Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/32/eaba2196/DC1

Supplementary Materials for

The narrative arc: Revealing core narrative structures through text analysis

Ryan L. Boyd*, Kate G. Blackburn, James W. Pennebaker

*Corresponding author. Email: r.boyd@lancaster.ac.uk

Published 7 August 2020, *Sci. Adv.* **6**, eaba2196 (2020)
DOI: 10.1126/sciadv.aba2196

This PDF file includes:

Supporting Online Materials A:
Narrative Corpora

The goals of this project were to determine whether a common pattern of language markers exists over the course of narratives and, if so, to identify the general shape of the patterns. To do this, 3 large, narrative-based corpora were collected from publicly accessible online sources (see Table S1). These data include the following types of narratives: Novels (*N* = 2,523; *M* word count = 73,954.75; *SD* word count = 50,446.22), Short Stories (*N* = 2,092; *M* word count = 3,782.70; *SD* word count = 4,353.39), and the Thematic Apperception Test (*N* = 14,419; *M* word count = 326.58; *SD* word count = 61.07). Each dataset showed considerable variation in length with the mean word count being much lower for the TAT corpus than for the Novels or Short Stories. One benefit of examining texts of such different length is that it allowed us to determine if comparable underlying structures emerged in narratives ranging from very short, non-professional narratives (i.e., amateur TAT stories) to traditional, long-form, professionally-composed narratives (i.e., Novels, Short Stories).

Table S1. Online Resources for Collection of Novels and Short Stories.

Website	URL
Project Gutenberg	www.gutenberg.org
East of the Web	eastoftheweb.com
American Literature	americanliterature.com
BookTrust	www.booktrust.org
Literature Network	www.online-literature.com
World English	www.world-english.org
Fiction Eserver Collection	www.fiction.eserver.org
Classic Reader	www.classicreader.com
Online English Library	www.englishlibrary.org
Tory Stars	www.storystar.com
Classic Short Stories	www.classicshorts.com
Thresholds	www.thresholds.chi.ac.uk
ABC Short Stories Radio Project	www.abc.net.au
Obooko	www.obooko.com
Lost Coast Review	www.lostcoastreview.com
Lumina	www.luminajournal.com
Feedbooks	www.feedbooks.com
NYC Midnight	www.nycmidnight.com
Words Without Border	www.wordswithoutborders
Kenyon Review	www.kenyonreview.org
Austin Short Story Contest	www.austinchronicle.com/short-story-contest
Ekerd Review	www.eckerd.edu/eckerd-review
Nice Stories	www.nicestories.com
Toast	www.the-toast.net
Bomb Magazine	www.bombmagazine.org
Short Story Me	www.short-story.me
Fictionaut	www.fictionaut.com
Carve Magazine	www.carvezine.com
ClarksWorld	www.clarkesworldmagazine.com
Flash Fiction Online	www.flashfictiononline.com
American Short Fiction	www.americanshortfiction.org
Sun	www.thesunmagazine.org
Manchester Review	www.themanchesterreview.co.uk
London Journal of Fiction	www.londonjournalofficion.com
Granta	www.granta.com
365 Tomorrows	www.365tomorrows.com
Every Writer	www.everywritersresource.com

Novels

A large corpus of texts were originally extracted from the 2010 Project Gutenberg DVD (http://www.gutenberg.org), then winnowed down to novels based on their official Library of Congress classifications, which produced the Novels corpus. All novels available from Project Gutenberg are distributed for free and exist within the public domain. All novels collected were written by authors living between 1789-1970. The novels were by authors such as Ambrose Bierce, Charles Dickens, Jane Austen, Herman Melville, Joseph Conrad, among hundreds of others. Exclusion criteria are discussed in the following SOM B section.

Short Stories

Short Stories were collected from online sources not restricted by copyright laws and freely available to the public. All short stories that we collected were written between 1819-2016 and were by professionally published, self-published, and/or anonymous authors. The short story corpus includes authors such as W.B. Yeats, Kate Chopin, Ernest Hemingway, as well as first time and budding writers of varying professional experience.

Thematic Apperception Test

The Thematic Apperception Test (TAT) corpus was collected from online users visiting the companion website for the book *Secret Life of Pronouns* (http://www.secretlifeofpronouns.com/exercises.php). The TAT was originally developed in the 1930s by Henry D. Murray and Christiana Morgan to catalog people's underlying motivations and personality through their use of words and themes (*43*). The test asks participants to view an ambiguous photo and to then compose a story based on that photo. All TAT narratives were based on a single drawing on the website that shows two people in a laboratory – i.e., McClelland's "women in laboratory" image (*44*, *45*).

Supporting Online Materials B: Other Corpora

Five additional datasets that varied in the degree to which they could be considered traditional narratives were examined to answer RQ2 and RQ3. From most to least narrative-like, respectively, these additional archival datasets included language from Romance Novels, Movies, New York Times, TED Talks and Supreme Court Opinions. To ensure consistency, data cleaning procedures applied to the Narrative Corpora were also applied to texts in these additional datasets, including the *a priori* word count criterion (i.e., WC $\geq$ 250). Basic information and descriptive statistics of each corpus are presented below.

Romance Novels

The Romance Novel corpus included 639 self-published books (word count M = 43,706.18; word count SD = 43,515.54) retrieved from Smashwords.com, a free e-book and online publishing platform with a catalog of over 400,000 books. We note here that, at the time research assistants compiled this corpus, the Smashwords publishing platform was publicly accessible; six months later, the publishing platform adopted a member sign-in requirement to access all novels. Research assistants only collected recent books written in English and that had corresponding reader ratings assigned. The ratings ranged from 1 to 5 stars and were averaged across raters for each book. Similar to other corpus collected, research assistants removed front end material, extraneous markings, and footnotes. Cleaning procedures resulted in text that only each story's content. For the purpose of authorship identification, a unique, anonymous identifier was assigned to each romance novel.

Although the ratings were highly skewed towards five stars, our research team classified the books into three categories: low ratings (with 3.9 or fewer stars, N = 149 books), mid-level ratings (for 4 to 4.9 stars, N = 348 books), or perfectly rated (5 stars, N = 142 books). This was done to answer RQ2. The decision to split the ratings into low, middle, and high ratings was to see if very high or very low ratings may have been substantially different from the others.

