

Why Computational Humanities needs falsifiable theories

Jean Barré, Nicolas Baumard, Jeanne Bollée, Alexandre Bluet, Jean-Baptiste Camps, Charles de Dampierre, Edgar Dubourg, Sylvain Estebe, Zhaolun Fan, Grégory Fiorio, et al.

▶ **To cite this version:**

Jean Barré, Nicolas Baumard, Jeanne Bollée, Alexandre Bluet, Jean-Baptiste Camps, et al.. Why Computational Humanities needs falsifiable theories. 2026. ⟨hal-05827443⟩

HAL Id: hal-05827443

https://enc.hal.science/hal-05827443v1

Preprint submitted on 20 Oct 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Why Computational Humanities needs falsifiable theories

Jean Barré[1*], Nicolas Baumard[2*], Jeanne Bollée[1*], Alexandre Bluet[2*], Jean-Baptiste Camps[3*], Charles de Dampierre[2,5*], Edgar Dubourg[2*], Sylvain Estebe[4*], Zhaolun Fan[2*], Grégory Fiorio[2*], Ze Fu[2*], Mingmeng Geng[1*], Ulysse Godreau[3*], Konstantin Henke[1*], Ruoting Liu[2*], Paco Maunoir[2*], Rowan Hall Maudslay[3*], Olivier Morin[2*], Suzanne Mpouli[3*], Hossein Samani[2*], Olga Seminck[1*], Valentin Thouzeau[2*], Ying Zhong[2*], Pascal Amsili[1†], Frédéric Landragin[1†], Théo Moins[3†].

1. LATTICE, CNRS, École normale supérieure–PSL, Université Sorbonne Nouvelle, Montrouge, France.
2. Institut Jean Nicod, Département d'études cognitives, École normale supérieure–PSL, EHESS, CNRS, Paris, France.
3. Centre Jean-Mabillon, École nationale des chartes–PSL, Paris, France.
4. Center for Humanities Computing, School of Culture and Society, Aarhus University, Aarhus, Denmark.
5. Bunka AI.

* Equal first-authors, in alphabetic order.
† Signing authors who will contribute to the revisions.

Corresponding address: ch.theory@proton.me

Computational humanities has powerful tools and abundant data, but too little research tests falsifiable theories. We argue for a shift: dedicating a share of research effort to theory-driven hypothesis testing, coexisting with descriptive scholarship and exploration. We review some past attempts at building theory-driven research programmes in the humanities and outside of them, some successful, some not. Successful turns towards theory-driven research do not diminish the relevance or importance of descriptive work but make it more focused. We show how emerging theories can supply predictive frameworks usable by computational humanities and outline practical steps for bridging gaps between theory-driven and descriptive research programmes.

Computational Humanities Has a Theory Problem

Over the past decades, the humanities have progressively started to adopt computational methodologies, yielding substantial results (e.g. Michel et al., 2011; Lansdall-Welfare et al., 2017; Neidorf et al., 2019; Assael et al., 2022). Much of this work, however, remains confined to exploratory data analysis: patterns are identified first and then interpreted *a posteriori*, and the focus often rests on description rather than generalisation. Nothing is lacking in terms of technical apparatus. Common methods range from principal component analysis and topic modelling to deep learning and the use of large language models, all of which are applied to corpora consisting of millions of books or digitised artifacts. What is lacking, however, is a theoretical framework that provides general explanations, makes falsifiable hypotheses and predictions, and integrates scattered pieces of knowledge into a global cumulative framework.

We refer to this style of research as "hypothetico-deductive science". What we mean by this is, roughly, the methods of the natural sciences as Hempel (1966) or Popper (1959) saw it. A theory generates hypotheses before any analysis is carried out, and these hypotheses can then be subjected to

experimentation and disproven. We know that this view of the natural sciences is contested (Hepburn and Andersen 2026). What matters, for our purpose, is that this view of science puts the emphasis where we want it to be: on theories, hypotheses, and the predictions they generate. Other views of the scientific method also stress the importance of theories and predictions; we could have called upon them as well. We do not claim that hypothetico-deductivism provides the one accurate theory for how the natural sciences work. We just see it as a useful methodological guide.

Most practitioners in the digital humanities use big databases to construct visualisations or to report summary statistics, not to test theories. A statistical model extracts structure from a corpus; a theory sits outside the corpus and states, before anyone looks at it, what structure should be found and what structure would refute it (Figure 1). Of course, quantification and big data are the natural allies of theory-driven science as they make it much easier to test theories systematically. The computational turn in the humanities is good news, but it only gets us so far.

The observation that humanities need more theory might be surprising: after all, theories abound in the humanities. But most of them are not what we have in mind here: they can accommodate a variety of outcomes. What is meant here is predictive theories, which are nomothetic rather than idiographic (Muthukrishna and Henrich 2019; Sobchuk and Youngblood 2026). That is to say, theories that are able to generate hypotheses that can be tested (Underwood, 2019) in broad accordance with the principles of hypothetico-deductive science (Hempel 1966; Popper 1959).

The share of theory-driven work should never account for 100% of humanities research. The insights gathered from exploratory data analysis as well as rigorous erudition are of paramount importance, and an integral part of the scientific endeavour: theories do not emerge from a vacuum, and usually require intimate knowledge of the objects to which they apply. Data analysis is crucial for generating hypotheses abductively, but theory-driven science requires another step in which those generated ideas are turned into predictions that can then be tested on new or independent datasets.

