

The Humanist in the Computer Lab

Joseph Raben

The attempts so far made to utilize computers in studying the humanities have resulted more in a fundamental analysis of the subject areas themselves than in any significant results. The problems of inputting and outputting, of determining what processes may appropriately be employed, and of developing new processes not derived from the customary numerical approaches which presently dominate computer thinking—all these have retarded what had been expected to be a dramatic advance into a new order of humanistic criticism. The most substantial accomplishments to date have been the rationalized lists of words (dictionaries, indexes, and concordances) for which the computer's capacity to sort rapidly without fatigue or error has accelerated production of these traditional aids to scholarship. A new breed of humanistic scholar now evolving—highly trained in the humanities and at the same time in those aspects of computer science genuinely relevant to his studies—will contribute to the creation of new programming languages specially designed for this work, assist in the training of others who follow, and help to guide computer-assisted instruction beyond the mechanistic mode in which it currently operates.

If, as we have been told, future historians will refer to ours as the Age of Analysis, an apposite illustration of the concern for probing the processes of the mind and of society may well be found in the activities of literary and linguistic scholars who have cast their professional careers in the mold of computer-aided research. Much of the drive for understanding through the close examination of accumulated details typifies a new species of scholarship, which in general parallels the trends we can discern in the physical sciences, the social sciences, and the management of all our institutions—including government. And as this trend was both the father and the child of the computer, so the student of verbal expression often finds himself involved in the challenges and the frustrations of electronic technology. On the whole, the encounter has been beneficial, even though dramatic results are coming more slowly than was originally expected.

167 *Raben : The Humanist in the Computer Lab*

And among the present benefits of this entry of computers into humanistic scholarship (or the reverse, the entry of that scholarship into the realms of the technocrat) has been a scrutiny of the basic materials of our discipline of a sort never before required.

Even the initial act of translating our verbal materials into a form comprehensible to machines intended almost exclusively for numerical processes has forced us to consider the most trivial aspects of these materials: how to indicate true capital letters when the only font available is all-capitals; how to include foreign accents and diacritics; how to indicate line-endings in poetry; how to represent non-roman alphabets and characters. Never before has there been such a need to discriminate the full stop that ends a sentence from that which marks an abbreviation, the apostrophe from the inverted comma, the question mark or exclamation point that ends a sentence from one that merely ends a parenthetical phrase within a sentence. Innumerable projects to explore the intellectual profundities of a literary corpus seem to have foundered in the shoals of irrational input conventions. At the other end of the process, the same complexities bedevil the computer-oriented literary scholar who, commanding a machine that will do almost anything it is told to, and nothing without being told, often appears lost in the maelstrom of dizzying possibilities and maddening impedimenta.

And in between, in the process for which all this anguish is suffered, what analytical processes will we order to be performed? Will the scholar be concerned with quantities and therefore concentrate on average words per sentence, average sentences per paragraph, average length of word, deviations from these averages, and similar numerical values? Many literary projects aided by computers are currently moving along these lines, perhaps because so many routines for this type of work have already been written in other disciplines and can easily be adapted. Sometimes these numerical exercises are ends in themselves; often they are intended to prove or disprove the authorship of one or more texts. Another approach has been to attempt to get at subliminal levels of meaning by grouping key words into semantic categories, either manually by the editor or automatically by the computer. The manual procedure creates the risk of inconsistency as the human mind runs through its cycles of attention and inattention, while the electronic approach goes too far

in the opposite direction, applying rigid instructions with little leeway to interpret the data in new and unexpected ways. Some of the surprise outputs from these unanticipated results of computer processing of literary texts have often been more informative than the results the investigator sought, for they have often forced him to think out problems he had never considered. For example, we may say that we are counting words, but what exactly is a word, that is, when has a word changed enough to be another word? Are *man* and *men* two words? What about the possessives, *man's* and *men's*? Do we include or exclude *mankind* when it is equatable with *men*? What about compounds containing *man* as an element: *man-about-town*, *man-at-arms*, *man Friday*? And then the derived forms: *manlike*, *manly*, *manliness*, *man-made*, *mannish*, *mannequin*, *man-of-war*, *man-size*, *man-slaughter*, etc. Clearly, at some point, we have departed from the central concept of *man*: a man-of-war is a ship, a mannequin is a woman. In trying to establish a line so firm that even a dull-witted computer can recognize it, many lonely scholars, who dreamed of putting the drudgery of their investigations on the back of a willing and industrious servant, are now wrestling with a whole new set of classification problems, and finding themselves frequently working back to fundamentals rather than forward into the territory they hoped to conquer.

