The international debate on humanities computing: education, technology and the primacy of languages[1]

Domenico Fiormonte
Università di Roma Tre, Dipartimento di Italianistica
Via Ostiense 236 - 00146 Roma
Last modiefied: 1/12/2006

Preamble

In this article I will limit myself to sketching out a brief (and by no means impartial) outline of the current debate on Humanities Computing from an Italian and, so to speak, "Southern-European" perspective. Since the discussion involves countless areas (ranging from theoretical research to applications, from education to the digitization of resources, etc.) I have selected, from among the many possible arguments, one which, at the moment seems to be a point of convergence of the practical and theoretical efforts of the discipline: the curriculum. I am convinced that behind this topic an underground river is flowing and nourishing the most delicate questions of our discipline.

The Italian situation and the debate concerning the theoretical status of HC

We have known (we have been saying so) for years, that the new technologies would have change the face of the humanities. In Rome, Tito Orlandi, Giuseppe Gigliozzi and Raul Mordenti began speaking of an 'epistemological turning point' at the beginning of the 1980s and have created a school based on this concept. Equally obvious was that this turning point would be accompanied by the creation of new professional figures, new jobs and new ways of teaching and doing research. A large number of students of the "Roman School" of humanities computing, (including myself) have eked out a living by selling their own skills outside the university: publishing houses, softwarehouses, broadcasting companies, professional training organisations. All of us had became aware of the importance of education (ranging from private educational institutions to the Confindustria[2]), all, that is, except our university. When it arrived, the *new economy* boom began involving those humanities faculties which were most exposed to the market, like Social Sciences and Communication Studies, but it only touched upon those who, for years, had been reflecting on computing from a linguistic, philological or literary point of view.

The "historical" event which struck at this stagnation and brought our skills into play is well-known in Italy: the last university reform launched by the center-left government coalition during 1996-2000. With the creation of the "laurea specialistica" in *Informatica Umanistica* (= IU), the segregation (at times, *self-segregation*) of this discipline came to an end. We now know that the humanities faculties of the main Italian universities, among which Rome "La Sapienza", Florence, Pisa, etc., will soon be activating the two-year course, and many others are ready to follow their example. In the meantime, the "compulsory" nature of credits for computing in the three-year course has brought about the chance for teaching contracts and this, despite all the problems (disorganisation, inadequacy of equipment, proletarisation of the

---

[1] An Italian version of this article was published in 2001 in the paper edition of the journal "Testo & Senso", n. 4 and 5, pp. 145-156 (see http://www.testoesenso.it).

[2] Confindustria = Confedaration of Italian Industries. See the online resource dedicated to the ICT job market created by this organisation: http://www.ict-job.it/.

teaching body) is far from being a negative development. Indeed, we should have the courage to admit that the Italian situation is in many ways an enviable one.

We are among the few European countries in which humanities computing is an "official" subject, provided at a national level and where a wide range of computing courses are present in all humanities curricula. This situation is due, I believe, to three main factors: 1) the rhetorical-humanistic tradition of the Italian school system, where the study of literature – at least up to the reform – has always occupied a role in the curriculum (and in the socio-cultural prestige of the country) which is unique in European systems. The digital revolution is a revolution in the treatment, elaboration and memorisation of data and information and it was inevitable that the linguistic-literary disciplines, centring on the text, would be those most affected by processes of digitization; 2) the fact that the birth of humanities computing coincided with the work of an Italian scholar, that is Roberto Busa SJ; 3) the growth, in Italy too, of the multimedia and distance learning markets. I believe that these factors are behind the spread of computing in the Humanities curricula, where the role of the computer has been recognised, (with varying degrees of awareness) in that it is a "manipulator of signs". Thus, never as today are we as near our goal: the official recognition of humanities computing as an *autonomous* discipline.