To give an order of magnitude, we could aim for a 10% pragmatic allocation target: roughly 1 in 10 papers targeting hypothesis testing, 1 in 10 citations referencing predictive frameworks, and 10% of our collective research effort dedicated to evaluating general explanations. This 10% agenda rules out two options: 0%, with no theory-driven work whatsoever, and 100% (or even a majority of) theory-driven research, where descriptive work would disappear.

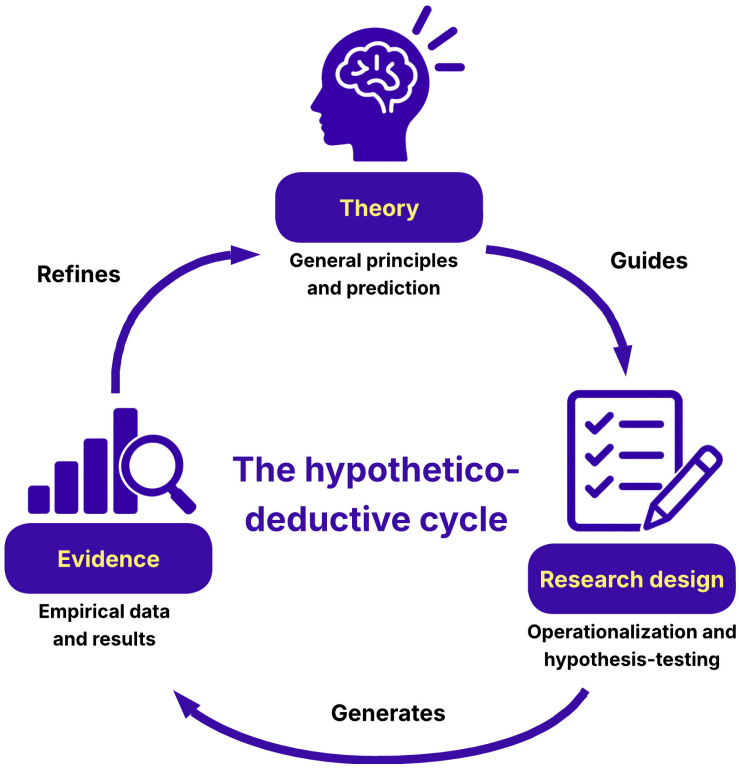


Figure 1. The hypothetico-deductive cycle. A theory external to the corpus (top) supplies general principles and predictions. The theory guides research design (right), where the prediction is operationalised into measurable outcomes and the corpus is built to test it rather than to maximise scale. Measurement generates evidence (left) capable of disconfirming the prediction. That evidence then informs a revision of the theory, which closes the cycle and raises new questions. The order of the steps is what matters: the hypothesis should be specified before the confirmatory test is conducted, and evaluated on data that were not used to formulate it. The theory specifies both what should be counted and what result would count against it (Hempel, 1966).

Nuance traps

Hypothetico-deductive science shocks many habits in the humanities. It tests simplified and abstract theories, built on operationalisation and assumption. These theories are the opposite of nuance, the opposite of thick description. They are not contextualised; they are risky, refutable, and often very general.

When Galileo experimented with inclined planes, he kept two features—distance and time—and threw away everything else. Was his approach nuanced? No: simplification was central to the enquiry (Smaldino 2017). From the standpoint of erudition, Galileo was sloppy. He failed to properly contextualise his observations, and was uncritical of his data (Koyré 1978). This attitude came with a cost: Galileo got his explanation of tides wrong because he misread the empirical literature. But overall, this cost was worth the benefit. Galileo would never have figured out a working theory of falling bodies with the tools of his time if he had been nuanced. The theory, of course, turned out to be wrong in important ways, but the more accurate theories that followed would not exist without Galileo.

The humanities, by and large, and for legitimate reasons, value exactly the opposite: they value nuanced, erudite, contextualised descriptions of data that treat generalisation with great caution. If these are the values one is trained in, it can be difficult to see the value of hypothetico-deductive

science when applied to questions in the humanities, especially given the long history of problematic attempts in that direction. To test scientific theories, data often needs to be simplified, and turned into a series of one-dimensional answers to simple questions. This can look like a failure to engage in deep, contextualised interpretation. But the "reductionism" that some decry is not a weakness of scientific inquiry, but its strength.

We endorse Healy's (2017) call against arguments based on nuance in the humanities. What Healy calls a "nuance trap" is an argument that does not refute a claim but simply asks its proponent to make it less general, or to mention some overlooked dimension of the problem. Calls for nuance always sound reasonable and moderate (therein lies the trap), but the result can be an avoidance of direct scientific debate. The result is close to 100% erudition, nuance and context—all valuable things—but close to 0% theory-driven research.

The humanities do not ignore abstract theorising, and they are not opposed to theoretical simplification per se; indeed, modelling is central to some approaches in digital humanities, for instance (e.g. McCarty 2019; Huitfield and Sperberg-McQueen 2008). Models, here, are abstract constructions that may describe the relation between a text and its transcription, or the ontology (in the computer science sense) of spatial relationships in literary novels. These models can clarify thought and also serve as a basis for coding better tools to describe data. What they do not provide is predictions that can be tested.

