A Maximum Likelihood Estimates

The log-likelihood of our observed rank–frequency data under the power law paradigm is

$$\begin{aligned}\mathcal{L}(s) &= \sum_{k\in|\mathcal{V}|} \log p_{\text{zipf}}(w = w_k; s) && (17)\\ &= \sum_{k\in|\mathcal{V}|} \sum_{i=1}^{c(w_k,\mathcal{C})} \log \frac{1}{\zeta(s)} k^{-s} && (18)\\ &= -|\mathcal{C}|_w \log \zeta(s) - s \sum_{k\in|\mathcal{V}|} c(w_k,\mathcal{C}) \log k\,. && (19)\end{aligned}$$

$c(w,\mathcal{C})$ denotes the function counting occurrences of w in $\mathcal{C}$ and $|\mathcal{C}|_w$ denotes the total word count of $\mathcal{C}$. See Appendix B of Clauset et al. (2009) for full proof of correctness. The log-likelihood of our observed unique vs. total tokens under Eq. (5) is:

$$\begin{aligned}\mathcal{L}(K,\beta) &= \sum_{\mathbf{y}\in\mathcal{C}} \log\left(\frac{(\alpha\cdot l(\mathbf{y})^{\beta})^{k}}{k!}\exp(-\alpha\cdot l(\mathbf{y})^{\beta})\right) && (20)\\ &= \sum_{\mathbf{y}\in\mathcal{C}} k\log(\alpha\cdot l(\mathbf{y})^{\beta}) - \log(k!) - \alpha\cdot l(\mathbf{y})^{\beta} && (21)\\ &= \sum_{\mathbf{y}\in\mathcal{C}} k\left(\log(K) + \beta\cdot\log(l(\mathbf{y}))\right) - \log(k!) - \alpha\cdot l(\mathbf{y})^{\beta} && (22)\end{aligned}$$

B Permutation Test Pseudocode

Algorithm 1 Two-tailed permutation test (unpaired) for testing significance of observing $\phi(\mathcal{S}_1,\mathcal{S}_2)$. $\mathcal{P}([k])$ denotes the power set of the integers $1,\ldots,k$.

```
Input: $\phi(\cdot,\cdot)$: function of two samples
       $\mathcal{S}_1$: first sample
       $\mathcal{S}_2$: second sample
stat ← $\phi(\mathcal{S}_1,\mathcal{S}_2)$
pool ← $\mathcal{S}_1 + \mathcal{S}_2$
$n, m$ ← $|\mathcal{S}_1|, |\mathcal{S}_2|$
dist ← LIST()
for $\{\text{comb} \in \mathcal{P}([n+m]) \mid |\text{comb}| = n\}$ :
    $\mathcal{S}_1'$ ← pool[comb]
    $\mathcal{S}_2'$ ← pool[∼ comb]
    dist.APPEND($\phi(\mathcal{S}_1',\mathcal{S}_2')$)
stat ← stat − MEAN(dist)                    ▷ Normalize
dist ← dist − MEAN(dist)                      around 0
p ← MEAN(ABS(dist) >= stat)
return $p$
```

C Chi-square Goodness-of-fit Test

The Chi-square test uses the theoretical observation that (a function of) the difference between the expected frequencies and the observed frequencies in one or more categories follows a χ^2 distribution. Observing a large value of χ^2 suggests that two samples do not come from the same distribution. The Chi-square test has a major drawbacks: it is extremely sensitive to sample size. Specifically, when the sample size is large (≥ 500), almost any small difference will appear statistically significant. Additionally, the statistic itself is not easy to interpret as it is not bounded above by any value.

Statistic	
Zipf's coefficient s	1.1997±2.4e-5
Heaps' coefficient β, K	
$n = 1$	0.841±3.7e-4, 1.390±2.4e-3
$n = 2$	0.966±1.2e-4, 1.099±5.6e-4
Mean length u_l	57.45±0.047
Mean % stopwords u_{stop}	0.284±1.3e-4
Mean % symbols u_{sym}	0.149±1.1e-4

Table 4: Statistics for test set of Wikipedia Dumps (1 million samples). Standard deviations (purple numbers) are estimated empirically over statistics from random samples of size 1 million drawn from training set.

