

# *Ten Thousand Journal Articles Later: Ethnography of «The Literature» in Science*

CHRISTOPHER KELTY

UCLA Center for Society and Genetics  
ckelty@ucla.edu

HANNAH LANDECKER

UCLA Center for Society and Genetics  
landecker@soc.ucla.edu

**Recibido:** 26.05.2009

**Acceptado:** 02.11.2009

## **1. FIRST, TWO STORIES**

1. A subfield of biology emerged in the 1960s called cell fusion. As with many developments in post-World War II sciences, the literature of cell fusion was bewildering for many reasons, not least because of its sheer volume – this work was being done in hundreds of laboratories, generating thousands of publications. Early papers on microinjection and nuclear transplantation were grouped under cell fusion even though they didn't necessarily have much to do with one another nor did they cite each other, even when included in the same volume of collected papers. The new cellular forms created by fusion led to an explosion of new words to describe them, making it difficult to determine when people were talking about the same thing referred to by different terms.

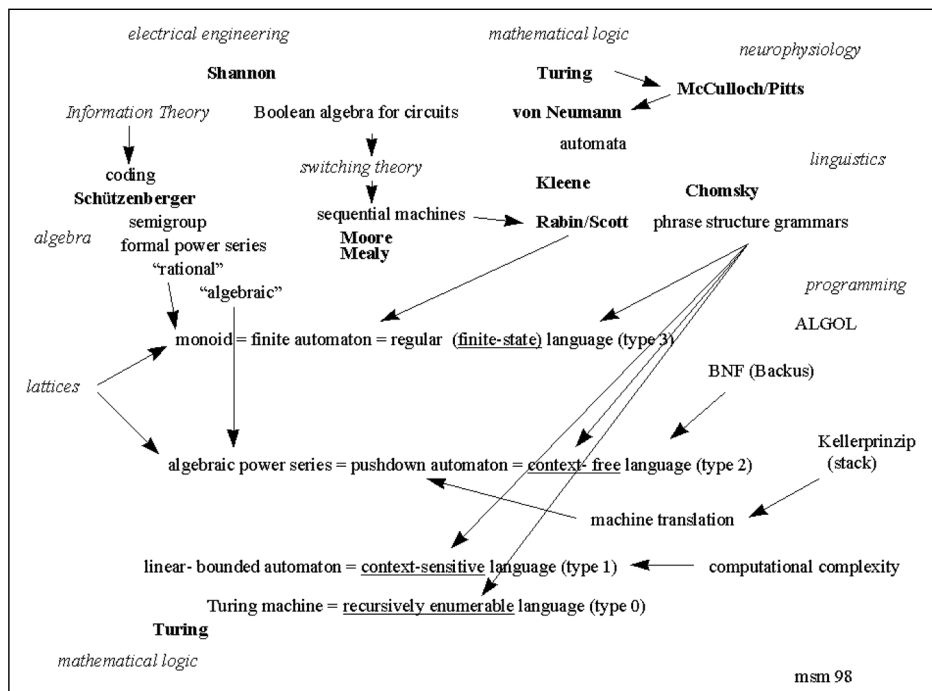
In addition, there were many papers cited and discussed in cell fusion literature that did not really involve fusing cells, which made it hard to determine what lay within the field of inquiry. While cell fusion seemed to clearly foreshadow the ideas and practices of both cloning and making transgenic organisms, there were no clear causal links to trace genealogical lines with. Scientists interviewed decades later had little to say about this work other than how quickly it was replaced by genetic engineering, or how it led to monoclonal antibodies, a useful biotechnical tool. Many younger scientists did not even know about it. You might even ask why it seemed to have any coherence at all, much less one that might necessitate historical analysis.

Here is one answer: Cell fusion marked the emergence of a new mode of approaching living matter which emphasized reproduction without sex and the internal compatibility of all life inside the cell membrane. Well before genetic engineering or cloning, the living cell as manipulable technical space was being established. What was going on in this literature was a gradual, cumulative process of redefinition of a core biological concept, that of *hybridity*. It was being redefined in practice, through the literal making of objects understood as hybrids, whose living properties came as a surprise to their makers. Cell fusion was just what it sounds like: the fusion, by artificial means, of different kinds of cells with one another. Much to the experimenters' surprise, the fused cells did not die in disorganized chaos, but instead their nuclei would merge, and the cells would even continue dividing with the chromosomes from both parent cells. Cells were fused across species boundaries, and extraordinary hybrids with the chromosomes of mouse and man, or hen and man, or mouse and rat became permanent cell lines.

These experiments made it clear that once the cell membrane was breached and resealed, whatever was on the inside of the cell was compatible across extreme biological difference. Experimenters spoke of functioning multi-part enzymes composed of rat subunits and mouse subunits, of human proteins produced in mouse cells – well before the language of genetic homology across species produced by large-scale gene sequencing. Biologists quickly moved on to juxtaposing all kinds of biological difference in the same cellular space. They would fuse old cells with young ones, malignant ones with normal ones, cells at different locations in the cell cycle, differentiated cells with undifferentiated cells. It didn't matter if cell fusion proper was used – early microinjection and nuclear transplantation were regarded as being of a piece with cell fusion exactly because the principle was to, in the words of one paper on microinjection, «turn the tissue culture cell into a test tube».

Cell fusion faded as quickly as it began, as the much more specific method of genetic engineering promised much higher precision in transferring genetic material between species. Journals founded less than ten years before to address the field changed their names or ceased publication. The significance and fleeting coherency of cell fusion as a field came not from what it turned into directly, but from a reconfiguration of ideas and practices of hybridity. The hybrid, rather than the rare oddity at the edge of species, such as the mule, came to mean the conjunction of biological difference in the same cellular space (Landecker 2007).