Movies

A corpus of 19,970 movie subtitles (word count *M* = 7,231.60; word count *SD* = 3,533.10) was generously provided to us by OpenSubtitles.org. The administrators provided us with their full collection of English-language movie subtitles in various formats that includes the subtitles for movies spanning several decades. The corpus is created, maintained, and updated by a massive volunteer-based crowd sourcing effort from people all over the world, working in several languages and performing translations across languages. The corpus provided by OpenSubtitles.org for this project consists only of subtitles in English corresponding to movies that were either originally released in English, or for international movies where the dialogue has been translated to English.

A number of inclusion requirements and data cleaning procedures were performed on the original subtitle corpus that contained 88,419 movie and television subtitles written in various languages. General information for each subtitle was collected from the IMDb API and was used for filtering and excluding observations from analyses. Additionally, all subtitles were manually examined to ensure that they had not been misclassified as English text. A total of 68,446 subtitles were removed as they did not meet the inclusion requirements. Data cleaning was performed on the remaining 19,970 texts and included removing all material not related to the actual language of the subtitles themselves. Information removed included text such as timestamps and font information such as italics notation, etc. In addition, scene release data embedded in the transcripts were removed where detected. The remaining text in each subtitle file represented only the dialogue between characters from the start of the movie until its completion.

New York Times

The corpus (*N* = 18,312) contains multiple types of articles, including editorials, features, opinions articles, world, U.S., and local news, letters to the editor, etc. The pieces were published between 1989 and 2017and were scraped from the publication's website. The overall mean word count for this corpus was 827.05 (*SD* = 490.91).

TED Talks

TED talks are highly rehearsed narratives generally based on the speakers' research or experiences that have clear implications for technology, education, or design. TED Talk transcripts were scraped from the official TED website (ted.com) during March of 2018. Only those talks for which English-language transcripts existed were collected (*N* = 2,226). The mean word count for this corpus was 2065.60 (*SD* = 895.78).

Supreme Court Opinions

A Supreme Court decision contains written opinions from one or more judges about the court's ruling. Although not defined as traditional narratives, these opinions offer the opportunity to explore whether or not narrative structure exists in other types of shared communication. The Supreme Court corpus includes 1,580 main, dissenting, and concurring opinions from 1954 to 1995, written by the following Supreme Court justices: Warren, Reed, Clark, Harlan, Frankfurter, Douglas, Brennan, Black, Stewart, Harlan, Goldberg, Fortas, White, Marshall, Blackmun, Powell, Rehnquist, Burger, Stevens, O'Connor, Scalia, Kennedy, Brennan, Thomas, Ginsburg, Souter, and Breyer. Supreme Court opinions were collected from https://public.resource.org/ and cleaned by research assistants at The University of Texas at

Austin. Extraneous markings and footnotes were removed from each Supreme Court opinion. The mean word count of each opinion is 3,259.20 (*SD* = 3,043.17).

Supporting Online Materials C:
Analysis of Language

Data cleaning

Several pre-processing steps were performed prior to language analyses in order to ensure high integrity and quality across all data. Text files not written primarily in English were removed. Where multiple texts were written by the same author, only a single, randomly-selected text from the given author was retained for analysis. Other measures were taken to limit duplicate text files or those that were not written by a single individual. For example, a series of algorithms were applied to the TAT data set to discern if any text file was posted more than once or included nonsensical language (e.g., type-token ratio analyses, common word frequency analyses). Steps were also taken to remove any identifying information when detected. In the case of Novels and Short Stories, identifying information was not removed if the work was published or self-published. In addition, words that were not a part of the story itself were removed from texts. For instance, front matter (e.g., contents, dedications, acknowledgements, etc.) and back matter were removed from Novels, Short Stories, and Romance Novels. Finally, texts that did not meet an *a priori* word count criterion (i.e., WC $\geq$ 250) were omitted from analyses. The word count minimum ensured that works comprised of less than 250 words did not influence the reliability of LIWC outputs, such that each narrative segment consisted of at least 50 words.

Text Analysis Method

Custom software based on the Linguistic Inquiry and Word Count method – specifically, LIWC2015 (*46*) – was used to quantify all language throughout all narratives. The LIWC approach analyzes texts for psychological information in an automatic manner using an extensively validated dictionary to count the number of function words (e.g., pronouns, articles, prepositions) or content words (e.g. affect, social, home) found in any given text. Output then provides the rate at which each psychological category of language is used relative to the size of the text (e.g., 5% pronouns, 10% cognitive process words).

The LIWC software and dictionary have been applied extensively for research used in academic, commercial, and clinical settings (*32*). However, there were a couple of potential issues that we identified within the LIWC2015 dictionary with regards to our study of narrative dimension shapes. First, the built-in LIWC2015 dictionary contains several words that overlap across multiple categories. For example, the word "never" is counted as both a *negation* and a *certainty* word in the default LIWC2015 dictionary – such overlaps can result in considerable non-independence between the narrative dimensions that we sought to study. Second, while the *cognitive processes* dimension of LIWC has been extensively validated, several high base-rate words belong to this category that did not clearly map onto the concept of *cognitive tension* that we sought to capture in the current work.

Given possible issues with the default LIWC2015 dictionary, we additionally created custom dictionaries that were derived from the default dictionaries that sought to address such issues. In these dictionaries, words that did not clearly belong to the narrative dimensions under study were removed, and two versions of our custom dictionary were created: one that minimized overlaps between narrative dimensions, and one that completely removed all overlaps across narrative dimensions. These custom dictionaries are available for viewing at:

https://osf.io/2ztvq/?view_only=3ec61cf65c08476f926624893d0d6fc3 https://osf.io/wpcx8/?view_only=3ec61cf65c08476f926624893d0d6fc3

However, we note that across all analyses, across all datasets, and across all segmentation strategies (see below), no version of the dictionary substantively changed the *shapes* of the narrative dimensions. All three versions – the built-in LIWC2015 dictionary, our custom dictionary *with* overlaps, and our custom dictionary *without* overlaps – were extremely consistent and resulted in identical conclusions to those reported. For the sake of consistency, we primarily report results from the “no overlaps” version of the dictionary, as this approach facilitates the simplest interpretation of each dimension.

Narrative Segmentation Strategy

Again, the overarching goal of the current study was to determine if psychologically relevant language unfolds across narratives in an objective, consistent fashion. In order to measure language patterns *across* a narrative, rather than simply describe each narrative as a whole, it is necessary to segment each narrative into multiple parts. The decision to rely on five segments was informed by theoretical and practical considerations.

