TECHNOLOGY

The Six Main Arcs in Storytelling, as Identified by an A.I.

A machine mapped the most frequently used emotional trajectories in fiction, and compared them with the ones readers like best.

By Adrienne LaFrance
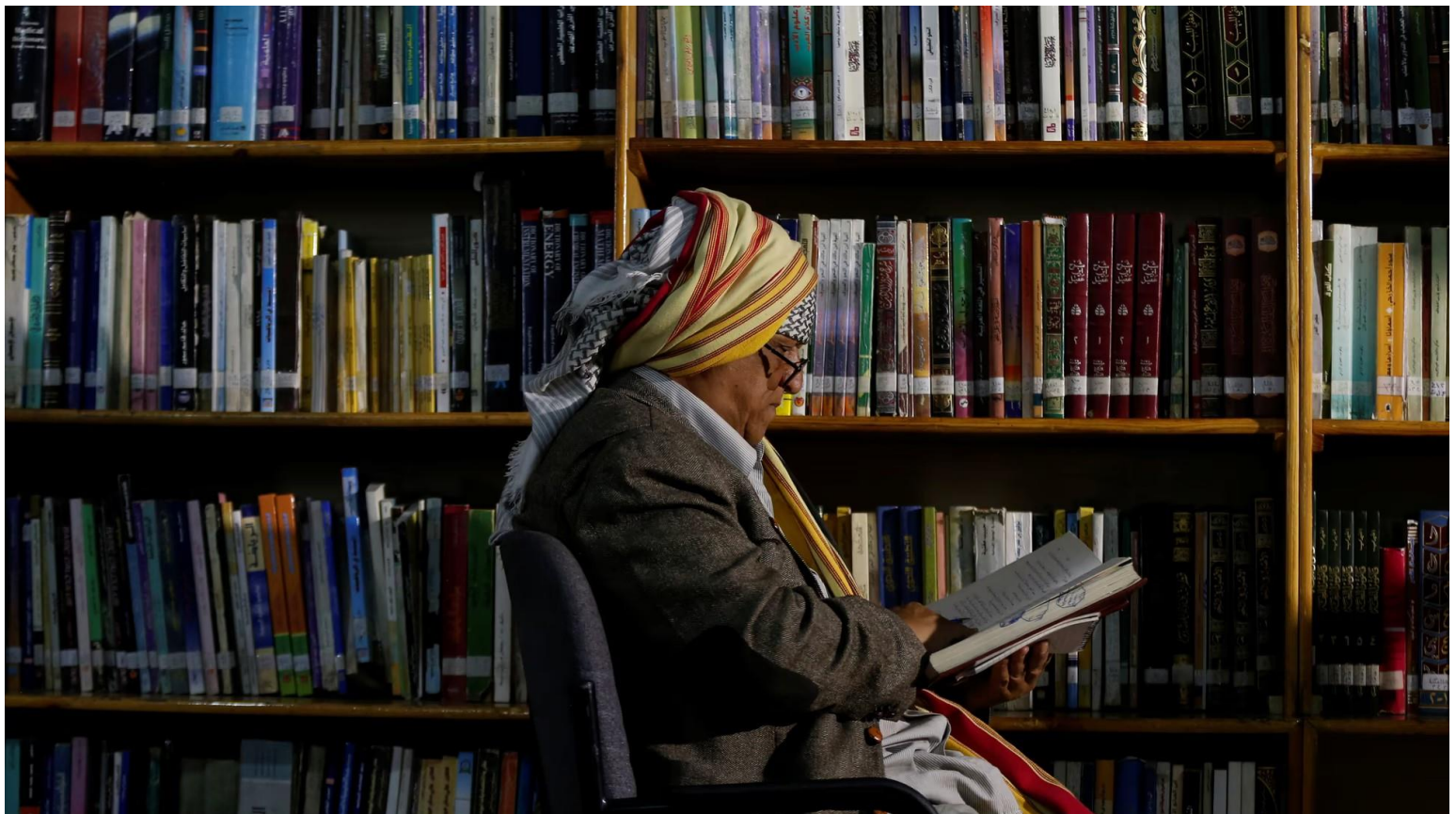
Khaled Abdullah / Reuters

"My prettiest contribution to my culture," the writer Kurt Vonnegut mused in his 1981 autobiography *Palm Sunday*, "was a master's thesis in anthropology which was rejected by the University of Chicago a long time ago."