"Theory" in the humanities comes in another, well-known form (e.g. post-structuralism, psychoanalytic criticism). A theory in this sense is a set of general claims. Theory-driven research rests on theories that are not just general claims. A scientific theory is a fully explicit and testable model of things, which can be refuted with data. "Theory" in the humanities sense, then, is a false friend, just like "models": in the humanities, "theory" is a near-synonym of "general claim". When we talk about theory, we talk about scientific theories that can feed the hypothetico-deductive cycle. When we talk about models, we mean models that yield testable predictions.

We have abstract models, most of which can't be tested; we have "Theory" but few theories.

How did a theory-driven approach benefit other fields?

In paleontology, fossil collections remained just that for a long time: assemblages of finding with no overarching theory to exploit them. Paleontologists tended to see their work as zoology and botany limited by the quality of the fossils. The acceptance of Darwin's theory did not immediately turn the field into a theory-driven enterprise. Before Darwinism was fully formalised and mathematicised in the mid-twentieth century, it made few specific predictions about systematics, the pace and tempo of evolution, or comparative morphology. This changed after the neo-Darwinian synthesis, and above all with the "paleobiology revolution" (Sepkoski and Ruse 2009): Simpson imported population genetics and reframed paleontological data around rates of change rather than sequences, and David Raup and Jack Sepkoski used sophisticated models of taphonomic biases to serve as null hypotheses against which hypotheses (such as the cyclicity of mass extinctions) could be tested (Perreault 2019). After this shift, paleontologists retained and refined their erudition and methodology, describing and naming specimens, correlating strata, and publishing fossil monographs. They also contributed to key advances in evolutionary biology, and indeed several influential hypotheses originated in paleontological work (e.g. Eldredge and Gould 1972; Jablonski and Shubin 2015).

In economic history, for most of the twentieth century the archive determined which questions could be asked. The transformation of the field came when a theory was adopted that was indifferent to period and place (Diebolt and Haupert 2019). This theory specified a small number of mechanisms

that affect economic systems: agents respond to relative prices, institutions determine the payoffs agents face, markets clear unless something identifiable prevents them, and so on. These mechanisms are claimed to hold for thirteenth-century grain markets and nineteenth-century railways alike. This shared theoretical core provided by economics establishes a set of constraints on admissible explanation. It makes quantities comparable in the first place: the standardised series (Maddison 2001; Allen 2001) rest on theoretical definitions of output and of a real wage, not on whatever the sources happened to record. It generates counterfactuals: Fogel (1964) could price a world without railroads because theory says what a transport cost is worth. It turns singular events into observations in a design: Acemoğlu et al. (2001) could treat colonial settlement as a natural experiment because theory specified what institutions were supposed to be doing. And it makes disagreement productive: two economic historians who disagree are disagreeing about a parameter or an identifying assumption, not about how to read a period. Such disputes can be settled.

These disciplines did not become scientific overnight. They went from being mostly descriptive and observational, prizing nuance and erudition above all else, to including some theory-driven work which set an agenda for the rest. In Boyer's terms (2003), there was a move from 100% erudition to something like 90% erudition + 10% (theory-driven) science.

Learning from previous attempts to make humanities more theory-driven

Attempts to make the humanities more theory-driven are as old as the humanities themselves. Yet these programmes either failed on their own terms or at the very least need an aggiornamento.

Case 1: The Annales school. The Annales historians rejected event-based narratives of history in favour of long time scales, serial data, and quantification. They built the archival infrastructure—price series, parish registers, probate inventories, tithe records—upon which much subsequent historical demography and economic history rests. The theoretical component, however, remained thin (Hunt 1986; Burke 2015). The key explanatory notion, *mentalités*, functioned as a residual category, describing collective attitudes without a model of the underlying psychology. The result was a powerful descriptive apparatus attached to explanations that could accommodate almost any outcome. Vovelle (1973) coded some 30,000 Provençal wills and dated the collapse of requested funeral masses to the middle of the eighteenth century. This was an exemplary piece of serial history whose explanation, a drift of *mentalités* toward dechristianisation, restated the curve rather than accounting for it. Ariès (1960) made the same move in reverse, inferring from iconography, dress, and pedagogical treatises that medieval parents withheld affection from their children, with no account of what would make parental attachment vary in the first place.

Case 2: Distant reading. Literary studies pursued a similar agenda with far greater technical means, including digitised archives of hundreds of thousands of volumes, and methods capable of tracking lexical, stylistic, and structural variation over centuries (Moretti 2005). The characteristic output, however, has been the description of trends rather than the testing of hypotheses. Theory was not completely lacking: Moretti drew his theories from quantitative history, geography, and evolutionary theory. But these models were rarely charged with making any prediction in advance: a graph showed a trend, and an explanation arrived afterward, imported from existing literary history rather than derived from a theory that could have been proven wrong. Disagreement concerns the reliability of measurement instead of the truth of competing explanations. The instruments are strong, but the hypotheses they are asked to test are usually absent.

Case 3: Genre theory. The structuralist school of Genette, Todorov and Schaeffer built the conceptual apparatus that literary studies still uses: intertextuality, architextuality, hypertextuality, paratext, transtextual relations, the fantastic as defined by the reader's hesitation, genres as regimes of

use rather than classes of objects, and so on (Todorov 1975; Schaeffer 1989; Genette 1992; 1997a; 1997b). This vocabulary survives because it captures real distinctions. The programme also produced substantive claims about literary history. Todorov held that the fantastic, as a literary genre, is bounded in time, arising with secularisation and dying around 1900 once psychoanalysis supplied a natural explanation for phenomena that the genre had described with imaginary means. Genette held that the lyric–epic–dramatic triad is a Romantic invention retrojected onto Aristotle. Schaeffer held that generic membership works by family resemblance rather than by necessary criteria, which entails that readers' classifications should be graded rather than binary. All these claims are testable. Genette and his colleagues never tested them. They worked from a few dozen works selected for their exemplary value, and never asked whether the claims held beyond that sample.