As briefly summarized in the main body of the text, no accepted theory or description of narrative suggests more than five unfolding features. Indeed, from Aristotle’s narrative essay, *Poetics*, to Joseph Campbell’s cross-culture work showing universal story arcs (i.e., Hero’s Journey) to Valdimir Propp’s analysis of Russian folklore, there is a central idea that a story must contain a beginning, middle, and end. In particular, the five segmentation strategy was inspired by the narrative theory work of German novelist Gustav Freytag (1894), who expanded on Aristotle’s beginning, middle, and end theoretical argument on narrative structure to build a universal plot structure: 1) exposition, 2) rising action, 3) climax, 4) falling action, and 5) denouement. Freytag’s theoretical narrative model pushed us to think about whether certain story components happened only at certain time points within the narrative or, rather, an ongoing set of psychological processes that transpire throughout a story.

The primary advantage in adopting this segmentation strategy is that it offers a simple and intuitive method for tracking changes in language over the course of a narrative. Using an equal sized segmentation strategy offers a uniform method that can easily be replicated with a wide array of texts, and does not require additional, more complicated coding procedures. Note that we have conducted additional analyses with 3, 4, 5, 6, 7, 8, 9, and 10 segments; results from these analyses are all comparable to the 5-segment approach. Illustrations of the narrative dimension shapes across each number of segments can be found on the OSF repository for this research under the "Segments vs Narrative Dimensions" folder.

More practically, both the TATs and Short Stories were relatively short and restricting word count to 250 words, or 50 words per segment, is a conventional limit to LIWC analyses used to yield reliable results (*47*, *48*). As such, we chose to segment all narratives across each corpus into 5 equally-sized chunks – this was done automatically during the language analyses. The result of this process was that for each psychological measure of language, 5 separate scores were generated for each narrative (e.g., rate of prepositions for Segment 1, rate of prepositions for Segment 2, and so on). As an example, consider a novel that consists of 10,000 words. Using the 5-segment strategy described above, the text would be automatically split into 5 chunks of 2,000 words. Each chunk would then be processed via the quantification schemes described earlier (e.g., the custom AON dictionary, built-in LIWC2015 dictionary, etc.).

Supporting Online Materials D:
Internal Consistency of Narrative Processes

The current study hypothesized that there are core narrative processes that naturally fluctuate over the course of a story: staging, plot progression, and cognitive tension. Given that the project's main goal is to explore (and potentially identify/describe) the general structure of narrative by tracking these processes, it is important to establish the basic psychometric properties of each narrative process in turn. This section provides reliability analyses of the processes themselves, descriptive statistics within and across corpora, and intercorrelations to better understand the measures that underpin all other analyses conducted.

Reliability

In establishing the basic psychometrics of a new measure, it is important to determine whether each measure exhibits some form of reliability/internal consistency *within* narratives. There is an interesting irony to such an analysis in that the current research is based on the assumption that language generally shifts as a narrative progresses. Importantly, however, the reliability assumption is not in conflict with the narrative shifting hypotheses. It is assumed that all authors have their own distinctive language fingerprints over the course of a story. Some authors may typically use a lot of articles and prepositions and their word use will be consistently evident at all points in the narrative. At the same time, it is assumed that authors in general will shift their word use over time. Indeed, these shifts are similar to the findings that posit the different effects of personality and situation in all behaviors, including language use. Most people are more formal in their language and demeanor in a courtroom than at a bar with friends. However, the most formal speaker in a group of friends will likely be the most formal in both the courtroom and the bar.

In simpler terms, even with the expectation that language will shift *within* any given narrative, prior work on the stability of language suggests that texts should exhibit stability when comparing psychological language rates *between* narratives (*49*, *50*). If an author tends to use staging language at rates higher than other authors during the first segment of a story, does the same author also tend to use staging language at higher rates during later parts of the narrative (again, relative to other authors)?

To determine if narrative processes were internally consistent across the five story segments, Cronbach's alphas using both relative term frequencies and one-hot encoded term frequencies were calculated for each narrative process across all primary narrative texts used in the project; results are presented in Table S2.

Table S2. Internal consistency of each narrative dimension across the 5 primary forms of traditional narratives (Studies 1 & 2).

	Internal Consistency			
Corpus	Staging	Plot Progression	Cognitive Tension	*M* Word Count (*SD*)
Novels	0.80 (0.58)	0.97 (0.68)	0.98 (0.88)	74,034 (47,816)
Short Stories	0.90 (0.53)	0.97 (0.47)	0.97 (0.63)	3,528 (3,150)
TAT	0.34 (0.06)	0.72 (0.14)	0.41 (0.29)	326 (60)
Movies	0.82 (0.68)	0.96 (0.75)	0.93 (0.64)	7,165 (3,199)
Romance	0.93 (0.60)	0.99 (0.27)	0.99 (0.72)	43,706 (43,515)

Note: Internal consistencies calculated from one-hot encoded terms (i.e., Kuder-Richardson Formula 20, or KR-20) are presented, with relative frequency-derived scores (i.e., Cronbach's alpha) displayed in parentheses.

The internal consistency of the reliability scores show that authors are generally consistent in the way they are using words across narrative segments. Note that the Movies and Novels evidenced the highest reliabilities and TATs the lowest. While such fluctuations are a common artifact of word count (*12*), we note that all metrics in these corpora range from average to high compared to those reported in prior research.

Supporting Online Materials E:
Narrative Process Intercorrelations

How are narrative processes related to each other? Pearson correlations, based on the overall scores of each process (i.e., the rate of each process in each narrative as a whole, rather than segmented), were computed to establish the basic relationships among narrative processes. As shown in Table S3, narrative processes were significantly correlated with one another; given the large sample size when combining across all 5 traditional story types (i.e., Novels, Movies, TATs, Short Stories, and Romance Novels), all correlations were statistically significant. The patterns of the relationships suggest several generalized relationships between each process within narratives. Particularly striking was the strong, negative correlation between staging and plot progression. This result demonstrates that authors tend to invoke each process in an inverted reciprocal manner, such that narratives with higher rates of staging engage in less plot progression (and vice versa; see Table S3). Similar correlations are presented on a by-segment basis (Table S4) – note that all correlations are statistically significant given the large sample size, however, the effect sizes are generally much weaker than the whole-text analyses reported in Table S3.