By then, he said, the thesis had long since vanished. ("It was rejected because it was so simple and looked like too much fun," Vonnegut explained.) But he continued to carry the idea with him for many years after that, and spoke publicly about it more than once. It was, essentially, this: "There is no reason why the simple shapes of stories can't be fed into computers. They are beautiful shapes."

That explanation comes from a lecture he gave, and which you can still watch on YouTube, that involves Vonnegut mapping the narrative arc of popular storylines along a simple graph. The X-axis represents the chronology of the story, from beginning to end, while the Y-axis represents the experience of the protagonist, on a spectrum of ill fortune to good fortune. "This is an exercise in relativity, really," Vonnegut explains. "The shape of the curve is what matters."

The most interesting shape to him, it turned out, was the one that reflected the tale of Cinderella, of all stories. Vonnegut visualizes its arc as a staircaselike climb in good fortune representing the arrival of Cinderella's fairy godmother, leading all the way to a high point at the ball, followed by a sudden plummet back to ill fortune at the stroke of midnight. Before too long, though, the Cinderella graph is marked by a sharp leap back to good fortune, what with the whole business of (spoiler alert) the glass slipper fitting and the happily ever after.

A rough map of the story, as measured by the titular character's fluctuating fortune. The precipitous drop represents the stroke of midnight; the leap at the end is the "happily ever after."
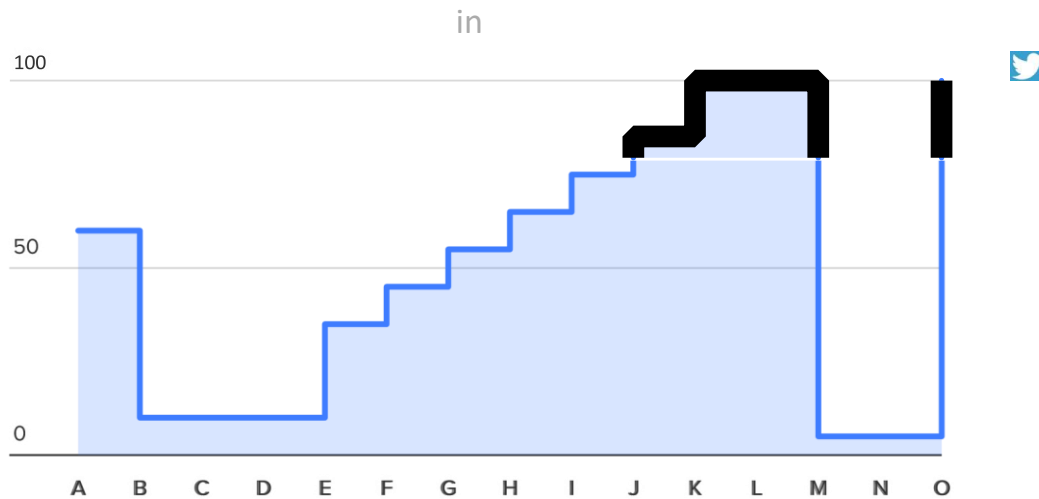
Inspired by Kurt Vonnegut's description of story mapping in Palm Sunday.

The Atlantic

| Get The Data | Embed

This may not seem like anything special, Vonnegut says—his actual words are, "it certainly looks like trash"—until he notices another well known story that shares this shape. "Those steps at the beginning look like the creation myth of virtually every society on earth. And then I saw that the stroke of midnight looked exactly like the unique creation myth in the Old Testament." Cinderella's curfew was, if you look at it on Vonnegut's chart, a mirror-image downfall to Adam and Eve's ejection from the Garden of Eden. "And then I saw the rise to bliss at the end was identical with the expectation of redemption as expressed in primitive Christianity. The tales were identical."

