DOCUMENT RESUME

ED 030 935 CG 004 177

By-Page, Ellis B.

Analyzing the Student's Natural Language for Guidance.

American Personnel and Guidance Association, Washington, D.C.; Connecticut Univ., Storrs.

Pub Date 31 Mar 69

Note-16p.; Paper was presented at the American Personnel and Guidance Association Convention, Las Vegas, Nevada, March 31, 1969.

EDRS Price MF-$0.25 HC-$0.90

Descriptors-*Communication (Thought Transfer), *Computational Linguistics, Computer Programs, *Computers, *Guidance, Linguistic Competence, Linguistic Patterns, *Linguistics

The author of this paper presents three arguments (philosophical, empirical, and linguistic) to make his point that the computer, far from being worthless with words, offers the brightest hope for the future management of the verbal processes so important in counseling and guidance. Philosophically, he argues, there is no deep support for bias against the machine, since, in any guidance situation, exact measurement must be taken by whatever means available. Computers can respond if there is insistence upon behavioral data rather than data concerning internal states, and operational definitions instead of idealistic ones. Empirically, the computer has proven itself valuable in many statistical demonstrations done by groups working independently of one another. The central linguistic problem appears to be in the area of transformational grammar or the relating of one statement to some transformed equivalent. Much work is currently being done in the area of approaches to meanings in the field of computational linguistics. Since counselors serve as information processors, and are presumably operating under "as-yet dimly-understood rules," the author feels that they can begin to make some practical use of the computer in language analysis. (Author/CJ)

ANALYZING THE STUDENT'S NATURAL LANGUAGE FOR GUIDANCE[1]

Ellis B. Page
The University of Connecticut, Storrs

In guidance, as in the rest of education, words are the coin of the realm. Our ethics and traditions in guidance normally restrict us to non-physiological procedures, and words seem to be almost all that we have left: We identify problems because of the words that students use. We analyze conditions through other words they use. We counsel in words, and we evaluate the effectiveness of our counsel through the consideration of more words. But we don't do any of this very much, or for many of our students, because there are too many of them, and we are too busy, and time is too short. This is all part of our traditional professional guilt.

True, the computer is beginning to help us in a number of important ways: It scores tests and questionnaires, and can make good analyses and predictions. Used properly, it can help us greatly with quantitative measures, can discharge class scheduling, grade reports, warning notices, absence accounting, and other routines. But for verbal interaction with students, the computer is apt to seem pretty far out. We cannot believe that it can help us much in a counseling interview, or can in any other way with our load of words. When it comes to conversation, we imagine that computers must be worse than useless.

The theme of this paper is that the computer, far from being worthless with words, offers us the brightest hope for future management of the verbal processes so important in guidance. To present this theme, I shall present three main kinds of argument: first, philosophical; second, empirical; third, linguistic.

I. Philosophical Arguments

The philosophical considerations will not take very much time here. The main arguments against a computer being used in conversation are of two sorts: (1) the alleged inability of the computer to handle conversation, and (2) the alleged immorality of the computer handling conversation.

Let us first dispose, if we can, of the immorality argument. One writer has been very critical of our own work because he believes that, in his words, reliability and dependability and objectivity are not appropriate goals in the evaluation of student themes.[2] Anyone trained in measurement must insist that such traits are very important, whatever techniques are employed, but it should not be worth space in a professional journal to argue this point.

A second argument to "immorality" is the claim that the computer cannot be swayed properly by human emotions and softness of judgment. The word "mechanical" is in this view associated with the notions of "coldness" and "unkindness."[3] On close inspection, this argument is seen as a variant of the argument to the inability of the computer, and should be so considered.

Concerning the alleged inability of the computer, the best discussion of the question of mental power is probably still the 1950 classic by the late Alan Turing,[4] which should be read by everyone interested in the question. The principal argument here centers, by fairly general agreement, around what Turing called "The Imitation Game," in which a judge is charged with the responsibility for telling whether answers to questions are generated by a man or by a machine. The usual consensus is that, if a machine can successfully imitate the man in this game, the machine can be said to "think" -- and consequently, be responsible for other duties ordinarily considered to require thought.