Inevitably, the type of classification which has accounted for the great bulk of literary data processing has that which does not require a fundamental principle but rather a traditional and arbitrary schema. In the making of verbal indexes and concordances, literary scholars have found an ideal combination of tradition and novelty: the compiling of alphabetical lists, with or without some context to suggest meaning and usage, has a history extending to the middle ages, while the tedium coupled with concentration once required to produce such lists through the manual sorting of paper slips has made this operation a natural candidate for computerization. Almost the entire history of this field of research can be observed in the efforts of Father Roberto Busa to produce an index to the works of Saint Thomas Aquinas. The justification of this monumental manipulation of 15 million words lies in Thomas' position as the principal shaper of modern Catholic philosophy, the author of a system widely regarded as the most influential and lasting in the history of Western thought.

Busa began his index before World War II on the traditional handwritten cards, shifted then to punch cards and sorters when they became available, moved on through the generations of computers both in Italy and the United States, and now—on the seven-hundredth anniversary of Thomas' death in 1274—has published the first volumes in pages generated by a computer-driven photocomposer.¹ For future generations of scholars in all corners of the world—to the extent that deeper insight into the thought of this philosopher can be gained by an examination of his vocabulary in rationalized form—this index will serve as a monument to Aquinas, to Busa the disciple who exemplified his belief that faith and reason could independently support each other, and to the wedding of a technological process and a purely humanistic end.

Much smaller in scope and in the time they have taken for execution are the numerous projects to create concordances to almost every conceivable body of literature: Elizabethan sonnets, *The Divine Comedy*, the plays of Eugene O'Neill, the poetry of Robert Frost, the history of Livy, *Paradise Lost*, the cuneiform of Ugarit, the prose of William Blake, and even the as-yet undeciphered writing of the lost civilization of Mohenjo-Daro in the Indus Valley. As crude as some of these concordances have been, they have served already in a number of ways: they have supported various types of analysis that could not otherwise have been performed, they have perfected techniques which are usable in producing future concordances, and they have (at least in recent years) contributed material to future analyses as we sharpen our vision of what the computer can do for us and develop techniques to make those visions reality. On the negative side, we must note that the rapid proliferation of concordances has had the effect of stifling the production of improved works under better technological conditions. Although we now have very good photocomposition capabilities, most libraries will be forced to make do with their current concordances, even when these are semi-legible offsets of computer printout, with no lower-case, often without any punctuation, and in general only a crude approximation of what such a book should and can be. Had we seen far enough into the on-line capacities of the computer, perhaps we should never have produced any of these books in their conventional form, not even the eminently useful concordances to Shakespeare or Marvin

Spevack (new spelling) and Trevor H. Howard-Hill (old spelling). To guarantee the utmost flexibility and the opportunity for each user, every time he consults the data, to structure it in the way most illuminating to him, all these texts should simply have been made machine-readable and supplied—on tape or over wires—whenever and wherever there was an interest in them. But cultural lag is nowhere more prominent than in promotion committees, deans, presidents, and trustees. In their eyes, the preparation of a text seems like secretarial work, but the publication of a book comes within the definition of scholarship. Thus the reflexive drive of the scholar to achieve conventional publication is reinforced by the social and economic pressures of the academic world, and another technological advance is held back by the world's inability to adjust immediately to the new conditions it has created.²

The only area where verbal data processing seems to be achieving permanent acceptance and adequate financing is the production of dictionaries. The specialized interests of students of Old English and old Scots are being met by long-term, partially computer-aided dictionary projects now in process. Where national pride generates governmental support—in France, in Italy, in Israel—very real progress can be measured. The acceptance of the computer as a legitimate adjunct of lexicography has reached the point where the editor of a dictionary who has no access to one or who for some other reason does not use one feels called upon to explain his adherence to completely manual methods. Since the value of dictionaries, unlike that of concordances and verbal indexes, is already well established, and since, furthermore, precise definition is at least as important to technology as it is to literary research, it is probably on this front that verbal processing will move forward most rapidly. It is for the production of dictionaries that national institutes are already being established, and the products of these institutes, publicized by the governments whose international reputation is thought to rest on the respect accorded their language, will carry the impression abroad that computers can help man to understand his languages and the works written in them as well, perhaps, as the scientific and social phenomena he is already studying through electronic analysis. In this area, then, we shall probably witness again the venerable battle between the idealists and pragmatists, the seekers of pure knowledge