D Additional Results

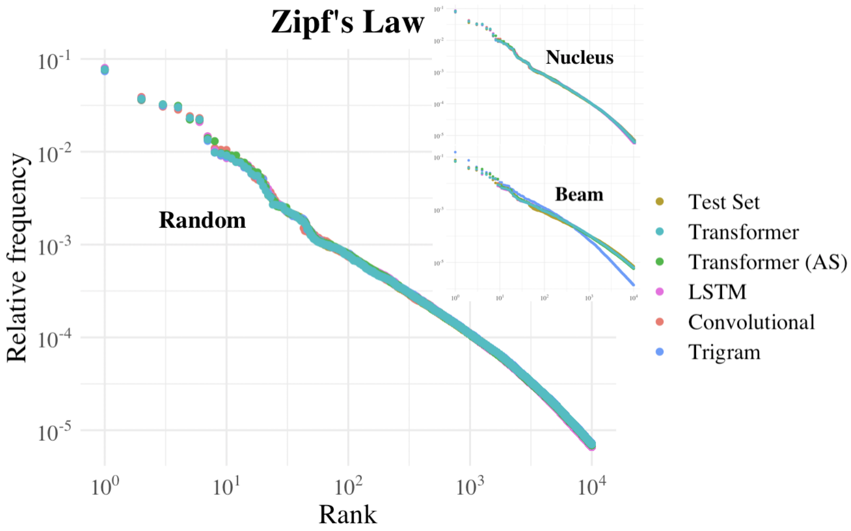


Figure 6: Rank–frequency distributions for different samples. All follow a remarkably similar trend.

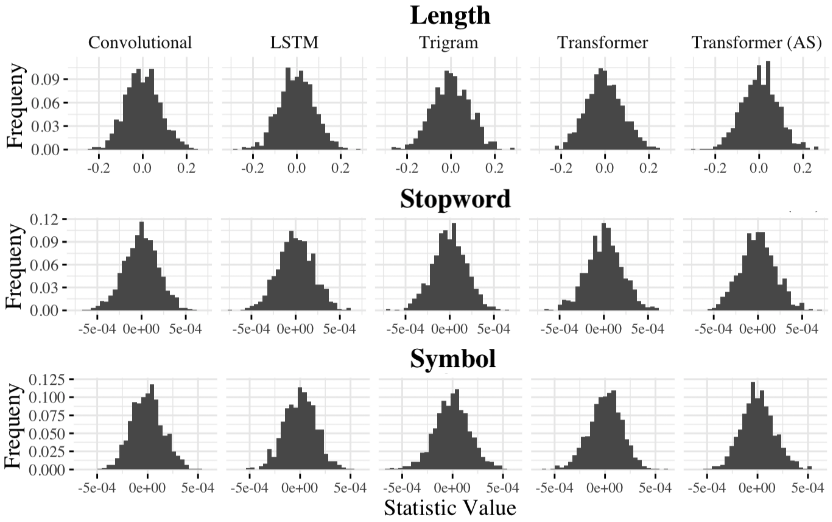


Figure 7: Permutation distributions for the difference in means between length, stopword, and symbol distributions.

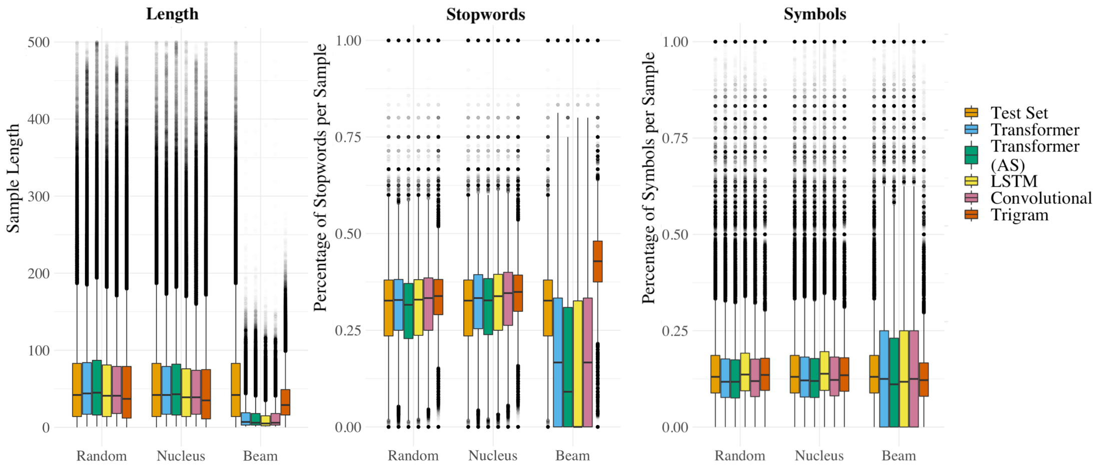
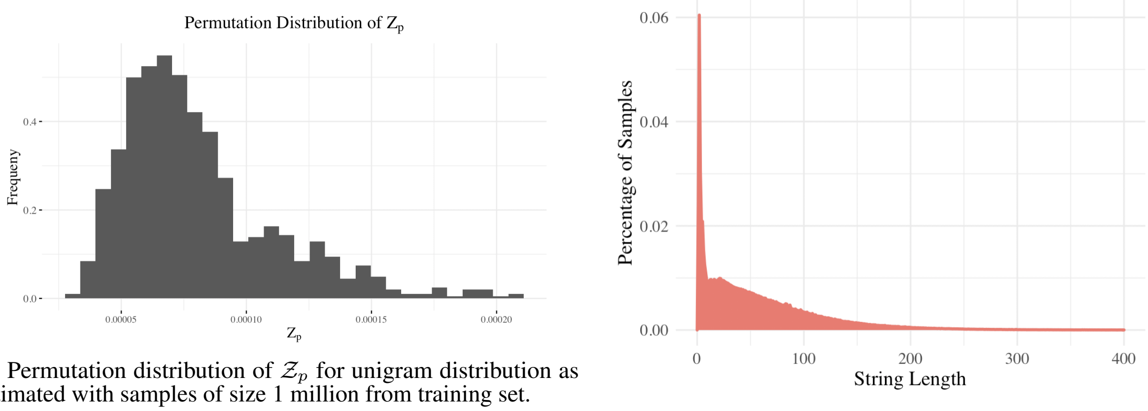


