Evolution of Modern Literature and Film

Oleg Sobchuk

sobchuk@shh.mpg.de

Max Planck Institute for the Science of Human History, Jena, Germany.

For submission to the *Oxford Handbook of Cultural Evolution*,
edited by R. Kendal, J. Tehrani, and J. Kendal.

Word count (excluding abstract and references): 5,165

Abstract

The evolution of complex narratives, such as fictional books or films, is a fairly new area of cultural-evolutionary research. This chapter, first, discusses a theoretical question: in what sense do narratives evolve? Then, it proceeds to describing how content-based, or hedonic, selection influences the evolution of narrative forms – and briefly, the role of several other evolutionary mechanisms, such as drift and accumulation of innovations. Finally, the chapter presents the methods used for studying the evolution of literature and film at a large scale: from manual coding to various computational techniques.

Keywords: cultural evolution; narrative; literature; film; hedonic selection; computational humanities.

Introduction

Why are some stories more viral than others? Why do some books become bestsellers while others become flops? Why do certain clichés repeat, again and again, from the ancient *Odyssey* to modern Marvel movies? We must answer these, and similar questions if we want to understand a key component of human culture: storytelling. Over the millennia of human history, stories evolved from short, simple forms – a legend, a myth, a folktale – to extremely complex ones, with convoluted plots, deeply realistic characters, and, in the case of films, stunning visual and sound effects. In many cases, we know *how* this transition happened, but we rarely know *why* it happened. We know a lot about the *history* of literature and film, thanks to the meticulous work of many literary and film scholars, but we know very little about the mechanisms behind this history. In other words, we know very little about the *cultural evolution* of narratives. Why?

Most likely, there are two reasons: a conceptual and a methodological one. Concepts-wise, it may be unclear in what *sense* books and films evolve, as it is clear with other cultural products. For example, with technologies. Many technologies become widespread because they are good at performing some practical function. And they usually improve at it over time: cars get faster and safer, microchips – smaller and more powerful. But books and films – how exactly do they evolve? What is their function, and are they becoming better at it? The second challenge is methodological. Studying cultural evolution, empirically, requires large amounts of high-quality data. Books and films, however, have a short record of being used *as data*. And for a good reason: manual collection of data from, say, Proust's seven-volume *In Search of Lost Time* or all 26 seasons of *Doctor Who* would be, at best, extremely time-consuming or, at worst, impossible.

This chapter describes how, in the recent years, researchers began overcoming these challenges. First, it clarifies in what sense narratives evolve: which theoretical concepts, both from the evolution theory and from the humanities, can be useful for understanding this process. Second, it discusses the empirical research on content-based, or hedonic, selection in the evolution of films and books. Finally, it showcases the methodological approaching that are used for studying the large-scale evolution of modern narratives.

Narratives evolve

In what sense do narratives evolve? Some scholars think that this sense is purely metaphorical. Literary critic René Wellek: "evolutionism is false when applied to literature, because there are no fixed genres comparable to biological species, which can serve as substrata of evolution. There is no inevitable growth and decay, no transformation of one genre into another, no actual struggle for life among genres" (Wellek, 1956: 661). Or, this "blindingly obvious" objection from Christopher Prendergast, fifty years later: "whatever their other properties, literary texts do not possess genes" (Prendergast, 2005: 59). Stories cannot breed, they cannot survive or get extinct, like biological species. Wellek, Prendergast, and many others agree that this metaphor can be interesting or even illuminating, but it remains a metaphor: a bridge between two *essentially different* phenomena.

Different, really? Many theorists of evolution have pointed out that evolution, as a concept, does not require genes (Mesoudi, 2011; Richerson & Boyd, 2005). It requires only three basic components: (1) a source of variation in a population, (2) a mechanism that prioritizes certain variants over others, and (3) a way of transmitting successful variants. The evolution of culture satisfies these three principles. Cultural information – that is, socially learnt information – has a source of variation (new cultural items are created by us all the time); cultural information and its physical embodiments – artefacts – may be more or less useful; finally, useful information and artefacts get transmitted within human populations. Note that evolution, in this sense (identical to the precise sense used in biology), does not assume the "inevitable growth and decay" mentioned by Wellek. There's nothing inevitable about cultural evolution. Like biology, culture is anything but inevitability: creative chaos, rather. This chaos, however, contains elements of order: not "laws", but loose regularities, tendencies, patterns. Studying them is the goal of cultural evolution research.

If cultural evolution isn't a metaphor, how shall we understand the evolution of literature and film, in a non-metaphorical sense? It is the evolution of *conventions*, or *techniques*, of storytelling: the how-to knowledge passed between the generations of writers and film producers. These conventions were given various names: more common ones – tropes, clichés, genres – and less common – motifs (Veselovsky, 1940), archetypes (Frye, 1957), formulas (Cawelti, 1976). One of them is "form" (Shklovsky, 1929): a recurring technique of storytelling. I will use it throughout this chapter, because it does not have demeaning connotations (like *cliché* or *trope*), is not tied to a particular theory (like *archetype* or *formula*), and is not bound to content alone (like *theme* and *genre*). Form is any technique of storytelling, thematic or stylistic.

