Seminar London 13 May 2000

Tito Orlandi

Ideas for a Theoretical Foundation of Humanities Computing

There was a time when people like myself was asked to explain the astonishing fact that number crunching computers could be useful also for textual devoted disciplines. This period in the history of the application of computers in the humanities is fortunately over. But my life became even worse: I was desperately trying to sell my colleagues the idea that there is more, much more, in humanities computing than collecting words in a text or calculating statistics in archaeological evidence, all without much effect. In recent times (but mainly in anglo-saxon environment!) attention has been increasingly brought on the essence of Humanities Computing, investigated from the point of view of teaching [1]and from the point of view of institutional organization [2]

The theoretical foundation of Humanities Computing may be discussed, and is in fact discussed, from three different points of view, which unfortunately are often mixed together, generating some confusion. One of them is the teaching methodology, which investigates whether it is convenient to teach some part of computer science to humanities student of all flavours, as a foundation of computer applications in the humanities. The second one is the academic organization point of view, which investigates whether it is convenient to institutionalise individual departments for the research and teaching of a special branch of computer science dedicated to the applications in the humanities.

In the present contribution I shall rather deal with the third, purely theoretical point of view, which investigates the epistemological and procedural foundation of humanities computing, thereby trying to establish whether in fact there exists such sort of foundation, and consequently there is a discipline which may be called “humanities computing”.

But before starting my reasoning, I shall recall some assertions which have been done in the two seminars mentioned above, in some discussions in the well known internet list called *Humanist*, and elsewhere, because they form the basis from which I shall try to add new elements, and in any case to show the solution of new problems.

---

[1] E.g. in the important conference, on “Is humanities computing an academic discipline?”, held under the auspices of the Institute for Advanced Technology in the Humanities (IATH), at the University of Virginia, Guy Fawkes Day 1999, cf. url: `http://www.iath.virginia.edu/hcs/.`

[2] E.g. in “a gathering of prominent individuals in the fields of computing and communications science, and arts and humanities research, sponsored by The Computer Science and Telecommunications Board (CSTB) of the National Research Council, in an attempt to explore the complexities of cross-disciplinary collaboration” = American Council of Learned Societies, Occasional Paper No. 41: *Computing and the Humanities*”, cf. url: `http://www.acls.org/op41-toc.htm`.

1. W. McCarty, Poem and algorithm: humanities computing in the life and place of the mind [3]

Creative expression and mechanical analysis

> I begin with the broad question of what happens to our understanding of the arts and humanities when we study them with the aid of a computer. I ask, what is the relationship between creative expression and mechanical analysis? What scholarly role can the algorithmic machine play in the life of the mind as practicing scholars live it, and how might this role best be carried out? [...] The effects of computing may easily be overemphasized, and often are, but we have good reason to suspect that fundamental changes are afoot.

Just a tool

> One still hears otherwise intelligent colleagues refer to the computer as "just a tool" or "simply a bunch of techniques", as if ways of knowing did not have much to do with what is known. [...] ...because the computer is a meta-instrument -- a means of constructing virtual instruments or models of knowing -- we need to understand the effects of modeling on the work we do as humanists.

Mediation of thought by the machine

> ... from the beginning of our field it has been quite clear that humanities computing is centred on the mediation of thought by the machine and the implications and consequences of this mediation for scholarship. As I suggested earlier, we need reminding, and are reminded by the cultural sea-change of which the computer is a most prominent manifestation, that our older scholarly technologies, such as alphabetic writing, the codex and printing, are technologies, and that they also shape our thinking.

2. Computing and the Humanities: Summary of a Roundtable Meeting [4]

(Anonymous, Introduction)

Shared intellectual interests

> Recent progress in information technologies in the humanities and computer science also encourages more programmatic collaboration. Going into the March (1997) meeting, it was understood that effective partnerships among computing and humanities professionals call for an understanding of the opportunities in both sets of disciplines: what are shared intellectual interests; which problems in one domain provide challenges for the other; where can collaborations prove particularly fruitful. That understanding is uneven at best today.

Workable implementation

> ... the difficulty of moving from a theoretical vision to workable systems design and implementation. A first step is to understand the "humanities process," analogous to various business processes.

(McCarty)

---

[3] Keynote speech for: “HumanITies”, *Information technology in the arts and humanities: Present applications and future perspectives*, The Open University Milton Keynes 10 October 1998.