2. The discipline of computer science did not exist before 1960. By 1970, several distinct subfields had emerged, each with its own publications, scholarly societies, techniques and practices and *very* large volume of literature. The 1950s therefore were thus a period of odd juxtaposition, creative re-mixing and re-labeling, and productive mixing across many disciplines, ranging from engineering and mathematics, to logic, to biology and neuroscience, to operations research



and game theory, to linguistics and cognitive science. Some research projects that would seem to be in obvious conversation with one another (such as the work on automatic machine translation and the work in linguistics inaugurated by Chomsky) employed different methods and theories; other work created surprising connections. Michael Mahoney (Mahoney 2002) has diagrammed some of these surprises that organized the seeming chaos of computer science before roughly 1965. As with cell fusion, an explosion of new words and new claims emerged in these disciplines, and many papers published in, for instance, the *Communications of the Association of Computing Machinery* or *Annals of Mathematics* seemed to be about neither computing machinery or mathematics. Contemporary computer scientists are rarely aware of the odd proleptic character of this early period, in which no one quite knew what the computer would become; instead, they often tell just-so stories of the origins of their subfields: artificial intelligence, complexity, algorithms and data structures, formal language/automata theory, etc. As with the case of cell fusion, one might well ask whether this early period had any cohesion at all.

One explanation is that this period marked a re-orientation of researchers towards *logic* as something highly plastic and instrument-like. Whereas logic in mathematics and philosophy prior to WWII was, with few exceptions, an intellectual pursuit implicitly organized around epistemology, or the limits of human

reason, after WWII, logic branches off, and becomes a distinctly technical affair, buttressed by inventions like Shannon's boolean logic of circuits, McCulloch and Pitts diagrams and models of logical nets in a brain, Chomsky's formal grammars, and Stephen Kleene's formalization of an algebra of regular expressions. The profusion of literature in this period is only tangentially related to the appearance of real computing machines; analyzing it reveals that across a stunning array of disciplines, people were creating new logical instruments, often couched in one or another disciplinary language and formalism (propositional calculus, an algebra of recursive functions, semi-group theory, or new ones like Chomsky's grammars and McCulloch and Pitts' nerve nets); citation analysis reveals, at least, the forms of direct borrowing that occurred, but does not reveal the context in which ideas about designing circuits started to seem like a good way to search for patterns in text (regular expressions) (Kelty n.d.), or ideas about designing algorithms became fodder for defining a formal language of biological growth (L-Systems) (Kelty and Landecker 2004).

Over time, some of these «logical instruments» have gone from being theoretical or mathematical research pursuits to core tools in computer science (as in the case of compilers, parsers and lexical analyzers in the 1970s); others have opened up whole new fields of mathematics, as in the case of formal language and automata theory; still others, such as the logical analysis of software for its verification and model checking (whether it will work and what it means for it to be «correct»), have seen fast growth and surprising failures along the way (MacKenzie 2001). Many new institutions re-configured themselves around these new tools and approaches, including the new computer science departments at places like Carnegie Mellon, Stanford, and MIT, but also sites like Bell Labs and the Stanford Research Institute. Logical instruments, much more than logic as such, came to be a kind of object that could be worked on and shared across disciplines, and could be transformed equally into mathematical formalisms, algorithms and new forms of software and hardware. The history of modeling in contemporary science may well be tied to the emergence of these new logical forms, visible in this early literature (Wise 2004). Indeed, it may be that by investigating the surprising ways in which researchers started to play with, transform, and operationalize logic, one can come-perhaps ironically-to understand something about the nature of scientific thought itself.

## **2. HISTORY OF SCIENCE AND THE «MASS OF FACTS»**

Both of these cursory stories are attempts to come to terms with the vast scale of post World War II science, technology and engineering; they are attempts at synthetic accounts of long-term, large-scale changes in thought and practice in the 20<sup>th</sup> century. By contrast, much work directed at science in this time period in the fields of sociology, history and anthropology of science and

technology is conducted on specific cases, people and places. There are many case studies and few synthetic works to give one ways to stitch together these case studies into some larger sense of what has happened, particularly since World War II, save by making grandiose generalizations about life, or information, or capitalism. We suggest here that this is not only a failing of contemporary analysis, but a feature of the period itself. There is, for example no *Origin of the Species* for 20<sup>th</sup> Century biology; there are certainly conceptual and practical advances with as much or more significance, but the scale and structure of 20<sup>th</sup> century science is such that a Darwin of the 20<sup>th</sup> century is not possible. By extension, the scale of science in the twentieth century, and into the present, creates a methodological challenge for historical work not answered by oral history, traditional archival techniques, or close reading of select texts.

How does one *fathom* the movements of science and technology in this period? What we propose here is a reconceptualization of historical reading and research practices regarding the published literature produced by twentieth century science, technology and engineering, a creative reformation of historical methods that takes account of a massive change of scale and kind in how scientific knowledge is circulated. Instead of case studies or close reading, what we propose is *highly specific empirical work on the general*, through developing a specific approach to the massive bodies of literature produced in this period, an approach we think of as treating *The Literature as an informant*. This approach is a deliberate anthropomorphization of this 'body' as something to be observed and engaged as something alive with concepts and practices not necessarily visible through the lens of single actors, institutions or key papers. We lay out the problem of the literature, and then discuss methodological approaches to working with it. Although an anthropomorphism might suggest unity, this is far from our intention; rather we want to indicate the notion of a complex body of processes, organs, functions, discourse and complex motives which it is possible to probe, dissect and experiment upon in a similar manner to the body and mind of an informant.

