Is Brevity the Soul of Wit? An Empirical Test Using 570,000 Reddit Jokes

Robin Young
University of Cambridge
Cambridge, UK

Abstract

The assertion that "brevity is the soul of wit" has become proverbial in discussions of effective communication. We test this hypothesis using 570,634 jokes from Reddit's r/Jokes community and a cross-platform comparison of 22.1 million posts across 10 subreddits. Longer posts receive more upvotes across all of Reddit, not just jokes, indicating the baseline length-score correlation reflects platform dynamics rather than humor-specific effects. However, we find a pattern unique to humor: the length-score correlation disappears entirely for top-performing jokes, while every non-humor subreddit retains a positive correlation for its best content. For genuine wit, length is irrelevant, which is a property that distinguishes humor from other content types tested. Temporal analysis reveals winner-take-all consolidation. As Reddit matured, long jokes' median scores dropped 71% while mean scores increased 6.8×, indicating most long jokes now fail while rare successes dominate.

1 Introduction

The relationship between brevity and effectiveness has been debated since at least Shakespeare (2016)'s *Hamlet*, where Polonius declares that "brevity is the soul of wit." This aphorism has become influential in writing instruction. Strunk and White (2000)'s foundational *The Elements of Style* enshrines conciseness as a core principle and in communication theory, where message length can function as a peripheral cue affecting persuasion (Petty and Cacioppo, 1986), but the specific case of humor has had relatively limited empirical attention. As such, we focus here on text-based jokes posted to a dedicated humor forum.

Understanding what factors contribute to humor success has practical implications for natural language generation, particularly humor generation systems (Weller and Seppi, 2020), which remain challenging despite advances in language modeling. If brevity indeed enhances wit, systems should optimize for conciseness. However, if the relationship differs from conventional wisdom, this informs both generation strategies and training data selection.

We test this hypothesis using 570,634 jokes from Reddit's r/Jokes subreddit spanning 2008–2020, using crowdsourced upvotes as a proxy for perceived humor, and a cross-platform comparison of 22.1 million posts across 10 subreddits. This dataset enables analysis at scale compared to typical humor studies (which use hundreds to thousands of examples), and provides ecological validity through real-world engagement metrics rather than elicited laboratory ratings. We operationalize "wit" as humor that succeeds with an audience, measured through community engagement, while acknowledging that Shakespeare's concept encompasses broader notions of cleverness and verbal dexterity. We focus specifically on text-based canned jokes posted to a dedicated humor forum and our findings should not be generalized to conversational, performative, or multimodal humor, where brevity may function differently.

Our findings reveal that longer posts receive more upvotes across all of Reddit, not just jokes, indicating the baseline length-score correlation ($r = +0.165$) reflects platform dynamics rather than a humor-specific effect. However, we identify a pattern unique to humor. The correlation disappears entirely for top-performing jokes ($r = -0.007$ for the top 1%), while every non-humor subreddit retains a positive correlation for its best content. Temporal analysis reveals platform maturation transformed the length-score relationship into winner-take-all dynamics, with implications for understanding selection bias in humor datasets.

Proceedings of the 15th International Joint Conference on Natural Language Processing and the 5th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, pages 254–261
November 6-10, 2026 ©2026 Association for Computational Linguistics

2 Related Work

Computational Humor Early work on automatic humor recognition used small datasets with hundreds of examples (Mihalcea and Strapparava, 2005). Recent advances employ deep learning (Chen and Soo, 2018) and pretrained language models (Weller and Seppi, 2019) on larger datasets. Theoretical frameworks for joke classification have drawn on script-based and incongruity theories to categorize humor types (Ahuja et al., 2018), though computational operationalization remains challenging. Humor generation has progressed from template-based approaches (Stock and Strapparava, 2005) to neural methods (He et al., 2019; Yu et al., 2018), though generating funny text remains challenging. Weller and Seppi (2020) introduced the r/Jokes dataset with 573k jokes, which we use here. However, their focus was detection rather than characterizing humor dynamics. Recent work has explored LLM capabilities for humor (Jentzsch and Kersting, 2023), but generation quality remains limited.

