

THE EMPEROR'S NEW CLOTHES*

In the fairy tale of Hans Christian Andersen, several scoundrels pass themselves off as tailors before the king and promise him a magnificent new wardrobe. As the story progresses, we learn that the clothing is non-existent, but the king, anxious to be sophisticated, up-to-date, and chic, accepts the word of the "tailors" that he has the latest in men's wear. His courtiers, even the citizenry, also wishing to be as sophisticated as the king, and wanting to be on his right side at the same time, also believe that the king is dressed in the highest fashion. Only when the king marches through the street is there a small voice raised in a sceptical way. A little boy looks, and says "the emperor has no clothes ! "

Now Andersen may or may not be trying to offer us a fable here, one to which we are to add a moral. I certainly do not wish to appear as a moralist. All of us have been, since we are for the most part academics, "tailors" at one time or another, some have even been emperors, and all of us were small boys once upon a time.

Following St. Paul's precept, we have put childish things away and gotten on with our business. I suppose most of us prefer the romantic optimism of the emperor and his entourage to the little boy's pessimism. I, perhaps

one of the few professing pessimists in the humanistic use of the computer, am pleased to have a chance to invade the sunny gardens of Eden in whose warm ambience so many projects now grace the academic scene.

Before I proceed to discuss the reasons for my pessimism I must, I'm afraid, say something about my own past, and, sinner that I am, make some apology for my ill-spent life. And so, at the risk of boring everyone, I shall discuss my own efforts as a computing humanist. I do believe that my comments may evoke that shock of recognition so beloved of the novelist and critic. If you will bear with me, I shall get beyond this quickly, and turn to what I see as the serious difficulties facing the scholar who wishes to enter the world of machines.

In 1963, when my home department of Classics at a major state university was faced with a financial crisis involving the support of graduate students, I came to the acting chairman with the brilliant idea that I could hire a research assistant for a project to put together a lexical and grammatical index to a short treatise of the late Latin author Cassiodorus which I had edited some years before. By hard experience I had been made aware of what sort of research grants succeed. I even had experimented in this area : the previous summer I had applied for a faculty summer grant to do what I really wanted to do, namely to do some reading in problems of Greek and Latin lyric that interested me. To no one's surprise (including my own) I did not receive the grant. Research with so practical an end in mind as my education would never be supported by anyone.

However that may be, I did apply for a research assistant to put together an index to the work of an author whom almost no one would ever read. And this time, again to no one's surprise, I received the grant. One of my

close friends in New York who had gone to work for IBM, supplied me with information on the KWIC index system. Since the computing center at my institution was located in the Physical Education building, I had no desire (being a sedentary type) to actually go there, and I sent my new assistant on his way with a copy of the text, and instructions to return with an index. Our computing center at that time had an IBM 709 computer and I expected no difficulty in placing an IBM program on an IBM machine.

The personnel of the computing center very kindly taught my assistant to operate a key-punch machine, and how to put the tabulating cards on the little drum (this took him about two months), and every time I spoke to him he told me that, while they knew nothing about KWIC indices, they saw no problem in his project.

Alas, why had I not remembered the impassioned words of Lucretius¹ :

“O miseras hominum mentis, o pectora caeca :
qualibus in tenebris vitae quantisque periculis
degitur hoc aevi quodcumquest ”?

By the end of the year, my student had punched out the text of Cassiodorus' *De anima*, one word per card, and had sorted them on a mechanical sorter alphabetically, and now delivered to me an alphabetized index printed on the reader-printer, and 6 boxes of cards in alphabetical order. Naturally, the text was no longer in the order of lines and sentences, and the cards totally useless for any other project.

I was happy for the student, however, for he had been able to stay in

graduate school for the year, and in the following year succeeded in obtaining a fellowship. Afterwards he went to Rome, where he studied literature and architecture, and promised me that he would never set foot in a computing establishment again. Needless to say, I still have the six boxes of cards in my office (one disadvantage of not moving from one institution to another), and I have made use of my index now and again. It is obvious that it is of not much use for any serious work, and could have been completed at a speedier rate. It is a matter of some pride to me that I resisted the impulse to publish it.