Table S3. Pearson Correlation Coefficients for Entire Narrative Corpus's Processes

Narrative Process	1	2	3
1: Staging	1.00		
2: Plot Progression	-.77	1.00	
3: Cognitive Tension	.20	-.02	1.00

Note. Total narrative texts N = 39,754. The general pattern of correlations held true across each corpus as well. Also note that while these correlations were derived from the non-overlapping version of the AON dictionary (see SOM C), the two other versions of the dictionary (overlapping allowed, default LIWC2015 dictionary) showed stronger intercorrelations between the Cognitive Tension dimension and the other two dimensions, particularly Plot Progression.

Table S4. Pearson Correlation Coefficients for Entire Narrative Corpus's Processes, by Segment

Narrative Process, Segment	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1: Staging, Segment 1														
2: Staging, Segment 2	-0.21													
3: Staging, Segment 3	-0.26	-0.22												
4: Staging, Segment 4	-0.29	-0.27	-0.22											
5: Staging, Segment 5	-0.27	-0.28	-0.28	-0.21										
6: PlotProg, Segment 1	-0.57	0.11	0.14	0.17	0.15									
7: PlotProg, Segment 2	0.11	-0.58	0.13	0.16	0.16	-0.18								
8: PlotProg, Segment 3	0.15	0.12	-0.59	0.13	0.17	-0.26	-0.21							
9: PlotProg, Segment 4	0.16	0.17	0.13	-0.59	0.13	-0.29	-0.28	-0.22						
10: PlotProg, Segment 5	0.15	0.17	0.17	0.12	-0.58	-0.27	-0.30	-0.29	-0.20					
11: CogTension, Segment 1	-0.13	0.02	0.04	0.04	0.03	0.20	-0.02	-0.05	-0.07	-0.06				
12: CogTension, Segment 2	0.02	-0.13	0.03	0.03	0.04	-0.03	0.18	-0.03	-0.05	-0.06	-0.20			
13: CogTension, Segment 3	0.04	0.02	-0.14	0.03	0.05	-0.05	-0.03	0.18	-0.03	-0.07	-0.26	-0.23		
14: CogTension, Segment 4	0.04	0.04	0.02	-0.14	0.03	-0.07	-0.06	-0.03	0.19	-0.03	-0.28	-0.28	-0.22	
15: CogTension, Segment 5	0.03	0.04	0.04	0.03	-0.15	-0.05	-0.07	-0.07	-0.04	0.22	-0.27	-0.28	-0.27	-0.22

Supporting Online Materials F:
Between-Narrative Standardization

One of the main goals of the current research is to better understand the “shape” that each narrative process takes across stories. In other words, if we find that the language of each process does shift significantly throughout narratives, in general, we seek to be able to visually and conceptually interpret the nature of such patterns.

As discussed in previous Supporting Materials sections, however, narrative processes showed high internal consistency (i.e., reliable *within* narratives) but considerable *between* narrative variation. This fact is underscored by the descriptive statistics reported in Table S5. Even upon visual inspection of this table, several clear differences exist both between narrative types and across segments. For example, the plot progression process increases appear to be strong and clearly increase for the TAT corpus. However, the magnitude of this trend (which increases from 31.13 to 34.27, i.e., an increase of 10% at the end relative to the beginning) is much larger than Novels (from 32.65 to 34.11, i.e., a 4.49% increase at the end relative to the beginning) and Short Stories (from 29.53 to 31.47, a 6.57% relative increase). As can be seen in Table S5, the TAT stories are associated with much larger changes for narrative processes than either Novels or Short Stories.

Table S5. Corpus (3) x Segment (5) x Narrative Processes (5) between-within descriptive statistics.

Corpus	Narrative Process	Segment 1 *M* (*SD*)	Segment 2 *M* (*SD*)	Segment 3 *M* (*SD*)	Segment 4 *M* (*SD*)	Segment 5 *M* (*SD*)
Novels	Staging	22.73 (2.22)	22.35 (2.25)	22.23 (2.26)	22.24 (2.31)	22.15 (2.33)
Short Stories	Staging	21.65 (3.75)	20.96 (3.75)	20.84 (3.91)	20.99 (3.95)	20.87 (3.90)
TAT	Staging	22.96 (5.43)	21.84 (5.39)	21.70 (5.40)	21.62 (5.50)	21.37 (5.53)
Novels	Plot Prog.	32.65 (3.97)	33.50 (3.96)	33.86 (4.01)	33.97 (4.05)	34.11 (4.12)
Short Stories	Plot Prog.	29.53 (6.09)	30.53 (6.11)	30.93 (6.14)	30.97 (6.24)	31.47 (6.21)
TAT	Plot Prog.	31.13 (8.03)	33.80 (8.46)	34.01 (8.60)	34.26 (8.65)	34.27 (8.66)
Novels	Cog. Tension	2.54 (0.63)	2.67 (0.66)	2.70 (0.70)	2.69 (0.70)	2.62 (0.68)
Short Stories	Cog. Tension	2.22 (1.14)	2.32 (1.15)	2.35 (1.19)	2.34 (1.17)	2.32 (1.17)
TAT	Cog. Tension	3.45 (2.62)	3.93 (2.78)	3.93 (2.77)	3.82 (2.75)	3.64 (2.73)

Note: Raw means are based on percentage of total words within segment and are calculated using non-overlapping version of AON dictionary.

Under different circumstances, the significant between-text variations might not be problematic. However, in the context of the current research and analyses, the magnitude of variability between narratives is highly problematic. For example, when attempting to determine the general *shape* of narrative process patterning, we are concerned only with the *relative* magnitude of narrative process change within texts rather than the *absolute* magnitude of change. For example, a narrative wherein the staging process decreases across 5 segments as a vector of {50, 40, 30, 20, 10} should be considered to have the same staging patterning (i.e., *shape*) as a narrative wherein the staging process exhibits a decrease of the same *relative* magnitude (e.g., {5, 4, 3, 2, 1}).

Accordingly, all narratives were internally standardized using a traditional z-scoring method, wherein the mean of each narrative process in each text is set to 0, with a standard deviation of 1. This procedure allows us to meaningful compare the narrative process *patterns* between each story while retaining the *relative* magnitude of change between each narrative's respective segments.