How exactly do narrative forms look like? Narratology, the discipline that studies storytelling, distinguishes between several "levels" of narration (the names of which may differ): the level of *story*: a bare-bones sequence of events and characters; the level of *plot*: a configuration of events, best suited for narration; and the level of *style*: verbal or visual presentation of plot. Each of these levels contains multiple conventions. Forms of the story level include large "building blocks" of narration: locations, events, characters (for example, a castle is a typical location in a gothic novel, a vampire – a typical character, killing a vampire with a wooden stake – a typical event). Plot forms are techniques of combining these story elements: for example, the same story can be told either in a chronological order, or it can have a more complex structure, with flashbacks and flashforwards. The level of style contains multiple formal conventions too: say, character speech can be presented directly, indirectly, or in a "free indirect" mode, which blurs the distinction between the speech of the character and the speech of the narrator.

Over the years, literary and film scholars created many taxonomies of narrative forms. Roland Barthes (1975) and John Cawelti (1976) described some of the basic elements and recurring patterns of the story level. Structural components of plot, such as the order of events or narrative perspective, were categorized by Boris Uspensky (1973) and Gerard Genette (1980). The elements of literary style – by Mikhail Bakhtin (1982), Ann Banfield (1982), and the members of "Group μ" (Dubois et al., 1981). Filmic forms – by Jurij Lotman (1976), Seymour Chatman (1978), David Bordwell (1985), and others. At the same time, there have been no large-scale attempts to assemble databases of literary or film forms, as it was done for other cultural phenomena, such as folktales (Uther, 2004) or songs (Lomax, 1968).

Even though many narrative forms may appear to be timeless, they all were invented at some point. For some forms, this moment of invention happened in the deep past. For example, many of the plot structures widespread in contemporary narratives, such as the Repetition-Break plot (J. Loewenstein & Heath, 2009) – the sequence of three similar events, two of which establish a pattern while the third event breaks it – originate in folklore, and so we do not know who exactly invented it and when. Other forms have a clear starting point. Take "invasion literature": a small genre popular from the 1870s to 1910s (Clarke, 1997). It began with a novella *The Battle of Dorking* (1871), which contained an innovative combination of forms: a first-person story (form #1) about a sudden invasion (#2) of your country, happening in the near future (#3). Following the success of the novella, hundreds of authors copied this distinctive formal blueprint.

The success enjoyed by narrative forms can vary. Some forms, like the invasion literature, have existed for several decades. Others – for centuries. What are the mechanisms behind this differential success? The next section covers one such mechanism.

Hedonic selection of stories

All stories compete for the attention of readers or viewers. This attention is limited, resembling a limited valuable resource: we have time and energy to consume only so many films or books. The result of this is what psychologist Colin Martindale (1990) suggested calling *hedonic selection*. Appealing stories and, broadly, narrative forms "survive", while boring stories and forms eventually get "extinct" – resembling Darwin's natural selection. Martindale wasn't the first to draw this evolutionary parallel. Sociologist of literature John Cawelti: "The process through which [literary] formulas develop, change, and give way to other formulas is a kind of cultural evolution with survival through audience selection" (Cawelti, 1976: 20). More recently, literary historian Franco Moretti explained the emergence of literary canons – critically recognized collections of "most important" books – as follows: "The slaughter of literature. And the butchers – readers: who read novel A (but not B, C, D, E, F, G, H, . . .) and so keep A "alive" into the next generation, when other readers may keep it alive into the following one, and so on until eventually A becomes canonized" (Moretti, 2000b: 209).

Hedonic selection, independently proposed several times in the humanities, is also present in the theory of cultural evolution, under the name of *content-based selection*, or content bias (Richerson & Boyd, 2005). Content-based selection means that a cultural item is selected due to its own properties – its "content", which includes all the thematic and stylistic properties of the item. This process is opposed to *context-based selection*, when cultural items get chosen not because they are appealing on their own, but because, for example, they are perceived as prestigious (e.g., we choose to read a book because it was awarded the Booker Prize) or popular (e.g., we watch a TV series because it has been watched by many of our friends). Content-based selection assumes that, other things being equal, the narratives and narrative forms that are more appealing will also become more successful: bestselling, and copied by other writers or filmmakers. But what exactly does it mean – to be appealing?

Let's take an example from film history. One of the most well-established trends here is the decrease of mean shot duration. In the early years of cinema, shots – the chunks of film between

two cuts – used to be, on average, 12–13 seconds long; but over the century the average shot length shrunk to around 4–5 seconds. **Figure 1** shows that this happened twice: first, in silent films, and, when sound films were invented, the decrease occurred again (Cutting & Candan, 2015). This trend was also found in the history of TV series (Butler, 2014). That is, an average film or TV series today has a much faster exchange of shots than in the past. This quickening aligns with several other tendencies in film history: for example, the amount of movement onscreen increased too (Cutting et al., 2011). Why are films getting faster? Psychologist James E. Cutting suggests that these trends reflect a selection process. Faster onscreen movement is more effective at keeping the audience's attention, since humans are wired to attend to moving objects. And so, in the process of film evolution, most films adopted the fast-paced style. This appealing technique became widespread in the population of films. In cultural evolution theory, such common (or even universal) psychological preferences are called cognitive attractors (see Miton, in this volume, for a review).