Social Media and Humor Engagement Research on content virality has identified factors affecting success, including emotional arousal (Berger and Milkman, 2012), timing (Muchnik et al., 2013), and social influence (Salganik et al., 2006). On Twitter, link and hashtag presence (Suh et al., 2010). Reddit-specific work has examined title effects (Lakkaraju et al., 2021), and linguistic features (Gilbert, 2013). At the intersection of humor and social media, Yang et al. (2021) use Facebook reactions as crowdsourced humor labels on 785k posts, demonstrating that platform engagement signals can serve as scalable humor proxies beyond Reddit. Zeng et al. (2024) incorporate social context graphs for humor recognition on Weibo, finding that engagement patterns differ systematically between humorous and non-humorous content. However, no prior work has examined how the relationship between content length and engagement evolves as platforms mature, or whether length-engagement patterns differ across content types.

Length and Engagement Research on the relationship between message length and success in social media has produced mixed findings. Gligorić et al. (2018) exploit Twitter's expansion from 140 to 280 characters as a natural experiment, finding that the added space changed writing style but had complex effects on engagement. Gligorić et al. (2019) extend this to establish causal effects of brevity on social media success. Jaidka et al. (2019) find that Twitter's character constraint shapes political discussion dynamics. Tan et al. (2014) use topic- and author-controlled natural experiments to isolate wording effects on message propagation. These studies employ quasi-experimental designs enabled by platform-specific affordances (Twitter's character limit change). Our work complements this line of research with a correlational approach on a platform without character constraints, focusing specifically on how length-engagement dynamics differ between humor and non-humor content.

3 Data and Methods

We use the r/Jokes dataset (Weller and Seppi, 2020), consisting of 1,064,928 jokes scraped from Reddit's r/Jokes subreddit (January 2008 - December 2020). We remove posts marked [deleted] or [removed] (492,065 jokes), filtering for valid content, and removing extreme outliers (< 20 or > 5000 characters). Our final dataset contains 570,634 jokes with valid scores.

Each joke includes: (1) title and body text, which we concatenate as the full joke; (2) upvote score, representing aggregated community assessment; (3) timestamp. Jokes range from 20 to 4,995 characters (mean=232, median=97) and scores range from 0 to 136,353 (mean=218, median=5).

We use Reddit upvotes as a proxy for perceived humor, following Weller and Seppi (2020). While upvotes are influenced by factors beyond humor quality (timing, visibility), they represent aggregated judgments from thousands of readers in natural consumption contexts. Within a dedicated humor forum, it is reasonable to assume that perceived funniness is a primary driver of this engagement. Therefore, while noisy, this metric reflects a joke's ability to succeed. The scale of our dataset (570k jokes) and temporal stability of findings (9 years) suggest the signal-to-noise ratio is sufficient for robust analysis. This approach provides ecological validity and scale infeasible with annotation.

We compute Spearman correlation between joke length (character count) and upvotes to measure the overall relationship. To examine functional form, we bin jokes by length into deciles and compute mean scores per bin. We assess temporal stability by repeating analyses within each year (2011-2019, excluding years with < 1000 jokes). To

test whether the relationship differs for successful jokes, we compute correlations within score quantiles (top 50%, 25%, 10%, 5%, 1%). All statistical tests use two-tailed significance at $\alpha = 0.05$.

4 Results

4.1 Positive Correlation

Contrary to Shakespeare's hypothesis, longer jokes receive more upvotes (Spearman $r = +0.165$, $p < 0.001$; Pearson $r = +0.046$, $p < 0.001$). Figure 1 shows a consistent increase across length deciles: the longest decile (626-4,995 characters) scores 6.1× higher than the shortest (20-53 characters), with mean scores of 473 vs 77 upvotes respectively.

The divergence between the Spearman correlation and the near-zero Pearson correlation suggests a highly non-linear nature to this relationship. A linear model fails entirely, as the trend is dominated by the power-law distribution of scores, a dynamic we explore further below. However, the effect size is small. Length explains only 2.7% of score variance, and Cohen's $d = 0.140$ when comparing the shortest vs longest quartiles, classified as negligible. This suggests that while length correlates with success, it is not a primary determinant of humor.

4.2 Top Jokes Differ

The length-score relationship varies dramatically by success level (Figure 2). For all jokes, $r = +0.165$, but this weakens progressively for higher-scoring subsets: top 10% ($r = +0.070$), top 5% ($r = +0.039$), and top 1% ($r = +0.010$, effectively zero). This suggests that while length correlates with average performance, it becomes irrelevant for viral humor, where quality dominates.