Supporting Online Materials G:
Modeling the Shape of Narrative Processes

There are multiple ways by which we can determine the patterns of narrative processes. One statistically pure way would be to identify specific mathematical functions (or statistical models) that capture the dominant pattern of narrative processes across each text's segments. Ultimately, our application of statistical modeling in this research serves two goals: 1) establish whether narrative processes show significant change throughout stories, and 2) create a mathematical function that *describes* narrative process patterning.

The most common and, arguably, simplest method for determining general patterns in data is by use of ordinary least squares linear (OLS) regression, where data from a single predictor is used to fit a straight line to a single criterion, attempting to minimize the sum of squared differences throughout. Such an approach is useful for understanding whether a criterion is increasing, decreasing, or relatively stable across varying levels of the predictor. An additional benefit to simple OLS regression modeling is that it is relatively robust to overfitting a dataset in the way that more complicated models (e.g., support vector machines) are capable of doing.

However, in the current research, we hypothesized that one of the narrative processes would follow a negative parabolic trend where the cognitive tension process category would start low, peak around the middle of a story, then end low again. Such a shape cannot be adequately captured using a pure linear regression. A simple polynomial variation on regression is able to adequately reflect such trends and is a better fit for modeling shapes that are suspected to be non-linear in nature. For this reason, a quadratic model (i.e., a 2^{nd} degree polynomial) is essential to accurately reflecting non-linear trends.

In order to determine the nature of narrative process patterning, we adopted an analytic approach that included both linear and quadratic regression modelling of each narrative process across each story, using the Narrative Segment as the predicting factor (or factors for quadratic models) and standardized narrative process scores as the criterions. By analyzing all narratives using both types of regression models, we are able to better determine the *shape* of each narrative process (i.e., linear or non-linear) if, in fact, the processes show consistent variation across stories.

Results from all modeling procedures found that each narrative process showed significant variation across stories using both linear and quadratic regression modeling. In all cases, the quadratic models accounted for significantly more variance in the data than their linear counterparts, suggesting that none of the narrative processes are best described by a linear increase or decrease. While a polynomial regression model does run a higher risk of data overfitting than a simple linear regression, we addressed this using a robust cross-validation procedure that is described in a later Supporting Materials section (see SOM H).

Supporting Online Materials H:
Cross-Validating Narrative Process Patterns

In the previous section, linear and quadratic models were built to explore the shape of narrative structure. Results suggested that quadratic functions were universally better suited to describing the narrative process shapes in our corpora. However, as noted above, a quadratic model has the inherent tendency to *overfit* a noisy dataset to a higher degree than a linear one, meaning that the significant ΔR^2 results may be an artifact of the modeling method rather than a true reflection of the narrative process patterns in the data.

To verify that the quadratic models do indeed better represent narrative process structures, we conducted a standard model competition experiment using a 10 by 10-fold cross-validation procedure (*51*). Under this procedure, each model's account of variance (i.e., R^2) is robustly estimated using a standard 10-fold cross-validation procedure. This process is repeated 10 times for each model, resulting in a total of 10 instances of 10-fold cross-validation being run, or 100 total fold calculations. The R^2 estimates from each fold are saved and compared to the complementary model using a standard between-groups t-test to robustly test whether one model does indeed account for significantly more variance than another.

For example, consider the case of the cognitive tension narrative process. In the previous section, we report that the quadratic model more accurately reflected the shape of this process throughout narratives in our Narrative corpora than did a linear model. To be certain of this and to avoid mistaking a quadratic model's ability to overfit data, we performed 10 separate 10-fold cross-validations of the linear model, resulting in 100 R^2 estimates for this model. The same process was repeated for the quadratic model, again resulting in 100 R^2 estimates. These 2 sets of R^2 estimates (1 set of 100 for linear models, 1 set of 100 for quadratic models) were then compared using a standard between-groups t-test. Should there be no significant difference between the R^2 estimates from these 2 models, we can conclude with confidence that the linear and quadratic models are equally able to summarize the true shape of the narrative process. In contrast, if there is a significant difference between the R^2 estimates from these 2 models, we can assume with confidence that the quadratic model indeed *does* more accurately reflect the true shape of the narrative process in our data compared to the linear model. Results for all 10 by 10-fold cross-validation analyses are reported in Table S6.

Table S6. T-tests Comparing Mean R^2 from Linear and Quadratic Models of Narrative Processes Across Corpora.

Corpus	Narrative Process	Intercept	x^2	x	Quadratic R^2	Linear R^2	ΔR^2	*t*	*p*
Novels	Staging	0.684	0.049	-0.409	0.041	0.033	0.008	5.607	< .001
Short Stories	Staging	0.457	0.040	-0.297	0.015	0.010	0.005	5.714	< .001
TAT	Staging	0.371	0.026	-0.218	0.013	0.010	0.002	7.005	< .001
Romance	Staging	1.047	0.066	-0.591	0.112	0.097	0.015	3.481	< .001
Movies	Staging	0.417	0.029	-0.245	0.016	0.013	0.003	8.758	< .001
Novels	Plot Progression	-0.957	-0.063	0.550	0.089	0.075	0.013	6.359	< .001
Short Stories	Plot Progression	-0.528	-0.025	0.268	0.037	0.035	0.002	1.478	0.141
TAT	Plot Progression	-0.625	-0.052	0.398	0.029	0.019	0.009	19.284	< .001
Romance	Plot Progression	-1.370	-0.101	0.826	0.160	0.125	0.035	6.798	< .001
Movies	Plot Progression	-1.205	-0.091	0.737	0.118	0.089	0.029	37.504	< .001
Novels	Cognitive Tension	-0.816	-0.093	0.611	0.038	0.009	0.030	27.841	< .001
Short Stories	Cognitive Tension	-0.231	-0.021	0.156	0.004	0.003	0.001	3.314	0.001
TAT	Cognitive Tension	-0.300	-0.039	0.243	0.006	0.000	0.005	31.500	< .001
Romance	Cognitive Tension	-0.667	-0.092	0.560	0.032	0.003	0.029	16.528	< .001
Movies	Cognitive Tension	-0.635	-0.070	0.468	0.023	0.006	0.017	53.404	< .001

Note: Intercepts and beta weights presented are drawn from the quadratic models.