[4] Cp. above, note 2.

Methodologies

> What jumps immediately into focus after five years of teaching humanities computing to graduate students at Toronto, and now undergraduates and postgraduates in London, is the importance of methodologies. When you teach humanities computing what immediately becomes obvious is that the only subject you have to talk about is the methodology.

Computing and the humanities not separated

> Although the roundtable was labeled "computing *and* the humanities" in recognition of the concerns McCarty raised, he cautioned against inferring from that wording that computing and the humanities are fundamentally separate. That inference, he explained, is "an illusion caused by a lack of historical perspective and perpetuated in the discipline-based structure of our institutions."

W. McCarty, "We would know how we know what we know" [5]

Philosophical training

> In the broad sense, philosophical questions naturally arise out of a machine that mediates knowledge and whose modelling of cognition reflects back on the question of how we know what we know. Philosophical training would seem a *sine qua non* because of its disciplined and systematic focus on logic and critical thinking skills, as well as a concern (as old as Plato, if not Parmenides) with how to interpret diverse representations of knowledge, including what philosophers and literary critics jointly refer to as hermeneutics.

Humanist Discussion Group, Vol. 12. [6]

Date: Tue, 19 Jan 1999 21:51:57 +0000 From: Willard McCarty <willard.mccarty@kcl.ac.uk

Computing not purely utilitarian

> The danger with the approach is, however, in the tendency to assume that computing mimics what we already do, that it is purely utilitarian. The falsity of this assumption would be proven if the sociological project were thoughtlessly undertaken, software then written and put out into the field, but it seems to me we can save much grief by prior thought about the questions we'd want to ask.

From: Francois Lachance <lachance@chass.utoronto.ca>

> I am wondering if the search for primitives should not begin with the language that contstructs the object of study.

Concept formation

From: Jim Marchand <marchand@UX1.CSO.UIUC.EDU

---

[5] In: *The Transformation of Science: Research between Printed Information and the Challenges of Electronic Networks*. Max Planck Gesellschaft, Schloss Elmau, 31 May -- 2 June 1999 = url: `http://ilex.cc.kcl.ac.uk/wlm/essays/know/`.

[6] Centre for Computing in the Humanities, King's College London Cp. url: `http://www.princeton.edu/~mccarty/humanist/`.

> If we wish to talk about first things in the humanistic endeavor, we might want to move the previous question as it were and start with concept formation, especially in this age of the computer. I have always maintained that it was I who first said: `The problem is not that computers may come to think like human beings, but rather that human beings may come to think like computers.' First said, who cares, but this is a chiasm of which we see the truth more and more every day. The Aristotelian (yes/no, exclusive either-or, etc.) frisst verheerend um sich, as my old professor used to say, every day. Most of the concepts of the humanities are not, however, Aristotelian and are not amenable to Aristotelian operations; at best they are of the nature of ideal types, and many are stippled spectrum, more-so/less-so, even porous.

The labour-saving myth

From: Willard McCarty <willard.mccarty@kcl.ac.uk>

> Nevertheless, when speaking to audiences of students and colleagues about computing, I still spend a few moments debunking the labour-saving myth. We know this myth to be silly; we know that only the dull, unimaginative scholar would not be inclined to do a better job with the time liberated from mechanical chores -- and get in a little gardening too? We also know that the computer does not so much save labour as change the nature as well as scope of what we labour at. But the myth persists and provides an easy excuse for policy-makers wanting to reconfigure the academic life along the lines of a factory model. It clutters up our colleagues' minds, making our job harder when we try to persuade them that humanities computing is a scholarly activity with tendencies to change the nature of that activity. So -- if I am not being too harsh -- it would seem to me that debunking the myth of labour-saving is a rather important task.

Research methods

> We must objectify our research methods before we can compute the artefacts we study, and in so doing we bring out into the open what has formerly been hidden from view. Hence as a direct consequence of humanities computing new areas for historians and philosophers to study are opening up. Am I wrong about this?
>
> Part of the problem has been the attitude in the humanities by which the physical bits and craftsmanship of research, its technology, are relegated to a lesser status.