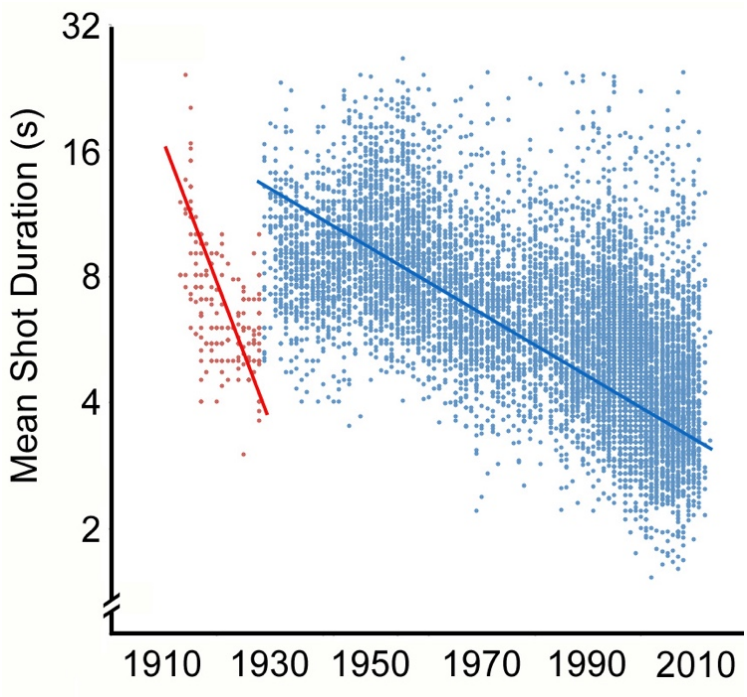

Figure 1. Mean shot duration in the history of films (Cutting & Candan, 2015), each dot being an individual film. Shots gradually become shorter – first, in silent films (red dots), then, in sound films (blue dots).

Attention is only a very basic contributor to hedonic selection. In the end, those stories survive that not only keep our attention, but also create pleasant emotional reactions: surprise, curiosity, suspense, laughter, pleasant fear, etc. Film scholar Ed S. Tan (1996) called films "emotion machines" – for their ability to trigger all these feelings. Fictional stories are technologies of

emotional engagement. And so, we might expect that those of them are selected that are better at this job. For example, one strong emotion, often exploited by stories, is curiosity. We feel the pleasant itch of curiosity when some information is hidden or incomplete – when we encounter what George Loewenstein (1994) called "information gaps". Later, psychologists demonstrated that some types of plots are more effective at evoking curiosity than others (Brewer & Lichtenstein, 1982; Hoeken & van Vliet, 2000). Plots with convoluted, non-chronological order of events (think of Christopher Nolan's *Memento*) are great at this, because changing the order of events automatically opens up information gaps: learning about an event that happened at fictional time t2, makes us wonder about the preceding event at t1. Does this mean that our shared preference for curiosity-evoking plots would drive film evolution towards increased number of non-chronological stories? Sobchuk & Tinits (2020) analyzed a corpus of mystery films, and found precisely that: film plots did become more convoluted and non-linear.

Not all the cognitive preferences are intuitive: we don't prefer only the beautiful and the interesting. Sometimes, in an oxymoronic way, "unappealing" content can perform rather well in cultural transmission. Take negativity bias: the fact that negative information is more noticeable and better memorized. In their large-scale study of books, Morin & Acerbi (2017) have found that, over the 20th century, positively-charged words (e.g., "love" or "beautiful") declined in frequency, while the proportion of negatively-charged words (e.g., "hate" or "agony") stayed the same: a finding that suggests negativity bias – even if explaining the exact causal mechanism here isn't easy (see also Brand et al., 2019). Even highly unpleasant emotions, like disgust, can be surprisingly appealing. Eriksson and Coultas (2014) have demonstrated that urban legends with more disgusting content (say, a story about a woman accidentally eating her own dog) are read more often and are shared more easily (see also Acerbi, 2019; Stubbersfield et al., 2017). One can easily find examples of such counterintuitive popularity of disgust in books and movies. Body horror – a movie subgenre – relies on depicting unpleasant-looking bodies: diseased, mutilated, mutated. A good example is David Cronenberg's *The Fly* (1986), showing a slow transformation of a man into a human-sized insect.

All these studies highlight that understanding hedonic selection – and cultural trends driven by it – is impossible without the understanding of human cognition. Our shared psychological preferences, such as the attention for quick movements or the enjoyment of curiosity-evoking information, exert mild, but constant, pressures on narrative forms, which, over multiple generations, result in visible historical trends. Fortunately, over the recent decades, several disciplines were built on the border of the humanities and the brain sciences, simplifying the much-

needed transfer of knowledge: empirical aesthetics (Wassiliwizky & Menninghaus, 2021), cognitive literary studies (Zunshine, 2015), neurocinematics (Hasson et al., 2008), and empirical literary studies (Kuiken & Jacobs, 2021). Further collaboration between cultural evolution and these disciplines will be crucial for advancing our understanding of the evolution of narratives.

On hedonic selection – two concluding remarks. First: this section was focused on the general psychological mechanisms of hedonic selection only because, by being general, they are easier to study. While some psychological preferences are fairly universal, many others are idiosyncratic. We should not underestimate the impact of one's life experiences and social context on their aesthetic taste. Second: hedonic selection is not the only mechanism making narrative forms succeed or fail. Let's discuss these other mechanisms.