Supporting Online Materials I:
Prevalence of Identified Narrative Process Patterns

One might ask how commonly each of the patterns occurs in our narratives. By definition, the patterns identified above for each narrative process must be dominant in order to drive the statistical models. However, "dominant" is a relative term – a coin toss series that results in 50.0001% of tosses landing as heads makes heads the "dominant" response, but not a particularly meaningful one.

A beta coefficient comparison strategy was used to compare the narrative shape of each observation against the normative narrative patterns across all corpora. For example, the normative quadratic model reported earlier for the cognitive tension process was as follows:

$$f_{(\text{Segment})} = (-0.06 \times \text{Segment}^2) + (0.41 \times \text{Segment}) + (-0.53)$$

A shorthand method for determining whether any given narrative follows this general pattern, then, is to create a narrative-specific quadratic model, checking to see if the beta coefficients corresponding to the Segment factor are in the same direction. For example, if a new story is analyzed and has a quadratic "cognitive tension" process model of the following:

$$f_{(\text{Segment})} = (0.12 \times \text{Segment}^2) + (-0.34 \times \text{Segment}) + (0.56)$$

As an aside, we know that the *shape* of the cognitive tension pattern in this text is, in effect, the opposite of that found in normative terms (i.e., this story has the shape of a *positive* parabola, whereas the normative model takes the form of a *negative* parabola). For these analyses, much like ones stated earlier, we are generally unconcerned with the *magnitude* of shape (as variations are expected between texts) so much as the general form of the shape itself.

In simple terms, then, the analyses reported were created using simple regression models for each narrative process for each text. When a narrative's model exhibited beta coefficients of the same sign(s) as those found and reported earlier in the normative models, it was flagged as having the same general shape as was found to be normative. Results of the distribution of matches is presented in Table S7 below.

Table S7. Percentage of Quadratic Shapes Matched by Novels, Short Stories, and TAT.

	Novels	Short Stories	TAT
	%	%	%
3 shapes	21.32%	20.46%	17.01%
2 shapes	24.14%	27.06%	28.29%
1 shape	20.45%	27.96%	30.57%
no shape	34.09%	24.42%	24.04%

Note. Beta coefficients were used to extract the number of texts belonging to each model and process.

Importantly, note that "%" values in the above table may also be seem somewhat misleading. For a quadratic model, however, a narrative must match both the x^2 and x beta coefficients, both of which may take on either a positive or negative sign – the likelihood of any narrative achieving this match by chance is:

$$50\% \times 50\% = 25\%$$

Related to these analyses is the question of *combined* narrative pattern dominance throughout the corpora. That is, what percentage of texts showed narrative process patterns that matched none, 1, 2, or all 3 narrative patterns? This question can be answered by performing analyses similar to those just described, albeit considering the match rate for all 3 narrative processes simultaneously. Given that several separate analyses up to this point have established a quadratic function as the proper descriptive model for all narrative processes, we performed this beta comparison analysis only in the context of quadratic models. The expected likelihood of any given narrative matching all 3 narrative process patterns simultaneously is quite low, particularly for quadratic models. The likelihood of randomly matching any single process' quadratic shape is, as detailed above, only 25%. The likelihood of matching 2 processes simultaneously, then, is 0.25^2 (i.e., .0625), and all 3 processes would be $.25^3$, or .0156 (i.e., 1.56%). Considering these particularly low expected random base rate frequencies, the percent of texts matching each narrative process' shape is indeed remarkably high – that is: 21.32%, 20.46%, 17.01% for Novels, Short Stories, and TATs, respectively, when an expected base rate by chance would be 1.56% for each genre.

Supporting Online Materials J:
Emotional Tone as a Narrative Process

Recent work by others has explored the emotional fluctuations across narratives in manners similar to those performed in this research, describing these fluctuations as emotional narrative structures (*52*). We performed a set of exploratory analyses parallel to the primary analyses performed in our current research using the Positive and Negative Emotional categories from LIWC2015 as a comparison to past work using our methods. Results from these analyses are discussed below.

On the surface, both linear and quadratic models explained a statistically significant amount of variance in the Positive and Negative Emotional Tone (see Tables S8 and S9). As seen in Figure S1 (top), the Narrative Corpora shows similar linear patterns found in the core narrative processes. However, as seen in Figure S1 (bottom), when broken apart by narrative type, the shape of emotion words becomes less apparent and does not follow clear linear or quadratic trends. For instance, positive emotion words in Short Stories fluctuates differently over the course of the narrative relative to the TAT corpus. Our results capture what other researchers have found when studying the behavior of emotion words in narratives. That is, emotion words do not appear to have a single structure across narratives. Indeed, this phenomenon was seen in the emotional arc research that identified six emotional arcs for stories (*13*). Our work supports this research by showing positive and negative emotion words fluctuate and take on several distinct patterns over the course of the story.

Table S8. Simple Regression Predicting Linear and Quadratic Trends for Positive and Negative Emotional Tone.

	Narrative Corpora						
	Linear Equation			Quadratic Equation			
	x	c	R^2	x^2	*x*	*c*	R^2
Positive Emotional Tone	.078	-.235	.015	.010	.016	-.163	.015
Negative Emotional Tone	.065	-.197	.011	-.038	.296	-.466	.016