New epistemology

From: "Prof. Roly Sussex" <sussex@lingua.arts.uq.edu>

> Willard is quite right. Especially in the "high" humanities the idea of "methodology" seems marginal:
>
> What is interesting about computational methods in language research - sorry, ONE of the interesting things - is that these methods are providing us with both a new methodology and a new epistemology. The notion of "data" is undergoing a reworking. Humanists are learning to interpret statistical reports on what our software says the text is doing. This whole process is tending to bring some areas of the Humanities closer to questions of methodology in other disciplines, and indeed to make the Humanities more scientific.

All the statements that we have quoted above are, in my opinon, not only right, but also exhaustive, when they describe the actual state of humanities computing as a discipline and as a research field: the misunderstandings, the problems, the resistences from a number of colleagues

and authorities. It is specially noteworthy the common idea the computers are introducing a radical change in the mental and methodological attitudes of humanities scholars. The survey shows that the best scholars in humanities computing share the appreciation of the theoretical frame inside which the problems should be collocated.

On the other hand, the very variety of expressions used to define the problems points to the lack of a firm structural organization of thought. I feel a sense of inadequateness, even disorder, in the overall change as presented by the same scholars. In fact, when they proceed to propose a definition of humanities computing, they tend to consider the products of computation, be they hardware (the Net) or software (applications like concordance programs or statistical packages), rather than the first principles of computing. This happens, I believe, because they do not want to discuss the fundamental question of *what is computing*, for fear that it raises sterile polemics and does not help in solving more pressing problems of immediate practical significance. On the contrary, it is my opinion that before defining humanities computing, we must propose a coherent definition of computing and also of humanities.

I do not want here to discuss the definition of humanities: I simply assume that we are dealing here with the group of disciplines commonly called humanities, or "arts and humanities", as different from the social sciences. In my opinion the difference consists in the historical approach, which is peculiar of the humanities, and not of the social sciences-as the Appendix to the Roundtable Meeting already quoted rightly states: [7]

> While the arts connote the arena of active creation or performance of works of human expression, the humanities comprise those areas of intellectual investigation that focus on the preservation, transmission, and interpretation of the human record. Their cousins, the social sciences, focus on the analysis of social forces and the modeling of laws underlying the formation of society and of the individual's relationship to society. The social sciences—economics, political science, sociology, psychology—rely more on observation of current social forces and analysis of data than on historical document and cultural object.
>
> One might broadly categorize the humanities as follows:
>
> the history and analysis of society and culture—its currents, events, and forms as measured and observed through objects and documents. Specialization is broadly by the type of activity studied or the means of analysis (history of law, labor, politics, science, economics), period (archaeology, various period studies, e.g. Medieval Studies, 18th-century studies), or geographical area (by country or other area, e.g. English History; Oriental Studies). Methodologies vary: narrative and thematic analysis, comparative studies, detailed data analysis.
>
> history and analysis of forms of thought and creative expression: philosophy; spoken and written language (literature and linguistics, drama); visual forms (visual art and architecture, cinema, theater); music; and dance. Approaches vary: biographical narrative, detailed internal and comparative analysis of works, broader studies of aesthetics and their historical development.

Coming to the problem of the definition of computing, I shall refer to the first sentence in our programme: "This one-day colloquium begins with the question of how we might best conceptualise the application of computing to the humanities", and I begin by noting that the application of computing is not the same as the application of computers. The computer as I see

---

[7] Cf. Computing and the Humanities: Summary of a Roundtable Meeting (cp. note 2), Appendix 2: *What are the Humanities*?

it, is not the type of the machine, of which the tokens are in front of us, on our desks or laps or palms, but the set of devices (not one device!) described by von Neumann, as the realization (we add) of the Turing universal machine, along the lines of the construction of the ENIAC, EDVAC, and the Mark I. [8]

The Turing machine is central in my approach to the problem of (humanities) computing, because it is the abstract, logic (I prefer to avoid the term "mathematical") model underlying every realization of a computer. Only an abstract, logic model can clarify the methodological problems raised by the meeting of the humanities with the computers. In other words, I am separating the concept of a "normal" machine, like the book or the typewriter or the calculator, from that of the universal machine, of the "automaton" *per se*. [9]

Such a view may of course be disputed, but if it is accepted, the next step is to realize that the computer may be used in two different ways: (1) to simulate the behaviour of another machine, because the computer can simulate any possible machine; (2) in its full capacity of computing machine, that is, for the peculiarity which distinguishes the computer from all other machines, which consists in the possibility to do "computation" as developed in the theory of recursive functions. [10]

The distinction is important, because it helps to establish why the application of computers raises methodological problems, and to what extent it does so. Because it seems evident that when the computer is applied in the humanities only so far as it simulates (does the work of) a "traditional" machine, then no new methodological problems arise, because there is no substantial difference from the traditional procedures, if not of speed and convenience.

