The Mavening: Analyzing the Stylistic Effects of Corporate Cost Cutting in Journalism

Benjamin LeBrun [1]   Andrew Piper [2]

[1]B.A. Computer Science and Linguistics   [2]Director, txtLAB@McGill
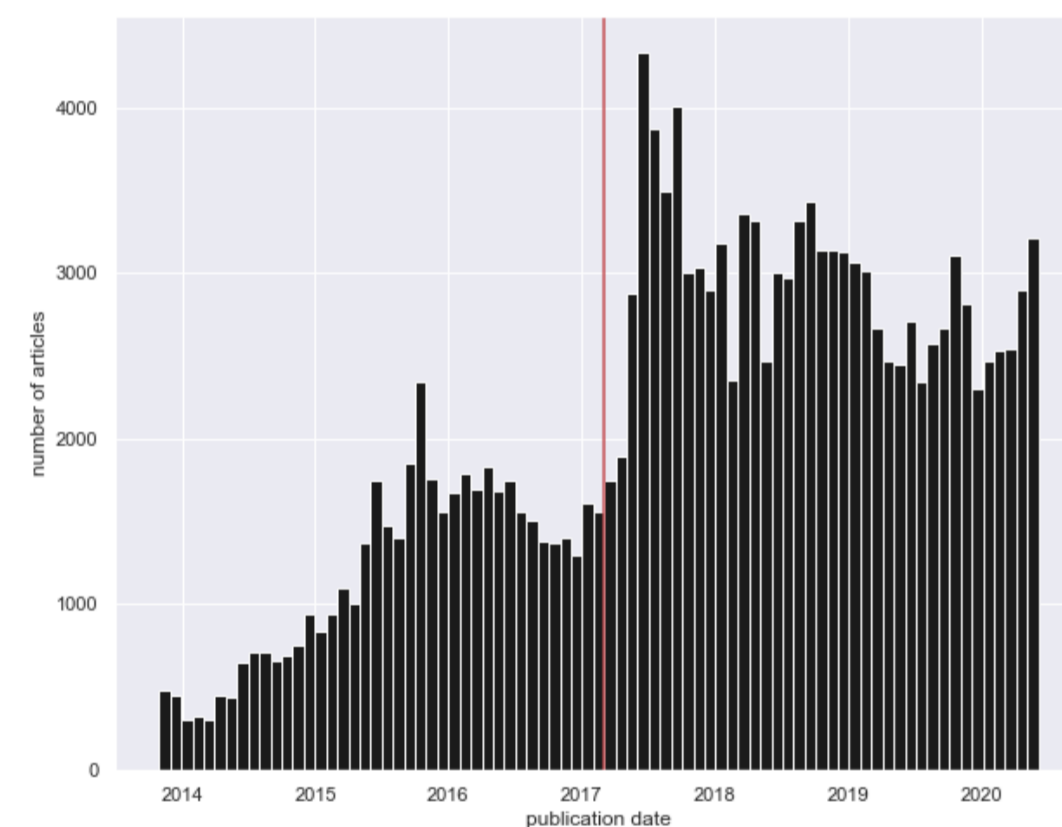
Introduction

Twenty-four hours after *Sports Illustrated* was purchased by the media conglomerate *The Maven*, nearly half of SI's editorial staff was laid off. This type of financial decision making, characterized by a dramatic reduction in staff and costs, has in recent years affected a number of prominent publications (e.g. *Newsweek*, *Sports Illustrated*, *LA Weekly*, *Deadspin*, *The Denver Post*). Dubbed *The Mavening*, this cost cutting trend has been hypothesized to have a negative effect on the journalistic content of affected publications. [1]