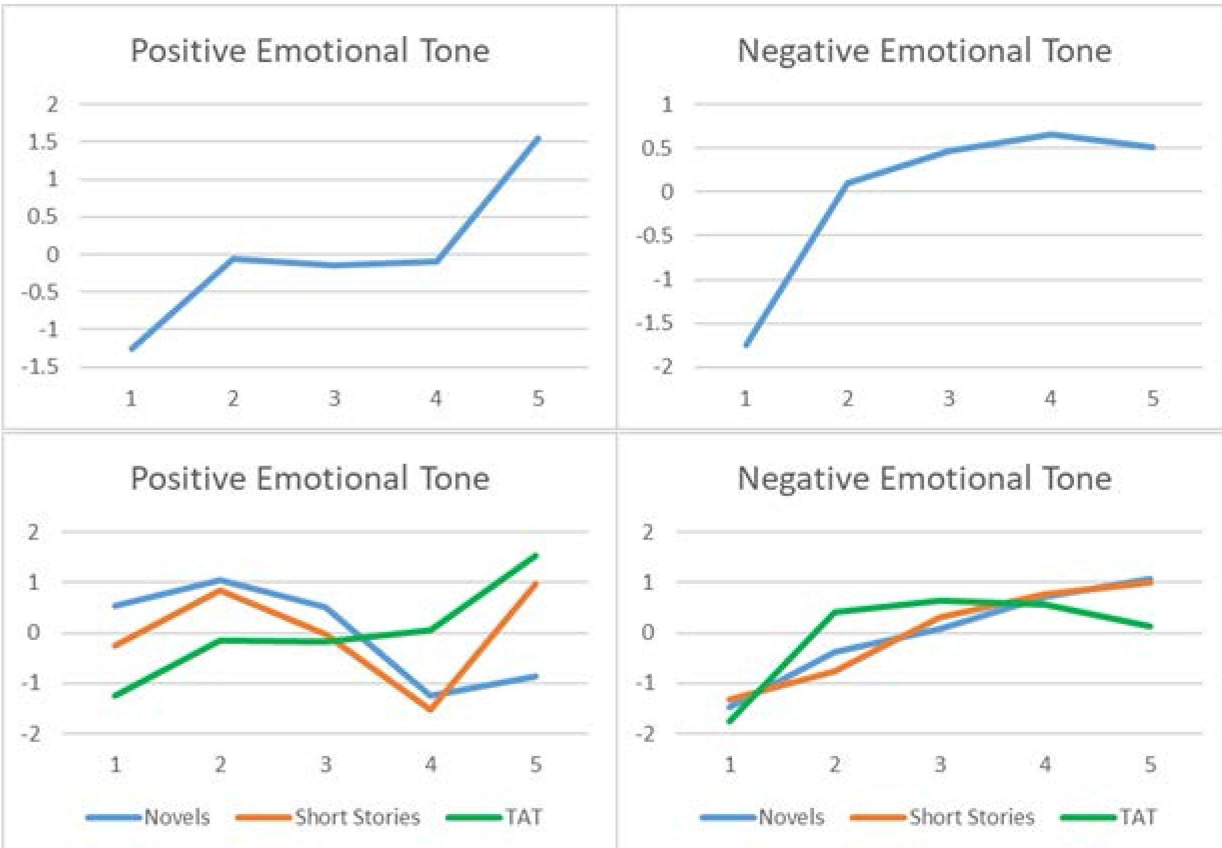
Note: All F-values for the Linear Model are significant at $p < .01$. All *F*-values for the Quadratic Model are significant at $p < .01$. All R^2 were significant at $p < .01$.

Table S9. Simple Regression Predicting Linear and Quadratic Trends for Positive and Negative Tone for Novels, Short Stories, and Thematic Apperception Test.

	Novels, Short Stories, and TAT						
	Linear Equation			Quadratic Equation			
	x	c	R^2	x^2	*x*	*c*	R^2
Positive Emotional Tone							
Novels	-.038	.114	.003	-.007	.009	.058	.003
Short Stories	.000	-.001	.000	-.041	.007	.047	.000
TAT	.110	-.330	.030	.013	.026	-.232	.031
Negative Emotional Tone							
Novels	.190	-.571	.090	-.028	.358	-.767	.093
Short Stories	.062	-.187	.009	-.009	.119	-.252	.010
TAT	.044	-.133	.005	-.044	.311	-.445	.012

Note: Novels: All *F*-values for the Linear Model are significant at $p < .01$. All *F*-values for the Quadratic Model are significant at $p < .05$, except for Positive Emotion which was not significant at $p < .05$. Short Stories: All *F*-values for the Linear Model are significant at $p < .01$, except for Positive Emotion which was not significant at $p < .05$. All *F*-values for the Quadratic Model are significant at $p < .01$, except for Positive Emotion which was not significant at $p < .05$. TAT: All *F*-values for the Linear Model are significant at $p < .01$. All *F*-values for the Quadratic Model are significant at $p < .01$.

Fig S1. Emotional Tone Patterns for all Corpora (top) and separately for Novels, Short Stories, and TAT (bottom).



REFERENCES AND NOTES

1. Aristotle, L. J. Potts, *Aristotle on the Art of Fiction: An English Translation of Aristotle's Poetics* (Cambridge Univ. Press, 1968).
2. J. Franklin, *Writing for Story: Craft Secrets of Dramatic Nonfiction by a Two-Time Pulitzer Prize Winner* (Plume, 1994).
3. N. Frye, *Anatomy of Criticism: Four Essays* (Princeton Univ. Press, 2000).
4. J. Campbell, *The Hero with a Thousand Faces* (Pantheon Books, 1949).
5. W. Labov, J. Waletzky, *Narrative Analysis: Oral Versions of Personal Experience* (John Benjamins, 1997), vol. 7.
6. V. Propp, *Morphology of the Folktale* (University of Texas Press, 1968).
7. A. C. Graesser, M. Singer, T. Trabasso, *Constructing Inferences During Narrative Text Comprehension* (American Psychological Association, 1994), vol. 101.
8. H. Li, A. C. Graesser, M. Conley, Z. Cai, P. I. Pavlik, J. W. Pennebaker, A new measure of text formality: An analysis of discourse of Mao Zedong. *Discourse Process.* **53**, 205–232 (2016).
9. J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, E. L. Aiden, Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).
10. K. C. McLean, M. Syed, M. Pasupathi, J. M. Adler, W. L. Dunlop, D. Drustrup, R. Fivush, M. E. Graci, J. P. Lilgendahl, J. Lodi-Smith, D. P. McAdams, T. P. McCoy, The empirical structure of narrative identity: The initial Big Three. *J. Pers. Soc. Psychol.* (2019).
11. P. S. Dodds, E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, C. M. Danforth, Human language reveals a universal positivity bias. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2389–2394 (2015).
12. A. C. Graesser, D. S. McNamara, M. M. Louwerse, Z. Cai, Coh-Metrix: Analysis of text on cohesion and language. *Behav. Res. Methods Instrum. Comput.* **36**, 193–202 (2004).
13. A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, P. S. Dodds, The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci.* **5**, 31 (2016).
14. G. Freytag, E. J. MacEwan, *Technique of the Drama: An Exposition of Dramatic Composition and Art; An Authorized Translation from the 6th German Edition* (Scholarly Press, 1960).
15. M. Mack, *The Norton Anthology of World Masterpieces* (Norton, ed. 5, 1985), vol. 1.
16. J. W. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us* (Bloomsbury, 2011).