Beyond Selection: Drift, Cumulation, Co-evolution

Is cultural success always meritocratic? It would be, if hedonic selection were the only force at play. But it is not. One of alternative, non-meritocratic, forces is drift, or random copying. In the case of drift, a cultural item is adopted (e.g., a book is read, or a film is watched) not because it is better than the competing items, but simply by chance. It's the opposite of selection. Drift is one of the better-studied mechanisms of cultural evolution – it is known to result in recognizable distributions, with few hyper-frequent items and a "long tail" of infrequent ones (Hahn & Bentley, 2003; for a critical take, see Leroi et al., 2020). Drift was shown to play an important role in many areas of culture – straightforward ones, like the adoption of baby names, where the selective forces are weak (can some names be much better than others, really?), and surprising ones, like music. Salganik et al. (2006) asked participants of an online study to listen to songs and rate them. The scholars wanted to find out why songs become popular. The experiment was run multiple times, and some songs were always rated well, others – always badly, which means that the hedonic appeal of songs does matter. But, interestingly, in different runs of the experiment, different songs would become "big hits", suggesting that *extreme* success is based on chance. The extreme inequality of success can be further exacerbated by the mechanisms of context-based selection, like conformity, which create the rich-get-richer dynamics: items that are already popular, become even more popular (Barabási, 2018).

Back to narratives. Do extremely successful books or films owe a part of their success to drift or conformist selection? The distribution of success of writers follows the same pattern of extreme

inequality. Few authors – like Stephen King or J. K. Rowling – are hyper-popular, while most others have much smaller success (Algee-Hewitt et al., 2016). This extreme inequality means that either drift or conformity (or other context-based mechanisms) are acting upon literary evolution, in addition to mere hedonic selection. However, to learn which exact mechanism plays the biggest role we would need to further explore these datasets with mathematical or agent-based models of evolutionary processes (as was done in Hahn & Bentley, 2003; Kandler & Shennan, 2013).

Or, another question: do narratives become more complex over the course of evolution? Many cultural products, evolving, also increase in complexity: contemporary cars are much more sophisticated than the Ford Model T. The theory of cultural evolution has a term for this: cumulative culture (Mesoudi & Thornton, 2018). It means that society, as a whole, evolves by producing individual innovations, often fairly minor ones, which are added to the shared pool of technologies, resulting in increasingly complex products. If fictional narratives are "emotion machines", as I argued above, shouldn't these machines also accumulate innovations? Or are narratives an exception, somehow? Elsewhere, I have shown that the classical detective novel evolved through accumulation of innovations made by many writers over more than a century (Sobchuk 2018). For example, Edgar Allan Poe invented the locked-room puzzle, Charles Felix was the first to include an image of a floor plan of the crime scene, and Wilkie Collins invented – or, at least, greatly popularized – the plot with a closed circle of suspects. These and many other formal inventions were eventually combined in the first modern crime novels, like Agatha Christie's *The Mysterious Affair at Styles* (1920). In another study, Tinits & Sobchuk (2020) found a similar cumulative evolution in film history. They analyzed the film production crews over a hundred years and found that they became increasingly complex: new jobs were invented and added to the repertoire of standard jobs in film production. This probably reflects the growing complexity of films themselves: new "components" of films, invented over time (such as 3D graphics) requires new crew members (3D artists).

Finally, the success or failure of narrative forms does not depend only on the shared psychological preferences. Literature and films *co-evolve* alongside other cultural phenomena: systems of production (e.g., Hollywood shapes the plots of films worldwide), political context (e.g., some political regimes tend to produce propaganda films and books), or technological breakthroughs (e.g., space exploration influenced the science fiction genre). All of these constitute a complex socio-technological landscape, the comprehensive map of which is yet to be drawn by historians. How exactly are narratives influenced by, or influencing, this landscape? We know very little about this. For example, Moretti (2000a) argued that the evolution of novels follows a predictable pattern:

their plots are copied from one national literature to another, but, to these plots, different local stylistic techniques are added. Or, Martins & Baumard (2020) showed that prosociality, measured as the proportion of cooperation-related words, increased in the dramatic plays prior to the English Civil War and the French Revolution. Both studies suggest a correlation between fictional narratives and their real-world context. But what is the causation here? Are fictional narratives mostly following the "reality", or are they influencing it in a major way? The question remains open.

Drift, cumulation, co-evolution: these are just three of many variables in the complex picture of the evolution of narratives. Since this research area is still very new, we have more questions than answers. To be able to answer them, we need suitable tools that would transform books and films into the something suitable for quantitative explorations. Book and films as data: our next subject.

Books and films as data

Big changes in the evolution of stories do not happen overnight. They happen over long stretches of time: years, decades, centuries. And so, a scholar interested in these changes must be able to somehow capture this grand, centuries-long process. The common approach to studying literature has traditionally been "close reading": comparing various texts in the search of intertextual connections. Such a search would result in lists of examples supporting a particular historical explanation. At times, close reading was used to back theories about literary or filmic evolution – like the early 20th-century theory about the co-evolution of "high-brow" and "low-brow" genres (Shklovsky, 1929; Tynianov, 2019): according to it, new generations of prestigious literature usually build upon the earlier forms of popular fiction, not prestigious fiction. However, close reading produces only anecdotal evidence – suitable for conducting nuanced case studies, but not for testing large-picture theories. And so, even the early scholars of literary evolution acknowledged the importance of moving beyond the anecdotal – towards quantitative research (Hankiss, 1972; Lotman, 1967; Yarkho, 2019).

Such large-scale quantitative analysis of trends in culture can take two different routes: (1) the analysis of *manually tagged features* from books and films; (2) the analysis of *algorithmically extracted features* from books and films.