and the producers of useful objects. In the instance of the new French national dictionary, for example, vast quantities of text were encoded to provide examples for the lexicographers to base their definitions on. I have not yet seen the dictionary, which has begun to appear, but I have been told that all this data is not available for other types of study. In other words, to get their dictionary out in a reasonable time and thereby justify the great investment of capital the government had been persuaded to make, the directors of this project seem to have forgone the possibility of utilizing the same data for further investigations, not only in France but wherever French is studied. This is the sort of problem we must continue to anticipate as work in this area grows in acceptance and perhaps even in glamor. As far as the realities of financing and support will allow, we must reasonably anticipate future needs. We must make every effort to assure that once a text has been made machine-readable, it will be available to every scholar who has a legitimate need for it. Let us assure that every product of a computer-aided study is as perfect in its appearance as the average conventionally printed book. Let all technology—computer, photocomposition, and others yet to come—produce results worthy to stand on the shelves of our libraries next to the works of the predecessor centuries.

In these ways, we can begin to see some of the positive benefits of the new conditions before the phrase “computer-assisted literary research” evokes in every hearer’s mind an image of ugly pages, filled with hard-to-read capital letters, and conveying only with difficulty the sense and tone of the original work it is intended to illuminate. The publishers are willing to produce the best books that the computers can be made to emit; the university administrators are authorizing courses specially geared for humanists who wish to learn how the new technology may help them; the only vocal opposition to this development seems based on mere ignorance and reflexive antagonism to progress (one bitter critic, having read that computer time cost \$300 per hour, assumed that this was the cost of keeping a data file in any form of memory—even a reel of tape). The balance between those who fear computers, whether that fear grows out of a genuine concern for the future of humanities research or out of shortsightedness and a longing for the familiar, seems to be shifting. With the gradual acceptance of computer-assisted literary research

has come an opportunity to move scholarship a real step ahead, into the twentieth and twenty-first centuries.

Like all historic developments, of whatever magnitude, this evolution will depend on the individuals who, as an aggregate of separate forces, give it shape and direction, thereby changing the techniques and perhaps even the substance of humanities research. At present, we seem to have too few scholars who perfectly blend a deep appreciation of literary values with a high skill in computer applications. Just as in the population at large we find a tendency for individuals to polarize either toward number skills and logic (so that, for example, they enjoy games like bridge and chess) or toward verbal skills (so that their pastimes are crossword puzzles and anagrams), so among the computer-oriented literary scholars there seems to be a group which tends generally toward technique for its own sake and seems to have few original ideas on how to apply it (these often have a mathematics background at some stage of their careers), while another group tends toward applications but with little skill at actual programming. So far, these two groups have limped along in tandem, like the cripple riding the shoulders of the blind man, but as the results have been less than spectacular, we must ensure that the next generation is equally well trained in both ends of this spectrum. As the computer becomes less mysterious, there will be an increasing reluctance to grant kudos to a man simply because he had a bright application and found someone able and willing to do the grubby computer work to get it into a practical form, much less to the technician who worked out a complex package of programs, but only for others to utilize. Within reasonable limits, both the user and the programmer will have to be the same person.

This realization is scarcely new; since the first involvement of literary scholars with computers the dispute over which skill should have priority has occupied many hours of talk and many pages of print. It is, in general, a specious argument, for a literary scholar clearly needs very little of FORTRAN or most of the other programming languages if he is doing truly humanistic research. If anything, too intimate a knowledge of the mathematical possibilities may seduce him away from his proper goals into counting for its own sake. The individuals who argue most for "rigor" in humanities research, even for a new "scientism," turn out usually to be those who have spent

more time in a love-hate syndrome with their programs than in the joys and despairs of literary analysis. Little of their work has concentrated on the subtle, the fragile aspects of verbal art. Having learned that the computer can manage certain objective manipulations, they satisfy themselves only with the objective aspects. The qualities which have kept a literary work alive, perhaps for centuries, are ignored in favor of others which are more amenable to processing techniques. If we have yet to see a major interpretive work issue from the efforts made so far, we need look no further for the cause than this unwinnable conflict between the advocates of programming and those of applications.