17. C. Chung, J. Pennebaker, in *Social Communication*, K. Fiedler, Ed. (Frontiers of Social Psychology, Psychology Press, 2007), pp. 343–359.

18. K. N. Jordan, J. Sterling, J. W. Pennebaker, R. L. Boyd, Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 3476–3481 (2019).

19. J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, D. I. Beaver, When small words foretell academic success: The case of college admissions essays. *PLOS ONE* **9**, e115844 (2014).

20. N. R. Norrick, *Conversational Narrative* (John Benjamins Publishing Company, 2000).

21. M.-L. Ryan, The modal structure of narrative universes. *Poetics Today* **6**, 717–755 (1985).

22. B. Kleim, A. B. Horn, R. Kraehenmann, M. R. Mehl, A. Ehlers, Early linguistic markers of trauma-specific processing predict post-trauma adjustment. *Front. Psychol.* **9**, 645 (2018).

23. J. M. Caspari, T. L. Raque-Bogdan, C. McRae, T. L. Simoneau, S. Ash-Lee, K. Hultgren, Posttraumatic growth after cancer: The role of perceived threat and cognitive processing. *J. Psychosoc. Oncol.* **35**, 561–577 (2017).

24. M. A. Khawaja, F. Chen, C. Owen, G. Hickey, in *Human-Computer Interaction—INTERACT 2009* (Lecture Notes in Computer Science, Springer, 2009), pp. 485–489.

25. D. G. Morrow, Spatial models, prepositions, and verb-aspect markers. *Discourse Process.* **13**, 441–469 (1990).

26. M. W. Lovett, M. Dennis, J. E. Newman, Making reference: The cohesive use of pronouns in the narrative discourse of hemidecorticate adolescents. *Brain Lang.* **29**, 224–251 (1986).

27. S. Kleijn, H. L. W. P. Maat, T. J. M. Sanders, Comprehension effects of connectives across texts, readers, and coherence relations. *Discourse Process.* **56**, 447–464 (2019).

28. F. Hartung, M. Burke, P. Hagoort, R. M. Willems, Taking perspective: Personal pronouns affect experiential aspects of literary reading. *PLOS ONE* **11**, e0154732 (2016).

29. J. G. Bohanek, R. Fivush, Personal narratives, well-being, and gender in adolescence. *Cogn. Dev.* **25**, 368–379 (2010).

30. P. J. Bauer, E. N. Stark, A. F. Lukowski, J. Rademacher, D. L. V. Abbema, J. K. Ackil, Working together to make sense of the past: Mothers' and children's use of internal states language in conversations about traumatic and nontraumatic events. *J. Cogn. Dev.* **6**, 463–488 (2005).

31. H. A. Murray, Thematic apperception test (Harvard University Press, 1943).

32. J. W. Pennebaker, R. L. Boyd, K. Jordan, K. Blackburn, *The Development and Psychometric Properties of LIWC2015* (The University of Texas at Austin, 2015), pp. 1–26.

33. U.S. adult fiction book sales 2017 | Statistic. *Statista*; https://www.statista.com/statistics/730316/adult-fiction-unit-sales/.
34. P. Regis, *A Natural History of the Romance Novel* (University of Pennsylvania Press, 2007).
35. IMDb | Help; https://help.imdb.com/article/imdb/track-movies-tv/ratings-faq/G67Y87TFYYP6TWAV#accurate.
36. B. Kovach, *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect* (Three Rivers Press, 2014).
37. Our organization; https://www.ted.com/about/our-organization.
38. R. A. Posner, Judges' writing styles (and do they matter). *U. Chi. L. Rev.* **62**, 1421 (1995).
39. R. J. Owens, J. P. Wedeking, Justices and legal clarity: Analyzing the complexity of U.S. Supreme Court opinions. *Law Soc. Rev.* **45**, 1027–1061 (2011).
40. D. Gentner, *Why Nouns are Learned Before Verbs: Linguistic Relativity Versus Natural Partitioning* (257, Illinois University, Center for the Study of Reading, 1982), pp. 1–86.
41. D. Gentner, Language as cognitive tool kit: How language supports relational thought. *Am. Psychol.* **71**, 650–657 (2016).
42. M. Tomasello, *Constructing a Language: A Usage-Based Theory of Language Acquisition* (Harvard Univ. Press, 2003).
43. D. C. McClelland, J. W. Atkinson, R. A. Clark, E. L. Lowell, *The Achievement Motive* (Irvington, 1953).
44. D. C. McClelland, *Power: The Inner Experience* (Irvington, 1975).
45. F. D. Schönbrodt, B. Hagemeyer, V. Brandstätter, T. Czikmantori, P. Gröpel, M. Hennecke, L. S. F. Israel, K. Janson, N. Kemper, M. Köllner, P. M. Kopp, A. Mojzisch, R. Müller-Hotop, J. Prüfer, M. Quirin, B. Scheidemann, L. Schiestel, S. Schulz-Hardt, L. Sust, C. Zygar, O. C. Schultheiss, Measuring implicit motives with the Picture Story Exercise (PSE): Databases of expert-coded German stories, pictures, and updated picture norms. *PsychArchives* (2020).
46. J. W. Pennebaker, R. J. Booth, R. L. Boyd, M. E. Francis, *Linguistic Inquiry and Word Count: LIWC2015* (Pennebaker Conglomerates (www.liwc.net), 2015).
47. E. Kacewicz, J. W. Pennebaker, M. Davis, M. Jeon, A. C. Graesser, Pronoun use reflects standings in social hierarchies. *J. Lang. Soc. Psychol.* **33**, 125–143 (2013).
48. R. L. Boyd, in *Data Analytics in Digital Humanities*, S. Hai-Jew, Ed. (Springer International Publishing, 2017), pp. 161–189.
49. R. L. Boyd, J. W. Pennebaker, Did Shakespeare write Double Falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychol. Sci.* **26**, 570–582 (2015).

50. J. W. Pennebaker, L. A. King, Linguistic styles: Language use as an individual difference. *J. Pers. Soc. Psychol.* **77**, 1296–1312 (1999).
51. I. Witten, E. Frank, M. Hall, C. Pal, *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, ed. 4, 2017).
52. M. L. Jockers, *Macroanalysis: Digital Methods and Literary History* (University of Illinois Press, 2013).