The three programmes failed differently. The Annales had data and no hypothesis, distant reading had instruments and no predictions, and structuralism had testable claims that nobody tested. But the experience of neighbouring fields and the availability of more testable theories could change this.

Theories of culture have become more testable

Today we have theories that are much more testable compared to those of previous programmes. These theories make more univocal predictions. They are often quantified. They make it harder to rationalise away anything we observe. One reason for this is increased formalisation, mathematisation, and operationalisation. Another is interdisciplinary contact between the humanities and fields like evolutionary biology or cognitive science. These research programmes exist in neighbouring fields, and similar endeavours could be easy to conceive for the computational humanities.

Cultural evolution

The evolutionary approach to the study of culture is based on the idea that cultural traits bear resemblance with genetic traits, and can be modelled in the same way (Cavalli-Sforza and Feldman 1981; Boyd and Richerson 1985). Cultural traits pass from brain to brain by imitation and teaching, they vary, and some spread further than others. Because that process is analogous to genetic transmission, the models and methods biologists built for genes can be imported and turned on cultural data. Several authors already used cultural evolution to call for theory-driven change in social science (Muthukrishna and Henrich 2019) or archaeology (Perreault 2019). Two imports are worth highlighting.

The neutral (or unbiased copying) model of cultural evolution is a bare-bones model of cultural transmission, inspired by population genetics but stripped of all biology-specific mechanisms (Neiman 1995; Bentley and Shennan 2004; Camps et al. 2026; Kandler and Shennan 2013). Agents transmit cultural traits to one another randomly. Computational models derived from this hypothesis correctly predict some striking aspects of cultural transmission that have puzzled researchers for a long time, such as its extremely inegalitarian outcomes. For example, when choosing a baby name, most people tend to choose a highly frequent name, and in assemblages of archaeological pots, a few signature decorations are more popular than all the rest combined. Few people take the hypothesis' assumptions seriously: in reality cultural transmission is not random, agents are not interchangeable, and transmission is not just copying. That is not the point: the point of the hypothesis is to provide a null baseline against which better hypotheses must be tested.

Cultural phylogenies, on the other hand, are methods that retrace changes in a family of cultural traits over time, assuming their transmission has the shape of a tree (Mace and Holden 2005; Jordan et al.

2009). Examples include language families and lineages of manuscript versions of a story. Humanists actually invented methods to reconstruct these histories before biologists did (Gray and Atkinson 2003; Currie et al. 2010; Camps and Cafiero 2015); today we in turn can benefit from the methodological refinements that biologists have brought to these techniques, from cladistics to Bayesian phylogenies. Phylogenetic methods force their users to make all their theoretical assumptions fully explicit: one must decide how fast traits change, whether they change randomly, how tree-like is cultural history, and so on. These methods can be used to test local hypotheses about a particular cultural history (e.g. "Did the Indo-European language family originate in Anatolia?"; "Is the Oxford assonanced *Song of Roland* closer to the common ancestor of all the tradition than the rhymed versions?"); they can also test more general questions about cultural evolution (e.g. "Does language transmission follow demographic lines or not?"; "Is the preservation of historical texts random, or biased towards innovation?"). Phylogenies and their results remain controversial but the controversy is productive: it forces people to make their disagreements explicit and it leaves room for future agreement based on empirical findings. Importantly, it raises the profile of descriptive research: a controversy over the phylogeny of Austronesian languages generates a lot of interest in knowing the full set of cognates for the word "canoe" in the Vanuatu archipelago.

One fruitful application of cultural evolution to literature was focused on the question of cohort-driven change. Does literature change because writers change their minds, or because writers die? Literary history has long organised itself around generations without any way of showing that generational membership has explanatory value. Planck's principle, borrowed from the sociology of science, states the claim in a form that can fail: new ideas prevail by the replacement of their opponents rather than by persuasion. Moretti predicted that genres follow generational cycles of twenty-five to thirty years, but left the prediction untested; a decade later Underwood tested this theory and found that detective and science fiction cohere far longer than the theory allows (Underwood 2019). A full test of the generational hypothesis requires an instrument that partitions a trend into demographic causes (Underwood et al. 2022), and evolutionary ecology supplies one: decomposition based on the Price equation splits the change in a population mean into the contributions of arrivals, departures, and change within individuals still present. Sobchuk and Beheim (2025) treat topic frequencies in 23,000 Anglophone novels (1850–1940) as the trait and authors' career spans as the demography. The prediction fails and they show that exits contribute least. They conclude that literature turns out to advance one birth at a time rather than one funeral, because entrances are synchronised (nearly all authors start young), while retirements are not.

