The Place of Character Recognition, Data Transmission and Document Handling in A.D.P. Systems

By J. B. Stringer

This paper presents a personal view, that is intended to stimulate discussion and a fuller appreciation of existing facilities in postal transmission, etc., before calling for new devices and services.

Introduction

What does an A.D.P. system consist of and why do we use it? It is an aggregate of machinery for automatically performing part of the information-processing functions of a business or government organization. The motives behind its introduction are of two kinds, economy and expediency. A.D.P. may be economic by being cheaper than the equivalent manual system, by being faster, or by providing additional facilities or information which would be impracticable in alternative systems. It may be expedient by avoiding embarrassment due, for example, to staff shortages.

Concept of an A.D.P. System

I like to think of an information-processing system as represented by a series of concentric circles. On the outside of the system is the general public, the board of directors and so on, in other words the originators and receivers of the information processed. In the centre is the part of the system concerned with purely mechanical operations such as repetitive calculations. In between, are various levels of preparation, editing, transcription and communication which form the link between the outside and the centre. The simplest mechanical installation possible would only replace the innermost area; this is typical of conventional punched-card systems. Increasingly complex systems would spread out to include parts of the intermediate areas. The situation in A.D.P. at the present time represents a fairly early stage of this process. In the limit, of course, the whole of the system is mechanized and the originators and receivers of information outside the system communicate more or less directly with the machines.

When this stage is reached, it is obvious that the communication will have to be in forms which are natural to the human originators of the information. Since the only natural human methods of communication in universal use are speech and writing, it is evident that the input organs of the machine must be capable of accepting one of these. Since the problem of speech recognition is not among the topics of this Conference, I will not discuss it further except to mention that there are, in my opinion, several valid reasons why speech is inferior to writing as a means of communicating with information-processing systems. Thus, in this ultimate stage, I foresee the written document performing the function of primary input to the mechanical system ("mechanical" here is used in its broadest sense). The first operation in the machine will be to transcribe the documents into the machine form. This may be magnetic tape or purely electrical signals, but the point is that to achieve the highest efficiency in the use of the machine, all operations should be performed on information in the medium most natural to the machine. Having performed the transcription, the document can, and should be, filed away, or in some cases thrown away.

Data Preparation and Transcription

Turning back to simpler systems such as punched-card installations, the commonest practice is, in fact, very similar to this, that is, punched cards are prepared from the original documents by hand, and then all sorting, calculating, etc., is performed on the cards, without further reference to the documents except by the auditors. It seems to me, therefore, that systems in which original documents in human language are directly manipulated, apart from simple transcription, do not meet the "spirit" of the laws of efficient A.D.P., however much they may lie within the "letter." I believe that documentation in its present form (that is, pieces of paper with writing on) will dwindle in importance and quantity as A.D.P. increases its scope.

The transcription process, of course, implies the use of document reading which, in turn, requires character recognition. There are two main classes in the range of character-reading devices; first, those which involve the use of magnetic reading, and secondly those which read the information optically. The first kind inevitably requires at least one manual transcription phase in which the information written in the ordinary way is read and reproduced in magnetic-ink form. I have never clearly understood what advantage can be obtained from such a system, since it seems to me that the manual operation might just as well produce marks in the machine language rather than in some hybrid form which is convenient neither to man nor machine. (This seems to me to be an example of a more general principle, that the origin of the machine language should be as early as possible in the preparation of the data, in order that the fullest advantage can be obtained from the A.D.P. system.) We have already heard what seems to me a powerful argument against the use of these special character forms, and for the provision of more directly coded systems. The argument against such a system that "the customer doesn't like it" seems to me somewhat weak for two reasons: first, "how do you know that he doesn't?" and secondly, "why not educate him to like it?" I find it surprising, in these times when the public is constantly being manipulated to change its living habits through the powers of advertising, that real and beneficial progress is being restricted by such arguments.

