Session *2:* CURRENT RESEARCH

QUESTIONS AND DISCUSSION

CANNON: I *see* that Professor Oettinger would like to be the first discussant.

OETTINGER: I would like to address the following question to Mr. Zarechnak. In the last issue of the Georgetown Newsletter a statement was made that an automatic translation system would be operating as a production system within a year. I would like to know whether this is the system that was described today, in which the mean number of errors is 13.4% and where the figures on individual errors are much higher, even as high as 30. 8%, for example, in the insertion of the preposition "of" ?

ZARECHNAK: I would like to answer that when we analyze the insertion of the preposition "of" in front of the noun, we count each occurrence in the output which is not properly executed, and we do not count the source which produces that particular occurrence. As you may well know, certain nouns in Russian may produce an insertion of "of", but there are also other reasons for inserting "of", and we count them all. Therefore, there is no contradiction in saying that a particular operation is 50% or even 80% a failure and then, toward the end counting all the outcomes and all the failures and seeing that the errors amount to 13%. I would like to remark that on the whole the percentage statement for machine translation is actually much more important for the person without analysis than for the quality of the product.

I find that when you read the machine translation output, you will be guided by the sentence as you read it, and you may or may not like it. If you know what the algorithms are by which that particular output has been produced, you will soon discover patterns of mistakes. Last summer we had Professor Summers, a chemist with us. He studied the chemical text output and after preliminary study he came to the conclusion that basically we are producing four types, or patterns, of mistakes. His observations were of use to us because they brought to our attention certain basic difficulties which had not been overcome at that time; some of them have not been even now. Perhaps I did not properly understand your question, but I do not see a contradiction in percentage as to the quality of the input.

OETTINGER: I would like to make clear what I mean. Far be it from me to attack a colleague on the question of mistakes. We all make mistakes; and the main purpose, perhaps, of such a meeting as this is to enable us to compare notes and analyze one another's mistakes, because even with the best goodwill one cannot always catch all of one's own. But I think there is a basic inconsistency between, on the one hand, the admission of mistakes -- which in a professional society is a perfectly reasonable and desirable thing -- and, on the other hand, the claim -- repeated time after time -- of a finished reliable product ready to go into production. These two things are simply not consistent, in my opinion.

BROWN: My colleague, Mr. Zarechnak, was referring to the output of tests essentially 8 months old; the prediction, which could conceivably be over-optimistic, refers to a time 12 months from now; and the mistakes to which he referred are essentially trivial. They look bad, but they are really trivial, logically. Why they should not be cleaned up in 20 months, and why this is an unreasonable claim, I do not know.

Now I have a question addressed to Dr. King. Why, having developed a new memory device, did you think it best to try also to develop an original computer, whose design is predicated on a certain theory of machine translation, and hook them up exclusively together? Why did you not hook the memory up to, say, an IBM 704?

KING: We do not see that the kind of things you do with the 704 could not be done some other way much more simply, but the fact remains that the 704 is available for the combination of a large memory and numerical processing. The other question about our developing a new type of machine is not relevant at all. I never mentioned anything about a new type of machine. Everything we have done so far has been done with a memory alone, and no further logic whatsoever. This apparently surprises you, but is a very interesting thing that memories are not quite so stupid as most people have thought them in the past.

SHERRY: I would like to address this question to Dr. King. He said that memory is the cheapest thing in the world. There seems to be some debate on this subject. I was wondering if he would have some

figures regarding this memory, perhaps the hardware cost, time involved in lookup, and also time involved in updating such a memory?

KING: I will answer those in reverse order. At the present time we could make a new dictionary in about half a day. This could be reduced to an hour if we really wanted to, but for our present purposes such speed is not very important. I might say that most of the time is taken in collating new material with the old material on the standard machine, whatever method is used. The random-access time to the memory now is around 25 milliseconds on the average. As to the cost of the machine, this is a question that is quite difficult to answer, because the cost depends on how many machines are going to be made. Since this is a general-purpose memory, with many other applications, the possibility is that many might be made. Nevertheless, the total cost is quite low. When I said memory was cheap, I was referring to the fact that basically this is a very low-cost memory. For instance, we are talking about production someday of words on the order of a mil per word, so that is a fairly good indication that the price is not fantastic. Half a billion bits is a very large number; in other words, if you have a large memory it is hard to fill it up, and to put in a lot of word pairs is not really eating away into any precious memory. Memory is always going to be available. If we should ever fill up the present memory, which I think would take many years and many dollars' worth of work, we can always change our coding to a Fano-Huffman code and double our capacity overnight. That is what I mean by saying that memory is cheap. Nobody has to worry about it.