In this poster, we present a preliminary analysis of the effects of this process on the linguistic style of mavened publications. We introduce our data set of 286, 215 articles collected from four publications and test for evidence of stylistic simplification following the hypothesized moments of transformation. Overall, we find a number of statistically significant stylistic changes that generally align with a trend of simplification, though not as systematically as expected.
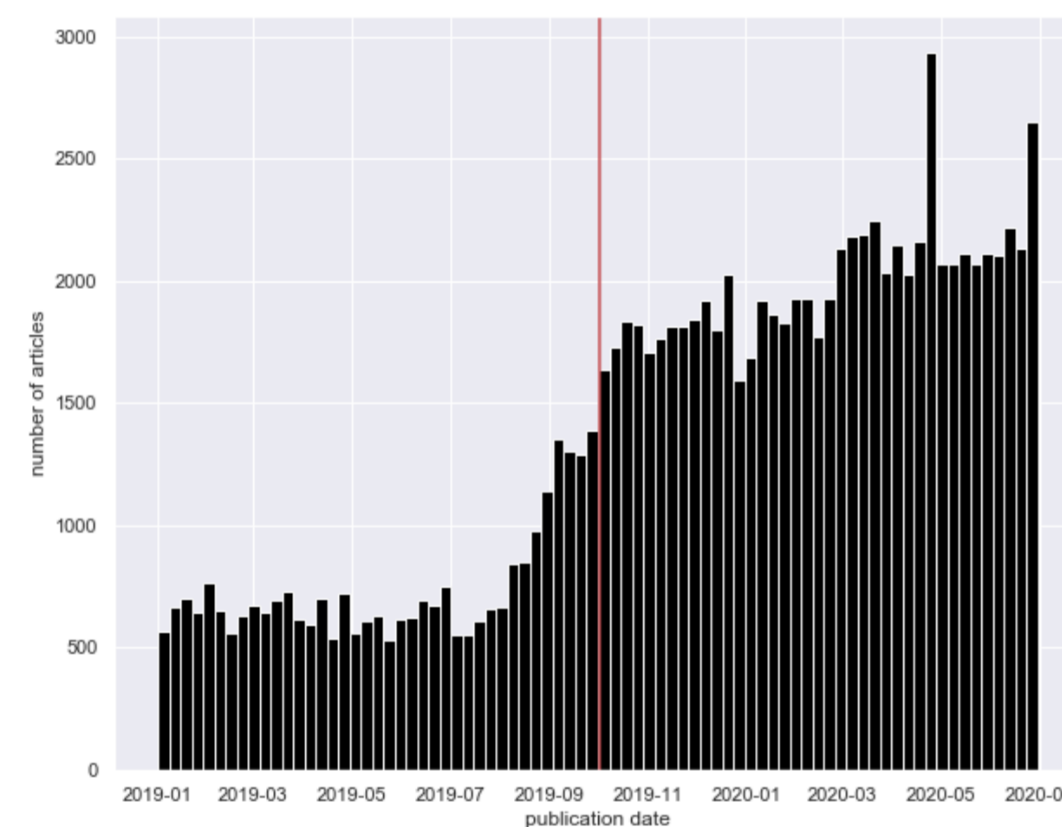
All the code for this project is available on GitHub at `www.github.com/benlebrun/mavening`. Contact: `benjamin.lebrun@mail.mcgill.ca`.
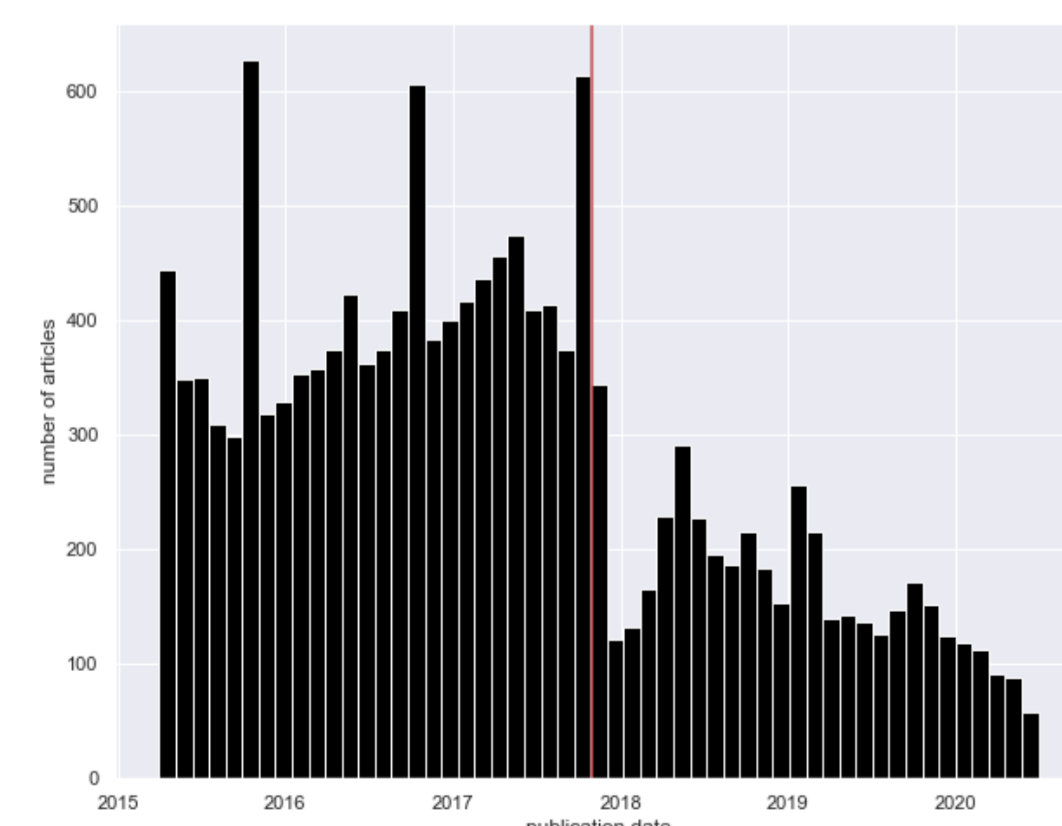
Data set