The following autumn, I attended a "Conference on Literary Data Processing" at Yorktown Heights, New York. Among the papers given was one which concerned a concordance generator which had been developed by the IBM research group there. Here was exactly the kind of concordance that I thought would be first-rate in the handling of prose texts. After the meeting I wrote to one of the authors of this program and asked for information on it, technical information useful in the implementation of the program. He referred me to the Proceedings of the Conference (published by the Modern Language Association),² but the article was not technical at all, and said absolutely nothing about *how* to generate such a concordance. I wrote again, and this time received a copy of the *IBM Systems Journal*³ which contained a more technical description of the concordance, basically the way the material is stored in the computer, but once again no practical method of employing the concordance in my work. I had by this time decided that I would try to concord the text of Cassiodorus' *Institutiones*.

Once again I wrote to Yorktown Heights, and this time a box of punched cards arrived together with a list of the program's statements in FAP. Now

at last I had what I needed. But no, it didn't turn out that way. In 1964 the computing center had purchased the machine of a competing company and the 709 was being phased out. Furthermore, I was told by the computing center personnel that the program I had was written for the 7094 and could not be run on the 709. In desperation I asked if the program could be "translated" for the new machine. Yes, I was told, at the cost of approximately \$8000 and even then there was no guarantee that it would run.

Fortunately, the wife of one of my friends was a systems programmer at the computing center, and, one day, as I wandered forlorn through the halls, she took pity on me and looked at the FAP listing. Within ten minutes she saw that by suppressing a handful of instructions which involved nothing central to the operation of the program, it could be made to run on the 709. But I would have to hurry, for the 709 was to be dismantled in a few months. I proceeded to get cards ready for processing and we tried out a pilot run. To our distress, the program did not give us full sentence context, but only the amount on each card. Once again I wrote to Yorktown Heights. I was told that the input for the program was to be on paper-tape and they were very sorry that this hadn't been pointed out. My friend's wife then sat down and in two days wrote a program blocking the records by sentences. We ran the pilot again. This time we lost most of the output. A check of instructions showed that the FAP program included early in its running an instruction read by the small computer which handled the printer as an end-of-file. We reprogrammed the output computer, and finally had our pilot run.

At last we were able to do a full-scale run on the last day of the operation of the 709 ! The program had several other options such as frequency-

listings, end-sorting, and so forth. We assumed that this was buried in the switches contained within the main program but we could not find them. I wrote again to Yorktown Heights. I received in return the necessary programs for the sortings on the tape, programmed for the IBM 1401. Unfortunately this program did not work, and so another systems programmer at my institution rewrote it for the 1401 and I then ran the program in our payroll office.

A final note on all this : I had sometime earlier, as a practice for learning FORTRAN, written a little program to do a word count. Just to try it out I used the same data I had used for the concordance. Much to my surprise I discovered that the word count offered by the IBM concordance and that of my program differed. Mine had been checked often enough so that I knew it was correct. Again I wrote to Yorktown Heights. The reply stated cheerfully that there were some bugs in the concordance generator, but after all it was intended for experimental use at Yorktown Heights only. They would, however, be working on a new and improved version, and I would have a copy of the new set of binary cards shortly. As you can already anticipate, I never did get that new set. Then three years ago, a computational linguist came to my campus to work in the linguistics department. He had been with IBM in California, and knew of the concordance generator. He wrote to California, they wrote to Yorktown Heights, and after six months I had a copy of the new deck. But, of course, by this time the 709 was long gone, and a neighboring university which had permitted faculty from my campus to use their 7094 had become so busy with its own work that this privilege could no longer be extended.