The resolution to this impasse lies in neither extremity, but (predictably) in a middle ground. Since none of the extant computer languages was written to handle verbal material *in a humanistic way*, something new is needed. Perhaps the computer manufacturers, finally recognizing the market potential in humanistic computer research, will modify one or more of the extant languages to serve needs beyond those of commerce and technology. Perhaps they will actually produce and support a new language designed for the needs of literary and linguistic research. Neither likelihood, however, is very great. What we can more reasonably expect is that a language like SNAP³ (probably as extended at the University of East Anglia) will grow in acceptance until it has been improved, simplified, speeded up, and in every way made suitable for our special needs. Like BASIC, this language will be interactive to stimulate new thoughts and responses. It will be easy for non-mathematicians to learn, but powerful enough in its higher levels to permit the rapid and efficient processing of the relatively large data banks customarily required for verbal studies. Its input and output conventions and its subroutines will be suited to the needs of such studies; for example, alphabetizing routines will not push all punctuation to the head of the list but will assign accented words to locations consistent with the practices of each language. Already, parts of such a language are being prepared piecemeal by individual programmers. When they are co-ordinated and when the gaps are filled in, computer-oriented literary research may begin to fulfil the golden promises that have kept so many hopes alive for so long.

A further desideratum is the expansion of opportunities for formal instruction. Although there are now at least several dozen university-level and post-graduate courses in computer techniques or applications for humanistic research available in America, Canada, and Western Europe,⁴ the student seeking such instruction is still likely to find that it is not offered on his campus. But administrators and faculty committees are slowly learning that this is not simply a basic skill, like typewriting, but an intellectually substantial learning experience, requiring all that a student can bring to it and sending him away with totally new concepts. Each year, new courses are being added to the roster, thereby encouraging other institutions to avoid cultural lag and offer such a course themselves. Not only individual courses, but also sequences of graduated instruction will ensure that all the potentials of the formal instructional mode will be exploited. And, of course, from these courses will come the next generation of instructors who will expand and sustain the effort to make technology serve the needs of the humanities.

While these new courses are slowly growing, we have had the benefit of several summer institutes. At Pisa, Illinois, and Kansas, groups of scholars have assembled to teach and learn computer techniques in literary and linguistic research. National governments have been relatively generous with faculty stipends, travel allowances, and student fellowships to encourage a good mix of scholars from varied backgrounds. The International Summer School at Pisa has been particularly successful in attracting people from many countries who have stimulated each other and the faculty they have studied under. As other summer institutes are now forming—at Göteborg and Minneapolis—we can anticipate that for the immediate future, these will be an increasingly popular means of helping established scholars to catch up and of training the younger scholars who will follow them.

In addition to formal instruction on as many campuses as can support it and annual summer institutes, there should also be, distributed geographically, a number of centers for advanced research. Probably the best arrangement would be to establish such centers in proximity to other research facilities with powerful computers and a variety of highly trained personnel. Their function should be twofold: to work on fundamental problems and to provide facilities

for selected individuals who are conducting their own research. Experience at such centers already operating at Cambridge and in Bonn should guide those entrusted with organizing new ones, while the lessons learned at the major facilities for new dictionaries—such as those at Edinburgh, Nancy, and Florence—should also be relevant. These centers would also be the logical repositories for the growing data banks of natural language that are being encoded, as well as nodes on any network that is to be established.

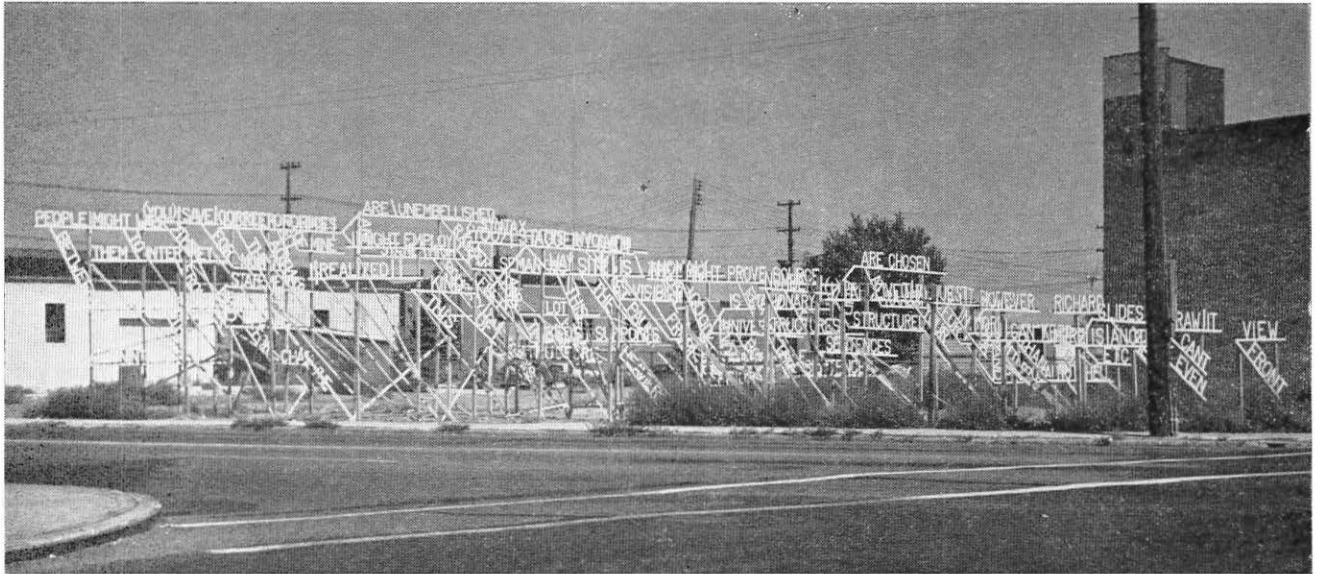
A final concern for the immediate future is computer-assisted instruction. Those who predicted an instant success with machine teaching have learned, like the early enthusiasts of machine translation, that the processes of learning are as complex as those of communication; indeed, they have much in common, especially their refusal to submit to simplistic analysis. Humanists trained to approach subjects which have no readily visible hierarchical structure may already have mastered the philosophy of multi-branched searching techniques that will bring computer-assisted instruction out of the drill-and-practice phase into the broad realm of true learning—that is, self-teaching. If humanists do not involve themselves in this new application, it will, by default, become the province of merely mechanical minds, a means of thrusting information into unwilling students and another triumph for technological impersonality over humanity. If humanists do not concern themselves with directing the future of computer-assisted instruction, they will have themselves to blame when only those factual aspects of a subject which most readily lend themselves to objective presentation drive out the intangible, the nuanced in our approach to humanistic learning.⁵