Now let us look at the international debate. The question autonomous discipline / curriculum, apart from the objective problem of the academic position of those who do research, constitutes the "battleground" because of the definition of the theoretical status of humanities computing. This has been explained clearly by Dino Buzzetti in a recent post on Idulist, where he distinguishes between two problems: the limited presence of philosophical disciplines in the humanities computing curricula and the alternative between those which Gino Roncaglia defined "Informatiche Umanistiche specifiche" and "Informatiche Umanistiche trasversali" ("specific" and "cross-discipline" Humanities Computer Science). Buzzetti comments:

> As I see it, both problems depend on the point of view taken concerning IU theoretical status*, intended as an autonomous discipline. (...) the difficulties emerging in the formulation of the table* depend on the lack of a specific discipline sector for IU in our system (...) As long as this continues to be the case, the alternative will not be between the "specific" IU and the "transverse" IU, but between the specific disciplines (historical, literary, etc.,) and computing *tout court*. With the result that on the one hand, the humanists will not be able to apply computing methods correctly, and on the other the computer scientists will not be able to suggest adequate methods for application to humanistic disciplines.[3]

In focusing on the question of HC/IU theoretical status, Buzzetti rightly recalls the contribution of ACO*HUM, an initiative concerning monitoring of European humanities computing financed by the EU in the three-year period 1996-99. [4] In the conclusions to the second chapter of the report of this project ("European Studies on formal methods in the Humanities") a tripartite division of course types is suggested:

1. *Humanities Computer Literacy*

2. *Humanities Computing* (which we might define as "Computing applied to humanistic disciplines", since we are referring to courses which make use of "computationally based methods (like database technology) or computationally dependent ones (like statistics) to gain specific results [...]");

---

[3] Messaggi sent to Idulist, http://linux.lettere.unige.it/mailman/listinfo/idulist, 7 May 2001.

[4] K. de Smedt, H. Gardiner, E. Ore, T. Orlandi, H. Short, J. Souillot, W. Vaughan (eds.), *Computing in Humanities Education. A European Perspective*, Bergen: University of Bergen – Socrates/Erasmus Thematic Network Project on Advanced Computing in the Humanities. Online version: http://www.hd.uib.no/AcoHum.

3. *Humanities Computer Science* (an autonomous discipline able to offer substantial contributions to computing – as in the case of computational linguistics, whose instruments are, today, part of the canon of computer science).

In itself this classification does not mean that there is a conflict. The three approaches can coexist inside the same institution. But it is also true that there is a tendency to concentrate on one or the other extremes of the spectrum, according to the type of educational structure and, naturally, the culture of each country. And the prevalence of one or the other cannot fail to reflect a different way of understanding the discipline. I recall that some years ago in Rome, during dinner, Lou Burnard, remarked ironically on the prosodic ambiguity of English in distinguishing the two possible interpretations of the centre he directs: *Humanities-Computing* Unit or Humanities *Computing-Unit*? Needless to say, it is also on this type of terminological ambiguity that the Anglo-Saxon "pragmatism" is based, as far our discipline is concerned.

'Continental' IU vs. Anglo-American HC?

I will return to linguistic questions later, but meanwhile we should note that the ACO*HUM type has been confirmed by two recent events: the CliP 2001 conference, held in Duisburg, in Germany, last December (http://www.uni-duisburg.de/FB3/CliP2001/) and the Malaspina colloquium "The Humanities Computing Curriculum" (http://web.mala.bc.ca/siemensr/HCCurriculum/).

Here we are dealing with very different events and it would not be a bad idea to begin with an historical outline. CLiP ("Computers, Literature and Philology") was created in 1998 and aimed at founding a small seminar *of* and *for* scholars of this discipline. The place and context are important: the School of European Languages and Cultures of the University of Edinburgh. The organiser was the Italian Department, supported by the Spanish and partly by the French. Some leading HC figures were invited to the seminar (Lou Burnard, Willard McCarty and Allen Renear, among others), but what was clear was the willingness to concentrate particularly on the world of Romance languages. And this willingness was confirmed by the places chosen for the seminars which followed: Rome (1999), Alicante (2000) and Duisburg (2001), where it was, once more, the Romance languages' sections which organised the event.