Mean length also increases with score threshold (all jokes: 232 chars; top 1%: 398 chars), but the correlation collapse indicates this reflects confounding factors (longer jokes may have more opportunities for quality) rather than length directly causing success at the top.

4.3 Winner-Take-All Over Time

Analysis across nine years (2011-2019) reveals consistent positive correlations (mean $r = 0.200$, std=0.063, all years positive), but dramatic distributional changes. For long jokes (> 400 chars), median scores decreased 71% (46 → 13 upvotes) while mean scores increased 6.8× (92 → 631 upvotes). Distribution skew increased from 2.0× to 49× (mean/median ratio).

Figure 3 illustrates this transformation. Early Reddit (2011) shows relatively normal distribution, where most long jokes received moderate success. Mature Reddit (2019) exhibits extreme power law: most long jokes score poorly, but rare successes achieve extreme scores. This winner-take-all pattern emerged as Reddit scaled from niche community (100k subscribers in 2011) to mainstream platform (20M subscribers by late 2020).

4.4 Cross-Subreddit Comparison

To test whether the positive length-score correlation is specific to humor or reflects broader platform dynamics, we computed length-score correlations across 10 subreddits using Pushshift data (Baumgartner et al., 2020) (22.1M posts, 2012–2018, same filtering as our primary analysis). The positive correlation is not specific to r/Jokes, we find that subreddits such as r/nosleep ($r = +0.418$), r/tifu ($r = +0.399$), and r/todayilearned ($r = +0.334$) show substantially stronger effects (Table 7 in Appendix E). This indicates the baseline correlation reported earlier reflects platform-wide dynamics rather than a humor-specific phenomenon.

However, the correlation decay for top-performing content is unique to humor. At the top 1% by score, every non-humor subreddit retains a positive length-score correlation, while r/Jokes reaches effectively zero ($r = -0.007$, $t = -0.61$, $p = 0.54$, $n \approx 7{,}938$; Table 8 in Appendix E). Against the non-humor distribution (mean $r = +0.077$, std$= 0.036$), r/Jokes is a significant outlier ($z = -2.32$). While Reddit generally rewards length across all content types, the best humor is uniquely length-independent.

5 Discussion

The cross-subreddit comparison provides context for the r/Jokes only findings. The overall positive correlation between length and upvotes in r/Jokes is not by itself evidence that longer jokes are funnier, but it is consistent with a general Reddit tendency to reward longer content, likely reflecting algorithmic sorting, self-selection, and effort signaling. However, the uniquely humor-specific correlation decay for top content suggests that for genuine wit, length simply does not matter. This property distinguishes humor from every other content type tested. Correlationally, the best stories, explanations, and opinions still benefit from length, but the best jokes

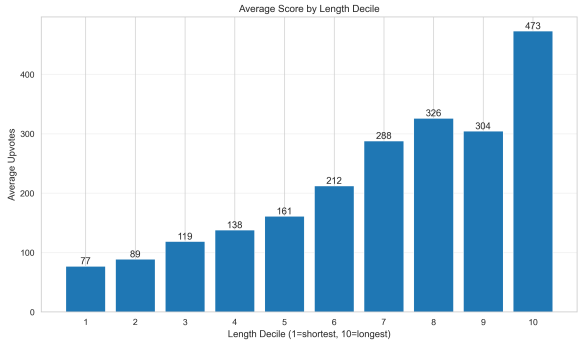


Figure 1: Average upvotes by joke length decile. Longer jokes receive higher scores. Each decile contains approximately 57,000 jokes.

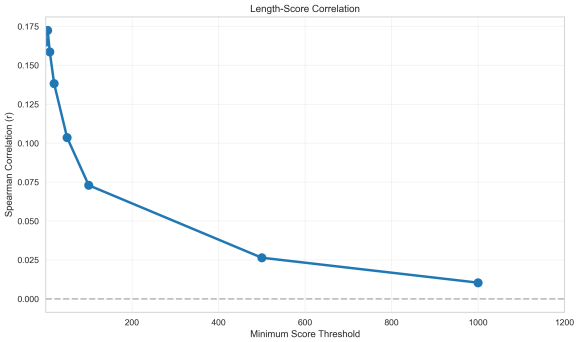


Figure 2: The length-score correlation weakens for high-performing jokes, approaching zero for the top 1%.

