Session 11:   EQUIPMENT

QUESTIONS AND DISCUSSION

A.F.R.BROWN: I would like to ask Mr. Blickstein if the IBM STRETCH is 100 times as fast as the IBM 704.   Is this true for a program in which no multiplication and no division takes place, and in which floating-point arithmetic is not used?   Machine translation does not involve any of these three.

BLICKSTEIN: The basic machine speed is the basic cycle time--logical instructions will be much faster.   This factor of 100 is a weighted average for arithmetic-type calculations.

OETTINGER: Professor Oswald concluded his remarks this morning by stressing the great need for automatic translation.   I agree, but I am a little bit concerned.   I want to know what is the quality, the reli-ability of the material produced by the machine, the degree of postediting required, and what is the level of training of the posteditors,   at what-ever rate it may be--at 100, 000 words an hour or 3 million words an hour-- and whether it be on the STRETCH or the LARK.

BLICKSTEIN: As much as I agree with you that this is definitely the essential question in the field, this is, frankly, neither my problem nor the problem of the session.   I think it is the linguists' problem.

MERSEL: We are not in the position to build special-purpose machine-translation machines today.   I do not want a special-purpose machine-translation machine today or tomorrow.   I do want machines,   such as those Colonel Kellogg spoke about,   which can read text and which will do things that cannot be done today; but, as the computer art has ad-vanced--and even when it was in the state of pencil and paper--we have been able to solve our problems on the machines available.   I think the important thing is that you want, not a machine specifically designed for your problem, but a general-purpose machine.   You want it not only because you do not know what form your problem is going to take later but also because you want, if your machine breaks down, to have other machines available in the country.   Furthermore, even if a special-purpose machine might be cheaper to build than the first model of a general-purpose machine, you have many more general-purpose machines.   The actual cost of running your problem on a general-purpose

515

machine will eventually be cheaper, just because of the mass production of the general-purpose machine. The cost per individual unit is cheaper on the bigger general-purpose machines.

EDMUNDSON: I would like to comment on a point raised by Colonel Kellogg about what to do when the print reader comes to graphs or equations.   The suggestion has been made in this Symposium to digitize. I would argue that we ought to retain them in analog form, not digitized, shunt them to a delay line, and feed them in at the end of the translation process at the appropriate point.   Of course, we need a clue to tell us that we are about to reach such an expression, and this will take further investigation.   I urge that equations and graphs be kept intact in analog form as opposed to digitizing and then reconverting to analog form. Most of this Symposium has dealt with problems associated with the central data processor--the analysis of the input language, the trans- formation rules, and finally, the synthesis of the output language. Colonel Kellogg's remarks were directed against the initial step; namely, getting the text to the central data processor.   It seems unfortunate, but not too surprising, that no-one at this Symposium has talked about the third problem--what happens after the central data processor. Here we have the problems of the users, the dissemination, and the switching circuits required for thousands of users.   This is precisely where the techniques that we are developing in information retrieval appear.   I hope that in a future MT Symposium we will be able to talk more about this third phase, which is on the far side of the central data processor, just as at this MT Symposium we have had an introduction to what hap- pens on the near side of the central data processor.

HAYES: I do not like special-purpose machines, even special-purpose general-purpose machines.   However, having gone on record for that, I should like to examine the design of a special-purpose general-purpose machine for machine translation. One might ask what the characteristics of such a machine might be.   First of all,   you will notice that I say "special-purpose general-purpose".   I define a general-purpose machine as one with, at least, a self-modification capability and a certain mini- mum number of commands.   A special-purpose general-purpose machine is such a machine with a special set of commands.

You might then ask what built-in operations,   micro-programmed or handled in any other way, might be most suitable.   I do not think this

516

question has been posed,   but I think that it is worthy of consideration. It was raised in my mind most immediately by the comments on MIMIC. Some of the charts shown there (any pseudo-command, as a matter of fact) give much the appearance of a special built-in command.   This has been done with data-processing machines intended for business processing, where a build-in sort command has been mechanized.   The National Cash Register 304 was based on this principle.   To go back to the bias I stated initially, I do not think that such machines are economic, but, disregarding that, it is a nice idea.   I   raise this question as a topic for consideration.   It is of importance to examine this,   because if you look at the growth of machines from the strictly arithmetic computers that some people still think of as computers,   through the addition   of index registers for better internal control,   the addition of decision-making operations of wider and wider capability,   and the addition of input-output capability which is now becoming quite vast,   it might be considered that the next generation of computers could reasonably have some operations of quite general purpose.   Machine translation might be one of the sources of growth in this; certainly this kind of command would be very desirable.

Another area of equipment development concerns the size of the external dictionary.   In answer to the question: How do technique and equipment interplay? I would say the external dictionary is one of the critical areas of interplay.   Given a billion-bit memory of certain characteristics,   you might consider techniques very different from those that you would use if you had an 8, 000-bit memory for your dictionary.   Certainly in this interplay the dictionary is a major question.   However, having a large dictionary--as has been pointed out--is not enough.   You have to be able to get into it; and whether it operates fast and is big is not really relevant. The question is: How is it organized?   I believe that the capability in a memory--a dictionary, if you please--of self-organization might be quite useful,   particularly if it could arrange itself in terms of some order of relevancy so that given a context the dictionary would become arranged in an order of relevancy such that the words likely to be of interest would be the ones most easily accessible to you.   I raise this as some-thing of interest.

I would like to make one comment concerning the question of output which Dr. Edmundson has raised--not with respect to the information

retrieval aspects, but more with respect to the physical output aspects. I would like to suggest the following, which I do not think has been seriously considered.   Monotype tapes are used for input, and H. P. Luhn has gone to some lengths to describe this process.   Why not use them for output?   This creates the possibility of producing a letter-press print at reasonable cost.   Chemical Abstracts spends a million dollars a year on printing.   Of that three-fourths is typesetting cost-- labor cost, if you will.   Monotype output would allow you to create a letterpress or a plate and produce mass copies now at reasonable cost, which is what Dr. Edmundson is talking about.

TOMA: I want to mention that I missed a few computers, and I think we have to take all of them into consideration since we are using general-purpose computers.   At the present time at Georgetown we have one major program for the IBM 705, which was not mentioned; and we have a smaller program for the RCA 501, and a smaller one for the S 2000, which is the only one of the three computers which was mentioned. The reason that I want to mention the IBM 705 or the RCA 501 is not that I particularly like those individual models, but that the IBM 7080 and the RCA 601 are coming out.   They are two very good computers for the future, which we certainly have to consider.

KELLOGG: I want to answer a question about when a scanner or a character reader might be available.   If you talk to engineers you get the statement, "Give us the money, and we will have it for you in nine months to a year".   Of course, when you want to get their names on a dotted line, this is always a little trickier; but this is the kind of period of which we are speaking.   The other point I want to mention is that at the moment, in trying to commit research and development money, we have two choices over the long range: to put it into either general-purpose or special-purpose computers.   Certainly we are reluctant, in our position, to look at the special-purpose machine unless some pro-gram indicates that there is a tremendous advantage in developing special-purpose MT machines.

MERSEL: I would like to ask a question of Dr. Ware, who returned from Russia recently.   What kind of computers are they using for their machine translation?