On the contrary, when the computer is applied in its full capacity of running algorithms, humanities are confronted with a radically new situation, for which there is no commonly recognized methodology. Something new happened in the field of epistemology when A. Turing proposed his famous paper, [11] because after it some of the rules which help to build our knowledge were changed in a basic way.

One of the observations often found in the quotations at the beginning of the present paper — and more or less a common opinion among the *cognoscenti* —refers to the fact that the use of computers may require (or sometimes produces) a change in our minds. I would say that the

---

[8] Rolf Herken (ed.), The Universal Turing Machine. A Half-Century Survey. II ed., Wien-New York, Springer, 1994. p. 135: Martin Davis, Mathematical Logic and the Origin of Modern Computers. - First Draft of a Report on the EDVAC 8.1. By John von Neumann, Contract No. W-670-ORD-4926. Between the United States Army Ordnance Department and the University of Pennsylvania. Moore School of Electrical Engineering, University of Pennsylvania. June 30, 1945. It can be read in B. Randell (ed.) The Origins of Digital Computers. Selected Papers, Berlin,Springer-Verlag, 1982.

[9] Cf. J.M. Brady, *The Theory of Computer Science*, London-New York, Chapman and Hall, 1977, p. 19-20; even better E.V. Krishnamurthy, *Introductory Theory of Computer Science*, London, MacMillan, 1983, p. 18-83.

[10] Hartley Rogers, Jr., Theory of Recursive Functions and Effective Computability, Cambridge, The MIT Press, 1987 - Hans Hermes, Enumerability, Decidability, Computability. An Introduction to the Theory of Recursive Functions, Berlin-Heidelberg-New York, Springer-Verlag, 1969.

[11] On Computable Numbers with an Application to the Entscheidungsproblem, P. Lond. Math. Soc. 42 (1936-7) 230-267.

Turing machine is in fact a way of thinking, the formal way of thinking, which might have remained restricted to the discipline of mathematics, had it not given birth, as a by-product, to the computers. Although some of the elements of the new methodology were present in many disciplines before the advent of computers, the systematic use of the Turing "scheme" is fundamentally altering in part all humanities disciplines.

I need not recall here the theory underlying the Turing machine, but I want to stress (1) the association of that theory with the theory of recursive functions (cp. H. Rogers, quoted above) and of formal languages; [12] (2) the fact that the mathematical model of which the Turing machine consists, is not only used to build the computers, [13] but dealing with the formalization of algorithms (i.e. methods to solve problems, also humanities problems) it is also fundamental in the automatic treatment of whatever problem. In this sense, the presence of computers is somewhat accidental and even not influent in the foundation of Humanities computing.

Two points seem fundamental for Humanities computing: (1) the formalization of data and procedures; and (2) the representation of data in input and output.

In order to be used in a proper way, that is, in order that it may give good results, or in any case the wanted results, the Turing machine dictates some conditions, and particularly it dictates the formalization of reasoning, and the formalization of data. If we accept this, we should see whether we have a good theory of formalization, and especially one which is valid in the field of the humanities. As often in such instances, everybody has an intuitive idea of what formalization is, but so far as I know there does not exist a consistent and generally accepted theory of it.

According to the tradition which goes from Leibniz through G. Boole to the contemporary logic, the concept of "formal" is not opposed to "substantial" as in Aristotle, or to "intuitive" as in Hilbert, but to "expressed in natural language". The question of the language is central, particularly in the Humanities, which are are not used to some sort of formal language, while the hard sciences after Galileo are used to the mathematical language. Gardin has recognized this in his reflections on Humanities computing.