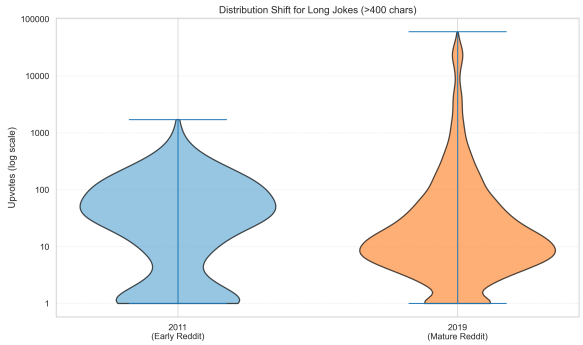


Figure 3: Distribution shift for long jokes (> 400 chars) from 2011 to 2019. Early Reddit shows relatively normal distribution (median=34, mean=78, skew=2.3×). Mature Reddit shows extreme power law (median=12, mean=588, skew=49×), indicating that most long jokes fail, but successful ones dominate.

do not.

The apparent vanishing effect for top jokes could have interesting implications for training humor generation systems. Standard practice in humor detection uses upvote thresholds to define successful jokes for training (Weller and Seppi, 2019, 2020), assuming highly-upvoted examples represent "good" humor. However, these examples come from the tail of a power law distribution as they are outliers, not typical successes. This may partially explain why humor generation remains challenging (Jentzsch and Kersting, 2023), namely that models learn from exceptional cases that may not generalize.

Additionally, the winner-take-all dynamics suggest length is not a reliable optimization target. While longer jokes have higher mean success, most fail. It is apparent that content quality matters far more than length for viral success, but the attribution of content quality remains a challenge.

The transformation from normal to power law distribution reflects broader patterns in digital platforms, where algorithmic amplification and attention scarcity create winner-take-all dynamics (Frank and Cook, 1995; Weng et al., 2012). As Reddit scaled, early success signals created positive feedback loops (Muchnik et al., 2013), raising the effective "cost" of long jokes: most fail, but the few that succeed do so because their quality justifies the reader's investment, producing the extreme mean/median split observed. This has methodological implications for social media research: raw engagement metrics should be interpreted in context of platform temporal dynamics, as comparing 2011 and 2019 jokes using raw upvotes would be misleading.

6 Conclusion

We test the claim that "brevity is the soul of wit" using 570,634 Reddit jokes and a cross-platform comparison of 22.1 million posts across 10 subreddits. The positive correlation between length and upvotes ($r = +0.165$) is not humor-specific as it exists across all of Reddit, reflecting platform dynamics rather than humor mechanics. However, we identify a pattern unique to humor, where the length-score correlation vanishes entirely for top-performing jokes ($r = -0.007$ for the top 1%), while every non-humor subreddit retains a positive correlation for its best content. For genuine wit, length is irrelevant.

Temporal analysis reveals winner-take-all consolidation as Reddit matured, and long jokes' median scores dropped 71% while mean scores increased 6.8×. This has implications for humor generation systems, which should focus on quality rather than length optimization, and for dataset

construction, as highly-upvoted examples represent power-law outliers rather than typical humor success.

Limitations

Our study has several limitations. Reddit upvotes are an imperfect humor proxy, influenced by timing, visibility, and community dynamics beyond pure funniness. Additionally, Reddit upvotes reflect both perceived quality and visibility, and we cannot fully disentangle these factors. Joke visibility is influenced by Reddit's ranking algorithm, which promotes content based on early engagement signals and may differentially affect posts of different lengths. It is possible that longer jokes receive more visibility through algorithmic amplification, or conversely, that the greater time investment required to read them suppresses casual engagement. View counts, which would allow normalization for exposure, are not available through Reddit's API or in the dataset.

The positive length-score correlation is consistent across nine years (2011–2019), during which Reddit's ranking algorithm underwent multiple changes, suggesting the pattern is not purely an artifact of any single algorithmic regime. Nevertheless, disentangling the contributions of humor quality, algorithmic visibility, and reader investment to the observed correlation remains an open question that would require platform-level data to resolve.