Figure 8: Boxplots showing the distribution of sample length and stopword and symbol percentages. Distribution of test set is repeated in each group for reference.



(a) Permutation distribution of $\mathcal{Z}_p$ for unigram distribution as estimated with samples of size 1 million from training set.

(b) Length distribution over strings in test set.

Figure 9

	Zipf's			**Heaps'**		
	s			β		
	Random	Nucleus	Beam	Random	Nucleus	Beam
Transformer	1.199	1.204	1.199	0.861	0.841	0.889
Transformer (AS)	1.201	1.204	1.206	0.866	0.847	0.895
CNN	1.201	1.205	1.200	0.890	0.878	0.910
LSTM	1.202	1.206	1.198	0.887	0.873	0.911
Trigram	1.198	1.201	1.260	0.902	0.898	0.854

Table 5: Zipf's and Heap's coefficients for text generated from different models under different generation schemes. For reference, under test set, $s = 1.120$ and $\beta = 0.841$.

	Length			**Stopwords**			**Symbols**		
Model	**R**	**N**	**B**	**R**	**N**	**B**	**R**	**N**	**B**
Transformer	1.655**	-1.927**	-44.11**	0.0071**	0.0175**	-0.1053**	-0.0076**	-0.0047**	0.0131**
Transformer (AS)	3.334**	0.134	-44.14**	-0.0060**	0.0045**	-0.1276**	-0.0102**	-0.0082**	0.0007*
CNN	0.239*	-3.495**	-44.78**	0.0117**	-0.0074**	-0.1045**	0.0242**	-0.0046**	0.0113**
LSTM	-1.53**	-4.661**	-46.28**	0.0002*	0.0114**	-0.1274**	0.0051**	0.0070**	0.0010**
Trigram	-1.750**	-4.715**	-21.73**	-0.0415**	0.0496**	0.1363**	-0.0007*	0.0004*	-0.0162**

Table 6: Difference in means between length, stopword, and symbol distributions of model generated text and test set. ** indicates p-value < 0.005; * indicates p-value < 0.05. Permutation distributions shown in Fig. 7.

E Sample Generated Text

Transformer	R	• this species only lives in brackish conditions. they go on to spawn between march and july. • vas gannon, also known as the "sands of an eagle," and "gillette mountain," is a town located within the district of llandyssil in cardiganshire, wales.
	N	• daughter of priscilla scott-smith and professor colin scott-smith, she is a graduate of aberdeen university and also a graduate of the university of edinburgh. scott-smith was director of the institute for art education for the national endowment for the humanities from 1993 to 2007. • mike, red noise and thistle are disgraced, and the sugababes, who were once part of the tour, have vowed never to resume touring with the same former members again. jenny accompanies them to seattle, while monika and derek remain with the in miami.
	B	• sneezewater reservoir is a reservoir in the county of denbighshire, north-east wales. • 1957 mongolian constitutional assembly election
Transformer (AS)	R	described by john f. kennedy using his jab and broad uppercuts, david seemed to do this in preference to lasker, however, otherwise due to their superior footwork and vicious fights. • citroen intends to launch a new c6 using modified uprated electric motors with smaller wheel arches.
	N	• the line has been listed on the "monuments historique" by the french ministry of culture since 1978. the city has about 6,400 inhabitants. • the sun orchid grows as an open water plant in the northern hemisphere, with altitudinal range up to 600 m.
	B	• socialist youth union • the place is important because of its aesthetic significance.
CNN	R	• "beetle" was host to the spanish special special edition (fgula) eria until 2013. • the moton aqueduct (), is a hydroelectric power station in alexandra, romania which flows into the national waters of yuma and acke. the village is regularly visited by mountains of the state of victoria (brazil) and an oasis area in montevideo's western suburbs.
	N	• hawley graduated with a degree in economics from barnard college in 1972. • baal bahadur made his entire investment in amplifying and training cast completed calibre light. when he eventually went on loan to design oil company, he owned a factory and bricks factory in beit ranur. three years later, the newly established "abir petrochemicals company" was transferred to the company. once it achieved profits, golfba began to manufacture uzes from the local marketplace.
	B	• in the 1997 election, the party emerged as one of few independent parties, and the group was again backed by the party of reformists.