The basis of the imitation game should be familiar to anyone in behavioral science: It is the foundation of scientific psychology, the insistence upon data concerning behavior, rather than internal states, and the insistence upon operational definitions, rather than idealistic definitions. The game is thus thoroughly consistent with our own philosophical roots, which depend on defined behaviors, rather than on inferred conscious attitudes. Given such a commitment, then there is nothing in principle illogical about machines thinking, or conversing, or for that matter, about machines feeling emotions and expressing personal opinion about a counselee's processes -- at least, to the degree we are willing to attribute these activities to human guidance personnel other than ourselves.

The fact is that, given suitable programs, the computer is today successfully playing Turing's game all the time -- though of course not yet with the exhaustive realism which would in all circumstances fool the judge. But consider a few of the machine's current successes in education: certain kinds of official letter writing, the scheduling of classes, the reporting of test performance, and so on. The extrapolation required to imagine successful playing of the imitation game is much less today, than it was in 1950.

The most comical recent example of the imitation game, successfully played, was reported in a newsletter of the artificial intelligence group of the Association for Computing Machinery.[5] You may know of the MAD DOCTOR program, written by Joseph Weizenbaum of MIT, which does converse with a client, and which assumes part of the role of a nondirective counselor. If a client at the terminal types, "I am worried about so-and-so," the computer will respond, "How long have you been worried about so-and-so?" In other words, the MAD DOCTOR is in the professional school (or the alleged school) of Carl Rogers. In Cambridge, a computer scientist at Bolt, Beranek and Newman had been recently working with the DOCTOR program, when a company vice president came to the machine. Dan Bobrow, whose terminal it was, tells what happened:

> The vice president found a note on the PDP-1 teletype that read: "TYPE ALL MESSAGES INTO THIS TELETYPE, AND END THEM WITH A PERIOD." This note was left there by one of the programming bums who had just succeeded in making run a LISP version of the DOCTOR program (based on ELIZA as developed by Weisenbaum). However, the VP thought that this note indicated that this teletype was linked with the one in my home, since I often used the computer all night. Therefore, he started typing to "me" but in fact was typing to this version of the DOCTOR program. The following conversation ensued:

VP:	DO YOU THINK THAT I CAN USE THE COMPUTER THIS MORNING?
Teletype Response:	WHY DO YOU ASK?
VP:	I MIGHT BE ABLE TO MAKE SOME ADDITIONAL TELCOMP SALES.

Teletype Response:	WHY AREN'T YOU SURE?
VP:	MY PERSPECTIVE CUSTOMERS HAVE NEVER SEEN A DEMONSTRATION OF THE SYSTEM.
Teletype Response:	DOES THIS MEAN VERY MUCH TO YOU?
VP:	OF COURSE IT DOES.
Teletype Response:	YOU SEEM VERY POSITIVE ABOUT THAT.
VP:	THIS IS RIDICULOUS.
Teletype Response:	WHAT DO YOU MEAN THIS IS RIDICULOUS?
VP:	IS THERE ANY REASON WHY I CAN'T USE THE COMPUTER NOW?
Teletype Response:	TELL ME MORE ABOUT YOUR PERSPECTIVE CUSTOMERS.
VP:	PLEASE DIAL ME UP ON 491-1850

Note that after the last remark the VP did not obey instructions, and left out the period. Therefore, of course, the computer didn't answer him. This so infuriated the VP, who thought I was playing games with him, that he called me up, woke me from a deep sleep, and said:

VP:	Why are you being so snotty to me?
Bobrow:	What do you mean why am I being snotty to you?

The VP angrily read me the dialog that "we" had been having, and couldn't get any response but laughter from me. It took a while to convince him it really was the computer.

So perhaps the Turing Test is losing remoteness faster than we think!

A more general answer to the objection concerning inability may be found in the early writings of computer scientists, and in a good number of writings of very recent date. This answer is based upon mathematical and logical foundations, and concerns what has been called the "universal machine." This is surely not the place to recapitulate the reasoning, but there are a number of excellent and extensive discussions of it available.[6] Briefly, it may be demonstrated that a very elementary machine, much simpler than our current computers, may perform tasks of any conceivable complexity. The sole exception to this open range of possibility is a certain sort of self-examination, which we cannot be sure the human can perform, either.