Written Data

Turning rapidly away from this controversial subject, let us now consider optical reading of written information. The challenge to the research worker is to produce a device that will read handwritten information at a cost competitive with the cost of, say, card-punch operators. A typical example of the kind of application envisaged would be the automatic reading of National Health prescriptions for pricing and accountancy purposes. You may say that this example is absurd and extreme and is not likely to be realized in the foreseeable future. I would agree. Nevertheless, it is worth remembering what the ultimate aim should be. To turn to more practicable limits I think that the first useful goal would be a device to read handwritten numerals in fairly well defined positions on the document with the same economic restrictions. I would insist that the hand-written provision is essential, for the following reasons. First, the reading of typed or printed documents implies the preparation of such documents. During this preparation the information must have been in mechanical form in some way, and it would seem logical that this is the stage to generate the machine language. Secondly, there are definite requirements for the reading of numerical information submitted by the general public (not all of whom own typewriters). An application of this kind which springs to my mind is the processing of twenty-five million tax-deduction cards every year by the Inland Revenue. As a guide to the economic requirements of the problem I would like to quote a few rough figures. The cost of key-punching cards (including verifications) works out at about £100 per million characters according to one estimate I have seen. A character-reading machine in order to match this cost would have to sell for considerably less than £25,000. Remembering that this machine must read handwritten numerals, and that the recognition error rate must not be significantly worse than that of verified key punching, in my view these figures show that, although the use of optical reading devices is not really "on" at the present time, it should become a practical proposition in the not too remote future.

Communications

All I have said up to now relates to the functional aspects of A.D.P. without regard to the more sordid practical matters such as geography. The basic fact here is that the customers of the system tend to be scattered over a large area whereas the machine tends to be concentrated in one place. This introduces a communication problem of a different kind from that so far discussed, namely a telecommunication problem. It is trivial to state that this problem consists of transferring information from one location to another, remote from the first, as cheaply as possible, with complete accuracy, and as speedily as is necessary to provide the required service to the users of the system. If the A.D.P. system is for the control of stocks of spare parts for defence weapons, it is evident that the requirement of speed is the most important. On the other hand, if the job is a weekly payroll it is only necessary to ensure that the computing system has the information in time to return the results before pay day, which, with a week's lying time, can give about two days' transmission time in each direction. Systems like the first of these examples obviously demand some form of data transmission over electrical or radio circuits. On the other hand, systems of the second kind, which form I think the vast majority of applications, present and future, could be accomplished with the technique called "threepenny-stamp-and-postman." No one would deny that data-transmission systems with ideal characteristics would be nice to have for all these jobs, but one must consider practical systems and the costs of these systems.

As I said before, it is the nature of an A.D.P. system that the communication is between each of a number (often large) of remote local stations and the central computing unit. This implies that the transmission terminal equipment should have an assymmetric form, one end of the link (at the local office) being very simple and cheap, whereas the end at the computing centre can be much more elaborate. Transmission is often required in both directions, which would seem to imply different kinds of error-control systems for the inward and outward traffic.

Error Control

I would like to digress at this point to make a few remarks about error control in data transmission. It has been estimated that the residual error rate in operations such as key punching of cards or paper tape, after verification and checking, is about one wrong character in about 250,000 characters. (This does not, of course, include errors in the original data, which may be considerably more frequent. Such errors are largely irrelevant to this argument if one accepts that the function of the A.D.P. system is to process the information it is given as accurately as possible, not to tell us what information it ought to have been given.) If the data transmission system is not to degrade seriously the overall performance, it should have an error rate at least one order better than that of the data preparation. Translated into binary terms this implies an undetected error rate of about one in 15 million bits. The question of automatic correction is one which I think generates considerably more heat than light. Error correction is costly however it is performed, but it seems likely that automatic forward correction, if effective, is considerably more costly than correction by retransmission. This is because the effectiveness of any given error-correcting scheme is dependent on the nature and distribution of the errors that arise. It is comparatively simple to devise a scheme which will cater for a particular error characteristic, but the fact is that very little is known about the actual characteristics that arise in practice, and I suspect that these factors change not only from route to route but also with time on any one route. Therefore any error-detection and correction system should be able to deal effectively with any distribution of errors and is thus unlikely to be simple or cheap. Furthermore, automatic retransmission involves fairly large amounts of storage equipment at the sending end, and is thus likely to be too costly for "inward" traffic. This leads me to the conclusion that, in most cases, error detection only will be automatic, and correction will be performed by retransmission, probably manually, and that schemes of automatic forward error correction will prove uneconomic.