Historians of science have spent much time in recent years elaborating the history of reading (Blair 2004; Daston 2004; Johns 2000; Secord 2000; Topham 2004). Through this work, reading practices have emerged as essential to understanding the reception of scientific texts and the generation of meaning in specific contexts, and these reading practices have been linked to transformations in publishing practices, allowing the reinterpretation, for example, of the Darwinian revolution as «an episode in the industrialization of communication and transformation of reading audiences». (Secord 2000:2), quoted in (Topham 2004:437) Much of this excellent work is part of the history of the book, much of it has been conducted on science in the 19<sup>th</sup> century or earlier; what happens to reading in the age of the ever more standardized journal article and the computer? What happens when a body of literature goes through a transition from something that one person might read all of, to something far larger than anyone can digest? What happens to reading and thinking after such a «bibliographic

transition»? Further, with these transformations, what has happened – or should happen – to the reading practices not just of scientists and popular audiences, but of historians of science and technology as well?

Common to the cell fusion and computer science stories recounted here is that they were both derived from a vast expanse of short papers sharing in the process of redefining some pretty fundamental concepts. This literature was produced in the diverse settings of many labs and institutions, by many people and experiments, and it was published in many different technical manuals, conference proceedings, journals and books – although books in latter half of the twentieth century occupy no place of privilege, giving way to the journal article as primary vehicle of science communication. One of the primary practical quandaries in deriving these stories was the basic one of where to look: what to read and how to read it, what to include and what to leave out, activity conducted within the double bind of how to determine a field of inquiry before its definition is even possible.

To rely on disciplinary or institutional boundaries in demarcating zones of close historical attention in this mass of literature is dangerous—at the very least because the question of the remaking of disciplines is itself one of the core questions one might ask, as in the cases of cell fusion or computer science. The knowledge and practices depicted in *The Literature* are themselves part and parcel of a remixing and reformatting of disciplines and scholarly identities that will only subsequently map onto the new fields that emerge. Individual scientists who lived through the events beat idiosyncratic routes through their times and rarely have a synoptic grasp of the literature they were immersed in at a discrete time point many decades before; as valuable as their perspectives are, they are extremely partial narratives overlaid with assumptions about what eventually proved important, or with what came in the years between event and interview. Institutional and individual archives similarly provide important but highly specific and partial insights about institutions and individuals, insights that are not necessarily additive as parts of a whole. Archiving practices themselves reinforce these partialities through the reliance on basic categories such as the archives of personal papers or institutions, which enhance findability even as they obscure connections and structures that might exist in *The Literature* as it lives and breathes.

To answer these challenges, we argue, requires an in-depth engagement with *The Literature*. We describe and objectify this entity in uppercase intentionally, to underline its existence as more than a bunch of individual works to be read; in fact, a defining characteristic of the Literature is its volume, a volume that goes beyond the physical capacities of any individual researcher to read and whose dynamics are not necessarily visible to any of its individual participants. Anyone who has typed a set of what they thought were specific keywords into Medline and gotten 723,000 hits and a polite query about narrowing the search will have a sense of the mind-boggling scale of this entity often referred to as a «body» of literature. Even earlier in the twentieth century, when there were

fewer scientists and fewer journals, the volume of the literature was huge<sup>1</sup>. This does not mean it cannot be systematically studied and interpreted, and we draw on the fields of cultural anthropology and literary history in delineating a mode of *ethnography* of the Literature as one approach to the reimagination of reading in doing historical work at mass scale. In suggesting a combination of quantitative and interpretative methods, we attempt to sail between the Scylla of citation indexing and the Charybdis of rhetoric of science, with their respective macro and micro approaches to textual sources. If, as we suggest, it is in the Literature that concepts and practices move and leave their traces, and that the Literature has its own dynamics and forces that are not the same as the individual production and consumption of discrete pieces of it, is there a way to quantitatively grasp these changes, and yet not abandon reading and interpretation?

The problem of studying literature at massive scale is not exclusive to the history of science. As Franco Moretti has observed of the history of the novel, most scholars in this field study only a minimal fraction of the literary field of any given time period:

A canon of two hundred novels, for instance, sounds very large for nineteenth century Britain, but it is still less than one percent of the novels that were actually published: twenty to thirty thousand, more, no one really knows – and close reading won't help here, a novel a day every day of the year would take century of so... And it's not even a matter of time, but of method: a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole. (Moretti 2005:4).

In arguing for what he calls «distant reading», Moretti asks what would happen if literary historians took their cue from the transformations in historical perspective grounded in the *Annales* school, and shifted attention from the close reading of supposedly representative texts – often selected using unclear criteria of their relative success or enduring nature – to a longer-term and larger scale of texts as a whole, over time. In this parallel shift to the *Annales*' call for shifting the historical gaze from the extraordinary to the everyday, he asks, «What literature would we find, in 'the large mass of facts'»? (3). Similarly, we may ask, what histories of science might be found there, and how?

<sup>1</sup> The growth in volume of the scientific literature over the twentieth century caused particularly marked anxiety at midcentury about the capacity of scientists' minds or memories to cope with the scale, and the dreaming of machines such as Vannevar Bush's Memex. Citations. Paul Otlet and the universal encyclopedia. HG Wells universal encyclopedia. Older dreams of a system of complete management of knowledge MM Slaughter and others? Bush's Memex adds the particular dimension of «traces» through the literature, like book marks or «histories» in a web browser, which could themselves be objectified and shared as a kind of knowledge. Similarity to the literature review, the guidebook.

### 3. BETWEEN CITATION INDEXING AND RHETORIC OF SCIENCE

Moretti's innovation has been to bring quantitative methods and the visualization devices of graphs, maps and trees to a field traditionally dominated by the close examination of a small selection of texts. He argues that you simply see different phenomena through these methods, things that are not visible or fathomable at the scale of the single work. Genre, for instance, is often approached by choosing a representative example. Through work with that example, the genre is defined as a whole. Moretti questions this use of the text as the representative object of knowledge for investigating genre in history: «Texts are certainly the real objects of literature (in the Strand Magazine you don't find 'clues' or 'detective fictio', you find *Sherlock Holmes*, or *Hilda Wade*, or the *Adventures of a Man of Science*), but they are not the right *objects of knowledge* for literary history». (Moretti 2005:76 original emphasis).