Cultural attraction, cultural production, and cultural ecology

In contrast to cultural evolution, the cultural attraction and cultural production approaches posit that cultural items are reconstructed in a non-random way, instead of being copied: minds attend to and remember some stimuli, and ignore and forget others (Sperber 1996; Claidière et al. 2014; Morin 2016). What a population ends up with is therefore determined by what human perceptual, linguistic and motivational systems find easy to process and hard to ignore. Artifacts, in this view, are not traits in circulation but products made by agents pursuing goals, and they can be analysed as behaviour (André et al. 2023; Baumard and André 2025; Safra et al. 2020). What a producer makes is what serves their ends given the audience they face, so the tools of behavioural ecology apply. Researchers can therefore identify the adaptive problem the artifact addresses, the constraints under which it is made, and the trade-offs it resolves. Because producers compete for a limited supply of attention, artifacts drift toward the forms audiences prefer, and can exceed the range those preferences evolved to handle. And because preferences are themselves calibrated by local conditions, such as affluence, mortality risk, or the reliability of the developmental environment, the same theory that says what people want also says how their wants should shift when conditions shift (Baumard and André 2025). This approach therefore predicts convergence across unrelated traditions, exaggeration over time

within traditions, and variations predicted by ecology—dated, directional predictions about the content of portraits, novels, songs, and so on.

Juvenile traits in faces. Pokémon, Hello Kitty, teddy bears, puppies in advertising: juvenile-looking forms succeed on a global scale, and people converge in their judgments about which ones are cute. Why is this the case? Here, insight from cognitive science can help (Garsmeur and Baumard 2026). Cognitive science accounts for this convergence in terms of a dedicated perceptual mechanism. The cue set described by Lorenz (1943) as the *Kindchenschema*—a large cranium relative to the face, large and low-set eyes, round cheeks, and a short snout—has since been isolated experimentally by parametrically morphing infant faces along that dimension. Glocker and colleagues (2009) showed that cuteness ratings and caretaking motivation vary monotonically with the manipulation and that activity in the nucleus accumbens scales with it. Because such a mechanism is exploitable by stimuli exceeding the natural range, cultural artifacts should drift toward exaggerated forms. The cue is measurable: facial landmarks yield a baby-schema index that can be computed at scale on portraits, illustrations, and animation frames. This permits directional predictions about cultural variation. For instance, where parental investment per child is higher—under lower fertility, extended schooling, and greater material security—depictions of children should score higher on the index, childhood should occupy a larger share of visual production, and non-human characters should be more strongly juvenilised (Garsmeur and Baumard 2026).

The shape of letters. Across historically independent writing systems, letter strokes over-represent the cardinal orientations relative to obliques, and symmetrical characters are more often symmetrical about a vertical rather than a horizontal axis. Both regularities can be derived from documented properties of the visual system. The primary visual cortex allocates a disproportionate number of orientation-tuned cells to vertical and horizontal contours, producing an “oblique effect" (Appelle 1972; Furmanski and Engel 2000): acuity, discrimination speed, and recall are all superior for cardinal orientations. Detection of mirror symmetry is likewise fastest and most accurate about a vertical axis, a specialisation expected in organisms for which the recognition of bilateral symmetry in conspecifics and prey is a recurrent task. Changizi and colleagues (2006) argued more generally that the stroke configurations of scripts match those found in natural images. Morin (2018) provided a corpus test, showing that orientation and symmetry axis anisotropies are present across a large sample of the world's writing systems and, critically, that they strengthen over the documented history of writing.

Applied fields can bridge science and erudition

Some fields in the humanities have begun to gravitate towards theoretical foundations, even if their theories are borrowed from external sciences, with relatively little internal theoretical debate. Stylometry is an example of such an applied field, where empiricism finally met theory. Initial forays were motivated by fairly specific questions: *What is the chronology of Plato’s dialogues? Did Shakespeare write the plays attributed to him?* Mendenhall (1887), a physicist, performed his ‘stylistic spectroscopy’ of texts, which had all the appearances of objectivity and quantification, with its painstaking computation of word-length distributions, and its surface similarity to physics, yet very little grounding in explicit hypotheses about how people write and what would make a style unique. Lutosławski’s *Principes de stylométrie* (1898) fleshed out a theory (or, rather, a set of definitions), while advocating for quantification of stylemes (e.g. function words) as a measurement of style. But for over a century, stylometry enjoyed empirical efficiency without a nomothetic framework. It neither tested its core assumptions, nor made falsifiable predictions. Even the notion that individual humans had a unique and measurable style was left untested for that time. Ultimately, its theoretical foundations were borrowed from psycholinguistics, and the work of scientists such as Pennebaker (2011), shifting the quantification of stylemes “from black magic to theory” (Kestemont 2014). Today we retrospectively understand why function words are successful authorial markers: they reflect

unconscious cognitive and emotional processes, varying from one individual to another. Stylometry is able to test hypotheses about the attribution of individual texts, or the robustness of stylistic markers across genres, even if its theory is not native, but borrowed from cognitive science. Stylometry provides an example of a bridge field that can both improve erudition (by establishing the authorship of individual texts) and serve as a test ground for cognitive linguistics theories, demonstrating how theoretical frameworks can serve all the aspects of the humanities.

What this looks like in practice

Most of the corpora we work on were assembled before we arrived, and a specialist usually has some sense of what the main series do. A researcher can therefore look at a trend, state the theory that predicts it, and report a successful test, with nothing in the finished paper to distinguish it from work done in the right order. This is not a problem peculiar to the humanities. Epidemiologists, astronomers, geneticists and applied economists all work on data they did not generate. We could benefit from their good practices.

Split the corpus, as is already done, for instance, in natural language processing. Divide the corpus before looking at any of it, explore a part freely, then design an analysis and run it once on another part of the corpus. This lets you discover a pattern and test it without the discovery counting as its own confirmation. The division can be random, or follow a line that makes the test harder: a later period, another tradition, another genre.