Vonnegut, in his ever charming way, was quite pleased with himself for making this connection. And 35 years later, his idea had resonated enough with a group of mathematicians and computer scientists that they decided to build an experiment around it. Vonnegut had mapped stories by hand, but in 2016, with sophisticated computing power, natural language processing, and reams of digitized text, it's possible to map the narrative patterns in a huge

corpus of literature. It's also possible to ask a computer to identify the shapes of stories for you.

That's what a group of researchers, from the University of Vermont and the University of Adelaide, set out to do. They collected computer-generated story arcs for nearly 2,000 works of fiction, classifying each into one of six core types of narratives (based on what happens to the protagonist):

1. Rags Riches to Rags (fall)

2. Man in a Hole (fall then rise)

3. Icarus (rise then fall)

4. Cinderella (rise then fall then rise)

5. Oedipus (fall then rise then fall)

6. to Riches (rise)

Their focus was on the emotional trajectory of a story, not merely its plot. They also analyzed which emotional structure writers used most, and how that contrasted with the ones readers liked best, then published a preprint paper of their findings on the scholarship website arXiv.org. More on that in a minute.

First, the researchers had to find a workable dataset. Using a collection of fiction from the digital library Project Gutenberg, they selected 1,737 English-language works of fiction between 10,000 and 200,000 words long.

Then, they ran their dataset through a sentiment analysis to generate an emotional arc for each work. "We're not imposing a set of shapes," said Andy Reagan, a Ph.D. candidate in mathematics at the University of Vermont and the lead author of the paper. "Rather: the math and machine learning have identified them."

They did this by training the machine to take all the words of the book, section by section, and measure the average happiness of a given bag of words based on how an individual word scored. The researchers assigned individual happiness scores to more than 10,000 frequently-used words by crowdsourcing the effort on the website Mechanical Turk. This portion of the research is fascinating in and of itself: The 10 words that people ranked as happiest were *laughter, happiness, love, happy, laughed, laugh, laughing, excellent, laughs,* and *joy.* The 10 words that people ranked as least happy were *terrorist, suicide, rape, terrorism, murder, death, cancer, killed, kill,* and *die.* (You can see how all the words ranked by visiting this site.)

There are several theories that say every story known to man can be reduced to one of just a handful of archetypes—a quest, overcoming the monster, rebirth, to name a few—but there's no consensus on what those stories are. In this case, researchers picked six from a mix of popular lists based on what shapes the computer identified most. And though the researchers were focused on a book's emotional arc—not the structure of its plot, per se—they found overlap in how plot points reflected emotional highs and lows as measured by the sentiment analysis.

While the plot of *Harry Potter and the Deathly Hallows* for instance, is "nested and complicated," they wrote, "the emotional arc associated with each subnarrative is clearly visible." (That said, emotional moments discussed briefly— the first kiss between Harry and Ginny, let's say—didn't register.)
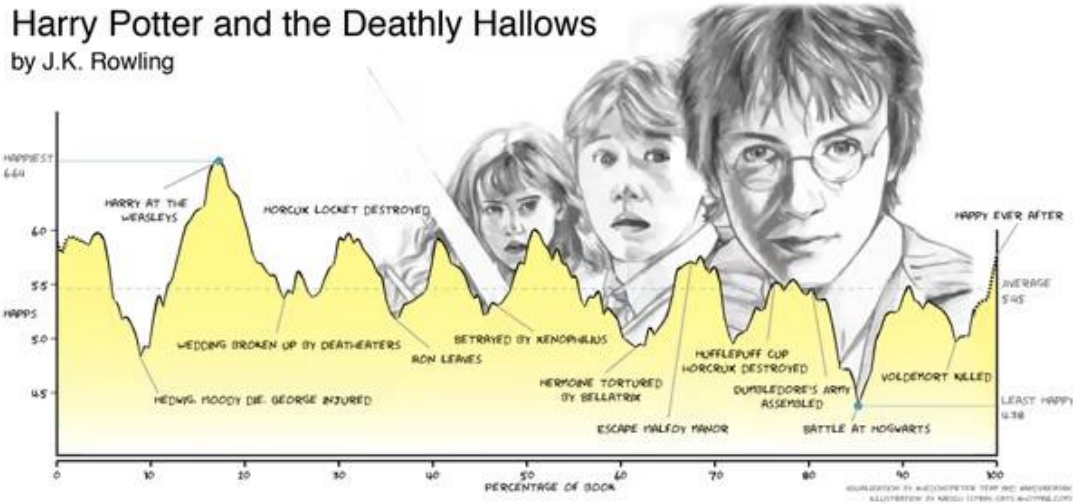