To build our data set, we scraped articles listed on each publication's respective site maps such that the periods before and after the transformational event are of equal lengths in time. In total, we collected $n = 286, 215$ articles from four mavened publications.
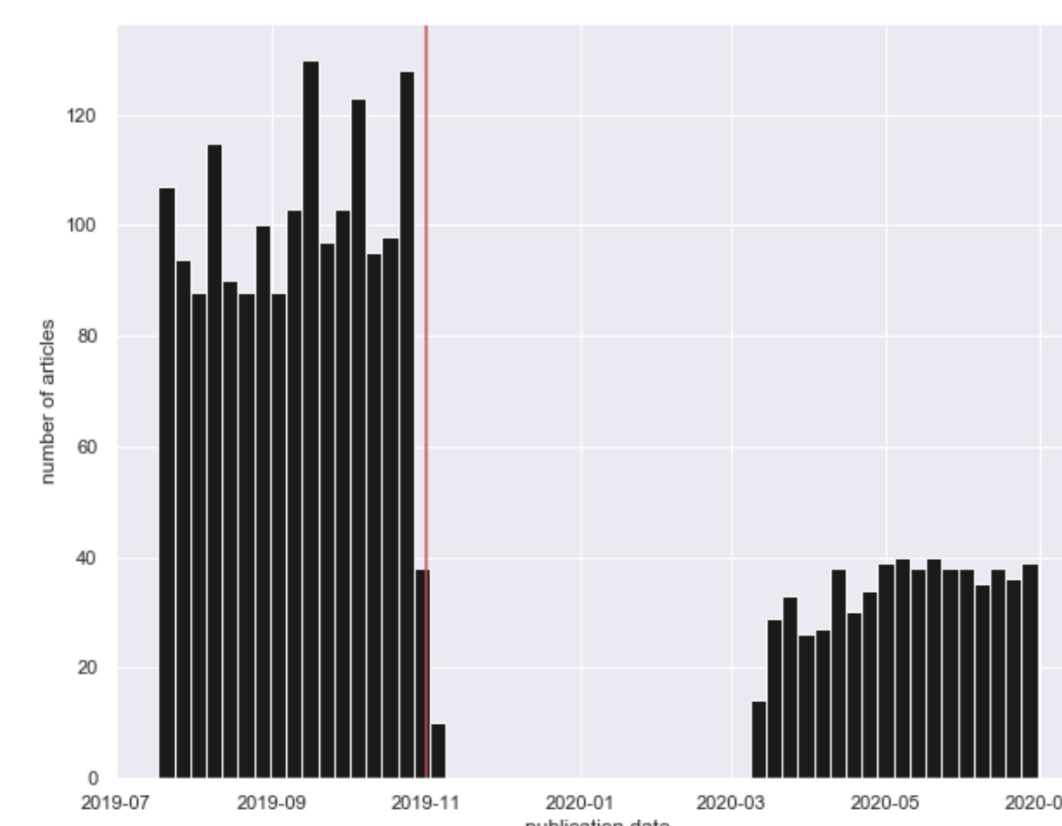