Manual tagging (or manual coding) is excellent for creating datasets tailored to a particular research question. As an example, let's consider a well-known study by Moretti (2000b), designed to answer a question: how exactly "clues" – the narrative form lying at heart of the detective fiction – were invented. Clues are pieces of information about a crime; using them, readers can try guessing the identity of the murderer before the investigator presents their solution. Clues transform reading fiction into a game, adding excitement and suspense. Moretti's hunch was that the invention of clues, this pivotal form for the genre, was serendipitous, resembling well-known accidental inventions in science (Yaqub, 2018). To find evidence for this, he manually coded 158 crime stories published in 1891–1900 on the pages of *The Strand Magazine*, famous for publishing the original Sherlock Holmes stories. Moretti classified each story according to its use of clues, into one of groups like these: absent (no clues used), present but not necessary (for example, when they are given at the very end of the story), necessary but not visible (the detective mentions that they exist, but the reader is not given the details), or decodable clues, which are *the* clues (**Figure 2**). Decodable clues began as a minor bifurcation on the tree of clue types. Invented early on by Conan Doyle, they are only rarely used by him (most Sherlock Holmes stories don't feature decodable clues), despite their clear potential for creating suspenseful storytelling. This suggests an invention, the full potential of which wasn't realized at first – and, at least in this sense, was a serendipity.

Similarly to this study, other scholars used manual coding to test their hypotheses about the large-scale historical patterns in literature (Gasparov, 1996; Paige, 2020) and film (Cutting, 2021; Sobchuk & Tinits, 2020). Manual coding comes with advantages and disadvantages. On the upside, it allows assembling fine-grained data: machine learning cannot identify "clues" in crime stories – at least, not yet. On the downside, it is labor-intensive, often leading to small sample sizes (158 short stories are many, but not that many). Despite the limitations, manual coding is important for creating large databases of cultural evolution, as illustrated by resources like D-PLACE or the Database of Religious History (for a review, see Slingerland et al., 2020).

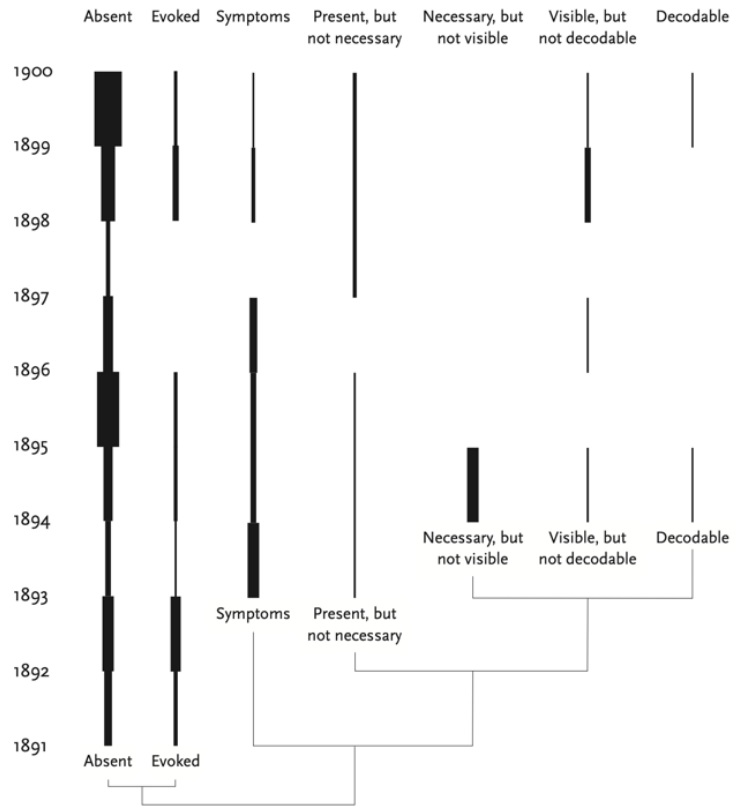

Figure 2. A dendrogram showing the presence of various types of clues in early detective stories (from Moretti, 2005). The thickness of the line shows the proportion of stories with a particular type of clues. In the 1890s, decodable clues – the centerpiece of modern detective fiction – were rare.

To overcome the main limitation of manual coding – laboriousness – many scholars, instead of manually collecting new datasets, use already existing manually collected data. Typically, these are crowdsourced datasets assembled on the websites for social cataloguing – creating own collections of items and rating them – such as Internet Movie Database (IMDb), MovieLens, or Goodreads. On these websites, users can rate films or books, give them tags, write reviews, add plot synopses, and so on. The amounts of these crowdsourced data are often staggering: for instance, as of September 2021, IMDb contained over 8.3 million unique titles and over 6.3 million user reviews (IMDb, 2021). This illustrates one obvious advantage of studying modern culture compared to ancient culture: today's readers or moviegoers actively engage with today's books or films, not with ancient papyri or medieval chivalric novels. Hence, such crowdsourced data is abundant, but its quality is uneven. For example, on Goodreads users can tag books by genre: potentially, useful data for researchers. The problem, however, is that an average user's idea of a genre may be very different from that of a literary scholar (e.g., much less granular or based on very different assumptions). This is why the most used type of data from these platforms is the most reliable one:

user ratings, which were shown to adequately reflect the cultural prominence of films (Wasserman et al., 2015) and books (Kousha et al., 2017).

