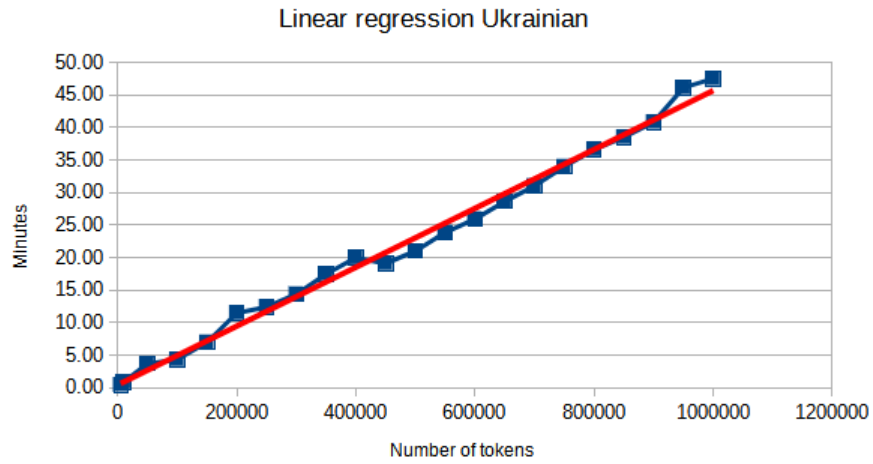


Figure 73: SM: Linear regression Ukrainian

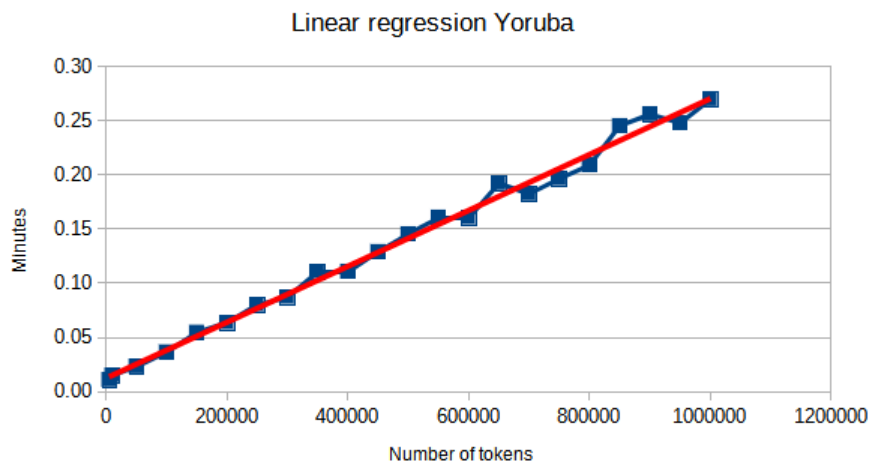


Figure 74: SM: Linear regression Yoruba

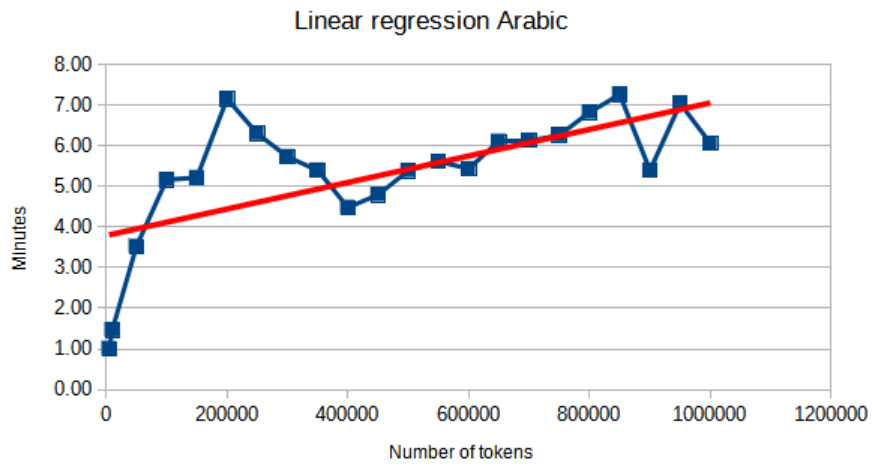


Figure 75: SM: Linear regression Arabic stanford

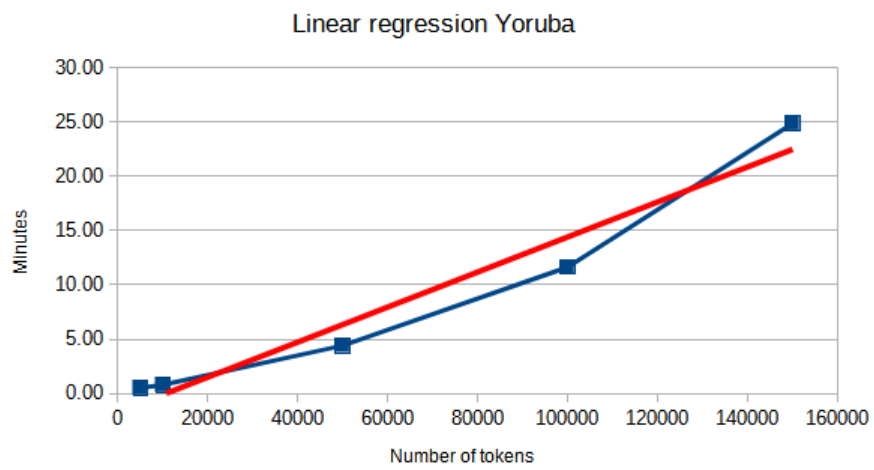


Figure 76: SM: Linear regression Hebrew yap