Quantitative visualization of a field of novels as a tree that represents both those branches that continue and those that end (borrowed from Darwin's *Origin of Species*), by contrast, visualizes genre as a 'diversity spectrum' (a term borrowed from the evolutionary theory of Ernst Mayr). An individual text can never represent the internal multiplicity of this spectrum; moreover, the choice of representative texts abandons almost all of the archive, whereas including the field in the tree makes them visible. «Instead of reiterating the verdict of the market, abandoning extinct literature to the oblivion decreed by its initial readers, these trees take the lost 99 per cent of the archive and reintegrate it into the fabric of literary history, allowing us to finally 'see' it» (Moretti 2005:77)<sup>2</sup>. Similar insights apply to the place of little-cited or obscurely published scientific papers, gray literature, or technical manuals in historical analysis, which are often ignored or simply overlooked because they didn't seem to have a big «impact» at the time – but «impact» might not be the right measure of historical value. The Literature is huge enough that it contains multiplicities; the world of scientific publication has its minor and major literatures, which contain stories that no one's telling, because they can't discern them.

Borrowing back from an analysis that itself borrows heavily from the history and philosophy of science is complicated here by the fact that history of science and science and technology studies are not naïve to quantitative study of scientific literature, a practice which became established with the theory and tools of citation analysis. Unlike novels, scientific papers are explicitly structured to connect to one another, primarily through the citation, the use of keywords and the standardized structure of format and organization.

<sup>2</sup> Using quantitative graphical methods you see *genres* quite differently – as temporary structures in the historical flow, «morphological arrangements that *last* in time, but always only for *some* time» (14); for a time – 25 to 30 years in his analysis – they function to structure «normal literature» (an analogy to normal science), until the form is no longer capable of representing the most significant aspects of contemporary reality (17-18). Cycles of genres, he argues, are exactly the kind of hidden thread of literary history that can be opened by the quantitative history of literature.



Any discussion of working with The Literature as a whole must begin with thinking about what citation analysis and its more recent quantitative companions can – and cannot – offer. Citation analysis begins with Eugene Garfield's work on the Science Citation Index in the 1970s (Garfield 1979), and continues in the 1980s with the growth in «bibliometrics» which saw a brief but intense uptake in Science Studies. Today the field of approaches is quite diverse, and the increasing availability of the scholarly literature on the Internet has extended tools beyond mere citation to include data mining and meta-data harvesting of diverse sorts. Indeed, whereas the field of literary analysis has effectively resisted any large scale application of these quantitative tools to the The Literature, science and technology are by contrast well studied by the fields of bibliometrics, informetrics, scientometrics and the analysis of science and technology indicators generally (Borgman and Furner 2002). While these fields share a filiation with the history of science and science studies broadly, their use in historical or social analysis has remained limited (but cf. Gingras 2009 for a revival).

All of these quantitative approaches can create more and less sophisticated diagrams of the Literature. They are not intended as interpretive tools. Indeed, Garfield noted that in contrast to organization by keyword or subject indexing, «The citation is a precise, unambiguous representation of a subject that requires no interpretation and is immune to changes in terminology (Garfield 1979:?)». For social scientists interested in following citations, «the intrinsic qualities of any given statement» are not of interest, because it is the citation itself that is meaningful. Bibliometric analyses of science in science studies quickly demurred, arguing that citations are in fact not so immune, but can be positive or negative, or perhaps indicate something else entirely. Out of this realization, Latour generated a complete method (in *Science in Action*) of analyzing science as a military contest over the consolidation of a fact, citations being merely one route to follow as claims proceed from conjecture to certainty to tacit knowledge. What is «social» in Latourian technoscience is the *collective* act of making and building associations that force acceptance of a statement as fact (Latour 1988). It was a point that Garfield also made:

What looked best about a citation index was the diversity of the insights it provided about the literature of a particular subject and the efficiency and stability with which they could be stated. By using author references to index documents, the limited ability of a subject indexer to make connections between ideas, concepts, and subjects was replaced by the far superior ability of the entire scientific community to do the same thing. This meant that a citation index would interpret each of the documents it covered from as many viewpoints as existed in the scientific community (Garfield 1979:9).

Thus it is the index that interprets, not the subject indexer (or, it should be said, the individual scientist). One of the tests that Garfield subjected the SCI to was a comparison between Isaac Asimov's version of the history of the discovery of the helical structure of DNA and one reconstructed using citations:

The citation analysis did more than just duplicate most of the account that Asimov had put together from a remarkable memory. It also added some insights into what happened by identifying 31 relationships and one event that Asimov did not mention...the relationships that a citation analyses shows among the components of a given body of work correspond very well to the relationships perceived by a scientist of Asimov's rank...a citation analysis can identify significant relationships and events that even a remarkable memory might forget, or that traditional techniques of historical research might miss (Garfield 1979:93).

The citation index not only captured what one («remarkable») brain could, but it was proof for Garfield of the index's distinctiveness that it could go beyond an individual human mind or memory.

Indeed, one of the great insights of citation analysis, and one reason why it is still useful is that it allows us to fathom the communal nature of scientific literature production, the status of science as a collective representation, and it allows us to fathom this collective representation by means other than individual interpretation or memory. In the contemporary language most commonly employed, citation analysis permits us to visualize scientific fields as networks implicit in citation. By making these networks explicit and visualizing them we can begin to analyze science using the «superior ability of the scientific community» to interpret it for us.