User ratings allow putting a number on the elusive phenomenon so important in cultural evolution: success. Measuring success is crucial for understanding various mechanisms of evolution, including hedonic selection or drift. For example, Liu et al. (2018) have used IMDb ratings to analyze the phenomenon of "hot streaks" – periods when one's success is much larger than average – in the careers of film directors, as well as scientists and artists. They found that hot streaks tend to be short and, most interestingly, seem to happen at random moments of one's career. Also, in a follow-up study (Liu et al., 2021), they found that, before the hot streaks, film directors use the "exploration" strategy – they experiment with various styles and genres – but after the hot streak they use the "exploitation" strategy: they produce works similar to those that are already successful.

The second approach to quantitative cultural analysis uses *algorithmic extraction of features*. It requires no manual tagging; instead, the necessary data points are "tagged" by algorithms. The digitization of literature and films, as well as the introduction of new techniques of text mining and machine learning, offers new possibilities for testing evolutionary hypotheses. It became possible to automatically analyze thousands of books and films – a practice that is often called "distant reading" (Moretti, 2013) and "distant viewing" (Arnold & Tilton, 2019): that is, looking at literature or films from afar, "reading" or "viewing" them without doing so in the literal sense. This extraction of useful information can take different shapes, depending on the algorithms applied.

The most basic technique of distant reading: extracting *word frequencies* of individual texts or authors. These frequencies form a unique "profile" of texts, which can be used for measuring similarity between texts, as it has traditionally been done in the discipline called stylometry (Roelli, 2020). Simple word frequencies were used to find an interesting evolutionary pattern: texts closer to one another in time are also more similar semantically (Hughes et al., 2012). A more complex type of algorithmically extracted features could involve building a *topic model*, which creates the clusters of semantically similar words – topics – based on word co-occurrences. Topic models have been used, for example, by Šeļa et al. (2022) to show that, in poetry, semantics co-evolves with the metrical structures: a poem about love will likely have a different meter (the sequence of stressed and unstressed syllables) than a poem about nature. Speaking more generally, capturing semantics is a key challenge for studying large-scale evolution of narratives, and several other approaches to it exist. One is *sentiment analysis*; in its common version, it involves finding emotionally charged words in text collections, using vocabularies of "positive" and "negative" words (Acerbi et al., 2013). A

different approach to capturing semantics is based on a technique called *word embeddings*. Here, the meaning of a word is based on the words that surround it in the text. Word embeddings can be used to capture all kinds of semantics, and they do not require vocabularies with pre-selected words. For example, using this approach, Garg et al. (2018) traced the long-term change in gender and ethnic stereotypes in literature. Logistic regression and other methods of *supervised machine learning* can also be used for studying the evolution of stories. Supervised learning involves a training dataset, which can be used, say, for teaching an algorithm to learn the distinctive features of some literary genre; afterwards, this trained algorithm can be applied to some historical dataset – to learn how deep are the historical roots of this genre. Underwood (2019) has used this approach, which he calls "perspectival modeling", to measure the lifespans of literary genres.

Conclusion

"One of the books that should be written as soon as possible […] is *The Origin of Literary Species*. If I remained alive, I would begin working on it right away" – wrote Boris Yarkho (2006: 302), a pioneer of quantitative literary studies. Yarkho did not have enough time to start this project – he died a few years after writing these words. But even if he remained alive, chances are that doing it "as soon as possible" would have been… premature. In the 1940s, there was no coherent theory of cultural evolution, no large datasets, and no powerful computational methods. Even a brilliant scientific project can fail, if carried out at a wrong time.

Has anything changed? 80 years after the unwritten *The Origin of Literary Species*, are we in a better position to study the evolution of narratives? Today, the necessary ingredients seem to be in place: the theory and the methods. We are witnessing a powerful joint effort at studying the basic mechanisms of cultural evolution. The scholars of cultural evolution are no longer lonely enthusiasts, like Yarkho: they are members of a quickly growing academic field with hundreds of papers appearing every year (see Gray & Watts, 2017; Youngblood & Lahti, 2018). Likewise, new datasets and new algorithms, suitable for studying the evolution of stories, are regularly released. This combination of theory and methods offers a promise – that a much better understanding of narrative evolution is soon to come.

References

Acerbi, A. (2019). Cognitive attraction and online misinformation. *Palgrave Communications*, *5*(1), 1–7. https://doi.org/10.1057/s41599-019-0224-y

Acerbi, A., Lampos, V., Garnett, P., & Bentley, R. A. (2013). The Expression of Emotions in 20th Century Books. *PLOS ONE*, *8*(3), e59030. https://doi.org/10.1371/journal.pone.0059030

Algee-Hewitt, M., Allison, S., Gemma, M., Heuser, R., Moretti, F., & Walser, H. (2016). Canon/archive: Large-scale dynamics in the literary field. *Pamphlets of the Literary Lab*, *11*.

Arnold, T., & Tilton, L. (2019). Distant viewing: Analyzing large visual corpora. *Digital Scholarship in the Humanities*, *34*, i3–i16. https://doi.org/10.1093/llc/fqz013

Bakhtin, M. (1982). *The Dialogic Imagination*. University of Texas Press.

Banfield, A. (1982). *Unspeakable Sentences*. Routledge & Kegan Paul.

Barabási, A.-L. (2018). *The Formula: The Universal Laws of Success*. Little, Brown.

Barthes, R. (1975). An Introduction to the Structural Analysis of Narrative. *New Literary History*, *6*(2), 237–272.

Bordwell, D. (1985). *Narration in the Fiction Film*. University of Wisconsin Press.