On the other hand, behind the Malaspina conference, is the international associations *milieu*, that is the *Association for Computers and the Humanities* (ACH) and the *Association for Literary and Linguistic Computing* (ALLC). These associations have had an annual forum for some time now, the united ACH/ALLC conference, which takes place in the United States one year and in Europe the next. Both associations have their official publications: *Computers and the Humanities* (CHUM) in the first case, and the historical *Literary and Linguistic Computing* (LLC) in the second (published by Oxford University Press). ACH and ALLC are also behind the *Text Encoding Initiative* (TEI) http://www.tei-c.org/), the international consortium which concerns itself with defining the standards for the SGML encoding of the Humanities tradition. These are prestigious publications and initiatives open to contributions from all countries and not rarely – above all in LLC – printed in non-English languages (i.e. German and French).

I mention this to dispel the suspicion that colleagues at the head of these organisations have not made great efforts in recent years to widen the geographical base of contributions and subscribers. However, a glance at the editorial boards is enough to verify the (legitimate) Anglo-American predominance: out of twenty components of the editorial council of CHUM four are from the United Kingdom and eight from the United States – and the two editors are both American. More European is the orientation of LLC whose editorial board, after the death of Antonio Zampolli, is made up of eighteen people: eight from UK, three from

Germany, one Canada, two from USA, one from Finland, one from Norway, one from Hungary, and one from Australia. As for TEI, things are more complicated.

In the last two years the group has been transformed into a consortium backed by the European Commission, as well as by the American National Endowment for the Humanities. At the moment of writing there are fifty five members and the board of directors is made up of twelve people (of whom five are American and three British). The situation of the editors of the Guidelines is not different. But in this case the balance of forces (or *capital stock*) is not reflected so much (or only) in political organs, but also in the choice of the examples, the real heart of the TEI, as far as arguments and norms are concerned. Let's look at one of the fundamental chapters of the *Guidelines*, n. 18 ("Transcription of Primary Sources"), edited by Matthew Driscoll.

We read of examples on texts by Chaucer, Shakespeare, Moore, D.H. Lawrence, Joyce, Browning, William James, etc., and certain (a few) Latin examples, German, Danish. But there is no analysis on the philological problems of documents of Italian, French or Spanish tradition. We will not speak of other important European languages, like Russian. Honest and universal as might prove the attempt to categorise textual phenomena (Driscoll himself warns us that "these recommendations are not intended to meet every transcriptional circumstance ever likely to be faced by any scholar"), I think I am justified in having doubts about a system which aims at standardisation depart from the cultural and linguistic limits of the editors, as far as their choices are concerned. (By any chance, is there a textual criticism independent of the texts it studies?)

We might object that the 'regional' translations of the TEI (the Italian or the Japanese for instance) incorporate examples in the languages of the translators. But, in my opinion, this does not impair the leading role of the original model. It is no heresy, I believe, to sustain that the specialisation and culture of the editors, mostly English-speaking, might have an influence on the perception of the textual phenomena. And we know that each encoding system serves specific interests of the text.

For the moment, those particular features of Romance languages, living or dead, are almost absent from the *Guidelines*, and it appears often difficult to open a dialogue with other (each other would say "wrong") schools of thought. I remember for example that during the CLiP 1999 Fernando Magán Muñoz, of the Centro Ramón Piñeiro di Santiago de Compostela, underwent heavy criticism on the part of TEI exponents for having presented an SGML encoding of Galitian-Portughese manuscripts which were alternatives to the system presented in the Guidelines (see http://til.scu.uniroma1.it/appuntamenti/magan.htm).