Conclusion

It would thus appear that data transmission is an expensive facility and is likely to remain so for some time since channels of inherently low error rate are expensive; and fairly elaborate and hence costly terminal equipment will be needed to achieve the performance figures quoted earlier. Data transmission is therefore unlikely to be competitive with postal or courier systems, except in the rare case where speed is essential. It is worth remembering that there are few places in Great Britain which cannot be reached within twelve hours by road or rail and that, with few exceptions at particular times in the year, postal delivery takes less than two days and in many cases less than one day. I calculate that a quarter of a million bits can be posted for sixpence in the form of punched tape, and the transmission is error free.

Before I sum up I would like to say that I am aware that I have been presenting a one-sided argument. These are entirely my personal views, and I have deliberately avoided presenting a balanced view in the hope that in this way more room will be left for discussion. The subject of A.D.P. is, I fear, plagued with fashion and fads; techniques are all too often adopted because they happen to be the latest thing, without sufficient regard for the real requirements of the system. One hears of inquiries for microwave high-speed data-transmission links from one end of the country to the other for use with service centre computers on research and engineering work.

In reply to the question implicit in the title of my talk I would say: use data transmission if you can't avoid it, use character recognition if and when a useful system becomes available, and don't handle documents!

Acknowledgement

Acknowledgement is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information from which these opinions have been derived.

Summary of Discussion

The Chairman, opening the discussion, said that Mr. Stringer had made some remarks on which, no doubt, some members would wish to comment; he was quite prepared to answer questions but would not object if somebody made a comment or criticism and somebody else from the floor answered.

Mr. L. T. G. Clarke (*Northampton College of Advanced Technology*) said that it was becoming increasingly difficult to move people about the country especially in the rush hour. Supposing it was said that everybody had to use such a transmission system as Mr. Stringer had been talking about, could he give any idea how much it would cost to put such a system throughout the country?

Mr. Stringer: What system?

Mr. Clarke replied that he referred to some form of data transmission system which involved telephone lines, including all forms of error correction.

Mr. Stringer said he was not qualified to answer that sort of question; there was no doubt that it would be very expensive.

Mr. L. R. Crawley (*Standard Telephones and Cables Ltd.*) said that a good deal had been said about the data transmission by messenger, but had they ever stopped to think about what happened if a messenger became involved in a road accident, or if the information were lost in the post.

If these forms of transmission were not duplicated there was the possibility of 100% loss of the information being transmitted.

Mr. Stringer said that he thought that in these cases the undetected error rate would be "extremely low".

Mr. R. O. Bennett (*T.S.U., H.M. Treasury*) said he would like Mr. Stringer to enlarge on his objections to the use of wide-band data transmission links for connection to a large central system, as in the case of remote research organizations sharing a common computer. What were the grounds for saying this was uneconomic and undesirable?

Mr. Stringer said that the particular example he had chosen was that where the cost of the transmission was expensive. One hundred miles of line cost many thousands of pounds. Secondly, between the central organization and the service computer it seemed to him the urgency of the data sent was not always so great; it did not always matter whether the research man got an answer this week or next; why not send it by post, which cost threepence?