Results may not generalize beyond Reddit's predominantly young, western, male-skewed demographic, or beyond English text-only jokes to other humor modalities such as conversational, performative, visual, or video humor. Cross-platform and multilingual studies would also strengthen external validity.

We measure correlation, not causation. While length correlates with success, we cannot conclude length causes higher scores. Controlled experiments (posting jokes of different lengths while controlling content) would be needed for causal claims, though such interventions raise ethical concerns about platform manipulation.

Finally, our analysis focuses on length as a single feature. Future work could examine interactions with other features (topic, structure, linguistic complexity) to provide more comprehensive understanding of humor dynamics.

7 Ethical Considerations

This research uses publicly available Reddit data accessed through the preprocessed dataset (Weller and Seppi, 2020). No personally identifiable information is included. All jokes remain anonymous, with only upvote counts and timestamps analyzed. We acknowledge that humor can reflect societal biases, but our focus on length rather than content minimizes potential harms.

References

Vikram Ahuja, Taradheesh Bali, and Navjyoti Singh. 2018. What makes us laugh? investigations into automatic humor classification. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pages 1–9, New Orleans, Louisiana, USA. Association for Computational Linguistics.

Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):830–839.

Jonah Berger and Katherine L. Milkman. 2012. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205.

Peng-Yu Chen and Von-Wun Soo. 2018. Humor recognition using deep learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 113–117, New Orleans, Louisiana. Association for Computational Linguistics.

Robert H. Frank and Philip J. Cook. 1995. *The Winner-Take-All Society: Why the Few at the Top Get So Much More Than the Rest of Us*. Free Press, New York, NY.

Eric Gilbert. 2013. Widespread underprovision on reddit. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, page 803–808, New York, NY, USA. Association for Computing Machinery.

Kristina Gligorić, Ashton Anderson, and Robert West. 2018. How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12.

Kristina Gligorić, Ashton Anderson, and Robert West. 2019. Causal Effects of Brevity on Style and Success in Social Media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23.

He He, Nanyun Peng, and Percy Liang. 2019. Pun generation with surprise. In *Proceedings of the 2019*

Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 1734–1744, Minneapolis, Minnesota. Association for Computational Linguistics.

Kokil Jaidka, Alvin Zhou, and Yphtach Lelkes. 2019. Brevity is the Soul of Twitter: The Constraint Affordance and Political Discussion. *Journal of Communication*, 69(4):345–372.

Sophie Jentzsch and Kristian Kersting. 2023. ChatGPT is fun, but it is not funny! humor is still challenging large language models. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 325–340, Toronto, Canada. Association for Computational Linguistics.

Himabindu Lakkaraju, Julian McAuley, and Jure Leskovec. 2021. What's in a name? understanding the interplay between titles, content, and communities in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, pages 311–320.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Lev Muchnik, Sinan Aral, and Sean J. Taylor. 2013. Social influence bias: A randomized experiment. *Science*, 341(6146):647–651.

Richard E. Petty and John T. Cacioppo. 1986. The elaboration likelihood model of persuasion. In *Advances in Experimental Social Psychology*, volume 19, pages 123–205.

Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856.

William Shakespeare. 2016. *Hamlet*, 2nd edition. The Arden Shakespeare Third Series. Bloomsbury Publishing, London.

Oliviero Stock and Carlo Strapparava. 2005. HAHAcronym: A computational humor system. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 113–116, Ann Arbor, Michigan. Association for Computational Linguistics.

William Jr. Strunk and E. B. White. 2000. *The Elements of Style*, 4 edition. Longman, New York.

Bongwon Suh, Lichan Hong, Peter Pirolli, and Ed H. Chi. 2010. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 IEEE Second International Conference on Social Computing*, pages 177–184.

Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 175–185, Baltimore, Maryland. Association for Computational Linguistics.

Orion Weller and Kevin Seppi. 2019. Humor detection: A transformer gets the last laugh. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3621–3625, Hong Kong, China. Association for Computational Linguistics.

Orion Weller and Kevin Seppi. 2020. The rJokes dataset: a large scale humor collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6136–6141, Marseille, France. European Language Resources Association.