All that has been said up to now reflects on the Malaspina conference. Only six out of thirty five papers were given by scholars from European universities: Willard McCarty (King's College, London), Susan Hockey (University College London), Dirk Van Hulle and Edward Vanhoutte (University of Antwerp, Belgium), Peter McKinney and Deneka MacDonald (University of Glasgow). But only two of these are from countries in which the main language is not English. Rather few for a conference which claims international status, at least in the title; perhaps "The Humanities Computing Curriculum in UK and North America" would have been less ambitious, but closer to reality.

More geographically oriented were the CLiP 2001 presentations (thanks also to the organisational efforts of Elisabeth Burr), with the participation of researchers and specialists from Spain, France, Great Britain, USA, Germany, Italy, The Netherlands, etc. The core of the seminar was dedicated to the European curriculum, with papers which take account of different national and local landscapes. Manfred Thaller's paper, concerning a new HC/IU curriculum which has been set up at the University of Cologne, establishes explicitly a dividing line between a conception of computing as *intellectual instrument* and computing as

application (a position which obviously legitimises its opposite, or the subordination of the humanities to computing). The curriculum set up in Cologne (*Historisch-Kulturwissenschaftliche Informationsverarbeitung*) is divided into a number of levels, from the two-year diploma to the doctorate. But what is it which makes it "fairly unique" compared to the more well-known British or American courses? As an explanation Thaller takes up once more the distinction between *Humanities Computing* and *Humanities Computer Science*:

> While the concept of Humanities Computing usually assumes that the Humanities apply mainly tools and occasionally concepts developed by others, the Cologne concept assumes that information as occurring within the Humanities has inherent properties […]. Business administration, according to the common (mis)understanding of computer science is much closer to it than the Humanities. If it still needs its own special brand of computer science, how much more so do the Humanities, which are much further removed from the common (mis)understanding of what computer science is all about? […]. Computer Science is not concerned *about computations in the numeric interpretation but with the general question of how information can be represented and how these representations can be processed. This definition does not reduce, but establish* the claim for a special brand of Humanities' computer science. After all, the Humanities have a very long tradition in the usage of complex, fuzzy and vague information, which is extremely relevant*to overcome the information glut much complained about. […]. That the Humanities in general, are much too timid at the moment to claim their proper relevance for the solution of the problems of an information society is something the confessing Humanities' computer scientist can only diagnose; he cannot be required to share that timidity.[5]

This idea is recapitulated in that passage of ACO*HUM where we see: 'We conclude that humanities computing courses are likely to remain 'a transient phenomenon, unless they include an understanding of what computer science is all about.'

It would be superficial – if not risky – to see a geo-political logic behind the ways we understand Humanities Computing. But neither can we deny the existence of both a "continental" and an "Anglo-American" approach. Of course, there are representatives of one and the other alignment on this as well as on that side of the Atlantic (and the English Channel), but the continental convergence is perhaps inevitable. It is the fruit, besides, of the slow but continuous process of integration of the higher education systems initiated with the Bologna Treaty in 1999.

Obviously, what we have said up to now has nothing to do with the quality of the Malaspina conference papers which represent still one of the highest points of the discussion about scientific and educational contents of Humanities computing. However, the only paper which appeals clearly to the European outline, that by Vanhoutte and Van Hulle, points out even more the difference between a conception of the discipline I would call *institutional* and another I would define (temporarily) *pragmatic-localistic.*

Institutions, rules, and the market

So we see taking shape those elements which come earlier in the process of the theoretical IU/HC divide: the cultural-institutional element and the political-geographical. On the one hand, there is an "island" system, fragmented, but with strong links to the territory, whose main feature is individual initiative. In the USA (and partly in the United Kingdom) universities can start up educational initiatives autonomously; anybody (if they have resources and academic consensus) can set up a Masters or a Doctorate. This liberty certainly makes the teaching market, and that of professional figures linked to it, more open, but it does not at all assure academic recognition.