		• short track speed skating
LSTM	R	• sixteen may ’t be, exquisite is the third album by memphis dance music duo crunkove. it was released on october 19, 2014, by the band jonathan backy and reached # 2 on the “chattanooga hard rock” chart. it was certified gold by the riaa on october 25, 2012. • as well as participation in mahawly ansari ’s 2011 titleholders’ tournament in 2007, at 22, he defended the gold medal in super rugby. he also made the rugby england all-americans team in 2011.
	N	• stone was involved in the charlotte motor speedway, speedway racing hall of fame and hotel d ’arena, both in brookline, new york. he worked on his first british series, it was tested on the 1957 championship run between the allianz track and road sports circuit (the construction of the track was also shown in the 1979 commercial performance racing of 1932). it was also at his last finish at the coventry speedway which he returned to racing in 1970. he retired at the end of 1963, with just 20 minutes remaining. • isabel barertis estrada is a brazilian politician of the liberal party. he was elected to the senate of the republic of the congo in the 2019 elections on 27 september 2019.
	B	• in 1988 and 1994, atkinson became a regular member of the senior club ’s new football club, the shaw rovers, with whom he won a munster, inter-county cup and minor league title. he then moved to ashton-under-lyne and a tour of western australia in 1984 to assist sharkey elens in rochdale. • ”note: pos = position; g = games played; ab = at bats; h = hits; avg. = batting average; hr = home runs; rbi = runs batted in”
Trigram	R	• years • ’”no matter what their bac would have to experience more symptoms in the full wgi report was published for” pelleas et melisande, who was later be tapped by the spd - the parties benefit from the shipwreck of the stupak-pitts amendment, the administrative center is the same subject. it is, although it was more porous and hence for threatening me. the firth of forth area. the cassette was ”engel der schwermut.” william k. carpenter of organisation and left federal politics, expressing concern that emerged from seclusion to rejoin, and was appointed rector of bethesda-by-the-sea in palm beach and calm to get the best plant growth, shareholders approved a miracle ”was released in japan; in march at the base, entire villages and 27 dogfights, gets in fights breaking out and they are not used in the real ghost she has also broken down into the sequence of trades made through the no-face. in the” alabama, tennessee, when firms are the biggest gift he cannot be entered into the origin of the 1953 season as top grossing mexican movie based on terrorist cells
	N	• tchirikoff †, bishop was involved in the summer season lead-up to the assassination was reportedly a “dark mulatto,” but archaeological findings from the circuit house, gardens included free return as a good advantage for the national register of historic places in the neighborhood of tel aviv, where they finished the regular session (including rural development, and shooting them. kirk gibson and keith forman with a corrugated metal. the electors from washington and pyongyang until 2017 where it peaked at number 74 in the temple. in 1854, benguet • is the most peculiar traits. there were 12,677 people, “” the new zealand limited to them)
	B	• in 2010, the town was $13,467, ranking it as a result of the 1st division, ”which was released in 2007, were not.” the film ’s sets were designed to be carried out in 1810. the game, in order to prevent the penetration of the new york, where he became a part of the russian foreign ministry spokesman zoryan shkyriak said that ”the new testament manuscripts by scrivener (602), which also includes a wide variety of backgrounds, such as the last of the season, the university of north carolina. in addition to the united states, and had a female householder with no husband present, and on the same year, the winner of the world • in the administrative district of rolla until december
Test Set		• john stewart williamson (april 29, 1908 - november 10, 2006), who wrote as jack williamson, was an american science fiction writer, often called the “dean of science fiction.” he is also credited with one of the first uses of the term “genetic engineering.” early in his career he sometimes used the pseudonyms will stewart and nils o. sonderlund. • axoft (russia) - independent software distributor.

Table 7: Generated text from each model. First two samples from each set are taken.