Philosophically, there is really no deep support for bias against the machine, or for thinking it cannot handle the language necessary to proper guidance of students.

II. Empirical Arguments

It is fine to talk of theoretical capacity, but what of real demonstration of language capability appropriate to the guidance responsibilities? Here we shall refer briefly to a few of the studies which have demonstrated high promise in related fields.

Besides the Weizenbaum work with the MAD DOCTOR and similar programs, there are a number of question-answering systems which exhibit interesting features. Some of these are described in Computers and Thought and elsewhere. David Tiedeman, Allan Ellis, and the other Cambridge men involved with Information Systems for Vocational Decisions, have related such work directly to the central problems of guidance: knowledge of the student, knowledge of the vocational world, and the interface between them.

Such systems are related to the field of information retrieval also. And the extensive work done in retrieval, in such systems as Gerard Salton's SMART, constitutes a major source of sophistication for eventual guidance packages. Some results are statistically very impressive.[7]

Statistical analysis forms the center of certain major lines of work. Philip Stone and others have conducted many content-analytic researches in the behavioral sciences using the "Inquirer" system. And others have now demonstrated the effectiveness of computers and statistics in analyzing the humanities.[8]

Statistical techniques are also at the heart of our own work in analyzing student essays. Beginning in 1965, we have tried to simulate the performance of a group of expert judges, evaluating student writing on a number of important variables. The first work was in English composition, and we showed that a computer, with a simple set of criteria but with a fairly advanced optimization strategy, could rate essays on traits such as <u>content</u>, <u>organization</u>, <u>style</u>, <u>mechanics</u>, and even <u>creativity</u> as well as could the usual experienced teacher of English. That is, the computer ratings resembled the expert <u>group</u> ratings, as much as did the individual member ratings from that group.[9]

More recently, workers at Connecticut have turned to the evaluation of subject-matter content in what the student was writing. For this they have turned to essay examinations in various disciplines, originally in courses in Western Civilization in large university classes. Some results were reported by John McManus and others at the recent meeting of the American Educational Research Association.[10]

The approach used may illustrate one sort of statistical technique and is consistent with other strategies of the project: Final examinations were collected for a large number of students of history. The portion dealing with short items was of particular interest, where students were asked to "identify" such terms as "manorialism" and "Thomas Acquinas." Their answers were punched into cards, sorted by item and randomly reordered anew for each of eight independent judges, graduate students in history chosen by the department chairman. They were supplied by the professor of the course with acceptable "key" answers for the items.

The performance of these judges was quite erratic, as measured by intercorrelations -- surprisingly, because one might presuppose that such short answers would be obviously "either right or wrong." For the four items the median interjudge correlation was about .45. There were only five "grades" which the judges could award, and for each item examined, there were some responses which indeed received all five grades. The typical number of grades was three, with more responses receiving four grades than receiving only two. That humanist who believes in subjectivity should be happy with the human performance on this one!

The statistical approach was to find those words which optimally separated the high-rated responses from the low-rated responses, to find those terms which occur in the best responses, but not in the others. These were then tested (cross-validated) against other responses not in the generating sample. Using this and certain other simple statistical variables, the computer was able (1) to determine which answers were "relevant" to the topic of the question; and (2) to determine which of the relevant responses deserved which grade. The performance was believed by

McManus to be slightly better than that of the usual human judge (although this comparison is a rather tricky one to make).

The main message of this beginning work is clear: The statistical approach, using a variety of techniques can reasonably evaluate the substantive content of student work, and not solely the essay traits formerly considered.

Still another statistical demonstration of computer effectiveness in a "soft" area has been made at Connecticut. The Torrance Tests of Creative Thinking (TTCT) is, of course, an attempt to establish a standardized measurement of "creativity," based upon a manual evaluation of student verbal responses to certain stimuli. Paulus and Renzulli simulated part of the TTCT, using statistical techniques rather similar to those of McManus. They produced a cross-validity correlation with criterion of .69 (which would become .75 corrected for unreliability of the criterion). This first effort, so far applied to only one subtest of the TTCT, was not yet up to human standards, but could become much cheaper and faster than the manual procedure, and has much room for improvement.

This portion of the paper has only pointed to a few of the empirical results which argue strongly for the eventual effectiveness of computer language analysis in the guidance process.