Citation Analysis is but one small corner of an increasingly large array of quantitative tools for analyzing texts, scientific and otherwise. Co-word analysis, natural language processing, and statistical analysis of texts can be used to dig into the structure of the language. Data mining is increasingly useful for analyzing the content of texts; the field of bioinformatics, for instance, which has relatively standardized nomenclature for chromosomes, genes, loci, mutations and single nucleotide variations is increasingly incorporated into the laboratory work of scientists through the use of software that maps a genetic sample in a lab to the relevant literature in databases like PubMed. Other tools allow the analysis of the «landscape» of patents and papers in scientific fields by drawing maps of the patent landscape (Aureka). Many of these tools are provided by Thomson-Reuters, the current owner of the Citation Index, and range in both cost and quality depending on the discipline or problem under study. Very few of these tools are likely to be of any extensive use to historians, given their more or less exclusive focus on the very recent and the already-digital. However, some of them may hold opportunities for exploration of The Literature in ways that citation indexing gestures toward. In addition to this, the rise recently of publicly shared bibliographic systems (CiteULike, Connotea and Zotero) provides both a new set of tools for mapping the literature, and a new mode of attention within which working scientists themselves come to understand the extent of and connections between the literatures they inhabit.

Of course, the great strength of these approaches is also their great shortcoming. Put simply, are they useful if one retains a commitment to interpreta-

tion? In Moretti's discussion of quantitative literary history, discussed above, he admits that the models he presents «share a clear preference for explanation over interpretation; or perhaps, better, for the explanation of general structures over the interpretation of individual texts (Moretti 2005:91)». The point, he says, is not to produce a new reading of this novel or that, «but the definition of those larger patterns that are their necessary preconditions (Moretti 2005:91)». But what if there was some middle ground to explore between the single text and the mass of texts? What if one remained committed to the idea that both scientists and observers of science can make interpretations, of better and worse quality, and that it is only by doing so that one can actually explore the transformation of meaning and significance in science? What if one wanted to go beyond the citation to read the papers in the vast networks generated by citation analysis? How should one then read? Are there reading and research practices for the Literature that combine a kind of topographic exploration of form and precondition, and the interpretative skills more traditionally associated with key papers or canonical texts?

Traditionally, the work of attention to the working of language within scientific activity has been in the realm of rhetoric of science. Rhetoric of science at first glance seems to be the necessary complement to citation indexing, since it is focused on language. Works in rhetoric of science are in general interested in the relationship between language and thinking, how the «structures and options available in language» lead to certain prepared lines of thought or argument (Fahnestock 1999). This has led to more work than anyone knows what to do with on metaphors in science.

There are both sympathies and tensions between the two tools, since rhetorical analysis is also focused on argument and persuasion. The two are alike in their attention to analyzing how scientific papers are effective in drawing and convincing readers. Unlike citation analysis, however, rhetoric is first focused on individual instances of speech or writing. The choice of paper on which to focus the tools of rhetorical analysis is not itself a subject of rhetorical analysis—it is usually pre-determined by other indicators like a paper's already established fame, its centrality in pedagogy, or its exemplariness of some other feature. It is the work of rhetorical analysis to identify general structures of persuasive speech in specific instances. While one can then track patterns of usage across many instances of scientific language, there is no necessary relation between papers. Can one conduct a rhetorical analysis not of a single paper, but of The Literature more generally? What persuasive or argumentative features obtain at the level of a mass of publications, rather than at the level of a single one?

Less formalized than rhetoric of science as technique or method in the history of science and science and technology studies is the practice of close reading. Published literature is extensively treated as «primary source» material alongside archival material and read with close attention, the form of that attention depending on the historian's interest. The grounds for which literature is included in the historian's catchment area for primary documents are often deter-

mined by assumed categories of relevance – reading everything a certain author wrote or laboratory produced, reading the papers commonly agreed upon as influential for further developments in the field – canonical works or «citation classics», reading along genealogies of pedagogy or mentorship between authors, reading landmark reference works or textbooks that were authoritative sources for large numbers of readers, just to name a few strategies which go more or less articulated in science historians' reflections on their own methods.

Close reading, like the rhetorical analysis of science, chooses texts for reasons that do not reflect their inclusion in a literature—they are important because they are written by a particular author, or are cited most frequently, or are generally included in collections of «key papers.» One cannot «close read» ten thousand papers, to be sure, but can one pay a similar kind of close attention to The Literature as such? It is to this question that we turn in the final section.

#### 4. THE LITERATURE AS INFORMANT

Thus we have three impulses feeding into an approach to the Literature, but their interrelation is not self-evident: (1) a commitment to the necessity and productivity of doing highly specific empirical work on the general – working at the level of the mass of publication as well as the individual instance, (2) a revived interest in the tools of citation analysis and more recent quantitative tools that give us interesting access to the collective representations of large numbers of scientific contributors, and (3) the insights of rhetoric of science or close reading, which give us tools of close attention to language and persuasion. This final section of the paper is dedicated to a re-imagination of the space between quantification and interpretation, at the intersection of the micro and the macro with large volumes of scientific publications, an approach we think of as The Literature as Informant.

One way to understand this thing, the Literature, is by analogy with the work of ethnographers. In much of socio-cultural anthropology, the central strategy of ethnography is to identify one or a few key informants; these informants are particularly valuable not because one observes what they do, or who they are, but because they themselves have a map of the culture, and an understanding of social norms, rules and processes which the ethnographer can use as a template through which to observe and probe the social reality around him or her. Key informants are essentially really good (native) anthropologists, and this accounts in part for why many in the field now refer to them as consultants or collaborators.