Without going into details, it is possible to show the essentials of a theory of formalization. It is built around a language, with few syntactic rules, which are formal both in the sense of being precise, and of being in first instance void of contents, but declared at the moment of their use. This permits both the experimental and the deductive procedures. And this is where humanities differ substantially from natural sciences, because such a language is not, or not necessarily, numerical, but what may be called a particular subset of the natural language. [14]

By way of example, "sgml" seems to me a good instance of a formal language humanities oriented. Contrary to what most people think, sgml is not simply a descriptive language of texts, but also of data bases, and ultimately of whatever kind of expression. But a more complex, and

---

[12] Cp. Thomas S. Sudkamp, *Languages and Machines*, Addison-Wesley, 1988.

[13] After the well known report by Von Neumann (above, note 8), and cp. M. Davis (above, note 8).

[14] Tito Orlandi (ed.), *Discipline umanistiche e informatica. Il problema della formalizzazione* (Contributi del Centro linceo interdisciplinare "Beniamino Segre" 96), Roma, Accademia Nazionale dei Lincei, 1997. Jean-Claude Gardin - Marie-Salomé Lagrange- J.-M. Martin, *La logique du plausible. Essais d' épistémologie pratique en sciences humaines*, Paris, Editions de la Maison des Sciences de l' Homme, 1981.

therefore more interesting but at the same time not yet developed language, is the one invoked, or we might better say evoked, by J.-C. Gardin for archaeology, literary criticism, and history. [15]

It is inevitable that logic, of the kind which one may prefer, from the Boolean formulas to the various paradigms of contemporary logic, is the main reference for the construction of those formal languages.

Analagous, but with substantial differences, is the theory of formalized representation of data. Computation requires the use of discrete data, which must be extracted from the continuous flow of the experience in our conscience, as Husserl has taught. Also the problems discussed in the frame of Semiotics must be taken into consideration. Data must be organized in models to represent the dynamic reality from which they are extracted, and therefore they must not be organized in a fixed structural composition. This is why the construction of abstract models for computation must conform to the General Systems Theory, which shows how to build dynamic structures which interact upon each other. [16]

The formalization of data depends on the extension of the recursive theory to non-mathematical entities. I refer expecially to what Hartley Rogers (quoted above, note 10) writes about “coding”:

> The partial recursive functions are mappings from integers to integers, and their algorithms carry us from notations for integers to notations for integers. The original, unrestricted, informal notion of algorithm concerns procedures with more general kinds of symbolic input and output; [omissis] We now ask: is there any way to include such broader non- numerical algorithms within our formal theory? Two approaches can be made to this problem.
>
> The first approach is as follows. Given a class of nonnumerical inputs or outputs, choose some fixed one-one mapping from this class into the integers. Henceforth, for theoretical purposes, identify each symbolic entity in the nonnumerical class with its corresponding integer "label." Such a standard mapping is called a coding, and the integers used as labels are called code numbers. The coding is chosen so that (a) it is itself given by an informal algorithm in the unrestricted sense; and (b) it is reversible; i.e., there exists an informal algorithm (in the unrestricted sense) for recog- nizing code numbers and carrying out the reverse "decoding" mappings from code numbers to nonnumerical entities. Furthermore, it is stipulated that a coding shall be used only when (c) an informal algorithm exists for recognizing the expressions that constitute the uncoded, nonnumerical class.
>
> [omissis]
>
> The second approach to a formal treatment of nonnumerical algorithms is as follows. The formal characterization of para. 1.5 is broadened to include directly, as inputs and outputs, expressions from wider "nonnumerical" classes. The Turing-machine characterization is especially convenient for this purpose. It requires only that the expressions of the wider classes be expressible as finite strings in a fixed finite alphabet of basic symbols (other than B and 1). The basic operations of Turing machines are extended to include printing and erasure of symbols from

---

[15] Cp. Note 14, and: Jean-Claude Gardin, Maria Novella Borghetti, *L’architettura dei testi storiografici*, a cura di Ivo Mattozzi, Bologna, Clueb, 1995. 215 pp.

[16] Husserl, Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, Tübingen, 1980- Lars Skyttner, General Systems Theory. An Introduction (Information System Series), Houndmills, Macmillan Press, 1996. Cp. PCP Principia Cybernetica Project: see `http://pespmc1.vub. ac.be`, on model, knowledge, etc.

this new alphabet. A nonnumerical mapping is then defined to be recursive (or partial recursive) if a Turing machine exists for carrying it out. After the formal characterization is so broadened, Parts I, II, and III of the Basic Result (para. 1.6) can themselves be modified and broadened to apply to this broader concept of recursiveness.