Preregister, already a standard practice in fields such as clinical trials, ecology, psychology**.** Write down what you expect to find, and how you will test it, before you run the analysis, and deposit the document on a dedicated platform, that timestamps it. You are anchoring your design choices while you still don't know which ones favour your hypothesis. The document should give the prediction, how it will be measured, how the corpus will be assembled and filtered, and which analysis counts as confirmatory.

State the mechanism, by saying what produces the effect that you predict. Writing systems become more legible over time because readers find some shapes easier to tell apart than others, and characters that are hard to distinguish get lost in copying. A mechanism of this kind commits you to more than the trend it was invented to explain. It says where else the effect should appear (e.g. in any graphic code transmitted by hand, including recently invented ones) and where it should not (in symbols that are never copied).

Why do hypothetico-deductive science, some of the time?

Theory-driven inquiry gives purpose to empirical research, it makes research generalist by default, and it generates meaningful disagreement.

Motivating data-collection and expertise. Predictive models give us expectations about what the data should look like, and allow us to be surprised and proven wrong. They also help establish criteria on which data should be collected, and how to approach unexpected results that seem to contradict previous scholarship. Far from threatening descriptive scholarship, they allow to integrate the granular findings of erudition into broader explanatory structures, cumulatively over time.

Generalist by default. Data-driven, empirical research can arrive at generalities, optionally and painfully. Scientific theories have a generalist scope by default. This changes everything. By looking for more general patterns or recurrences, by asking nomothetic questions, we gain the opportunity to compress information in mentally tractable statements. In a 0% theory-driven world, one has to

apologise for putting forward a general claim. In a 10% theory-driven world, the only way to refute a general claim is to offer a better one: dismissing general claims as speculative is not enough. A world without theory is a sea of confetti. In a world with shared theoretical frameworks, any line of research has the potential to inform the rest.

Theories lead to meaningful disagreement. Explicit, falsifiable predictions establish clear criteria for confirmation or refutation. When a model fails, it highlights specific parameter errors or invalid assumptions, turning academic disagreement into a productive engine for theoretical refinement. The most common reason to dislike theories is that most of them are wrong, and not just wrong, but wrong in a simplistic, arbitrary, one-sided, tone-deaf way. Yet, a clearly stated claim, with positions articulated and lines drawn in the sand, is the best start to a meaningful debate, even if that claim turns out to be false. In a field that acknowledges theory-driven research, the way to engage with theoretical work is to criticise it on its own terms, by refuting its predictions. The alternative to meaningful disagreement is nuance traps, respectful avoidance, and chloroformed debate.

Ultimately, establishing a modest portfolio of theory-driven research does not require abandoning the practice of rich, nuanced, contextualized descriptions that the humanities take justifiable pride in. Instead, dedicating even a fraction of our collective effort to systematic hypothesis testing ensures that computational tools move beyond mere description to advance our understanding of human culture.

Acknowledgements

We thank Emmanuelle Bermès, Florian Cafiero, Yoann Dupont, Alessandro Gnasso, Brenna Hensley, Paola Peratello, Bianca Picariello, Nicolas Pommier, Emma Montecchiari, Thierry Poibeau, Peter Stokes, for discussing previous versions of this paper. This work has received support under the Major Research Program of PSL Research University “CultureLab” launched by PSL Research University and implemented by ANR with the reference ANR-10-IDEX-0001.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. ‘The Colonial Origins of Comparative Development: An Empirical Investigation’. *American Economic Review* 91 (5): 1369–401. https://doi.org/10.1257/aer.91.5.1369.

Allen, Robert C. 2001. ‘The Great Divergence in European Wages and Prices from the Middle Ages to the First World War’. *Explorations in Economic History* 38 (4): 411–47. https://doi.org/10.1006/exeh.2001.0775.

André, Jean-Baptiste, Nicolas Baumard, and Pascal Boyer. 2023. ‘Cultural Evolution from the Producers’ Standpoint’. *Evolutionary Human Sciences* 5: e25. https://doi.org/10.1017/ehs.2023.20.

Appelle, Stuart. 1972. ‘Perception and Discrimination as a Function of Stimulus Orientation: The “Oblique Effect” in Man and Animals.’ *Psychological Bulletin* 78 (4): 266–78. https://doi.org/10.1037/h0033117.

Ariès, Philippe. 1960. *L’enfant et la vie familiale sous l’Ancien Régime*. [Nachdr.], Dépot legal 1975. Points Histoire 20. Éd. du Seuil.

Assael, Yannis, Thea Sommerschield, Brendan Shillingford, et al. 2022. ‘Restoring and Attributing Ancient Texts Using Deep Neural Networks’. *Nature* 603 (7900): 280–83. https://doi.org/10.1038/s41586-022-04448-z.

Baumard, Nicolas, and Jean-Baptiste André. 2025. ‘The Ecological Approach to Culture’. *Evolution and Human Behavior* 46 (3): 106686. https://doi.org/10.1016/j.evolhumbehav.2025.106686.

Bentley, R. Alexander, Matthew W. Hahn, and Stephen J. Shennan. 2004. ‘Random Drift and Culture Change’. *Proceedings. Biological Sciences* 271 (1547): 1443–50. https://doi.org/10.1098/rspb.2004.2746.

Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process.* Culture and the Evolutionary Process. University of Chicago Press.

Boyer, Pascal. 2003. ‘Science, Erudition and Relevant Connections’. *Journal of Cognition and Culture* 3 (4): 344–58. https://doi.org/10.1163/156853703771818109.