Mr. J. J. Sharp (*IBM British Laboratories*) said that he must take issue with Mr. Stringer about this last statement. He did not think it was true that if one was remote from one's computer it was easy to operate. If one was remote from the computer he thought one would like to feel that it was just behind a brick wall and one could have a console and input and output of one's own, so that the computer could be operated as if it were there. The business of sending cards through the post and running them through the machine did not work easily on non-routine jobs. His Company had

had recent experience of this. In this case the remote operator was on the telephone all the time, because, for instance, the cards were not going through quite right, and the process was jammed. He was not at all sure that there was any need for a telephone connection of that sort. If the information was remote there should be some such system as was described by Dr. John Carr III where there was a network of computers with similar consoles, connected by U.H.F. links, scattered over Southern Carolina in the United States. He thought there was a need for this kind of application of teleprocessing.

Mr. Stringer said that this was a point on which Mr. Sharp and he were in disagreement. He felt that there were only two reasons for access to a computer, to test a program or to run a program. If one wanted to run the program when it was tested, that was all right; when the program was sent to an operator, he or she passed it through the computer and then sent the result back. That was true whether one was in the next room or a thousand miles away. To test a program, one did not go to the computer when one felt like trying a new program for a couple of hours. One had to work within a schedule, and the chances of getting the computer just when one wanted it were very remote. One had to wait; in any case one might as well wait the extra time for the test; he did not see that it made any significant difference.

Mr. Sharp agreed with Mr. Stringer. In the generation of present computers there were no multi-program services. One of the advantages of the multi-program service was that it did not have schedules and there was some advantage in this. He thought the argument could be advanced that as they could use computers without taking time on the whole machine, they could do other processes.

Mr. Stringer said it seemed to him that even with multi-programming the last thing one wanted to do was to put an untested program into the middle of the production.

Mr. Sharp said that surely with multi-programming one could come in on any one's program on one's own corner of the machine.

Mr. W. S. Ryan (*G.P.O.*), emphasizing that he was speaking personally, said that while he accepted Mr. Stringer's point that we should not handle documents, this was not the whole story. He would remind the Conference that the purpose of computers is to compute and, in commerce, this meant the maintenance of central records. These records are, of course, amended when a transaction has taken place, but during the interval between a transaction taking place and the central record being amended this latter record was not a true record. Decisions based on it were, therefore, suspect and could lead to a wrong course of action being taken, with a subsequent loss of money. (This was not the only explanation for management making wrong decisions, and he offered for the contemplation of the Conference the thought that computers would enable management to make wrong, as well as right, decisions more quickly!)

What he was suggesting was that whenever a record was out-of-date the possibility of not making money, if not positively losing it, was present and he thought it would be a good thing if every executive had inscribed above his desk the phrase, "Records out of date mean waste of money."

That this was true had been demonstrated in the talk given on the IBM SABER system which showed that up-to-date records, made possible by data transmission, had produced financial benefit to the airlines concerned. As a further illustration he quoted the profitable short-term money market which depended on the institutions concerned knowing accurately their financial position. All this supported the case for rapidly up-dating records and he felt that this was the essential argument in favour of data transmission. He thought that the SABER system showed what could be done and the talk given by Mr. Wright on the STL machine suggested that it could be done cheaply.

Mr. Ryan felt in the light of what had been said in the previous three days that an ultimate concept of business communication could now be formulated. The ultimate, as he saw it, was that as soon as a transaction had taken place the relevant data would be passed to a computer which would up-date the central record. The data-bearing documents, if any, need not leave the point of transaction and, as far as he could see, character recognition had no place in this concept.

All this must be done at a price which made it profitable and he felt that this represented a challenge which all the organizations concerned, who were represented at the Conference, were well able to meet. Computer manufacturers would need to produce cheap, large-capacity, random-access stores; peripheral machines would be needed in very large numbers and must be inexpensive; means of transmitting data accurately and cheaply was the challenge to the telecommunications industry and in this context it was perhaps pertinent to remember that the average telephone line was available for 24 hours a day, but was used only for a few minutes—this spare capacity invited exploitation.

Mr. Stringer said he had thought the speaker said that there should not be a document at the origin. He agreed that if possible there should not be a document at the origin of the data; unfortunately, they lived in a world where few people in a transaction had access to data transmission, nor were they likely to have. Many of the documents and much of the information which was required would arrive from the general public, not just "you and me" but the man in the street who did not have access to data transmission. He sent in a form saying he wanted to reclaim such and such income tax, or pay such and such gas bill, and he did this in the form of a document, and to this extent there would always be documents; they were here to stay, it was the irreducible minimum as a speaker had said that morning.

