

C 82

# ECHOS DU MONDE CLASSIQUE CLASSICAL VIEWS

XXX, N.S. 5, 1986

No. 1



THE UNIVERSITY OF CALGARY PRESS  
CLASSICAL ASSOCIATION OF CANADA  
SOCIÉTÉ CANADIENNE DES ÉTUDES CLASSIQUES

ISSN 0012-9356

## THE COMPUTER AND CLASSICAL RESEARCH<sup>1</sup>

I am all too well aware that writing and speaking about technology puts me in the precarious position of appearing as an expert. Nothing could be less true. It is rather as if someone who had been knocked down by an automobile were to be dragged out from under the wheels and asked to speak about the development of motor vehicle technology. I am a computer user, still in the early stages of taking advantage of the capabilities which I know to be there, with my reach for what I think will be coming always a bit beyond my grasp.

I propose to organize my remarks into six parts: (1) the existence of machine-readable files providing basic resources for research; (2) the access to such resources for the individual scholar; (3) the scholar's appropriation and organization of these resources; (4) the process of formation of hypotheses and their testing; (5) writing; and (6) dissemination of the results. For each of these, I shall try to assess where in classical studies computers can help now, may be able to help some day, or will probably never play an important role.

First, resources. The classics are comparatively well off in the availability of literary sources for computer use. Over fifty million words of Greek (approaching 85% of the extant corpus) now exist in machine-readable form at the Thesaurus Linguae Graecae at the

---

<sup>1</sup> This paper is printed here substantially in the form in which it was given orally to the annual meeting of the Classical Association of Canada in Montreal, May, 1985. I am grateful to Pierre Brind'Amour for the invitation to take part in the session on classics and the computer of which it was a part. Earlier versions were given to a colloquium at Yale University sponsored by NERCOMP (March, 1984) and a meeting of the Northeast Association for Computing and the Humanities (New York, May, 1984).

University of California, Irvine. A lesser portion of the much smaller corpus of Latin is available from the American Philological Association's Repository of Machine-Readable Texts, and a major project to complete the recording of Latin texts in digital form is about to begin with the support of the Packard Foundation. We may reasonably expect that within a few years the large majority of what any student of classical antiquity would want in the way of literary texts will be available. A project at Duke University has made a start on the Greek papyri, and one at the Institute for Advanced Study on Greek inscriptions. There are still substantial areas in the documents left untouched, notably Latin inscriptions, but progress has been good.

For bibliography, the situation is the reverse. *L'Année Philologique* is built on a particular conception of bibliography espoused by its French creators; it is meant to be read, citations, abstracts and all. Indexing is sketchy and not likely to be increased. The French principals are opposed to any automation of the creation of the bibliography (as opposed to composition). The existence of a well-done and influential printed bibliography has stood as a wall against proposals to have an automated one. And yet the latter would offer some weighty advantages to the user: quicker availability is the most obvious and the least important. Cumulative character is the greatest mechanical convenience to be gained; intellectually, the ability to do sophisticated subject searches and to ask questions about the bibliography not envisaged by the abstracters is the greatest opportunity.

For objects, we have had some progress. The *Lexicon Iconographicum Mythologiae Classicae* has accumulated gigantic files about objects and representations in the course of its work. The American Office of this project, at Rutgers, is currently planning to turn these files into an automated database accessible through microcomputers. If this project is successful, it will show the way to what can be done in the future with other areas of art and archaeology. A few smaller projects in Europe, mainly focused on vase-painting, are also developing automated resources.

All of this development, which is partly quite recent, suggests that classicists will soon have available to them a large part of their primary source material in machine-readable form. The gap in bibliography may to some extent be alleviated by library-based systems, as the Research Libraries Information Network (RLIN) and OCLC become more widely available for individual searching and as individual libraries bring their own online catalogues into existence and use. Nonetheless, I believe that such resources, which do not record articles, and which have fairly crude subject indexing, are no substitute for the development of a proper online bibliography for the discipline.