III. Linguistic Arguments

The biggest limitations of the statistical methods described are not really in the field of statistics. Rather, they are in the raw material provided to the statistical programs, in the descriptions of the student language which the computer is able to generate. With better descriptions, we may expect increasingly effective statistical strategies, for statistics is a very effective and versatile computer tool.

When we speak of descriptions of language, we must depend on our knowledge of language structure, hence on linguistics. However, we do not need to restrict ourselves to the techniques and theories of conventional linguists, although their work may be an important starting point for our research.

When a computer is given the task of such description, the words will usually be looked up in dictionaries in the computer, and the strings will be recoded in some higher-level description. The entry words, those actually used by the student, are often called "terminal symbols," because they are the end (or at times the beginning) of the supposed generative process. The higher-level translations of these terminals may take on an almost limitless variety of forms. Some will be close to the ordinary "parts of speech," so that the terminal hit may be recoded as N for noun, or VT for transitive verb, and so on. Obviously, many terminals may take on a number of such grammatical descriptions, just as hit may be two different parts of speech, and may have a number of definitions within one of those parts (a hit may occur in baseball or in theater).

Parsing systems have been implemented for computer analysis of language, and some of these have been tried out, in a tentative way, on examples of student writing. In most current versions, the program tries out each different grammatical designation for each word, and tracks out each grammatical tree generated by the rules of the grammar stored in the computer. These systems can produce parsings which are extremely detailed and rich, and usually better than would ever be executed by 99% of the users of English. Unfortunately, however, they produce ordinarily far more parsings than are wanted, and automatic techniques for telling which are valid are not well developed.

One of the best current systems is that of Professor Susumu Kuno of Harvard, who kindly parsed for us 50 sentences from high school essays. These sentences were independently judged to be "grammatical" or "not grammatical," and we were interested to see whether the Kuno-Oettinger program was useful in predicting the human evaluation. From the crossbreaks table, the discovered contingency coefficient was .48, supporting strongly the relationship between human and machine judgment. However, for each of the "grammatical" sentences, the program produced over 23 parsings on the average, and the best we could hope for is that just one of these would be the "correct" parsing. And how do we tell the difference?

Even given a perfect description of the syntax of a sentence, however, still harder problems lie ahead. For how are we to recognize the relationship of one parsed sentence to another? For instance, consider two sentences: (1) Columbus discovered America; and (2) The Western Hemisphere was first explored by Columbus. We would probably consider them acceptable alternatives to a single question. But how is the machine to learn this equivalence?

In a way, this assault on "meaning," more direct than the word approaches described earlier in this paper, is the central problem in linguistics. Thus this work shares with linguistics a concern for what is called "transformational grammar" --

techniques for relating one statement to some transformed equivalent (as in active-passive verb changes). But it also has a great concern for problems of dictionary ("America" and "Western Hemisphere"), and of background knowledge, and of logic.

These problems are connected with much exciting material being treated currently in the fields of computational linguistics, and of artificial intelligence. Some of the most intriguing approaches to meaning can be read about in a new book entitled Semantic Information Processing.[12] In this Minsky has drawn together a number of studies, mostly by young and recent Ph.D.s. A few of the feats can be briefly recounted: The system by Raphael performs some "understanding" on a limited subset of English, and makes intelligent responses to questions. A sample dialogue follows:

Human:	EVERY BOY IS A PERSON
Teletype:	I UNDERSTAND
Human:	A FINGER IS PART OF A HAND
Teletype:	I UNDERSTAND
Human:	EACH PERSON HAS TWO HANDS
Teletype:	THE ABOVE SENTENCE IS AMBIGUOUS. PLEASE RE-PHRASE IT.
Human:	THERE ARE TWO HANDS ON EACH PERSON
Teletype:	I UNDERSTAND
Human:	HOW MANY FINGERS DOES JOHN HAVE?
Teletype:	THE ABOVE SENTENCE IS AMBIGUOUS. BUT I ASSUME (HAS) MEANS (HAS AS PARTS). I DON'T KNOW WHETHER FINGER IS PART OF JOHN.
Human:	EVERY HAND HAS 5 FINGERS.
Teletype:	. . . I ASSUME (HAS) MEANS (HAS AS PARTS). I UNDERSTAND.