Mr. Wright's demonstration was mentioned. He was very interested and intrigued by this, but, nevertheless, he was rather suspicious of it in the sense that it seemed that they were dealing with the general public and local offices in their hundreds and thousands: the amount of data which was transmitted, he suggested, would be quite astronomical for this device. He sympathized and agreed with Mr. Wright's object in making use of existing common facilities and that the transmission device was very cheap and good, but it was painfully slow. The maximum rate which Mr. Wright was able to obtain was two or three digits per second on an average, which was not very useful.

SABER was all right for the purpose for which it was designed. This was one of the applications in which speed counted minute by minute, but for the others it was a question of time. Let us secure up-to-date records by all means, but how up-to-date did this mean? Up to the minute would not pay; would up to the day be enough? Two days or even a week in many cases would be sufficient. It seemed to him that situations did not change so much in a day or two, that one could not afford in the majority of cases to be that much out-of-date on a situation. If some bodies could at present have information as little as six months out-of-date, they would be very pleased!

Mr. W. E. Norman (*IBM*) said that there was one point which affected small companies as well as large ones, and that was the question of security. If the defence of the West depended on an early-warning network over Western Europe, including England, one wondered whether we were doing our share to build up that network. Telecommunications in America were developed in the field by the independent telephone authorities and their customers, who provided the experience on which the Government could draw, just as research groups in the 1930's had provided the radar knowledge and know-how here in Britain on which the Government was able to draw in 1939 and 1940. But if industry did not now request data-transmission facilities, the G.P.O. would hardly be justified in building up the high-speed data transmission network upon which we might one day depend.

It was a great encouragement, therefore, to hear Mr. W. S. Ryan of the G.P.O. give such a clear lead to the meeting and show that the G.P.O. was well aware of this. One wondered if the invitation would be taken up by industry and commerce.

Mr. Ryan, still speaking personally, suggested that the lead should come from the executives of industry, who were finally responsible for money matters. What he would like to see was a generation of executives who would insist that when they arrived at their desks in the morning they should have available a statement of the state of business at the close of the previous day. Such a generation would say to the computer and machine manufacturers and the telecommunications industry, "This is what we want. Provide it!" This is a challenge which all concerned would accept and be capable of meeting. Having had that generation of executives, Mr. Ryan wanted a new generation which would say that at five minutes past five they would want to know the state of business up to five o'clock the same day.

Mr. Stringer thought Mr. Norman was reading rather more into what he had said than he had intended and certainly more than he personally believed. He did not think that development of data transmission for defence purposes would be hampered by any slowness in development of communications in their application to business. If the defence departments wanted it, they would get it; nothing was more certain than that. But before people started putting money, which, after all, was their money, shareholders' money and companies' money, into schemes which might never pay off, they should know what they were doing.

Mr. C. P. H. Marks (*Ministry of Aviation*) took up the point of the presentation of information which should be on the executive's desk at 5.5 p.m., and said this was of no value whatever if he could not at that moment do anything with it. Information, whether accurate or inaccurate, referred to a situation at a particular instant or at some period of time before then. It had no intrinsic value, unless some action resulted, and that was something of which one tended to lose sight. If it was accurate and timely enough for a particular purpose and could be got at a reasonable price, it was valuable. What should not be done was to pay an exorbitant price for extremely up-to-date or accurate information unless there was a positive and specific advantage in doing so. He was afraid this fact was not always recognized.

Mr. Stringer said that what the executive would probably like to have at five o'clock was not the state of the business today but a forecast of the state of the business at 5.5 p.m. tomorrow!

Mr. W. E. Norman (*IBM*) wished to answer Mr. Stringer's suggestion that money should not be put into something which would not pay off. If it was left to others to take risks and make experiments, they might have to spend money, but they would gain experience which would never be acquired in any other way. The G.P.O. would have no reason for building up an efficient data transmission network if industry did not prod the G.P.O. to make the facilities available.