We all know that primary material and bibliography, important though they be, are not the end of the trail. We need the actual works to which the bibliography leads us. Here, I think that it will be the next generation which will enter the promised land. There is at present a powerful impetus for the recording through some means other than paper of our vast heritage of the written word, namely the disintegration of paper in a large part of our library holdings. Preservation of the contents of these books and journals--for only a few can survive physically--is to my mind the most urgent task of this decade. This preservation work is leading to the creation of substantial bodies of material in microform, and it is likely that this material will in the next round be transferred to optical disks, which can be manipulated by computers. It seems likely that in another generation there will be vast amounts of scholarship available in very compact form and accessible from microcomputers. For most of us, however, this is the stuff of the future. We may help it along, but it will not change our working habits very soon.

Now to the second point: how will we get access to all these resources? There are several possibilities, all with major drawbacks at present. First, some of them can be used over telephone lines to a large central computer. RLIN operates this way, as do most large bibliographic databases like the MLA bibliography. Large numbers of commercial vendors of online information in the Sciences and Social Sciences also work this way. For the user, the drawbacks are twofold: the deregulation of long-distance telephone service in the

US and the breakup of AT&T have led to a rapid growth in telephone charges, which were once relatively stable; and more importantly, the user pays for each consultation. Books, once bought, may be consulted freely by their buyers without additional charge. An online database keeps charging you for access to the same body of information. For the occasional bibliographic search this may not be too uneconomical, but for constant access to textual material and for most other purposes, it is hard to accept it, at least for those who do not normally have grant money available for routine research. On the other hand, we must remember that three-quarters of the costs of operating a large library come from salaries, maintenance, utilities and the like. Much of this comes from the cost of acquiring and cataloguing books; but a significant part of it comes from collection maintenance and public services. Books do keep costing something after they have been catalogued.

Alternatively, one can buy copies of some databases (or at least acquire long-term licenses to use copies). So it is with the TLG Greek texts. But there are problems here, too: the user must make sure to get all subsequent releases of corrections and additions from the vendor. Moreover, the quantity of material is large enough to demand substantial computer resources from the user or the user's institution. Not many of us can afford as yet to own a 500 megabyte hard disk to store all of Greek. That may well come with optical disk drives, rendering the individual user with a microcomputer and an optical disk drive fairly independent. But experience suggests to me that the demand for huge volumes of information will grow faster than the means of satisfying it on an individual level. By the time 500 Mb costs \$1,000, we will all want ten times as much storage capacity. And in any case, we are probably some years away yet from this kind of golden age. Optical disk systems are coming onto the market rapidly and prices are dropping fast; but the cost of creating the master optical disk from machine-readable data on tape or other disk format is large, and it must be paid again and again for different optical disk formats, of which there is no standardization yet. For the moment, one has various choices ranging from large microcomputer systems with good-sized and expensive hard disk systems up to use

of mainframe computers on a timesharing basis. The choices are complex and I cannot discuss them in any detail here.

When we come to my third stage, the organization of collected material, things become still more complicated. It is true that it will be increasingly possible to "download" both textual and bibliographical material from large databases into our own files. But to copy is not to organize, any more than a pile of photocopied articles gives one the basis for a book. The organization of collected material involves intellectual judgments, the recognition of significant facts and relationships. Now this is one of the areas in which I think the computer has the potential to be most useful; at the same time, it is one in which the burden on the scholar is very great. Scholars will create their own databases, bodies of material on whatever subject they are working, organized in a fashion which they decide. But it is up to us to decide what to select and how to arrange it.

What strikes me in this prospect is partly the new and partly the old. To be sure, instead of making index cards--which I was always too lazy to do--I will have made a database, electronic index cards which can be reshuffled in any order I wish, and to which I can add at will. It will be vastly easier to correlate many variables rather than just a few, and I ought to be able to ask much better questions.