Lilian Weng, Alessandro Flammini, Alessandro Vespignani, and Filippo Menczer. 2012. Competition among memes in a world with limited attention. *Scientific Reports*, 2:335.

Zixiaofan Yang, Shayan Hooshmand, and Julia Hirschberg. 2021. CHoRaL: Collecting humor reaction labels from millions of social media users. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4429–4435, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zhiwei Yu, Jiwei Tan, and Xiaojun Wan. 2018. A neural approach to pun generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1660, Melbourne, Australia. Association for Computational Linguistics.

Zeyuan Zeng, Zefeng Li, Liang Yang, and Hongfei Lin. 2024. Leveraging social context for humor recognition and sense of humor evaluation in social media with a new Chinese humor corpus - HumorWB. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10393–10402, Torino, Italia. ELRA and ICCL.

A Threshold Sensitivity Analysis

Our primary analyses use thresholds of 20–5,000 characters for joke length filtering, >400 characters for the "long joke" category, and a minimum of 1,000 jokes per year for temporal analysis. We justify these choices and demonstrate robustness to alternative thresholds.

The 20-character minimum removes posts that are predominantly remnants of deleted content such as titles without joke bodies where the punchline was removed (e.g. residual text after `[deleted]`

or [removed] filtering). The 5,000-character maximum removes 545 posts (0.05% of the dataset). The 1,000-joke yearly minimum excludes years with insufficient data for stable correlation estimates in 2008 (63 jokes), 2009 (185), 2010 (431), and 2020 (339).

Table 1 and Table 2 show that the Spearman correlation remains stable across all reasonable minimum and maximum thresholds, ranging from $r = +0.14$ to $+0.17$. The direction never reverses, and effect sizes remain consistently small (Cohen's $d < 0.20$) across all split points tested (Table 3).

Min Length	n	Spearman r	Mean Score
0	608,910	+0.165	197.5
10	581,498	+0.166	205.6
20	**570,634**	**+0.165**	**217.6**
30	556,498	+0.165	221.2
50	498,271	+0.159	236.9
75	413,652	+0.149	262.6
100	341,908	+0.152	288.5

Table 1: Sensitivity to minimum length threshold (maximum fixed at 5,000). **Bold** indicates the threshold used in the paper.

Max Length	n	Spearman r	Mean Score
500	520,870	+0.144	188.7
1,000	558,283	+0.159	207.8
2,000	567,577	+0.165	215.9
3,000	569,569	+0.166	217.2
5,000	**570,634**	**+0.165**	**217.6**
10,000	571,029	+0.166	217.7
50,000	571,179	+0.165	217.7

Table 2: Sensitivity to maximum length threshold (minimum fixed at 20). **Bold** indicates the threshold used in the paper.

Split	Short ($\leq$)	Long ($\geq$)	Δ Mean	Cohen's d
P10 vs P90	35 chars	524 chars	+305.8	0.189
P20 vs P80	46 chars	332 chars	+236.1	0.154
P25 vs P75	53 chars	266 chars	+200.3	0.140
P33 vs P67	63 chars	192 chars	+148.2	0.109

Table 3: Effect sizes across different quartile/percentile splits. All values fall in the "small" range ($d < 0.20$) by conventional standards.

B Word Count Robustness

To verify that our findings are not an artifact of the character count operationalization, we replicate all core analyses using word count. Table 4 summarizes the comparison. All findings are substantively identical. The Spearman correlation, effect size, and correlation decay pattern are consistent across both metrics.

Metric	Char Count	Word Count
Spearman r	+0.165	+0.168
Pearson r	+0.046	+0.048
R^2 (Spearman)	2.7%	2.8%
Cohen's d (Q1 vs Q4)	0.140	0.142
Correlation decay by score quantile:		
All jokes	+0.165	+0.168
Top 50%	+0.148	+0.151
Top 10%	+0.070	+0.072
Top 5%	+0.039	+0.041
Top 1%	+0.010	+0.011

Table 4: Findings replicated with word count. All results are substantively identical to the character count analysis.

C Full Yearly Statistics

Table 5 reports per-year Spearman correlations for both length metrics. All years with $n > 100$ show positive correlations for both character and word count. Years with fewer than 1,000 jokes (2008, 2009, 2010, 2020) are included here for completeness but excluded from the temporal analysis in the main paper due to sample size instability.

