REVIEW

COMPUTERS IN THE HUMANITIES

J. L. MITCHELL, EDITOR
University of Minnesota

Edinburgh University Press
University of Minnesota Press
Minneapolis 1974

Note: The table of contents appears in AJCL, Microfiche 14, frames 67-69, and summaries of several contributions appear on the same fiche.

REVIEWED BY EDITH J. HOLS AND DAN BURROWS
University of Minnesota, Duluth 55812

Computers in the Humanities is a selection of papers from 115 presented at a conference on computers and the humanities, 1973, at the University of Minnesota in Minneapolis Its value is both as a bank of ideas and as a cross section of the very broad area at the confluence of those two disciplines The variety of interests displayed here is an indication of the considerable breadth of opportunity for further exploration Its editor hopes it will be "appropriate for use in such courses in 'Computers and the Humanities' as are now found in many major American and European universities " As an idea book for such a

course it should serve rather well Certainly it has little competition

Scholars in the humanities are only beginning to see the computer as the greatest thing since the invention (discovery?) of the stylus, and recent extensions into languages more agreeable to the soft sciences and to the arts have made more attractive a continuing growth on several fronts. Concordances and indexes are growing in number and in informatory powers, large banks of texts are being created, and new and imaginative uses of such stores are being attempted More sophisticated methods of analysis are being devised, and programs are being developed which perform increasingly complex tasks Much of this work needs to be done only once, so in order that efforts not be duplicated it is urgent that information on work completed or in progress be publicized. The Minnesota conference and the book which has come from it share some things that have been done

The collection is a mixed bag in more than one sense. Disciplines represented include music, art, archaeology, literary analysis, dialectology, language history, lexicography, and Roman history Papers vary in length and in readability

> With this example, we used a multiplicative application probabilities model which was far more consistent with the data than a non-application model, as measured by a chi-square comparison of predicted versus observed frequencies (D Sankoff and P Rousseau, p. 7).

> lyric poetry tends toward introversion and internal movement, towards travel through Shelley's 'caverns of the mind,' that 'thought can with difficulty visit' (C. Martindale, p 57)
>
> Automating poetry is, on the whole, a fairly harmless activity (R. W. Bailey, p 283)

And, inevitably, the papers vary in importance

The only serious fault we find in the collection is that editorial comment is insufficient This is especially true in the section, 'Art and Poetry,' where titles and appropriate credits are given, one or two completed designs shown, but no textual advice on the nature of the programs used

We have chosen to survey some of the papers herein from two points of view, first as a list of accomplishments, and second as a source of inspiration for new accomplishments Readers outside the field, who are aware that something is going on, but who are not quite sure what, will find that the computer is being used most in doing the things it can do best, that is, indexes and concordances. But its use is being extended to more complicated tasks, its symbology is being extended to new alphabets These are some of the things which are being done

THE COMPUTER AS WORKHORSE

The computer serves best as a workhorse, doing things that are at least tedious and time-consuming, sometimes impossible, for the unassisted human mind P Bratley, S Lusignan, and

Francine Ouellette, in "JEUDEMO a text-handling system," describe just such a system, remarking that "the computer's main contribution to literary endeavour is in the provision of concordances, word-indexes, and rather unsophisticated statistics." The text processing system they describe is JEUDEMO, designed to perform "typical jobs " The system allows for several types of texts, including scripts with several actors, scripts which can be subdivided, for example, into several acts. Output ranges from vocabulary lists to a Key Word in Context (KWIC) listing, which gives the researcher a good idea of the use of a word within the text. JEUDEMO is designed to " meet some of the basic requirements of the user from the humanities, enabling him to realize some rather sophisticated text processing operations with a high degree of computational efficiency and comparatively little effort, thus freeing him from much routine work and allowing his creativity to be applied at a much more fruitful stage " The writers instruct the reader in some detail in the use of the system, including the use of commands, and including an illustration of a command sequence Options available in the program are given and illustrated

BEYOND THE CATALOG

But we can go beyond the catalog Sara R Jordan's METQA, described in "A computer program that learns to understand natural language," structures and adapts its own memory to reflect experience According to the author, most computer programs

that attempt to teach natural languages use dictionaries and built-in linguistic information But METQA "learns" over time by the usage of each word In terms understandable by the non-specialist, Jordan gives a detailed account of the memory structure developed by the program Briefly, the trainer makes his input, without spaces. The computer responds either with an equivalent, or with a question ("What is it?") If questioned, the trainer gives an answer, which is stored and assigned to one of a system of nodes, connected by labeled links Here machine acquisition of language leans in the direction of human acquisition of language

A reminder that the machine has limitations in that respect is struck by R W Bailey in the liveliest paper of the book, "Computer-assisted poetry the writing machine is for everybody Bailey finds that mechanical techniques for the production of poetry work very well, but do not produce poetry The strategies he describes are of constructions based on typical poetic patterns, without artistic direction An example,

> Furtive is mahogany
> And delirious are the shadows of its pants (p 288)

The computer manipulated by the artist is another matter Last in the book, captioned and credited but not explained, are four examples of computer art Ruth Leavitt's "Computer Graphics' and "SPLAT. A computer language for artists," by D Donohue and J Skelton, are better than capable demonstrations of art work

generated by the artist-in-command, the computer as tool. Unfortunately, space was severely curtailed and explanations of the programs were not included.