(a) *Newsweek*: $n_{pre} = 47,238$, $n_{post} = 114,021$

(b) *Sports Illustrated*: $n_{pre} = 29,221$, $n_{post} = 77,487$

(c) *LA Weekly*: $n_{pre} = 11,076$, $n_{post} = 4,253$

(d) *Deadspin*: $n_{pre} = 2307$, $n_{post} = 612$

Figure: Histograms of collected articles. The vertical red line indicates the date of the transformative event.

Acknowledgements

Modelling

Mavening events are characterized by significant losses in professional editorial staff and, in most cases, an increased demand for content. Given these circumstances, **we hypothesize a simplification in the linguistic style of mavened publications post transformational event**. Thus, when comparing the periods before and after the mavening event, we expect to observe a decrease in each of the seven dimensions of style outlined below.

Lexical Diversity

- We use the **type-token ratio (TTR)**, defined as the number of unique lemmas (*types*) expressed as a proportion of all lemmas (*tokens*), to measure the **lexical diversity** of a given article.
- A decline in the type-token ratio indicates a reduction in authorial vocabulary.

Grammatical Variation

- To measure **grammatical variation**, we calculate the information theoretic **entropy** [5] of a given article's part-of-speech (POS) bi-grams. For instance, *big tree*, an adjective-noun bi-gram, is represented as (JJ, NN).
- A smaller value of entropy indicates less variation in grammatical structure.

Semantic Cohesion

- *Cohesion* refers to the characteristics of a text that help the reader mentally connect ideas. [2]
- To measure **semantic cohesion**, we use the Universal Sentence Encoder [3] to generate semantic representations of sentences, and take the **mean cosine similarity** of a given article's set of sentence bi-grams.
- Articles with lower mean similarity scores are considered less semantically cohesive and thus less readable.

Descriptiveness

- *Descriptiveness* captures the degree of modification surrounding entities and actions in text. We use the **rate of modifiers**, defined as the number of modifiers normalized by word count, to measure the **descriptiveness** of a given article.
- Lower rates of modification are considered to be indicative of less detailed representations of entities and actions.

Concreteness

- *Concreteness* refers to the degree to which the concept denoted by a word (here, nouns) refers to a perceptible entity (e.g. *watermelon* vs. *justice*). Words low in concreteness are considered *abstract*. To measure the **concreteness** of a given article, we take the **mean concreteness score of nouns**, obtained from a data set of lemma concreteness scores. [4]
- We consider a decline in concreteness to be indicative of less sensory based event description.

Novelty

- The **surprisal** [5] of a word given a context (here, the preceding word) can be thought of as the amount of novelty (or information) transmitted by that word. Words with low probability of occurrence given a context (e.g. *power ballads*) will obtain higher surprisal scores than probable ones (e.g. *movie star*). We measure the **novelty** of a given article as the **mean surprisal** of word bi-grams.
- A decline in novelty indicates reliance on more familiar forms of expression.