Imagine my surprise recently to learn that the concordance generator had turned up again at NYU's Institute for Computer Research in the

Humanities.⁴ I wrote to the author of the article on the concordance (one of the co-authors of the original version) asking if the "bugs" in the program had been removed and for some information on its present availability. I wrote several months ago, and, although promised a reply by the Editor of the IRCH Newsletter, I have so far heard nothing.⁵

In the computing business I suppose one becomes accustomed to this kind of trivia, but it does make you very suspicious of those articles that appear again and again in the *New York Times* and elsewhere about the great results to be expected from literary data processing now and in the near future. I might mention in passing that in September 1964, the conferees in Yorktown Heights were told

- a) that machine translation was a success, and
- b) that the optical scanner was only a year away from being a successful and economically feasible reality.

But enough of my own experience. Although I could add many other examples, I think you can begin to appreciate why I don't believe the millenium has arrived quite yet.

There are three areas of computer operations that need attention. I call these : machines, people, applications. Please forgive me for not using the terms "hardware" and "software" so loved by computer manufacturers. Somehow I've never gotten used to the idea that a programmer is "software", and I think it is essential even in our darkest moments that we continue to regard programmers as part of the same human race as ourselves.

The problems that confront the humanist when he deals with machines are

many. First, it needs to be stated that it is an axiom of this business that as soon as you have a nice program to run, the machine at your institution will probably be altered or removed. And, most often, the program will be totally incompatible with the new equipment.

Furthermore, the various computer manufacturers have different clientele in mind, and a large university, in choosing a machine, must also take into account a variety of users. At the moment, my own university is considering a move to a third-generation machine, and, as a member of the policy committee of the computing center, I have had an opportunity to see how difficult it is to come to a reasonable decision. At first, we felt that we didn't know enough about the various interests on the campus. A survey was made (and an excellent survey, too) and suddenly we found that perhaps we now had too much information and we could not digest it all. It is eye-opening to realize what a gap exists between the humanities user and the physical scientist. For the latter, as one of them said, "speed is the name of the game". What he wants is the fastest machine available, all other things being equal. I couldn't care less about this aspect, and I suppose I can be so cavalier since my computer time is paid for by the university and none of my projects is of great length. But I do care about customer service, about the availability of help in determining the kinds of peripheral equipment I can use, and the availability of various languages for the machine.

Let me point out that the last can be quite essential, and it isn't the name of the language alone that matters. FORTRAN as it exists for the present computer on my campus is a far more powerful instrument for scientific research than the FORTRAN available on other machines; to use SNOBOL however, which this company does not supply, our systems people must

rewrite what is sent to them from Bell laboratories.

By far the greatest difficulty humanists have with computing machinery is that they came late into the business. When computing began, most campuses had a very small, easily operated kind of machine (say, something in the IBM 600 series). In addition, there weren't very many people at the university using the machine, and so you could tinker with the machine, run your programs, get results, run them again, learn how to mount tapes, put paper in the printers, and so on. One learned a great deal about the operation of these devices in this way. But now the machines are highly complex, the number of users is so large that the only open-shop time may be from 4 to 6 in the morning, and the turn-around time for any job can run to a full day or more.

With some knowledge of machines and peripheral equipment, you may find that your project doesn't require a computer at all. Some small projects are needlessly complicated by the use of a computer, projects which the user might more easily have brought to a successful conclusion by the use of other means or equipment.

Do not be misled by the constant statements you hear about the speed of computer operations. Professor David Packard made this point very well in the preface to his *Livy Concordance*.⁶ It took only two hours of computer time to "do" the concordance – it is difficult even to guess how many hours it took to get the text ready to be entered. This point actually leads into my next topic, what I have called "people", and deals with programmers and programming.

Most of the projects that humanists undertake are one-shot affairs. Say

you are interested in certain phrases in a particular author, prepositional phrases in Catullus, for example. Once the print-out is done and has been handed to you, your work has begun. It may take you a good year or two to put together the observations you want to make on it. But you may never use that program again, and, like many humanists, you may never deal with a topic concerned with Catullus again. What you need to ask yourself before committing the work to a computer orientation is whether you would not accomplish your work more quickly and more efficiently in some other way, perhaps using Wetmore's index⁷ and your own jottings from previous readings. Even if the corpus of your author is much larger, you must consider the time it will take you to learn a programming language, to get the text into machine-readable form, to write, test, and debug the program. For a scientist the whole matter is so very different. He may very well use the same program year after year with new data. The equations he has to solve may be so involved that only the computer can find solutions. Only one of these solutions may be of any use to him. It is very unlikely that a humanist will be doing a project expecting results of this kind.