Burke, Peter. 2015. *The French Historical Revolution: The Annales School, 1929-2014*. Revised and Updated second edition. Polity.

Camps, Jean-Baptiste, and Florian Cafiero. 2015. ‘Genealogical Variant Locations and Simplified Stemma: A Test Case’. In *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, edited by Tara Andrews and Caroline Macé. Brepols.

Camps, Jean-Baptiste, Julien Randon-Furling, and Ulysse Godreau. 2026. ‘On the Transmission of Texts: Written Cultures as Complex Systems’. *PNAS Nexus* 5 (7): pgag207. https://doi.org/10.1093/pnasnexus/pgag207.

Cavalli-Sforza, L. L., and M. W. Feldman. 1981. *Cultural Transmission and Evolution: A Quantitative Approach.* Princeton University Press. https://doi.org/10.2307/j.ctvx5wbt8.

Changizi, Mark A., Qiong Zhang, Hao Ye, and Shinsuke Shimojo. 2006. ‘The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes’. *The American Naturalist* 167 (5): E117–39. https://doi.org/10.1086/502806.

Claidière, Nicolas, Thomas C. Scott-Phillips, and Dan Sperber. 2014. ‘How Darwinian Is Cultural Evolution?’ *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1642): 20130368. https://doi.org/10.1098/rstb.2013.0368.

Currie, Thomas E., Simon J. Greenhill, and Ruth Mace. 2010. ‘Is Horizontal Transmission Really a Problem for Phylogenetic Comparative Methods? A Simulation Study Using Continuous Cultural Traits’. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1559): 3903–12. https://doi.org/10.1098/rstb.2010.0014.

Diebolt, Claude, and Michael J. Haupert, eds. 2019. *Handbook of Cliometrics: With 131 Figures and 73 Tables*. Second edition. Springer Reference. Springer.

Eldredge, Niles, and Stephen Jay Gould. 1972. ‘Punctuated Equilibria: An Alternative to Phyletic Gradualism’. In *Models in Paleobiology*, edited by Thomas J. M. Schopf. Freeman, Cooper.

Fogel, Robert W. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Johns Hopkins Press.

Furmanski, Christopher S., and Stephen A. Engel. 2000. ‘An Oblique Effect in Human Primary Visual Cortex’. *Nature Neuroscience* 3 (6): 535–36. https://doi.org/10.1038/75702.

Garsmeur, Alexis, and Nicolas Baumard. 2026. ‘A Theory of Cuteness as Developmental Investment’. *Behavioral and Brain Sciences*, August 12, 1–61. https://doi.org/10.1017/S0140525X26106918.

Genette, Gerard. 1992. *The Architext: An Introduction*. Edited by Jane E. Lewin. University of California Press.

Genette, Gerard. 1997a. *Palimpsests: Literature in the Second Degree*. Edited by Channa Newman and Claude Doubinsky. University of Nebraska Press.

Genette, Gerard. 1997b. *Paratexts: Thresholds of Interpretation*. Edited by Jane E. Lewin. Cambridge University Press.

Glocker, Melanie L., Daniel D. Langleben, Kosha Ruparel, et al. 2009. ‘Baby Schema Modulates the Brain Reward System in Nulliparous Women’. *Proceedings of the National Academy of Sciences* 106 (22): 9115–19. https://doi.org/10.1073/pnas.0811620106.

Gray, Russell D., and Quentin D. Atkinson. 2003. ‘Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin’. *Nature* 426 (6965): 435–39. https://doi.org/10.1038/nature02029.

Healy, Kieran. 2017. ‘Fuck Nuance’. *Sociological Theory* 35 (2): 118–27. https://doi.org/10.1177/0735275117709046.

Hempel, Carl G. 1966. *Philosophy of Natural Science*. Prentice-Hall.

Hepburn, Brian, and Hanne Andersen. 2026. ‘Scientific Method’. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/archives/fall2026/entries/scientific-method/.

Huitfeldt, C., and C. M. Sperberg-McQueen. 2008. ‘What Is Transcription?’ *Literary and Linguistic Computing* 23 (3): 295–310. https://doi.org/10.1093/llc/fqn013.

Hunt, Lynn. 1986. ‘French History in the Last Twenty Years: The Rise and Fall of the Annales Paradigm’. *Journal of Contemporary History* 21 (2): 209–24. https://doi.org/10.1177/002200948602100205.

Jablonski, David, and Neil H. Shubin. 2015. 'The Future of the Fossil Record: Paleontology in the 21st Century'. *Proceedings of the National Academy of Sciences* 112 (16): 4852–58. https://doi.org/10.1073/pnas.1505146112.

Jordan, Fiona M., Russell D. Gray, Simon J. Greenhill, and Ruth Mace. 2009. 'Matrilocal Residence Is Ancestral in Austronesian Societies'. *Proceedings. Biological Sciences* 276 (1664): 1957–64. https://doi.org/10.1098/rspb.2009.0088.

Kandler, Anne, and Stephen Shennan. 2013. 'A Non-Equilibrium Neutral Model for Analysing Cultural Change'. *Journal of Theoretical Biology* 330 (August): 18–25. https://doi.org/10.1016/j.jtbi.2013.03.006.

Kestemont, Mike. 2014. 'Function Words in Authorship Attribution. From Black Magic to Theory?' *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 59–66. https://doi.org/10.3115/v1/W14-0908.