Typically, in the anthropology or sociology of science, one would treat the scientist or engineer as the informant for a given field of scientific or technical practice. But what we are proposing is to do the reverse: treat *the Literature as informant*. This deliberate anthropomorphization is intended to upend the assumption of people as the privileged carriers of culture and the authoritative

maps of that culture. Here we suggest instead starting with the *body* of the Literature that through analysis can lead to the scientists, not the scientists to the literature. Often in ethnographic or historical work there is an assumption that interviews and archives are somehow closer to reality than published papers, and it is therefore more authentic, or at least more recognizable as *research*, to have accessed the experience of the person, or documents not processed for public consumption. Published papers are, as they say, *secondary* sources, except for those select key ones singled out for primary status and close, careful reading. In our proposal, the body of published Literature is the primary informant and the interviews and archival material are secondary instances of the individual processing of that literature.

From our own experience, we see (at least) three specific modes of ethnographic reading made possible by this imagination of the Literature as the work to be engaged. All of these modes are possible with single texts, but take on different dimensions when done at scale and informed by the quantitative tools of working at the large scale: (1) following narratives of material action, (2) attention to emplotment, and (3) attention to problematizations.

*Narratives of material action.* With citation indexing (and data mining more generally) on the one hand and rhetoric of science on the other, one is stuck between a mode of understanding the Literature either as a dynamic, networked and quantifiable mass phenomenon or as a disconnected but patterned welter of statements; between a materialist method in which only the quantifiable citations or links matter and an idealist method in which it doesn't matter what scientists are actually doing. Both methods are focused on *argumentation*. Or to put in the language of the philosophy of science, both are methods focused on the *context of justification* at the expense of the *context of discovery* (Fagan 2007). The Literature, however, might actually be more important as part of the latter than the former. This is most visible in the ways materials and methods sections become the medium of both discovery and justification. This is especially true of the life sciences, as Fagan demonstrates, when the elaborate setting out of the methods in all their glorious and seemingly impenetrable detail is both a description of the experiment and an implicit stating of the argument. A particular configuration of tools, materials, living organisms and inscriptions itself becomes a kind of *statement* in response to other statements of similar form throughout The Literature. One can think of it on analogy with citation practices: scientists «cite» not only other papers, but specific methods and materials for accomplishing an experiment or staking a claim about life. Thus, if one is interested in the intrinsic qualities not necessarily of any one paper, but in the intrinsic qualities of the conversation happening among them, then it is necessary to observe The Literature not as a debate but as a *depiction*.

In the case of cell fusion, for instance, the Literature can be productively read for how it depicts human action on matter. For all the attention in history of science on the importance of material practices, there remains the inescapable fact that much practice is described in textual form – far from being a paradox, or a

difficulty, it means that one can *read for practice*; one can even imagine an intellectual history of practice, as practices and their assumptions contain concepts of how humans should act upon the natural world<sup>3</sup>. Quantitative tools can be used to map the movement of techniques carried out under the sign of hybridization, with citational patterns compared to other modes of connection – word usage, co-authorship or co-publication, geographical location, model organism, material specimen or experimental system; these tools can also track shifts in scientific attention and production. As discussed above, representation of publication subsets as trees rather than citational lineages can show those things that appeared and disappeared, allowing one a sense of what perished as well as what led to other developments. These maps can then produce subsets of literature in which the materials and methods sections can be closely read, not necessarily for the content of any one paper, but for the conversation happening between them. It is important to reiterate that it is the concept of working on the Literature as an entity with its own dynamics that go well beyond simple citational connection. The methods and materials sections of these papers depict patterns of practice that develop and move at a scale beyond individual action, but do not necessarily map onto citation practices. Quantitative methods used in a comparative manner, are only the beginning, but they are still essential because they allow the comprehension of movement at scale.

A similar case can be made in the story of computer science, where the Literature depicts new configurations of logic and matter; it gives form to logical abstractions both in new modes of representation (state diagrams or cellular automata, for instance) and in new modes of expression (algorithms programmed in software)-all of which are borrowed or «cited» in a similar fashion. This act of depiction, it is worth repeating, is never in just one paper, but in the conversation among many. In the case of computer science, quantitative methods can identify patterns in the choice of formalisms, the borrowing of terminology and in some cases, the type of software and operating systems forming the basis of the work, but interpretation is necessary to read for the conversation, as these things are very rarely explicit in the papers themselves-and often it is the surprising forms of borrowing that are either not clear (or not systematic) simply in the citational patterns which are the most interesting.

*Reading for the plot.* Historical or interpretative approaches developed for scientific literatures of the nineteenth century and before do not necessarily become defunct with the rise of the journal article, although it may become harder and harder for the humanistically-inclined scholar to see the nine-page article «The Histone H3 Lysine 27-Specific Demethylase Jmjd3 is Required for Neural Commitment» or «A heuristic for the Stacker Crane Problem on trees which is almost surely exact» as *literary* in the same way as the still eminently readable texts of Thomas Huxley or the writings of Pascal.

<sup>3</sup> For the phrase «intellectual history of practice» we are indebted to Charis Thompson.

However, once we re-imagine the Literature as an entity alive with internal conversation, with cross-referencing and foreshadowing, with patterns of attention and elision, can some of the approaches developed with books be productively transposed and transformed for use with this differently scaled corpus? If one takes the collective expression, the mass of literature as «the work» to be analyzed, we might better appreciate a multitude of papers as a work of scientific observation, practice and theory analogous to the *Origin of Species*, i.e. a synthetic work with particular emplotment and narrative structures visible in it. Clearly this deliberate anthropomorphization cannot be approached with the concept of the individual «author». The Literature in science begins to look more like a Hollywood studio film, which can nonetheless be read with the tools of literary analysis.