As a consequence, most computer programs for humanists must be custom-made. Computing centers are naturally reluctant to get involved in such programming, because the time involved cannot be written off economically. A programmer who writes a statistical program, even of a very sophisticated kind, may well expect that it will be used by a number of sociologists and maybe up to half of the new Ed. D's from the education department. This is the kind of thing that directors of computing centers, who are giving service to a large number of people, must keep in mind.

Can the data be properly prepared ? I have heard it said by classicists

bemused by the computer that if we had all the texts in Latin (or Greek) on tape, we could discover many interesting things. I couldn't agree more. But what do you do when you realize that a large number of texts have never been properly edited? Professor Packard's concordance of Livy is a work that I doubt anyone will want to do over in the near future, but we must recognize that we have no thoroughly edited text beyond Book 35, and the orthography of the editions used by Professor Packard varies. To see what differences there are between the text used before Book 36 and that used after the end of Book 35, I took a quick glance at a book for which there are both editions, namely Book 31. I compared the opening lines of that book in McDonald's OCT text of 1965 and Weissenborn-Mueller's Teubner text of 1890 (in the 1930 reprint) and discovered several differences immediately. Orthography: *omnis, tris W-M; omnes, tres McD.* Text Reading: 31.1.4: *octoginta septem W-M, duodenaginta McD.* Weissenborn-Mueller's reading is, of course, not included in Packard's concordance (which leaves out variants). Differences in numbers, as every text-critic knows, are one of the commonest variants in MSS. I wonder if the uniformity of Professor Packard's concordance may not be misleading to students and scholars who are not text-critics. I also find the orthography of *tris/tres* very troublesome. I haven't figured out how many *i*-stem nouns in *-es* exist only in the plural, but I doubt that many of us would remember to look for accusative examples of this sort beyond Book 35 other than under *tres*. Professor Packard did go to the trouble of including all Latin "u"'s under the vocalic symbol (in spite of Weissenborn-Mueller's distinctions). Being a creature of habit (and disliking the lack of distinction of this sort in OCT's in general), I am as disturbed by this in the Livy Concordance as I am to find in the new Oxford Latin Dictionary the words "Baucis, Bauius, Bauli" in that order. I prefer Lewis and Short's "Baucis, Bauli, Bavius."

We are probably worse off in Greek, for it is hard to think of what text to use for a concordance of Euripides. And I leave out of the discussion entirely what kinds of concordances we actually want. I should much prefer to have had Professor Packard's concordance giving full sentence context, but then the text could not have been printed. Still, if one had Professor Packard's tapes of the text, one could write a program to call for the specific words one wanted in full sentence context and suppress all the rest. In concordance making in particular, it seems to me that we are holding to old methods. I hasten to add that I do not hate printed books and I am *not* a McLuhanite. The Livy Concordance is a fine work, but there is a lot of noise in there for someone who wants only to check the meaning of a single verb in Livy.

Should one take the trouble to learn one of the programming languages? I have read up on various projects outlined in *Computers and the Humanities* and have written to the scholar who purports to be in charge to find out exactly what he is doing. Often it turns out that he isn't doing much more than key-punching the text, and hasn't the foggiest notion what "his" programmer is doing with the material. Computing centers by and large have been scientifically oriented since their inception. Most scientists do (and prefer to) write their own programs. Computing centers have usually had consultants who can help you with specific problems in your program, but they aren't very happy about doing all your programming for you.