Koyre, Alexandre. 1978. *Galileo Studies*. Edited by John Mepham. Harvester Press.

Lansdall-Welfare, Thomas, Saatviga Sudhahar, James Thompson, et al. 2017. 'Content Analysis of 150 Years of British Periodicals'. *Proceedings of the National Academy of Sciences* 114 (4). https://doi.org/10.1073/pnas.1606380114.

Lorenz, Konrad. 1943. 'Die Angeborenen Formen Möglicher Erfahrung'. *Zeitschrift Für Tierpsychologie* 5 (2): 235–409.

Lutoslawski, W. 1898. 'Principes de stylométrie appliqués à la chronologie des œuvres de Platon'. *Revue des Études Grecques* 11 (41): 61–81. https://doi.org/10.3406/reg.1898.5847.

Mace, R., and C. Holden. 2005. 'A Phylogenetic Approach to Cultural Evolution'. *Trends in Ecology & Evolution* 20 (3): 116–21. https://doi.org/10.1016/j.tree.2004.12.002.

Maddison, Angus. 2001. *The World Economy: A Millennial Perspective*. Development Centre Studies. OECD. https://doi.org/10.1787/9789264189980-en.

McCarty, Willard. 2019. 'Modelling the Actual, Simulating the Possible'. In *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*, edited by Julia Flanders and Fotis Jannidis. Routledge.

Mendenhall, T. C. 1887. 'The Characteristic Curves of Composition'. *Science* ns-9 (214s): 237–46. https://doi.org/10.1126/science.ns-9.214S.237.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, et al. 2011. 'Quantitative Analysis of Culture Using Millions of Digitized Books'. *Science* 331 (6014): 176–82. https://doi.org/10.1126/science.1199644.

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

Morin, Olivier. 2016. *How Traditions Live and Die*. Foundations of Human Interaction. Oxford University press.

Morin, Olivier. 2018. 'Spontaneous Emergence of Legibility in Writing Systems: The Case of Orientation Anisotropy'. *Cognitive Science* 42 (2): 664–77. https://doi.org/10.1111/cogs.12550.

Muthukrishna, Michael, and Joseph Henrich. 2019. 'A Problem in Theory'. *Nature Human Behaviour* 3 (3): 221–29. https://doi.org/10.1038/s41562-018-0522-1.

Neidorf, Leonard, Madison S. Krieger, Michelle Yakubek, Pramit Chaudhuri, and Joseph P. Dexter. 2019. 'Large-Scale Quantitative Profiling of the Old English Verse Tradition'. *Nature Human Behaviour* 3 (6): 560–67. https://doi.org/10.1038/s41562-019-0570-1.

Neiman, Fraser D. 1995. 'Stylistic Variation in Evolutionary Perspective: Inferences from Decorative Diversity and Interassemblage Distance in Illinois Woodland Ceramic Assemblages'. *American Antiquity* 60 (1): 7–36. https://doi.org/10.2307/282074.

Pennebaker, James W. 2011. 'The Secret Life of Pronouns'. *New Scientist* 211 (2828): 42–45. https://doi.org/10.1016/S0262-4079(11)62167-2.

Perreault, Charles. 2019. *The Quality of the Archaeological Record*. The University of Chicago Press.

Popper, Karl R. 1959. *The Logic of Scientific Discovery*. Hutchinson.

Safra, Lou, Coralie Chevallier, Julie Grèzes, and Nicolas Baumard. 2020. 'Tracking Historical Changes in Perceived Trustworthiness in Western Europe Using Machine Learning Analyses of Facial Cues in Paintings'. *Nature Communications* 11 (1): 4728. https://doi.org/10.1038/s41467-020-18566-7.

Schaeffer, Jean-Marie. 1989. *Qu'est-ce qu'un genre littéraire?* Poétique. Éd. du Seuil.

Sepkoski, David, and Michael Ruse, eds. 2009. *The Paleobiological Revolution: Essays on the Growth of Modern Paleontology*. University of Chicago Press.

Smaldino, Paul E. 2017. 'Models Are Stupid, and We Should Love Them'. In *Computational Social Psychology*, edited by Robin Vallacher, Stephen Read, and Andrzej Nowak. Routledge.

Sobchuk, Oleg, and Bret Beheim. 2025. 'Does Literature Evolve One Funeral at a Time?' *Proceedings of the Royal Society B: Biological Sciences* 292 (2040): 20242033. https://doi.org/10.1098/rspb.2024.2033.

Sobchuk, Oleg, and Mason Youngblood. 2026. 'Cultural Evolution – of the Arts'. *Evolutionary Human Sciences* 8: e19. https://doi.org/10.1017/ehs.2026.10048.

Sperber, Dan. 1996. *Explaining Culture: A Naturalistic Approach*. [Repr.], 1. publ. 1996. Blackwell.

Todorov, Tzvetan. 1975. *The Fantastic: A Structural Approach to a Literary Genre*. Edited by Richard Howard. Cornell University Press.

Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. The University of Chicago press.

Underwood, Ted, Kevin Kiley, Wenyi Shang, and Stephen Vaisey. 2022. 'Cohort Succession Explains Most Change in Literary Culture'. *Sociological Science* 9: 184–205. https://doi.org/10.15195/v9.a8.

Vovelle, Michel. 1973. *Piété Baroque et Déchristianisation En Provence Au XVIIIe Siècle: Les Attitudes Devant La Mort d'après Les Clauses Des Testaments*. Plon.