Mr. Stringer said that it seemed a disservice could be done to progress in rushing in with a half-baked system and half-baked ideas and finding it would not work. It might be ten years before one tried again and this was what sometimes did happen.

Mr. F. Shore (*Automatic Telephone and Electric Co. Ltd.*) said that some of Mr. Stringer's earlier comments were by way of giving a specification. He said it cost a fortune to transmit a set of characters, whereas it would cost only sixpence to post the tape—but tape might be months late or even lost, and the same applied whether it went by bicycle or aircraft! In a well-designed transmission system one was aware that transmission had commenced and if anything was missed one did something about it. If Mr. Stringer was prepared to pay for the equipment to any specification, it could be provided, but at a cost; equipment is designed to the requirements of the user, and he did not accept the argument that people were rushing in and spending money on expensive equipment without considering what it was going to save.

In dealing with industry, he found himself dealing with departments who were conscious of every penny which they spent, and if a system was not going to produce a saving in money or in time they were not interested in it. To give one example, recently a certain organization was considering the use of telex as against relatively lightly loaded data transmission, for the particular volume of data which they wished to transmit over the public network and at the error rate they required. The cost per 1,000 characters by data transmission was 18d.; the total cost including overheads and every single consideration taken into account for telex transmission was 21·5d.; he did not think 6d. per roll of tape through the post took into account the overheads. As to the case of answers being required quickly, there were many instances where an answer was required very quickly for very good reasons. Instances were quoted, including a case of a company stocking a large variety of merchandise and distributing it to many places; that organization could cut down its stock holding, which was money lying idle. There was a big extension in credit buying in this country and sooner or later we would rarely need to handle money; people would have cards with a code number and when they went to buy something the retailer would punch a few holes and from a central organization would come the credit rating for the amount the retailer could safely let the customer buy.

Another good example of a possible application would be the Post Office banking system. There appeared to be nothing to stop a dishonest person from forging these documents, going into the Post Office, drawing out some money and going into another and drawing more; but if the postal clerk had rapid access to the central office and could get an answer saying how much money was available, this could not happen. Supporting a previous speaker, and in answer to the cynics who contend that a postal service is adequate in the computer age, one is reminded of a quotation from Edmund Burke—"All that is necessary to assure the triumph of evil is that good men do nothing."

Mr. D. A. Young (*Elliott Bros.* (*London*) *Ltd.*) said that, firstly, he would like to emphasize that what he now said was his own opinion and nothing to do with Elliott's, as his company had not yet started a project in character recognition.

He was very pleased to hear someone emphasize the importance of really versatile equipment which was capable of reading most reasonably legible kinds of hand-writing as well as typescript. He would like also to agree with Mr. Stringer in saying that reading of hand-written characters was not as far off as some people had suggested.

It would be possible to devise an ACR equipment which employed *innate* specific analysis of character patterns without the use of a computer. An innate specific analysis mechanism implied the use of a comparatively small number of recognition criteria. Otherwise one would require to build into the machine all the criteria that the perception technique derived and stored automatically, thus nullifying the advantages of the innate system. There were probably not many ways in which one could derive such a set of criteria, particularly in view of the need to represent them in terms of logical hardware. One of the ways in which this could be done, however, was to classify each character under a semantic definition, which delineated the features and limits of the semantic set of character patterns of which it was a member. All block capital B's, for instance, or all lower-case b's, had certain features and limits of pattern which enabled them to be distinguished from all other sets of features and limits of pattern. By a lengthy process of pattern analysis and computer simulated experiments in discrimination, it was possible to provide definitions of the semantic features and limits of all upper- or lower-case character patterns, whether typescript or disjointed handscript. (Those interested in further details of this procedure were referred to an article by the present speaker in the January 1960 issue of *Electronic Engineering*.)