The lack of full professorships in Humanities Computing proves that the problem of the autonomy exists, even in the countries where the link between technological innovation and education is vaster and historically rooted. Neither an immense educational market, nor the

---

[5] I quote from the online abstract e: http://www.uni-duisburg.de/FB3/CLiP2001/abstracts/Thaller_en.htm.

flexibility of the educational system are enough to knock down the "walls" between the academic disciplines and guarantee the quality of the education.

Even in the United States and the United Kingdom, those who want a place as lecturer / faculty have to face up to the departments and with the equilibrium of the forces inside them. The case of the *Composition* subject is a good example. Towards the mid-Eighties in the United States from the English departments the teaching of writing began to separate itself. This discipline, over time, forming an autonomous sector, is still today the Cinderella of English departments and struggles constantly to find spaces and places.

But who can deny that it is precisely because of compulsory composition courses for all first-year students that Humanities faculties are able to top up their budgets?

> Central to the dilemma is the question of how work with technology fits into traditional categories of research, teaching, and service. [...] we are caught between tradition and transition, attempting to evaluate a technology and practice with which we have inadequate experience, and which keeps evolving as we watch. Our department has accepted that the best we can do is to openly recognize, first, that many fine teachers and bright researchers are doing good and interesting academic work with computers; second, that in the course of time, as members of our department and colleagues, friends, and acquaintances at other institutions do more work with computers, the whole field of computer-related activity in English studies will take shape for us; and, third, that through argument, conversation, and compromise, a consensus will develop as to how that activity is to be evaluated and rewarded.[6]

These words are a confirmation of the fact that even the USA is far from a legitimate and automatic recognition of the importance of computing in the education of the humanist. It might seem almost paradoxical that "backward Europe"[7] is, with respect to this specific problem, at an advantage. Indeed, this equivocal mixture of institutional rigidity and desire for *Information Technology* is forcing the educational system to recognise the specificity of certain key subjects and disciplines. In Italy, the introduction of new three-year and two-year degrees (3+2, as required by the Bologna Treaty), clashes with the old division of discipline groups. There are subjects, and courses have been set up, which do not have teachers – and amongst these, indeed, we find IU. While this does not reduce the defects in our system (e.g. the detrimental discipline barriers), neither does it show that the only way to open up spaces to innovation is individual contracting, according to the American model.

Conclusion: the "added value" of languages

José Antonio Millán, philologist and lexicographer is a net analyst who left university some years ago to dedicate himself entirely to electronic editing and the study of digital textuality. His latest book[8] is a valuable source of information and data, but above all it is a clear political manifesto of the "digital south" of the world. Spanish is the third language on the net and its prestige is clearly not to be compared to Italian, but it is precisely for this reason that the analysis of the *technolinguistic* powers carried out by Millán is even more worrying.

Thanks to a number of different investigations, carried out in Italy[9] too, we know that US organizations preside over, directly or indirectly, the main institutions which run the Web. What is clear is the anti-democratic nature of these organisms, as in the case of ICANN (*Internet Corporation for Assigned Names and Numbers*), the centre which is in charge of assigning Internet domains. This is a billion dollar commerce controlled by the USA through

---

[6] S. Lang, J. R. Walzer, K. Dorwick, *Letter from the guest editors*, "Computers & Composition", 17, 2000, p. 1.

[7] See the data of the technological gap: http://www.eto.org.uk/eustats/graphs/93-99.htm.

[8] J. A. Millán, *Internet y el Español,* Madrid: Fundación Retevisión, 2001.

[9] See *I Signori della Rete. I mondi di Internet, le frontiere delle telecomunicazioni, la guerra dei bottoni*, "I Quaderni Speciali di *Limes*", 1, 2001.

a business-government cartel. To get an idea of the dimensions of the business, we might mention that "the domain business.com has been sold for ten billion liras and the www.tv rights have reached more than a hundred billion"[10]. However, the question is not only economic, it is also political. For this reason too, the European Union has decided to ask for the setting up of an *.eu*, domain, which ICANN has hindered in every way. Full administrative control of the domain would, in fact, mean submitting to European legislation – concerning brand protection, competition, etc. – every act carried out under it. After a great deal of international pressure in 2000, what some called the first rigged "virtual elections in history" took place , but despite the protests – also on the part of associations and American NGOs – ICANN is still solidly in American hands.