Reading Difficulty

- We use the **Flesch reading ease** score to model **reading difficulty** according to the frequency of syllables, words, and sentences in a given article.
- A higher score is associated with a decrease in reading difficulty. Note that due to the inverse nature of the score, we report an increase in flesch as a decrease in reading difficulty.

Results

We consider differences between the pre and post distributions of a given metric to be significant when $p < 0.05$ and $|d| > 0.2$, where d is Cohen's d effect size. We denote significant differences in **bold**. We also report the **common-language effect size** f, which indicates the probability that a score sampled at random from the pre distribution will be greater than a score sampled from the post distribution.

Stylistic dimension	d	f	change	d	f	change	d	f	change	d	f	change
Semantic cohesion	-.09	.53	none	**.45**	**.38**	decrease	**.35**	**.41**	decrease	.11	.47	none
Lexical diversity	**.29**	**.42**	decrease	.11	.46	none	.16	.45	none	-.16	.55	none
Grammatical variation	-.08	.52	none	**-.34**	**.58**	increase	**-.21**	**.54**	increase	**.27**	**.44**	decrease
Concreteness	-.15	.54	none	**.22**	**.43**	decrease	**.43**	**.38**	decrease	**.22**	**.44**	decrease
Descriptiveness	.09	.47	none	**.41**	**.39**	decrease	**.31**	**.42**	decrease	**-.39**	**.59**	increase
Novelty	.12	.47	none	.14	.45	none	.11	.46	none	**-.22**	**.56**	increase
Reading Difficulty	**-.38**	**.60**	decrease	**-.32**	**.59**	decrease	-.03	.51	none	**.27**	**.42**	increase
	Newsweek			*Deadspin*			*Sports Illustrated*			*LA Weekly*		

Summary

We find significant change in each of the seven dimensions studied. The observed changes generally align with our hypothesis of stylistic simplification, although not as neatly as expected. Across stylistic dimensions, we can partition our results into three categories:

- **Metrics exhibiting change that aligns with an overall simplification of linguistic style.** All cases of significant change in concreteness, semantic cohesion, and lexical diversity decrease. This offers evidence to suggest that periods post transformational event are characterized, to varying degrees, by a **decrease in sensory based event description**, less **semantic continuity**, and a **smaller authorial vocabulary**.
- **Metrics exhibiting change that generally tends towards stylistic simplification.** While not all significant changes in descriptiveness and reading difficulty map neatly onto our hypothesis, the majority decrease, indicating a trend towards **less precise description of entities and actions** and **decreased reading difficulty**.
- **Metrics exhibiting change that does not align with stylistic simplification.** The majority of instances of change in grammatical variation and novelty increase rather than decrease. We therefore do not find evidence to indicate that the post mavening period is characterized by reliance on familiar forms of expression and repetitive grammatical structure.

Across publications, we find three trends of stylistic change. Observed changes in *Sports Illustrated* and *Deadspin* generally adhere to our hypothesis of simplification, whereas stylistic trends in *LA Weekly* tend in the opposite direction, and *Newsweek* appears to exhibit comparatively more stylistic continuity.

We conclude by noting that some of the effects being captured may align with other stylistic priorities that need further investigation. In an attempt to control for these effects, our next step is to compare stylistic change in mavened publications to those of a publication which has not undergone a transformational event. We also intend on investigating topical change.

References

1 B. Curtis. The mavening of sportswriting, Oct 2019.

2 A. Graesser et al. What readers need to learn in order to process coherence relations in narrative and expository text. 2003.

3 D. C. et al. Universal sentence encoder, 2018.

4 M. Brysbaert et al. Concreteness ratings for 40 thousand generally known English word lemmas. *Bahaviour research methods*, 2013.

5 C. E. Shannon. A mathethetical theory of communication. *Bell System Technical Journal*, 1948.