Brand, C. O., Acerbi, A., & Mesoudi, A. (2019). Cultural evolution of emotional expression in 50 years of song lyrics. *Evolutionary Human Sciences*, *1*. https://doi.org/10.1017/ehs.2019.11

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, *6*(5), 473–486. https://doi.org/10.1016/0378-2166(82)90021-2

Butler, J. (2014). Statistical Analysis of Television Style: What Can Numbers Tell Us about TV Editing? *Cinema Journal*, *54*(1), 25–44.

Cawelti, J. G. (1976). *Adventure, Mystery, and Romance: Formula Stories as Art and Popular Culture*. University of Chicago Press.

Chatman, S. B. (1978). *Story and Discourse: Narrative Structure in Fiction and Film*. Cornell University Press.

Clarke, I. F. (1997). Future-War Fiction: The First Main Phase, 1871-1900. *Science Fiction Studies*, *24*(3), 387–412.

Cutting, J. E. (2021). *Movies on Our Minds: The Evolution of Cinematic Engagement*. Oxford University Press.

Cutting, J. E., Brunick, K. L., DeLong, J. E., Iricinschi, C., & Candan, A. (2011). Quicker, Faster, Darker: Changes in Hollywood Film over 75 Years: *I-Perception*. https://doi.org/10.1068/i0441aap

Cutting, J. E., & Candan, A. (2015). Shot Durations, Shot Classes, and the Increased Pace of Popular Movies. *Projections*, *9*(2), 40–62. https://doi.org/10.3167/proj.2015.090204

Dubois, J., Edeline, F., Klinkenberg, J.-M., Minguet, P., Pire, F., & Trinon, H. (1981). *A General Rhetoric*. The Johns Hopkins University Press.

Eriksson, K., & Coultas, J. C. (2014). Corpses, Maggots, Poodles and Rats: Emotional Selection Operating in Three Phases of Cultural Transmission of Urban Legends. *Journal of Cognition and Culture*, *14*(1–2), 1–26. https://doi.org/10.1163/15685373-12342107

Frye, N. (1957). *Anatomy of criticism*. Princeton University Press.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gasparov, M. L. (1996). *A History of European Versification*. Oxford University Press.

Genette, G. (1980). *Narrative Discourse: An Essay in Method*. Cornell University Press.

Gray, R. D., & Watts, J. (2017). Cultural macroevolution matters. *Proceedings of the National Academy of Sciences*, *114*(30), 7846–7852.

Hahn, M. W., & Bentley, R. A. (2003). Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society B*, *270*, S120-123. https://doi.org/10.1098/rsbl.2003.0045

Hankiss, E. (1972). The structure of literary evolution. *Poetics*, *5*, 40–66.

Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008). Neurocinematics: The Neuroscience of Film. *Projections*, *2*(1), 1–26. https://doi.org/10.3167/proj.2008.020102

Hoeken, H., & van Vliet, M. (2000). Suspense, curiosity, and surprise: How discourse structure influences the affective and cognitive processing of a story. *Poetics*, *27*(4), 277–286. https://doi.org/10.1016/S0304-422X(99)00021-2

Hughes, J. M., Foti, N. J., Krakauer, D. C., & Rockmore, D. N. (2012). Quantitative patterns of stylistic influence in the evolution of literature. *Proceedings of the National Academy of Sciences*, *109*(20), 7682–7686.

IMDb. (2021). *Press Room*. IMDb. http://www.imdb.com/pressroom/stats/

Kandler, A., & Shennan, S. (2013). A non-equilibrium neutral model for analysing cultural change. *Journal of Theoretical Biology*, *330*, 18–25. https://doi.org/10.1016/j.jtbi.2013.03.006

Kousha, K., Thelwall, M., & Abdoli, M. (2017). Goodreads reviews to assess the wider impacts of books. *Journal of the Association for Information Science and Technology*, *68*(8), 2004–2016. https://doi.org/10.1002/asi.23805

Kuiken, D., & Jacobs, A. M. (Eds.). (2021). *Handbook of Empirical Literary Studies*. Walter de Gruyter.

Leroi, A. M., Lambert, B., Rosindell, J., Zhang, X., & Kokkoris, G. D. (2020). Neutral syndrome. *Nature Human Behaviour*, 1–11. https://doi.org/10.1038/s41562-020-0844-7

Liu, L., Dehmamy, N., Chown, J., Giles, C. L., & Wang, D. (2021). Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nature Communications*, *12*(1), 5392. https://doi.org/10.1038/s41467-021-25477-8

Liu, L., Wang, Y., Sinatra, R., Giles, C. L., Song, C., & Wang, D. (2018). Hot streaks in artistic, cultural, and scientific careers. *Nature*, *559*(7714), 396–399. https://doi.org/10.1038/s41586-018-0315-8

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Loewenstein, J., & Heath, C. (2009). The Repetition-Break Plot Structure: A Cognitive Influence on Selection in the Marketplace of Ideas. *Cognitive Science*, *33*(1), 1–19. https://doi.org/10.1111/j.1551-6709.2008.01001.x

Lomax, A. (1968). *Folk Song Style and Culture*. Transaction Publishers.

Lotman, J. (1967). Literaturovedenie dolzhno but naukoj. *Voprosy Literatury*, *1*, 90–100.

Lotman, J. (1976). *Semiotics of Cinema*. The University of Michigan.

Martindale, C. (1990). *The clockwork muse: The predictability of artistic change* (pp. xiv, 411). Basic Books.