How these major advances are to grow from the present facilities and personnel is a question that will occupy many of us for some time to come. Without the prestige of scientific or technological research—yet requiring much the same kind of expensive equipment, released time, and support personnel—computer research in the humanities can expect to live under pioneer conditions for perhaps another decade or generation. Until many archaic attitudes have changed, the practitioners of this new art will need to comfort each other with only slowly growing acceptance and support from the larger world of traditional scholarship. But in that long wait for recognition, they

will be able to reach out to colleagues, to exchange information and techniques, sharpen their opinions through the growing number of journals and newsletters devoted to one or more areas of this scholarly field.⁶ In the end, they will have developed for themselves and for those who follow them, another means of understanding and responding to the inherited wealth we call the humanities.

1. (Stuttgart: Friedrich Frommann Verlag).
2. For a more detailed discussion of this alternative, see my article "The Death of the Handmade Concordance," *Scholarly Publishing*, I (October 1969), 61-69.
3. Described in Michael P. Barnett, *Computer Programming in English* (New York: Harcourt, Brace & World, 1969).
4. For listings of such courses, see Edmund A. Bowles, "Towards a Computer Curriculum for the Humanities," *Computers and the Humanities*, VI (September 1971), 35-38; Leila de Campo, "Computer Courses for the Humanist: A Survey," *ibid.* (September 1972), 57-62.
5. For the most recent survey of computer-assisted instruction in the humanities, with an extensive bibliography, see John R. Allen, "Current Trends in Computer-Assisted Instruction," *Computers and the Humanities*, VII (September 1972), 47-55.
6. Current activities in computer-aided humanities research are reported regularly in *Computers and the Humanities* (1966-date); a comprehensive annual bibliography is published every spring.

This article has been reprinted with kind permission from the inaugural issue of the *ALLC Bulletin* of the Association for Literary and Linguistic Computing; Professor R. A. Wisbey, Chairman (Department of German, University of London, King's College, London WC2R 2LS).



Sentence Structures was created by Robert Cumming (227 South Shaffer, Orange, Calif. 92666) and constructed with the cooperation of Richard Koshalek on a vacant lot in St. Paul, Minnesota. Mr. Cumming has commented: "From the letters I had written describing the idea of *Sentence Structures*, I selected about twenty sentences of varying lengths and complexities. The lot I selected was wedge-shaped, so I staggered the individual structures from the simplest back to the most complex at the widest portion of the lot. I cut the letters from $\frac{3}{8}$ inch plywood using a standard one foot high gothic stencil so they could be spray-painted onto the plywood and cut out. There were about 1,200 letters (150 letter E's), which were spray-painted a bright yellow. I made blueprints of each of the structures since a construction firm had to install them. I knew in advance that from a distance the twenty or more structures would overlap and be unintelligible. To sort out the meanings, a person would have to walk among the structures to unscramble them. Then the structures would reveal the thinking that created them. . . ." Photograph: Eric Sutherland. Walker Art Center.