Here what is essential is the realization of units of phenomena, and their expression by means of definite symbols. And here we are led mainly by two theoretical disciplines: phenomenology and semiotics. The first helps us in studying the delicate process through which we pass from the immediate perception of a continuous stream of phenomena in our conscience, to the assumption of individual facts or acts or things which constitute the reality as an objective complex of data. [17]

Semiotics help in building a coherent system of symbols to express the individual objects of reality by means of other objects, which *stay for* them, and preserve, for what is possible, the relations which we pose between the objects of reality. [18] But a simply sequential structure of symbols, as it is suggested by a theory of languages assumed only in its basic meaning, in a sense a naive one, is ultimately not sufficient to cover a reality which, also when the passage from the continuous to the discrete level is granted, reveals itself very complicated. Therefore it is necessary to have recourse for one piece of reality to a number of formal structures, which may form what is called a *model*.

We avoid here too abstract a discussion, and proceed to the description of two "cases in point", which will illustrate the effect of formalization on the methodology of humanities disciplines.

Two cases in point.

1. Archaeology

Though it is not often recalled, archaeology was one of the earliest disciplines in which computers were experimented, [19] and one of the first problems discussed was how to store archaeological data in a magnetic memory. Still common is the idea that this is an intuitive passage, consisting in typing in the computer the names of the objects or some of their features, to be processed by data base management systems, without further preoccupations. Different was from the beginning the position of Jean-Claude Gardin, who, after studying the problems of the "encoding for the analysis" of archaeological material, [20] eventually considered the representation of data (called by him *compilation*) as one of the two essential tasks of archaeology, the other being the *explanation*. Such task is to be accomplished by means of a special kind of language, particularly suited to the scientific description of objects: "What distinguishes a scientific compilation from the prose of a collector is that it resorts to a representational language that is supposed to have higher merits, both theoretically, for the

[17] Husserl, cp. note 16.

[18] Winfried Noth, *Handbook of Semiotics (Advances in Semiotics),* Indianapolis, Indiana University Press, 1990,.

[19] Cp. Especially the series "Codes pour l'analyse..." by J.-C. Gardin, Paris CNRS, 1976-78

[20] Cp. above, note 19.

accumulation of scientific knowledge, and practically, for the handling of information." [21]

François Djindjian partially accepts this statement, but observes that it is valid especially from a data-base perspective. For broader computer analysis he defines encoding as the choice of those features in an artifact that make possible various kinds of data processing: "Les approches décrites... mettent en évidence différents types de codage de la description. ... S'il est sûr que certains codages sont moins bons pour des raisons variées, il faut consdérer qu'il est nécessaire de rechercher des codages adaptés à des morphologies particulières mais aussi à des variabilités de morphologie particulière." [22]

The encoding procedures imply the conscious individuation of discrete elements in a continuous universe (cf. above), and this is precisely the kind of formalization required by the Turing machine; in less abstract words this is recognized by theoriticians of archaeology like Albert Spaulding: "The first concept, the artifact, provides the class of entities with which archaeology is concerned (...) The ability to recognize artifacts implies the ability to recognize those attributes of artifacts which represent patterned human modification, and we are brought to a fundamental collectivity, the set of culturally meaningful attributes bound together by their occurrence on a single object, the artifact." [23]

Choice and (formal) description are broadly accepted as the essence of encoding in view of an automatic analysis, of data, but how this should be done, what are the possible criteria to establish their correctness, it is not commonly recognized. Here is where semiotics helps, or even may be decisive (cf. above). We start from the observation that the distinction between archaeological and non archaeological objects depends on the kind of information which they can convey to us. In this sense the objects assume a communicative function, and become the symbols of a cultural "meaning", of a dynamic reality which cannot be expressed by mere digital symbols, but also by the accompanying competence. This is expressed, in a computing environment, by an interpretive language (therefore a piece of software) which can reconstruct the relations between the objects, matching them with the relations between the symbols.

It seems to me that the process of representation of the archaeological objects by means of digital symbols and their subsequent management by means of computer software may be synthesized as follows: [24]

1) the first step is to recognize the character of *message* (from the past) which the material objects assume when they are considered as archaeological evidence. They acquire a meaning in connection with the culture of the people who have produced them. At this point the objects are somewhat transformed into signals or signs, discrete units which convey information. But it is also to take into consideration the well known semiotic doctrine that signs as such are mental

---

[21] Jean-Claude Gardin, *Archaeological Constructs*, Cambridge University Press-Editions de la Maison des Sciences de l'Homme, Cambridge-Paris, 1980, p. 38.