Gillian Beer, in her groundbreaking work on Darwin, nineteenth century evolutionary narrative in fiction, and science and fiction more generally has argued that «fundamental narrative hopes and fears may be renewed and reshaped by the implications of current scientific theories», and urges readers to look for them in literary form, not explicit content: «these reinterpretations of significant story may be expressed in the ordering of narrative rather than declaring themselves in the language of the text (Beer 1996:185)»<sup>4</sup>. In fact, she says, they may be figured in *disarrangements* of forms of storytelling. This insight goes for scientific literature as much as for fiction. Reading across a large number of journal articles for emplotment can give one a better sense of the emergence and disappearance of disciplines, styles of reasoning or collective projects related to national goals or the commercialization of science.

*Attending to problematization.* All research articles pose themselves as contributions to problems larger than the article itself. The definition, implicit or explicit of the problem to which each article is conceived as a (partial) solution is perhaps least visible in the citations, since the more accepted certain theoretical paradigms or modes of work become, the less necessary it is to cite specific cases. Problematization can also be an interpretive act on the part of the analyst: looking for ways in which articles array themselves around a particular problem to be solved in the future, as well as looking for ways that articles reinterpret past work as a resource for new problems. This is particularly clear in the case of the development of computer science in the 1950s, when many of the most important articles pitched themselves as solutions to one kind of problem, but were picked up subsequently as resources for very different problems. McCulloch and Pitts famous article «A Logical Calculus of Ideas Immanent in Nervous Activity» is a most illustrative case—explicitly posed as an attempt to model the structure of the human brain, it was re-oriented by Von Neuman for the construction of a computer, by Stephen Kleene for the creation of an algebra of recursive

<sup>4</sup> Gillian Beer, «Translation or Transformation?» in *Open Fields: Science in Cultural Encounter*, p. 185.

functions, by the field of «Neural Nets» as a technique for solving problems, and by cybernetics as an emblem of interdisciplinary creativity. This small scale example can be recapitulated at the large scale, as in the case of cell fusion, where the characterization of this era of cell fusion as a profound rethinking of hybridity is a problematization posed by the analyst to make sense of a body of research otherwise seen by contemporary scientists as merely prologue to genetic engineering or monoclonal antibodies. Reading for problematizations can be understood, perhaps, as a contribution to historical epistemology in an era when reading practices, and hence definitions of understanding, must confront the massive amount of literature being produced in the 20<sup>th</sup> century.

## 5. CONCLUSION

These three modes of approaching The Literature could be supplemented by many others, derived perhaps from questions already clear in the history of science. Such an approach can equally be of value in terms of how to find and approach scientists for interviews. Scientists today live inside The Literature. It is their milieu, their culture, and it is what in many ways gives meaning to the details of what they do. Some scientists have partial and narrow understandings of The Literature, and are primarily concerned with their small part of it. Others have a more synoptic view. Asking scientists for a map of The Literature might be seen as one kind of interpretation or reading of the literature as a work. Especially for more contemporary work, «reading» may be more a practice of checking in with PubCrawler during morning coffee and browsing the abstracts brought up by a certain set of pre-decided keywords<sup>5</sup>.

Returning to scientists or engineers with multiple readings of the Literature in hand can also provide a good ethnographic starting point for a conversation, a probe of the way scientists today see the literature of the past or present, and a way to challenge them on their own interpretations of how their work is emplotted within changing fields and disciplines, or how it relates to different problematizations. To take yet another example, recent ethnographic and science studies work on the fields of nanotechnology and synthetic biology could benefit from careful analysis of the literature in these fields and its reconfigurations in response to funding, new challenges or threats, discourses of risk and responsibility, or demands for solutions to new problems from clean energy to pharmaceuticals.

At one level what we propose here is simply a demand for a historical, or historicizing version of quantitative research which goes beyond simple models

---

<sup>5</sup> PubCrawler is a free online service that alerts subscribers to new database entries to PubMed and GenBank that match their research interests. A subscriber enters a set of research interests upon signing up, and gets a daily message about new additions to these databases that match them. There are many similar automated «alerting» services now available for different databases, news sources, and disciplines.



of growth or transition. The Literature contains (necessarily partial) accounts of what people did and how they did it which are repeatedly re-interpreted by scientists, engineers and others who observe it. The Literature can be mined not just for data, but for narratives of material action, taking the whole mass of The Literature as «the work» to be analyzed. Such an approach necessarily demands interpretive skills: one must delimit the literature in a meaningful way (the field of cell fusion cannot be defined only by those papers which claim to be about cell fusion, and yet it cannot include everything); one must understand the difference in concepts, not only in terminology (i.e. cellular automata and cell fusion have nothing to do with each, or if they do, it must be something more than the terminology in use); one must understand not just biographical and institutional contexts for The Literature, but conceptual and technical contexts as well.

In short, there is interpretive work to be done on the *mass* not only on the individual level. We have no *Origin of Species* for the 20<sup>th</sup> century, but with the insight of citation indexing and a feeling for interpretation, one can better appreciate a multitude of papers in The Literature as an analogous work of biological observation, practice and theory.

The work of analysis in the two stories that open this paper is impossible without the insights of citation indexing because it demands an approach to the literature as a work of collective representation. Similarly, rhetorical analysis is useful in understanding some aspects of the articulation of this new hybridity such as its relationship to more traditional sexual breeding. However, an anthropology of science, or a history of science, that aims beyond the case study demands a different relationship to the literature than these two tools provide by themselves. An ethnographic sensibility for the literature allows interpretation back into the space opened by citation indexing, and adds an awareness of material practice to analysis afforded by rhetoric of science. It gives new meaning to the idea of sitting down with the literature, situating individual scientists as secondary sources as, alongside the ethnographer, they too process the mass of literature that forms their intellectual and practical environment.