MECHANICAL PROBLEMS

This book includes accounts of some attempts to solve problems peculiar to computer use and problems which arise on extensions of computer use. One is the need to feed alphabets other than Roman into the computer.

K. L. Su, in "The creation of a set of alphabets for the Chinese language," shows a set of symbols which can be combined to represent nearly all Chinese characters. The symbols chosen by Su are still larger in number than the number of characters necessary for other languages: the keyboard will have 256 keys, including 210 symbols, 26 English letters, 10 numbers, and 10 notations and punctuations. But without some such breakdown the tens of thousands of Chinese characters could not be used at all The author admits a "slight loss of readability" but "not of any grave consequence."

"MUSTRAN II: foundation for computational musicology" (J. Wenker) describes a system of notation which can be used in recording, and subsequently in reproducing, a musical score.

Another mechanical problem, the necessity that products of research be available and usable in many contexts, is broached by D. Sherman in "A common structure for lexicographic data." He suggests that a standard structure would help solve many

problems in exchange of data. The common structure that he finds and uses in data records for Webster's Seventh Collegiate Dictionary, is a revision of Machine Readable Catalog (MARC) which is already in use in libraries of the United States and England. Sherman's WEBMARC permits the addition of phonetic information

Problems of dimension in catalog assignment are solved to some degree by D. D. Fisher in "An information system for the Joint Caesarea Maritima (Israel) archaeological excavations." In answer to the necessity for precise recording of archaeological finds he catalogs complex structures three-dimensionally. A fairly uncomplicated data base kept on punch cards helps keep information of finds during an excavation in a usable form, with a precision not easily achieved. The data base can later be used to compare artifacts and to work with scientific shapes to aid in chronological studies.

The computer has increased what is for other reasons a serious problem--the paper surge. W. P Cole, in "Computer-output microfiche in the Catalog of American Portraits" comes up with the inevitable solution--microform. The method he describes for outputting on microfiche (COM-fiche) involves no paper output Copies for duplicate sets are said to be comparatively low in cost.

SOME DIRECTIONS FOR COMPUTERS IN THE HUMANITIES

As an idea book; Computers in the Humanities must give some leads to future work. The best example of this and one of the

most worth-while papers in the collection, is the editor's own study of "The language of the Peterborough Chronicle." His work with the Chronicle is a continuing effort which is attempting to do two things: to establish a definitive text, and to make contributions to a grammar for Old English. What makes his paper especially interesting is the full and clear account of the way he is going about it.

CONCORDANCES AND DICTIONARIES

The obvious and immediate task for the humanities is the assemblage of concordances and dictionaries. Two kinds of program are presented in Mitchell's book 1) the program which simply gets the information out and prints on order, and 2) the program which uses the information in some kind of analysis.

It is not surprising that a concordance and a dictionary of Shakespeare would be an early choice. M. Spevack, H. J. Neuhaus, and T. Finkenstaedt describe the operation of "SHAD a Shakespeare dictionary." At the time of their writing, the dictionary was being prepared with the use of an already existing concordance (Spevack's Complete and Systematic Concordance to Shakespeare), the magnetic tapes Urwesen (containing all of Shakespeare), a Computer Dictionary (CD) composed of entries from the Shorter Oxford English Dictionary (SOED), and LEMCA, a semi-automatic process of lemmatization. SHAD will have, according to the writers, "almost unlimited possibilities for presenting questions and so eliciting still more information." Examples of the

information which may be given for the word gasp include these things: The word occurs six times, contains no grammatical ambiguity, is composed of one free morpheme, it functions as a nominal, is not a variant spelling, in uninflected. As a type, it has eight tokens (two as a verb). It is classed as a noun in SOED. Its status is common, it is Germanic in origin, Old Norse, it first appeared in print in 1577. It occurs four times in history plays, five in plays of the 1590's, it is spoken by both men and women, by both major and minor characters. It is always directly preceded by one of these adjectives. last, latter, latest. Such information is only typical, not exhaustive.

Another project simple in structure but large in scope which can be attempted only with the aid of a computer is E J. Jory's word index described in "New approaches to epigraphic problems in Roman history " Jory has completed a word index of one volume of an already published sixteen-volume set of Latin epigrams. He proposes as an extension of his index a central data bank containing all the information about every single known Latin inscription.

LIMITS OF THEORY

It is probably not possible to set limits to the cataloging functions of the computer, and perhaps not necessary What is needed is an approximate line between what the computer can do best and what the human mind alone can do best. Bailey demonstrates rather convincingly that the computer cannot, unassisted,

write poetry. What he did not show, but that we believe can be shown, is that the poetry writing programs he criticizes are at least useful in understanding forms in poetry

Establishing limits is a matter of cumulative results from uninhibited projects.