But what has not changed, it seems to me, is the nature of the intellectual activity involved. And here we move into the hypothesis and testing stage. The task of imagining the past, of inventing stories or pictures which explain our evidence, is the same, a leap of intuition. A stupid, unimaginative scholar faced with an electronic database will be just as stupid and unimaginative as faced with a shoebox full of 4x6 cards. He may produce a greater volume of idiocies, freed of much of the drudgery of setting them down. That is a problem to which I shall return later. But the nature of the task, and what makes good scholarship good, will not alter.

Testing hypotheses will not either. The actual labor will be simplified by the ability to interrogate a database at will and from many different points of view. Often enough, with manual methods, one may be too lazy to check out a possibility which seems intriguing

but hard to verify or disprove, or perhaps which seems only a long shot and not worth the effort or time. With electronic means, we shall be encouraged to try. But once again, the judgment involved in deciding whether the evidence supports or disproves a hypothesis will be the key element.

One incidental problem which deserves a moment's attention is the future of all of these personal databases. I suppose most of you have had the experience of dealing with the *Nachlass* of a dead colleague or teacher, those file folders full of notes, those boxes full of file cards. I have always found this a wrenching task. I know perfectly well that neither I nor anyone else will ever use those card files; and yet I simply cannot bring myself to pitch them out. And they can take up a lot of space. Or, a different problem: I know that a colleague in Paris has collected all of the attestations of shepherds in the papyri in his card file. I would really like to have a look through that. If I know him really well (and papyrology is, to be sure, a pretty friendly field), I might look through it with his permission next time I am in Paris. But not even a close friend is about to mail me his file box, let alone copy the contents onto another set of cards. Am I to redo his work? What a waste of time!

In twenty years' time, there will be thousands of these electronic file boxes in existence; all of us will have them. How will they be preserved? Who will have access to them? When I publish a book with a chapter on shepherds in fourth century Egypt, will it come with a note saying that interested readers may acquire the database on which it is based by sending me \$5.00 and a stamped, self-addressed envelope in which I will mail a copy of a floppy disk? Will I deposit it in my library and list it in RLIN, so that any interested party can send an electronic message to my university asking for a copy of the database, like a sort of extended Interlibrary Loan? Or will I keep it private? I have no answers to any of these questions, nor yet to what you will do with your teacher's database on some subject which doesn't interest you very much. But they deserve serious thought: some of my librarian friends are already worrying about what scholars may expect of them in this area.

About the next stage, writing, I doubt much need be said here. The advantages of electronic editing and formatting are becoming obvious and are addressed by others in this session. The subject is technical.

Printed books are not going to go away for a long time; in fact, I think they will always be with us and our descendants, despite visionaries' babble to the contrary. They are simply the most convenient and pleasant form in which to read, an activity about whose survival I have no doubts. For the near term, it matters a great deal how we can get scholarly books printed at reasonable cost in order to disseminate the work we produce at such a cost in time and effort. One method of keeping costs down which has played a large part in the last two decades is the use of camera-ready copy produced by the author (or even by someone else) at much lower cost than typesetting. I suspect that this will become the norm rather than the exception as people get more sophisticated facilities at their disposal. One other point about dissemination needs making. We have heard a great deal about electronic publishing, how articles and books will be available online, not printed. It is already happening in some of the Sciences and a bit in the Social Sciences. I do not believe it will make any large inroads in the humanistic and historical disciplines very soon. Why? Partly because of cost. Storage of full text is still expensive, and only a fairly high level of usage charged by the second will justify this expenditure. Most works in the humanities will not get this level of paid usage, and if they do, they will be popular enough to justify printing costs.

Another factor will, and in my judgment should, slow the conversion of our means of dissemination from print (or photographic film) to machine-readable form. That is the international character of scholarship. A few of our Western European colleagues have computers and are at stages of development comparable to our own; most are not. In the Eastern European countries, development is far behind, and in the Third World, all of this is barely imaginable. If we create significant original works of any sort which are available only in machine-readable form, we will be cutting our colleagues in many countries off from access to what we are doing. Perhaps in



some fields this will not matter, but I do not see how such arrogance can be tolerated in any field which is as international as ours. For quite a while to come, therefore, we must remain sensitive to the needs of the whole scholarly community, and not only to the wishes of those as rich as ourselves.