Hedonometer / Andy Reagan / Kirsch

All in all, "Rags to Riches" stories represented about one-fifth of all the works analyzed. This isn't surprising. It's easy to think of examples of such tales in classic literature. The canons of Charles Dickens, Edith Wharton, and Jane Austen are arguably defined by them.

"The 'Rags to Riches' emotional arc embodies a story that we all love to believe in, widely popular in the American dream itself," Reagan said. "It's a story of hope and fairness, where regardless of beginning in bad times, with effort things will get better and eventually result in good fortune."

In this case, the prototypical example, according to the researchers, is Lewis Carroll's *Alice's Adventures Under Ground*—which would later be published as *Alice's Adventures in Wonderland*. An 1890 novel by the writer Olive Schreiner, *Dreams*, was another clear match for the "Rags to Riches" model. For both stories, the computer found a near-identical match to "Rags to Riches" with few if any connections to other kinds of emotional arcs. Here's how the top 20 stories that fit the "Rags to Riches" mode appear on a graph in their paper:
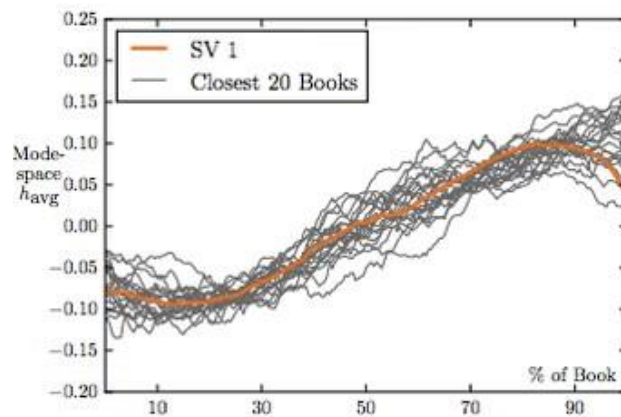
"Rags to Riches" may be popular among writers, but it isn't necessarily the emotional arc that readers reach for most. The categories that include the greatest total number of books are not the most popular, the researchers found. They examined total downloads for all books from Project Gutenberg, then divvied them up by mode. Measured this way, "Rags to Riches" is eclipsed by "Oedipus", "Man in a Hole" and, perhaps not surprisingly, "Cinderella," all of which were more popular. Reagan told me he and his colleagues now plan to analyze how different arcs are sequenced together in a single story, like in the Harry Potter example above.

Eventually, he says, this research could help scientists train machines to reverse-engineer what they learn about story trajectory to generate their own compelling original works. Already, there are competitions for story-writing bots. (Incidentally, I attempted a similar experiment and it didn't exactly go as planned.)

"This is an active area of research," Reagan says, "and there are a lot of hard problems yet to be solved. In addition to the plot, structure, and emotional arc, to write great stories, a computer will need to create characters and dialogue that are compelling and meaningful."

Vonnegut, of course, always made it sound easy. Consider how he describes the "Man in a Hole" narrative, which is characterized by M.R. James's *Ghost Stories of an Antiquary*—or pretty much any 22-minute sitcom: "Somebody gets in to trouble, gets out of it again," Vonnegut once said in a lecture. "People love that story. They never get sick of it."