The greatest difficulty in learning one of the programming languages is that they are so badly taught. Most instructors waste a large amount of time right at the beginning talking about the internal operations of the computer. Instead of exhibiting his talent in the complications of

internal floating point representation and double precision, the teacher might show his students first how to operate a key-punch, a reproducer, an interpreter.

The manuals available for learning programming languages are not much better. Since the author and the publisher know that the book will be out-of-date within a short time, no effort is made to proof-read the book properly or even to set the text in a sensible order. If the book contains problems, you can be sure that some are incorrectly stated, and if the answers are given, ten to one several of them are incorrect, too.

My final topic under this heading may well have belonged under machines and not people, but it seems to me that sort of no-man's land where major trouble often arises. I am thinking of the entire matter of input/output. I have my sincere doubts about the coming of the optical scanner, and, in any case, for medieval MSS with which I work, it would be of no use. I don't think the nature of the input device is all that important anyway. One can always find fault with punch cards, paper tape, magnetic tape, and so on. Rather, the problem for humanists is that they want tricky things that most computing centers can't give them, because these centers are not set up for humanists. For example, the majority of users find the FORTRAN keyboard on the standard key-punch and the FORTRAN chain on the printer sufficient. Yet it would be quite unfair, I think, to ask a computing center to go to the expense of putting a upper-lower case print chain on-line for your two-minute run. Notice that I'm not even talking about more expensive types of output devices. We are just going to have to realize that we shall need to do our printing of tapes somewhere else, just as we do with our books and articles. It is possible that we can find another location on our campuses, but until we can show that we are

major users of computer time, we shall just have to solve our problems by using ingenuity in place of large grants of money.

Lastly, I come to the topic of applications. Americans, and now Europeans are beginning to pick up this idea, have the notion that if something can be done it should be done. This is especially true if it involves any kind of technology.

What kind of projects are suitable for computers ? To my mind, very few humanities projects can be handled by the computer in an efficient and worth-while manner. I add "worth-while" to the sentence, because the computer can do many things, but they are not necessarily *utile* or even *honestum*.

The use of the computer can reveal something about certain kinds of scholarly research. The computer is a merciless instrument. The great German wit of the 18th century, Lichtenberg, once stated the following aphorism : "a book is like a mirror; if an ass peeps in, don't expect an angel to look out." One needs only to revise this slightly to fit the use of computers. The computer is a magnifying mirror : ask the computer a stupid question, it will invariably give you an even stupider answer. The funny thing about it is that sometimes the answer is so stupid people begin to believe it is clever.

The computer is essentially a device that handles quantities. It does have a logic, but it does not yet seem to be the logic that can deal with literary problems. What the scholar needs to keep in mind is not whether the problem *can* be reduced to one involving quantities, but whether once the problem has been so reduced, its very nature is not altered.

As an example of the kind of nonsense that can result from the manipulation of data, I offer the following quotation from the latest issue of *Computers and the Humanities* (3 [1969] 256). The author is discussing the application of “content analytic and statistical techniques” to *Paradise Lost*. “The Fallen Angels and Satan were grouped together consistently across all six sets of data. Since those critics who take the view that Satan is the hero rarely praise the Fallen Angels, evidence is produced which makes their opinion less tenable. In those analyses of the poem representing portions of the text before the Fall, Adam is grouped with God the Son, the Good Angels or God the Father. After the Fall, however, Adam is grouped with Eve. Objective verification of his fall from a state of grace is thus obtained.” The paragraph goes on, but I shall spare you the rest. Notice the scholarly *hubris* of this unfortunate : *objective verification* of Adam’s fall is obtained – by the use of a computer !

Computer studies which quantify literature without regard to meaning present a false view of that literature. For example a scholar who develops a binary-octal code for the scansion of Latin hexameters and then works up some statistics on the frequency of fourth-foot dactyls clearly has never taken the trouble to *write* any Latin verse himself. He would soon discover, if he did, that you don’t write dactyls and spondees; you write words which have a particular flow. If you are anything of a poet the thought is more important to you than the quantities, although you must have those right, too. Poets don’t slot words in so crude and mechanical a fashion.