Having devised a way of obtaining the discrimination criteria, it was necessary to design a system of logic which could embody them as an innate specific analysis mechanism. This would be closely linked with the hardware for sensing the patterns, and each would to some extent determine the nature of the other.

In this connection it was an interesting feature of ACR that useful ideas could sometimes be gleaned from otherwise quite unconnected disciplines such as zoology and psychology. Dr. Stuart Sutherland, for instance, of the Institute of Experimental Psychology at Oxford, had been making a study of the way in which the octopus recognized shapes. There was considerable evidence, both anatomical and behaviouristic, to show that the animal used some form of innate specific analysis of patterns in terms of functions of their horizontal and vertical projections and, probably, one other parameter. The sorts of patterns which had been presented to the octopuses in learning experiments were not unlike alphabetical characters, and it was from the results of these experiments that Sutherland had been able to propose a discriminatory mechanism which must be similar to, or the same as, that used by the octopus. J. Z. Young had, at the same time, put forward a more detailed explanation of the functioning of the octopus's visual "logic," as evidenced by histological and neurological data, which, in itself, may ultimately prove helpful in the understanding of our own human, optical pattern recognition and reading processes. There was, for instance, a particularly interesting layout of retinal light-sensitive cells which suggested a system of recognition logic that might be of assistance in the design of artificial pattern recognition systems. In general, since handwriting and typescript have evolved to be read by the human visual system, it was reasonable to assume that workers in the field of artificial pattern recognition might study that system, in detail, to their advantage.

Mr. J. W. Freebody (*T.S.U., H.M. Treasury*) said that Mr. Stringer had been deliberately provocative that afternoon to stimulate discussion and had certainly succeeded in doing this. Mr. Stringer was on his staff and there was not the slightest need for the secrecy implied by the Chairman concerning the fact that Mr. Stringer was a member of the T.S.U. Furthermore, Mr. Stringer was free to say anything he liked, provided he made it clear, as indeed he had, that he was expressing personal views and not those of H.M. Treasury.

There were a number of points on which he would take issue with him, for instance when he said that one should have machinery automatically to read documents such as doctors' prescriptions. He fully supported the need for scientific advance, and it was right that some scientific effort should be devoted to the investigation of problems of this sort. If machines were found not to be needed, or were uneconomic for reading handwriting, there was no doubt that a great deal of useful knowledge having other applications would result from the research. He agreed with what had been said, that the thing to do was to simplify the problem. It was difficult enough for humans to read each other's handwriting at times, so why confront an electronic device with the great difficulties of this problem if it can be avoided? Why not generate the data at the first point of its creation in a machine language form, so that it was simple to handle by computer; the need for automatic character recognition would then be avoided.

He had felt frustrated by the lack of progress in the United Kingdom in the design and production of data-transmission systems. In the British Post Office, it had been necessary for several years to buy substantial quantities of error-correcting equipment, for use on radio circuits, from overseas, because it could not be obtained from British industry. The Post Office had itself to undertake the development and research work to evolve the system which it required, and this eventually was further developed and the equipment was produced by a British contractor. He would like to see the United Kingdom foremost in this field, which was one of the most important fields of progress for the future. We had the scientific, engineering and manufacturing skills, and if we could not lead the world then at least we should try to become level with others; this would be an important advance. A good deal of progress had been made in recent years, but we were far from the system needed, and the standards of accuracy in data transmission fell short of what was required in many cases.

Speaking of A.D.P. systems in general, one must aim at equipment designs which were capable of giving reliable service *every* day for twenty-four hours a day. Our aim must be to secure truly integrated systems, in which most of the documentation which at present plagued one's life would disappear and most of it could be handled by machines. One should then only have to read the things which required decision.

Mr. Stringer's main appeal to us seemed to be "Do not be hoodwinked into thinking that data transmission must of necessity be a part of your A.D.P. system at the moment. In some cases physical transport of the data, e.g. by postal service, may be preferable." He would urge Governments, commercial users and industry to get together on the subject to work out system needs and standards. As far as codes were concerned there was need for some rationalization; that was difficult but there were now hopeful signs of some successful thinking and collaboration in this field.