[22] Francois Djindjian, *Méthodes pour l'archéologie*, Colin, Paris, 1991, p. 99.

[23] A. C. Spaulding, Some Elements of Quantitative Archaeology, in: F. R. Hodson (etc., eds.), Mathematics in the Archaeological and Historical Sciences, Edinburgh 1971, p. 4.

[24] Detailed critical discussion of the problem in T. Orlandi, *Formalizzazione dei dati, semiotica e comunicazione*, in: Archeologia e Calcolatori 7 (1996) 1247-1258,

processes, not material elements (distinction between signs and “sign vehicles”).

2) The second step is to recognize the structure of the objects/signs. This has to be done having in mind two different *competences*: the one of the people who have produced the objects/signs, and the one of the scholar. In fact it is often the case that the objects/signs have meaning independently from the intention of those who have made them, but depending on the aims of investigation from the part of the scholars.

3) The establishment of a code (a set of symbols) which permits the representation the elements of the structure of the objects/signs. The code must be provided itself of a structure which may match the structure of the objects/signs. We are referring here to a sort of scientific language, as those discussed by Gardin. [25]

4) The establishment of a binary code which permits to transfer the representation of point (3) into a computer memory.

5) The construction of a model which is the semiotic correspondent to the structure of the objects/signs and may be submitted to computational processes, in order to investigate the cultural processes which have produced the objects. The model will contain the logic rules (under which we assume also the statistical procedures) as proposed by Gardin. [26]

2. Philology

From the beginning, the editorial activity, like Humanities in general, was conceived as an art rather than a science. It was based more on the competence of the scholar, on his culture and education, than on a set of theoretical principles and rules of applicability. Even when the necessity of such a theory was partly acknowledged, what was proposed was only some abstract recommendations—as in Boeckh: the verification of historical, linguistic, and stylistic congruence of the text—or some practical suggestions—as in Lachmann: how to compare the *lectiones* in the manuscripts, and establish a genealogy of the manuscripts.

The documents in which ancient texts are transmitted—the manuscripts in their various forms—were of course considered as objects to be studied in depth in their historical and material execution, but were always described as a medium to be received by the mind of the scholar, who would critically assess its content and passed it over to the readers. The essence of the documents themselves, and their qualities, were not investigated, as also those of the text, at least until the rise of Semiotics.

When the task of the human editor is in part assigned to an *automaton*, the documents must become part of the automatic process, therefore they must be submitted to formalization, which is a transforation much more complicated and delicate than is generally assumed. In fact, it is generally believed that the transformation of the sequence of graphemes into a sequence of binary units is all that is required. This is far from true. In a manuscript or book the disposition and other aspects of the graphemes (size, color, etc.) have a great importance, because they signal to the reader additional meanings, which the graphemes, taken only as representative of alphabetical letters, do not convey.

---

[25] Cp. above, note 16.

[26] Ibid.

When we want to store a text in a magnetic memory in order that it be managed by an automatic device, and not simply restituted to the attention of a reader, the electronic text must be constituted not only of a sequence of graphemes (i.e., the sequence of symbols representing graphemes), but also of symbols representing the extra-graphematic features of the text itself. The reader acknowledges, more or less consciously, the value of the extra-graphematic features and signs. In a computer, the system of the binary symbols (bytes) must contain all relevant meanings in a completely explicit way.

In the conventional environment (book reading) the text is assumed as a unitarian phenomenon with different features, each of which is taken into consideration in different times, according to the problems actually considered by the reader or scholar: content, language, orthography, layout, are generally indicated as "the text". Such assumption can work only in the conventional environment, but it is not correct in the computational environment, where a text cannot be an object of automatic procedures without the representation of the reader's competence, which is essential to the comprehension of the text. The text exists only as the meeting of the document and the reader.