We can be harsher with other kinds of studies, those that truly lend themselves to computer analysis. Gilbert Highet in a recent review in *Computers and the Humanities*⁸ justly complains when a list of proper names in Latin verse is defective and incomplete.

Although certain types of computer studies have proved to be feasible, do such studies deserve the highest priority ? No one can deny the usefulness of concordances, but is this the time for scholars to be doing such lexicographical work ? Those of us who teach in America know that many of our classes read their texts through the Victorian spectacles of late 19th century annotated editions. Would it not be better for scholars to consider the less faddish, but more substantial work of commentaries on such neglected authors as Ovid and Tacitus ?

I realize that there can be a number of counter-arguments. After all, someone may say, hasn't Professor Raben refined conclusions concerning the intellectual affiliations between Shelley and Milton ? And hasn't Professor Raben said that "... we can demonstrate of the computer : it will deliver unscathed any literary work we surrender to it." ?⁹

Let me quote in reply the English classicist, L. P. Wilkinson :¹⁰ "... For a century now a great part of the energy of scholars, and of the space in learned journals, has been devoted to tracing the traditional element in classical poems. This activity takes various forms. First there is the search for verbal reminiscences. A man with a remarkable verbal memory, such as must have been possessed by Zingerle, who collected the imitations in Ovid, will naturally want to exercise his gift, though there are now electrical machines almost capable of doing the same work. I am not sure that the results are of much value, and they may even do harm. Men of similar gifts in antiquity soon applied themselves to Vergil : "Perellius Faustus furta contraxit," says Suetonius (*Vit. Verg. 44*). *Furta* – the word indicates the danger If we had available all the literature the poet had read, we should be able to determine the extent of his originality. But this is a matter of biographical rather than critical interest. Ancient poets were not

themselves much interested in originality of this kind, while few modern readers of *Hamlet*, for instance, busy themselves over the extent to which Shakespeare has improved on previous treatments of the story One cannot help feeling that our energies ought now to be concentrated, first of all on the study of the poet's own environment ... and then on the appreciation of the poems themselves as works of art making an immediate impact on ourselves. Accumulation of irrelevant knowledge may even dull that impact. In the eighteenth century men read the classics with less knowledge than ourselves; but one has the uncomfortable feeling that they *got more out of them.*"

This is not intended to be a Luddite plea on my part; the machine is in the garden and we have to live with it. But I think we ought, with Simonides, not blame anyone, but rather remember, "ton ... elithion apeiron genethla." 11

Indiana University.

James W. HALPORN

FOOTNOTES

* This paper was delivered, in slightly different form, at a symposium, "Latin Literature and the Computer," held at Pennsylvania State University on May 3, 1969.

1. Lucretius *De rerum natura* 2.14-16.
2. *Literary Data Processing Conference Proceedings* September 9, 10, 11 – 1964, New York 1964.
3. K. F. Scharfenberg, P. H. Smith, Jr., R. O. Villani, "A Concordance Generator," *IBM Systems Journal* 3.1 (1964) 104-111.
4. Philip H. Smith, Jr., "The State of the ICRH Concordance Generator," *ICRH Newsletter* 4.5 (1969) 1-4.
5. I have recently learned that the Indiana University Research Computing Center is preparing a similar concordance generator for the CDC 3400/3600.
6. *A Concordance to Livy*, Cambridge, Mass. 1968.
7. M. N. Wetmore, *Index Verborum Catullianus*, New Haven, 1912.
8. Gilbert Highet, rev. of Donald C. Swanson *The Name in Roman Verse*, *Computers and the Humanities* 3 (1968) 119-122.

9. Joseph Raben, "Computers and Literary Studies," *Data Processor* 10.1 (1967) 12.
10. L. P. Wilkinson, "Greek Influence on the Poetry of Ovid," in *L'influence grecque sur la poésie latine de Catulle à Ovide*, Vandoeuvres 1956, 240-243.
11. Simonides fr. 542, 37-38 PMG.