All of this is not to say that electronic dissemination will not have a considerable effect on work in classical disciplines along with others. It is not formal publication which will be affected, I think, so much as informal consultation of colleagues and private circulation of unpublished work. This off-the-record activity is likely to be greatly facilitated by electronic mail. Once again, though, we need to keep in mind that we should not wind up by default limiting ourselves in maintaining these networks of colleagues to those who have electronic facilities.

We turn now to think about the implications of the changes--and of those areas in which there will not be much change--in these stages of scholarly work. The most obvious thing, perhaps, is that we must all learn new skills. I do not think this means that in addition to Greek, Latin, French, German and Italian (I put this list from an American point of view) we must also learn several computer languages. But we will have to learn how to make these machines do what we want; and for all that one hears about the user-friendly machines of the future, I am persuaded that the capabilities of technology will always run a generation or two ahead of the capabilities that have been made easy to use. The next generation, to be sure, will grow up with computers and find these skills more natural than we.

Secondly, a lot of drudgery ought to disappear. In the first six months that I used a computer, I could have done the same work much more easily without it. But after that, I have been spared countless retypings or hours with whiteout fluid or scissors and paste. And that is only the start; the electronic resources, both systemic and personal, which I have described will spare us enormous amounts of time and effort on menial chores. There are still some who regret this fact, as if these chores built character, but my own impatient temper has rather been soured countless times by drudgery.

Thirdly, many types of so-called scholarly work, which in reality only put on paper the results of thoughtless drudgery, should disappear. Mechanical concordances, for example: what continuing role will they have? When you can frame a sophisticated question about the occurrence of a word in a particular phrase and in conjunction with other phrases, why should you spend time looking through lines in a printed concordance? Possibly one can justify their existence by the serendipitous discoveries they sometimes make possible. But anyone can produce his own with straightforward software on his own system. And why should anyone get tenure for producing them? Most bibliographies can go the same way; uncritical extracts from other works, they can be compiled by a computer in a few seconds or a few minutes, and they will be just as useless in that form as when compiled by a human drudge. Compilation without thought, in other words, has lost whatever purpose it once had. That will be bad news for some, but not for the world of learning. With drudgery reduced and thought focused, the imagination to frame hypotheses and recognize significance, and the shrewdness to distinguish what can be disproved from what can be supported with evidence, will be able to stand out conspicuously.

Fourthly, new types of investigation, involving many more correlations of different factors, will become possible. Others are more visionary than I about the possibilities, but they are unquestionably real. Even if the nature of human thought does not alter, the range to which it can apply it itself will.

Finally, I think the technological revolution will in the end result in a democratization of access to information and of ability to do scholarly work. We are all aware of the results of the ending of the exceptional economic circumstances of the 1960's: great libraries unable to keep up with what is published, mediocre and poor libraries unable to transform themselves into good working collections, scholars of ability teaching in institutions with heavy teaching loads and poor research resources, vastly reduced academic mobility. All of these factors have produced a well-recognized generation of lost scholars. Various programs try to bring them in summers to the resources they need, to keep them involved in their professions. But nothing would

work like replacing the library of East Nowhere State College with a copy of that of (let us say) Yale - occupying a lot less space, to be sure, and not quite so easy to browse through, but conveniently consulted in machine-readable form. For those of us who teach in or manage great universities, such a prospect is not without its threats, for our inherited capital of research resources is one of the biggest factors in making the institution more distinguished and attractive than most others. But if this means that we must live by our wits, and distinguish ourselves by intelligence and daring, that might not be such a bad thing, for us and for society.

COLUMBIA UNIVERSITY

ROGER S. BAGNALL