Year	n	r (char)	r (word)	Mean	Median
2009	185	+0.184	+0.135	6.6	0
2010	431	+0.303	+0.297	18.8	4
2011	4,057	+0.240	+0.245	42.9	10
2012	12,928	+0.302	+0.305	94.0	19
2013	16,789	+0.298	+0.301	109.9	17
2014	32,285	+0.215	+0.217	95.4	4
2015	72,842	+0.153	+0.155	98.7	2
2016	99,169	+0.141	+0.143	163.8	3
2017	103,696	+0.145	+0.149	284.0	5
2018	106,032	+0.165	+0.168	310.9	6
2019	121,818	+0.140	+0.142	264.3	6
2020	339	+0.132	+0.117	189.9	3

Table 5: Per-year Spearman correlations (character and word count) with descriptive statistics. Horizontal rules separate years excluded from the main temporal analysis ($n < 1{,}000$). All years show positive correlations.

D Winner-Take-All Dynamics Full Temporal Detail

Table 6 extends the analysis of Section 4.3 to all years with sufficient data, showing the intensification of winner-take-all dynamics. The mean/median ratio for long jokes (>400 characters) increases monotonically from 2.3× in 2011

to 49.0× in 2019. Short jokes (<100 characters) also exhibit increasing skew, but less dramatically. The 400-character threshold for "long" was chosen as an approximate boundary above the 75th percentile (266 chars); the pattern is not sensitive to this choice.

Year	Category	Mean	Median	Mean/Med.
2011	Long (>400)	78.1	34	2.3×
	Short (<100)	30.4	4	7.6×
2012	Long	171.5	51	3.4×
	Short	48.2	10	4.8×
2013	Long	202.9	48	4.2×
	Short	59.7	9	6.6×
2014	Long	179.6	13	13.8×
	Short	55.3	2	27.6×
2015	Long	211.6	7	30.2×
	Short	59.1	2	29.6×
2016	Long	342.9	9	38.1×
	Short	96.3	2	48.2×
2017	Long	531.6	13	40.9×
	Short	160.2	4	40.0×
2018	Long	652.4	14	46.6×
	Short	157.8	5	31.6×
2019	Long	588.4	12	49.0×
	Short	132.4	6	22.1×

Table 6: Mean/median divergence for long (>400 chars) and short (<100 chars) jokes across years.

The long/short mean ratio (L/S column) remains relatively stable at 3–4× across all years, indicating that the *relative* advantage of long jokes is consistent even as the distributional shape changes dramatically. The winner-take-all effect is thus primarily a within-category phenomenon that long jokes as a class maintain their advantage, but the variance within long jokes explodes as the platform matures.

E Cross-Subreddit Analysis

To assess whether the length-score correlation is specific to humor, we computed Spearman correlations between post length and score across 10 subreddits using Pushshift data (Baumgartner et al., 2020) (22.1M posts, 2012–2018), applying the same character count filters as our primary analysis.

Subreddit	n	Spearman r
r/nosleep	124,947	+0.418
r/tifu	259,890	+0.399
r/unpopularopinion	183,036	+0.339
r/todayilearned	1,440,474	+0.334
r/Jokes	793,844	+0.300
r/LifeProTips	380,256	+0.248
r/WritingPrompts	595,527	+0.096
r/explainlikeimfive	1,217,341	+0.078
r/Showerthoughts	2,843,163	+0.040
r/AskReddit	14,274,314	−0.045

Table 7: Length-score Spearman correlations across 10 subreddits. The positive correlation is not specific to r/Jokes.

Subreddit	All posts r	Top 1% r
r/tifu	+0.399	+0.132
r/Showerthoughts	+0.040	+0.126
r/todayilearned	+0.334	+0.094
r/WritingPrompts	+0.096	+0.090
r/nosleep	+0.418	+0.079
r/AskReddit	−0.045	+0.062
r/LifeProTips	+0.248	+0.055
r/explainlikeimfive	+0.078	+0.038
r/unpopularopinion	+0.339	+0.016
r/Jokes	+0.300	−0.007

Table 8: Correlation decay from all posts to top 1%. Every non-humor subreddit retains a positive correlation for its best content. Only r/Jokes reaches zero ($t = -0.61$, $p = 0.54$).