In a computer environment it is not possible to propose a unifying solution to the problem of building a text, as in printing, but a sort of communicative area in which the separate plansor points of viewand their interconnection be present. It is necessary to formulate a theory and praxis which translates the multiform aspects of the "text" into a binary representation which may be integrally and correctly understood by the *automaton*, and to achive that goal it is also necessary to define a formal definition of *text*. Efforts in this direction have been made after a while, independently from the text digitization problems, inside the Semiotics and Structuralism disciplines. [27] They set the general frame, but unfortunately are too bound to the content of the texts, and their meaning, rather than to aspects like orthography and layout, to be completely satisfactory. In my opinion the best inspirations that we can find today come from the *(General) Systems Theory*, and its pragmatic counterpart, the *theory of models.*

One may recall that the Systems Theory was born from Cybernetics, therefore it is in tune with computing activities and computer applications, though of course it is not necessarily restricted to them. It tries and gives a methodological vision of complex phenomena, formed of elements not exactly definable in a consistent way, which act upon themselves producing different results according to the specific behaviour of such elements..[28]According to this theory, in order to configure the text as a system, it is necessary to distinguish its components, which may themselves be systems, processing input data received from one or more other system, and producing an output, which may be processed by one or more other system.

According to this view, a text cannot be any more represented in a computer memory by means of a "static" file, but something more complex and dynamic is required, which is called a *model*. We shall briefly try [29] to delineate here the scheme of such a model.

---

[27] E.g. cp. C. Segre, *Avviamento all'analisi del testo letterario,* Torino, 1985. – R.-A. de Beaugrande, W. Dressler, *Einfürung in die Textlinguistik,* Tübingen, 1981.

[28] Lars Skyttner, *General Systems Theory. An Introduction*, Houndmills etc., 1996.

[29] For a more detailed, although not at all exhaustive or conclusive exposition, cp. Tito Orlandi, Ripartiamo dai diasistemi in: I nuovi orizzonti della filologia. Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici, Conv. Int. 27-29 maggio 1998, Roma 1999 (Atti dei Convegni lincei, 151), p. 87-101 - TestI, modelli, e

The *model* that we suggest here by way of example, to show how a formalized theory of the text may work, is constituted of the models of the (sub)systems that we recognize in the text. We assume the text to be a particular way, the linguistic way, to convey a message. The first step will be to process the message in order to formulate its content in a linguistic form, and therefore

1) the first model to introduce is the one which represents the process of what may be called the abstract (mental) linguistic *system* (according to the Systems theory, cp. above), which may be defined as the ability to form senteces in a given language (cp. .the Saussurian *compétence*). The system, and the model, process the message by means of a vocabulary and of the grammatical rules, and produce the sentences in their abstract form, that is, before being pronounced or written.

2) The second model represents the process of the *abstract (mental) graphic system*, which processes the abstract sentences in order to produce the abstract graphic representation of the same sentences.The process is done by means of the set of graphemes and by the orthographic rules, which match the "sounds" with the graphemes, and the product is still an abstract one, because of the difference between graphemes and glyphs. [30]

3) The third model represents the process of the *concrete graphic system*, which processes the abstract graphic sentences in order to produce a layout of glyphs in a real (material) writing surface, normally a page of a book or manuscript. The process is done by means of the set of glyphs (fonts of different shape and size; types of writing; etc.), the layout rules or tastes, and the competence of the use of the surface.

In conclusion, I hope to have shown what new methodological concepts and procedures computation is introducing in Humanities, especially for what concerns the formalization of problems and data representation. On the other hand, it is easy to realize (a) that part of the humanities was "computed" well before computers were used, and (b) that even where the computer is used as it were a "common" machine, it imposes some constraints on the form of data, which did not exist before.

The reflection on, and clarification of all these fundamental issues seems both necessary and urgent, as it is, as a consequence, the foundation of an independent scientific discipline, Humanities computing, which studies the problems of formalization and models, crossing all humanities disciplines (linguistic, literature, history, archaeology, history of art, history of music), but which none of them can fully develop by itself.

---

sistemi. I nuovi orizzonti della filologia, in: Ecdotica, critica testuale, editoria scientifica e mezzi informatici elettronici, Conv. Int. 27-29 maggio 1998, Roma 1999 (Atti dei Convegni lincei, 151), p. 87-101.

[30] Cp. e.g. Groupe Francophone des Utilisateurs de Tex (GUTenberg), Cahier 20: *Multilinguisme et codage des caractères ; d'Ascii à Unicode*, Mai 1995, `url:http://www.gutenberg.eu.org/pub/gut/publications/cahiers.html#Cahier20.`