This is the background, but Millán's book takes up and at the same time widens the scenario by analysing closely the production and spread of all Internet technologies concerning language. There are many products and services which derive from these technologies, all of them of a strategic value and all in "alien" hands: operational systems, search engines, intelligent agents, distance teaching and training, electronic commerce, the copyright industry, etc., Each of these areas presupposes or stimulates specific research sectors. They range from automatic translators to syntactic parsers, from terminological databases to software for speech recognition, etc. Even though the estimate for the burden of linguistic technology for each product and service analysed is low (see Table 1), the result is astonishing: calculating a catchment area of 61 million Spanish speakers, the annual business turnover is estimated by the author at something like 31 billion pesetas, or almost 91 million euros (Millán 2001: 148-149).

---

[10] F. Vitali, *Icann, il nome del dominio americano*, "I Quaderni Speciali di *Limes*", p. 10.

Product/ service	Peso in linguistic technology
Electronic Commerce	0.01
Copyright industry	0.01
Tourist information services	0.03
Operating Systems	0.05
Distance Education	0.07
Wordprocessing	0.10
Teaching material for Spanish as SL	0.10
Information services (non-touristic)	0.10
Editorial platforms	0.20
Search engines	0.30
Information managers	0.50
Intelligent agents	0,80
Didaction software	from 0.80 to 0.90
Terminology assistant	0.90
Translation software	0.90

Table 1. Linguistic technologies and products: weight per product or service (in ascendine order). In the list the author purposely leaves out the technologies voice recognition, dealt with in ch. 9 (Millán 2001: 134).

Millán concludes by saying: "while the nets are the highways of digital goods and services flows, the technologies linked to users' languages are the compulsory tolls" (Millán 2001: 140). Thus, at the roots of economic, social, political primacy we do not find "brute" technology, but rather the control which this guarantees over languages. Presiding over both has become a succulent business, but it is not only this: not investing in this sector would mean, in the not too distant future, being forced *to pay to be able to use one's own language*.

I will now come back to my original question. The problem of the theoretical status and the curriculum of IU is to be faced, according to me, within a European multilinguistic and multicultural framework. Our continent has material borders on which other languages are spoken. Non-existent in the United Kingdom or Australia, limited in the USA (immigration and linguistic borders are two very different concepts), these borders, their peoples and their languages have long been marginalised, if not repressed; but they are now a great wealth: perhaps the only real wealth of the old continent. Computing technologies are applied to the semiotic universe: texts and documents are not set aside from the languages in which they are made up and the latter constitute, to use a fortunate image of Millán, the "added value" of the Internet (understood here as synonymous with and a paradigm of the new technologies). To whom do we entrust the task of running this *value*? Who has to discuss the standards, rules, if not the variegated Humanities Computing community?

We know well what ignoring this responsibility would mean. At a time when western cultures are setting about translating (and transforming) their own knowledge into digital formats, I believe that the need, today as five centuries ago, is that of elaborating a new *paideia*: thus the creation of a European curriculum to *train the trainers* gains more credibility.

Personally I am convinced that only a common educational tool – flexible and respectful of different cultures – can defend us from that which by now is shaping up, thanks to the alliance with technology, as a possible loss of *linguistic sovereignty*. Perhaps one day it will not be an anti-global heresy to demand, together with the right to linguistic self-determination (as it happens in modern state constitutions), regulations which protect languages (and texts) against exploitation, perhaps through the creation of a special copyright, a patent for the industrial use of goods which are revealing themselves the *goldmine* of the new economy.