Human: HOW MANY FINGERS DOES JOHN HAVE?

Teletype: . . . THE ANSWER IS 10.

And so on. In this dialogue with Raphael's program, the human is clearly teaching the computer about the world, or at least about certain kinds of symbolic relationships representing the world.

In the same book, Thomas Evans demonstrates a program which identifies visual analogies involving geometric figures, and passes an intelligence test on them at about the 15-year-old level. And Ross Quillian generates an impressive associational net for word meanings. And John McCarthy describes what he calls a "program with common sense."

All of these "semantic" programs share a common concern with representing reality in symbolic strings, and then with performing information retrieval, and logical inference, about the data in these strings. Space is too limited to describe these efforts in any detail, but the alert guidance worker must see parallels with the thought processes of his profession.

From all these approaches, logical, empirical, linguistic, there seems to be a clear message for us: We are after all, in our roles as counselor, guidance expert, or student, serving as information processors. We are presumably operating under rules, although we have still only a dim understanding of the rules, and we shall never understand the rules completely, any more than we shall understand completely any other part of the world around us. Yet we can begin to make some practical use of the computer in language analysis, still using only a portion of the present knowledge, let alone what we shall know in the future about language and its use. Surely, guidance practice in the future must benefit from such analysis. Words are indeed our game, and the computer can help us play it.

Footnotes

1. Paper presented at a Symposium entitled "Computers, Natural Language, and Guidance," at the Annual Meeting of the American Personnel and Guidance Association, Las Vegas, March 31, 1969. The Symposium was chaired by David V. Tiedeman.

2. Kinley Roby, "A Voice Against the Computer Correction of Themes." The Maine Teacher, 1966, 27, 1, 19-20.

3. Those interested in the question of whether computers can be "emotional" may enjoy the book edited by S. S. Tomkins and S. Messick, Computer Simulation of Personality. New York: Wiley, 1963. But the more important general argument is discussed elsewhere, as noted in this paper.

4. A. M. Turing, "Computing Machinery and Intelligence." Mind, October, 1950, 59, 433-460. Reprinted in E. A. Feigenbaum and J. Feldman, Computers and Thought. New York: McGraw-Hill, 1963. Pp. 9-35.

5. D. Bobrow, "A Turing Test Passed." ACM-SIGART Newsletter, December 1968, No. 13, 14-15.

6. Perhaps the most readable is by Robert Korfhage, Logic and Algorithms (New York: Wiley, 1966). But a more thorough treatment is by Marvin Minsky, Computation: Finite and Infinite Machines. Englewood Cliffs, N.J.: Prentice-Hall, 1967.

7. For instance, see G. Salton, "The Evaluation of Automatic Retrieval Procedures: Selected Test Results Using the SMART System." American Documentation, July 1965, 16(3).

8. See P. Stone, D. Dunphy, M. Smith, D. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. Cambridge, Mass.: M.I.T. Press, 1966. Pp. 651. For the humanities, the latest work is edited by L. Dolezel and R. Bailey, Statistics and Style. New York: American Elsevier, 1969.

9. Some early results from Project Essay Grade were first described by E. B. Page, "The Imminence of Grading Essays by Computer." Phi Delta Kappan, January 1966, 238-243. The most complete account of the work through early 1968 is found in E. B. Page and D. H. Paulus, The Analysis of Essays by Computer. U. S. Office of Education Final Report on Project 6-1318. Storrs, Conn.: Bureau of Educational Research, 1968. Pp. 269.

10. J. F. McManus, E. B. Page, and D. H. Paulus, "A Computer Grading of Essay Examinations in History." Paper read at Annual Meeting of the American Educational Research Association at Los Angeles, February 6, 1969. Other related work will be finished soon by D. Marcotte and S. Wirth. A new research project at Connecticut is planned to press further into such evaluation of substance than any previous work.

11. D. H. Paulus and J. S. Renzulli, "Scoring Creativity Tests by Computer." Gifted Child Quarterly, Summer 1968, 79-83. A regional grant from the U.S. Office of Education is permitting them to expand the research through the next two years.

12. M. Minsky (Ed.), Semantic Information Processing. Cambridge, Mass.: M.I.T. Press, 1968. Pp. 438.