Martins, M. de J. D., & Baumard, N. (2020). The rise of prosociality in fiction preceded democratic revolutions in Early Modern Europe. *Proceedings of the National Academy of Sciences*, *117*(46), 28684–28691. https://doi.org/10.1073/pnas.2009571117

Mesoudi, A. (2011). *Cultural Evolution*. The University of Chicago Press.

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B: Biological Sciences*, *285*(1880), 20180712. https://doi.org/10.1098/rspb.2018.0712

Moretti, F. (2000a). Conjectures on World Literature. *New Left Review*, *1*, 54–68.

Moretti, F. (2000b). The Slaughterhouse of Literature. *Modern Language Quarterly*, *61*(1), 207–228.

Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.

Moretti, F. (2013). *Distant Reading*. Verso.

Morin, O., & Acerbi, A. (2017). Birth of the cool: A two-centuries decline in emotional expression in Anglophone fiction. *Cognition and Emotion*, *31*(8), 1663–1675. https://doi.org/10.1080/02699931.2016.1260528

Paige, N. D. (2020). *Technologies of the Novel: Quantitative Data and the Evolution of Literary Systems*. Cambridge University Press. https://doi.org/10.1017/9781108890861

Prendergast, C. (2005). Evolution and Literary History. *New Left Review*, *34*, 40–62.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution* (pp. ix, 332). University of Chicago Press.

Roelli, P. (2020). *Handbook of Stemmatology History, Methodology, Digital Approaches*. De Gruyter. 10.1515/9783110684384

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, *311*, 854–856. https://doi.org/10.1126/science.1121066

Šeļa, A., Plecháč, P., & Lassche, A. (2022). Semantics of European poetry is shaped by conservative forces: The relationship between poetic meter and meaning in accentual-syllabic verse. *PLOS ONE*, *17*(4), e0266556. https://doi.org/10.1371/journal.pone.0266556

Shklovsky, V. (1929). *O teorii prozy*. Federatsija.

Slingerland, E., Atkinson, Q. D., Ember, C. R., Sheehan, O., Muthukrishna, M., Bulbulia, J., & Gray, R. D. (2020). Coding culture: Challenges and recommendations for comparative cultural databases. *Evolutionary Human Sciences*, *2*. https://doi.org/10.1017/ehs.2020.30

Sobchuk, O. (2018). *Charting artistic evolution: An essay in theory* [Thesis, Unversity of Tartu]. https://dspace.ut.ee/handle/10062/62406

Sobchuk, O., & Tinits, P. (2020). Cultural Attraction in Film Evolution: The Case of Anachronies. *Journal of Cognition and Culture*, *20*(3–4), 218–237.

Stubbersfield, J. M., Tehrani, J. J., & Flynn, E. G. (2017). Chicken Tumours and a Fishy Revenge: Evidence for Emotional Content Bias in the Cumulative Recall of Urban Legends. *Journal of Cognition and Culture*, *17*(1–2), 12–26. https://doi.org/10.1163/15685373-12342189

Tan, E. S. (1996). *Emotion and the Structure of Narrative Film: Film as Emotion Machine*. Lawrence Erlbaum.

Tinits, P., & Sobchuk, O. (2020). Open-ended cumulative cultural evolution of Hollywood film crews. *Evolutionary Human Sciences*, *2*. https://doi.org/10.1017/ehs.2020.21

Tynianov, Y. (2019). Permanent Evolution: Selected Essays on Literature, Theory and Film. In *Permanent Evolution*. Academic Studies Press. https://doi.org/10.1515/9781644690635

Underwood, T. (2019). *Distant horizons: Digital evidence and literary change*. The University of Chicago Press.

Uspensky, B. (1973). *A Poetics of Composition*. University of California Press.

Uther, H.-J. (2004). *The types of international folktales: A classification and bibliography, based on the system of Antti Aarne and Stith Thompson*. Suomalainen Tiedeakatemia, Academia Scientiarum Fennica.

Veselovsky, A. (1940). *Istoricheskaia poetika*. Khudozhestvennaia literatura.

Wasserman, M., Zeng, X. H. T., & Amaral, L. A. N. (2015). Cross-evaluation of metrics to estimate the significance of creative works. *Proceedings of the National Academy of Sciences*, *112*(5), 1281–1286.

Wassiliwizky, E., & Menninghaus, W. (2021). Why and How Should Cognitive Science Care about Aesthetics? *Trends in Cognitive Sciences*, *25*(6), 437–449. https://doi.org/10.1016/j.tics.2021.03.008

Wellek, R. (1956). The Concept of Evolution in Literary History. In *For Roman Jakobson* (pp. 653–661). Mouton.

Yaqub, O. (2018). Serendipity: Towards a taxonomy and a theory. *Research Policy*, *47*(1), 169–179. https://doi.org/10.1016/j.respol.2017.10.007

Yarkho, B. I. (2006). *The Methodology of Scientific Study of Literature*. Jazyki slavianskih kultur.

Yarkho, B. I. (2019). Speech Distribution in Five-Act Tragedies (A Question of Classicism and Romanticism). *Journal of Literary Theory*, *13*(1), 13–76. https://doi.org/10.1515/jlt-2019-0002

Youngblood, M., & Lahti, D. (2018). A bibliometric analysis of the interdisciplinary field of cultural evolution. *Palgrave Communications*, *4*(1), 1–9.

Zunshine, L. (Ed.). (2015). *The Oxford Handbook of Cognitive Literary Studies*. Oxford University Press.