Another kind of limitation, boundaries of a theory, may be discovered at any time. For example, semantic field theory is an attractive method of organization of meaning. The difficulty is that, extended beyond certain neatly arranged categories, where distinctions are clear, the method becomes unmanageably complex. E. R. Maxwell and R. M. Smith ("A computerized lexicon of English") describe a process which may stretch field theory far beyond its present limits, or which may show it to be so severely limited as to be of little value. Beginning with an undefined set of concepts called primitives (sameness, difference, motion, space, etc.), each word is processed, continuing then through finer degrees of differentiation which ultimately define it and distinguish it from all other words. Hit "would be defined by primitives: an event involving a motion against an object . . . by another object, the two objects being different " It would then be differentiated from such other words as punch, jab, and slap.

STYLISTICS

Beyond cataloging, the most urgent need for the computer in the humanities is in stylistic studies. We have already mentioned the establishment of definitive texts. Several studies

recently completed promise to make determination of authorship more certain. One is W. M. Baillie's "Authorship attribution in Jacobean dramatic texts," in Computers in the Humanities.

Baillie investigates two texts of Shakespeare s and two of Fletcher's, looking for automatic distinctions in the dialog of the two writers. EYEBALL lists part-of-speech and function categories wherein the investigator finds that several variables, including descriptive adverbs and complements, achieve a differentiation success rate of seventy percent or better. From the information he forms a graph that shows some consistency in discrimination between the two playwrights.

Baillie is aware of at least some of the difficulties in establishment of authorship. 1) The plays are not totally consistent. In certain scenes the distinction is negative, that is, the norms of one playwright are replaced by the norms of the other. Such scenes are not destructive to the credibility of the method, but rather they open up new questions of style and authorship. 2) The project is based on the assumption that a writer has an identifiable style, an assumption which is yet to be proved. 3) In a play each character may have a style distinct from others and from the author's personal style. That personal quality may or may not be given to any one or more characters and may have only an indefinite effect on the style of all of the characters. 4) Any stylistic study in breadth must assume that over a period of years a writer's style will change, that he may revise an earlier writing and so create a problem in authorship. All of

these difficulties must be considered, but not as insurmountable. It is just such complexities which make the use of the computer urgent and probably essential.

Another who attacks the problem of authorship is J. R. Allen, in "On the authenticity of the Baligant episode in the Chanson de Roland". His method is different: he makes a judgment on the basis of the frequencies of high frequency lemmata. In six out of nine cases the Baligant episode varies significantly from the others. Allen comments that this does represent a clear stylistic difference, but not necessarily a difference in authorship, since it may be the result of revision and a reflection of stylistic change in the single author. Poet A ≠ Poet A + X years.

Any stylistic study is conducted in a vacuum unless two items are in comparison, for example Writer A in Poem X vs Writer A in Poem Y, or Writer A vs Writer B, or Writer A vs Writer A + ten years. Eventually norms of a sort will be established. D. Ross, in "An EYEBALL view of Blake's Songs of Innocence and Experience" establishes standards of "simplicity" against which other poets' work can be measured. He makes a stylistic comparison of the two poem sequences, Songs of Innocence and Songs of Experience, making use of EYEBALL, which produces for him an "augmented text," noting for each word syllable length, grammatical category and function, and location in clause, sentence, and text. He finds the style of both groups of poems to be much alike.

Different from all these studies is that of C. Martindale, who considers categories of thought, as represented by vocabulary items, and in "The semantic significance of spatial movement in narrative verse: patterns of regressive imagery in the Divine Comedy," he finds evidence of stylistic progression. He makes use of the Regressive Imagery Dictionary (RID) which lists 3647 words "divided into thirty-six categories designed to tap primary process and secondary process thinking " His measures are secondary process thought, i.e., logical, reality-oriented, and goal-directed, in contrast with primary process thought, which is more primitive, free-associative, and sensation- and drive-oriented.

One last thing we note here is a direction toward a new meeting place between disciplines. I. A. Morton, in "Analysis of tonal music at the level of perception," describes a system of tonal analysis which deals with harmonic relationships between chords in similar contexts. Chords which are performatively identical may be ambiguous as to key until the ambiguity is resolved in an ensuing pattern of sounds. Morton insists on the psychological reality of music,

> . . . much of the 'meaning' in music is contained in a network of abstract relationships which is conceived in the mind of the composer and communicated to the auditor through a sonic description of it. The musical score, then, is but a visual encoding of the sonic image of the network. The network is reconstructed in the mind of the auditor . . .

Everything that Morton says in his paper about the perception of musical relationships has an echo in perception of speech relationships. The concept of music as being somehow parallel to speech in competence is not new, but his analysis of musical relationships draws in some verifiable points for comparison.

Here we have a renewed realization of what this collection of papers is really all about, a meeting of minds at the water hole.