Mr. Stringer concluded by saying that many of the things he had said he had said with his tongue in his cheek, and it had had the desired effect of encouraging discussion. This he found most gratifying, since he considered that the discussion was more important than what he had said in his paper.

The Chairman said he was sure it would be agreed that Mr. Stringer's provocative address was the note on which to close the conference. He would like to say "Thank you" for coming along in such numbers and making the Conference such a success.

The Conference then terminated.

Correspondence

To the Editor,
The Computer Journal.

Sir,

I was very interested in the recent articles of Brooker and Morris (1960, 1961) on an "Assembly Program for a Phase Structure Language." Their scheme is very general, but it suffers from the disadvantage that it is not possible to use statements as parameters in statement definitions—only one level and not a hierarchy of statements is allowed. This restriction could be serious if their program were used to translate a language such as ALGOL (Woodger, 1960).

In their articles, Brooker and Morris distinguish carefully between *phrases* and *statements*. The format or *syntax* of each is defined, but the meaning or *semantics* of a phrase is defined only when this occurs as part of a statement.

If the distinction between phrases and statements is removed (the term *class* being used to include both concepts), a simpler and more powerful scheme can be developed. We construct a dictionary of class identifiers giving the syntactic definition of the associated classes. Some of these classes will also be defined semantically and then the dictionary will contain a reference to this definition also. This semantic definition corresponds to the statement definition of Brooker and Morris. It will be obeyed interpretively by the translation routine to produce the compiled program. One can now easily allow any class to appear as parameter in the syntactic and semantic definitions. The definition of general types of statement (such as the ALGOL conditional statement) is now much more convenient.

The analysis record constructed by the expression recognition routine will now include, at the end of the record of each subclass that is defined semantically, a reference to this semantic definition. This will be used later by the translation routine.

As an example I give the definitions of [GE] which would result in instructions being compiled, which when obeyed would set the accumulator A equal to the current value of [GE].

Syntactic definition : [GE] = [± ?]T, [GE] ± T
Semantic definition : [GE]
→ 1 if [GE] ≡ [GE/1] ± T
Let [GE] ≡ [± ?]T
A = [± ?]T
END
1] [GE/1]
A = A ± T
END

As syntactic definition of a conditional statement we might take:

Syntactic definition : [Unconditional statement] = [jump N], [Y = [GE]]

Syntactic definition : [Conditional statement] = if [GE/1] ϕ [GE/2] then [unconditional statement]

which includes both the forms:

If $y > 0$ then jump N

and

If $a^2 + b^2 > a/b$ then $x = a^2 + b^2$.

The semantic definition is easy to construct once elementary conditional jumps have been defined. It would take the form

semantic definition : [conditional statement]
Let [conditional statement] ≡ if [GE/1] ϕ [GE/2] then [unconditional statement]
ts1 = [GE/1]
ts2 = [GE/2]
jump α_3 unless ts1 ϕ ts2
[unconditional statement]
$(\alpha_1 + \alpha_3) = \alpha_2$
END.
(here α_3 contains the number of an unused label).

The scheme suggested here seems to be a little simpler logically, as well as being more general, than that of Brooker and Morris, but no account has been taken of the efficiency of translation. On a particular machine it might be found that one method can be programmed much more efficiently than the other, and that method would then be chosen.

J. M. Watt.

Computer Laboratory,
The University,
Liverpool 3.
3 March 1961

(*See* p. 176 *for authors' reply*.)

References

Brooker, R. A., and Morris, D. (1960). "An Assembly Program for a Phrase Structure Language," *The Computer Journal*, Vol. 3, p. 168.

Brooker, R. A., and Morris, D. (1961). "Some Proposals for the Realization of a Certain Assembly Program," *The Computer Journal*, Vol. 3, p. 220.

Woodger, M. (1960). "An Introduction to ALGOL 60," *The Computer Journal*, Vol. 3, p. 67.