## REFERENCES

- BEER, Gillian (2000): *Darwin's Plots: Evolutionary Narrative in Darwin, George Eliot, and Nineteenth-Century Fiction*, 2<sup>nd</sup> Ed. Cambridge, UK: Cambridge University Press.
- BEER, Gillian (1996): *Open Fields: Science in Cultural Encounter*. Oxford [England]: Clarendon Press.
- BLAIR, Ann (2004): «An Early Modernist's Perspective». *Isis* 95:420-430.
- BORGMAN, Christine L., and Jonathan Furner (2002): «Scholarly communication and bibliometrics». *Annual Review of Information Science and Technology* 36:2-72.
- BUSH, Vannevar (1945): «As We May Think». *The Atlantic*, 176(1): 101-108.
- CAMERON, Brian (2005): «Trends in the Usage of ISI Bibliometric Data: Uses, Abuses, and Implications». *Portal: Libraries and the Academy* 5:105-125.

- COJA-OGHLAN, Amin, Sven O. KRUMKE, and Till NIERHOFF (2006): «A heuristic for the Stacker Crane Problem on trees which is almost surely exact». *Journal of Algorithms* 61:1-19.
- COLLIER, Stephen J., Andrew Lakoff, and Paul Rabinow (2004): «Biosecurity: Towards an Anthropology of the Contemporary». *Anthropology Today* 20:3-7.
- CREAGER, Angela, Elizabeth Lunbeck, and M. Norton Wise (2007): *Science Without Laws: Model Systems, Cases, Exemplary Narratives*. Princeton: Princeton University Press
- DASTON, Lorraine (2004): «Taking Note(s)». *Isis* 95:443-448.
- DE SANTA, F., M. TOTARO, E. PROSPERINI, S. NOTARBARTOLO, G. TESTA, G. NATOLI (2007): «The History H3 Lysine-27 Demethylase Jmjd3 Links Inflammation to Inhibition of Polycomb-Mediated Gene Silencing». *Cell* 130: 1083-1094.
- FAGAN, Melinda B. (2007): «The search for the hematopoietic stem cell: social interaction and epistemic success in immunology». *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 38:217-237.
- FAHNESTOCK, Jeanne (1999): *Rhetorical Figures in Science*. New York: Oxford University Press.
- GARFIELD, Eugene (1979): *Citation Indexing-Its Theory and Application in Science, Technology, and Humanities*. New York: Wiley.
- JOHNS, Adrian (2000): *The Nature of the Book: Print and Knowledge in the Making*. 1st ed. University Of Chicago Press.
- KELTY, Christopher. «Logical Instruments: Regular Expressions, Artificial Intelligence, and Thinking about thinking». in *Towards an Archaeology of Artificial Intelligence*, edited by Stefano Franchi and Francesco Bianchini. En prensa.
- KELTY, Christopher, and Hannah Landecker (2004): «A Theory of Animation: Cells, L-systems, and Film». *Grey Room*:30-63.
- LANDECKER, Hannah (2007): *Culturing Life: How Cells Became Technologies*. Cambridge, MA: Harvard University Press.
- LATOUR, Bruno (1988): *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- MACKENZIE, Donald A (2001): *Mechanizing Proof: Computing, Risk, and Trust*. Cambridge, MA: MIT Press.
- MAHONEY, Michael (2002): «Software as Science – Science as Software.» in *History of Computing: Software Issues*. Springer Verlag.
- MCCULLOCH, Warren S. and Walter Pitts (1943): «A logical calculus of the idea immanent in nervous activity». *Bulletin of Mathematical Biology* 5: 115-133.
- MORETTI, Franco (2005): *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- RABINOW, Paul (2003): *Anthropos Today: Reflections on Modern Equipment*. Princeton: Princeton University Press.
- SECOND, James A. (2000): *Victorian Sensation: The Extraordinary Publication, Reception, and Secret Authorship of Vestiges of the Natural History of Creation*. Chicago, Ill: University of Chicago Press.
- TOPHAM, Jonathan R. (2004): «A View from the Industrial Age». *Isis* 95:431-442.
- WISE, M. Norton, ed. (2004): *Growing Explanations: Historical Perspectives on Recent Science*. Duke University Press.

## RESUMEN

El artículo explora las dificultades que, para la historia de la ciencia y la tecnología del siglo XX, supone la lectura de un gran número de artículos científicos. Se argumenta que hay tensión entre dos estrategias distintas: la lectura en detalle de los artículos clave y el análisis de citaciones y otros métodos automáticos de rastrear la literatura. El desafío que supone alternar estos dos métodos y la necesidad de mantener una aproximación interpretativa cuando nos enfrentamos a una extensa literatura (miles o decenas de miles de artículos) pueden ser abordados aproximándonos a la literatura científica como si de un informante se tratara, en el sentido antropológico. Caracterizamos a la literatura científica como capaz de describir el cambio en los conceptos y en las prácticas científicas y técnicas. El artículo explora esta estrategia y propone centrarse en asuntos tales como los materiales y métodos (y las prácticas que conllevan), así como en los problemas a los que éstos contribuyen o intentan solucionar.

## PALABRAS CLAVE

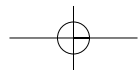
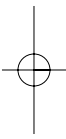
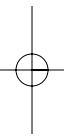
Historia, ciencia del siglo XX, metodología, análisis de citación, retórica científica, interpretación.

## ABSTRACT

This paper explores the difficulties of reading large numbers of scientific papers when doing the history of 20th century science and technology. It argues that there is a tension between two modes: that of close reading of key papers and that of citation analysis and other automated modes of mapping the literature. The challenge charting between the two and of maintaining an interpretive approach when confronted with a large body of literature (thousands, or tens of thousands of papers) can be met by approaching The Literature as a kind of informant in the anthropological sense. We characterize the literature here as something which can depict the movement of scientific and technical concepts and practices. The paper explores this approach, and suggests ways to track such things as materials and methods sections, the mode of employment they employ and the problematizations they propose or participate in.

## KEYWORDS

History, 20th century science, methodology, citation analysis, rhetoric of science, interpretation.



# TEXTO CLÁSICO

