AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
WASHINGTON 25, D. C.

SRI

6 September 1963

Mr. Ray Solomonoff
Rockford Research Institute
140 1/2 Mount Auburn Street
Cambridge 38, Massachusetts

Dear Ray:

Our old friend, Bar-Hillel, in the midst of four papers on algebraic linguistics and machine translation given at the NATO Advanced Summer Institute on Automatic Translation of Languages in Venice, Italy, July, 1962, has the following statements to make about inductive inference:

"But some people do talk about letting computers discover rules of grammar or expand an incomplete set of rules fed into it, by going over large texts and using 'induction'. But let me repeat, this talk is quite irresponsible and 'induction' is nothing but a magic word in this connection. All attempts at formalizing what they believe to be inductive inference have completely failed, and inductive inference machines are pipe dreams even more than autonomous translation machines."

I enclose copies of the pages in which he expands this argument.

Any comments?

Sincerely,

Harold

HAROLD WOOSTER
Director
Information Sciences Directorate

Enclosure
a/s

general remains to be seen, but the time has come for those interested in the mechanical determination of syntactic structure, whether for its own sake, for MT or for other applications, to get out of the self-imposed straitjacket of immediate constituent grammars and start working with more powerful models, such as transformational grammars.

Let me illustrate by just one example: one of the best programs in existence, on one of the best computers in existence, recently needed 12 minutes (and something like $100 on a commercial basis) to provide an exhaustive syntactic analysis of a 35-word sentence [47]. I understand that the program has been improved in the meantime and that the time required for such an analysis is now closer to one minute. However, the output of this analysis is multiple, leaving the selection of the single analysis, which is correct in accordance with context and background, to other parts of the program or to the human posteditor. But there are other troubles with using immediate constituent grammars only for MT purposes. In his lecture to this Institute, Mr. Gross gave an example of a French sentence in the passive mood which could be translated into English only by ad hoc procedures so long as its syntactic analysis is made on an immediate constituent basis only. The translation into English is straightforward as soon as the French sentence is first detransformed into the active mood. A grammar which is unable to provide this conversion, besides being scientifically unsatisfactory, will increase the difficulties of MT.

In the time left to me I would like to return to what is perhaps the most widespread fallacy connected with MT, the fallacy I call, in variation of a well known term of Whitehead, The Fallacy of Misplaced Economy. I refer to the idea that indirect machine translation through an <u>intermediate language</u> will result in considerable to vast economies over direct translation from source to target language, on the obvious condition that should MT turn out to be feasible at all, in some sense or other, many opportunities for simultaneous translation from one source language into many target languages (and vice versa) will arise. I already once before discussed both the attractiveness of this idea and the fallaciousness of the reasoning behind it. Let

of the situation is indicated and already is gaining ground, if I am not mistaken. More and more people have become convinced that the inadequacies of present methods of mechanical determination of syntactic structure, in comparison with what competent and linguistically trained native speakers are able to do, are not only due to the fact that we don't know as yet enough about the semantics of our language — though this is surely true enough — but also to the perhaps not too surprising fact that the grammars which were in the back of the minds of almost all MT people were of too simple a type, namely of the so-called immediate constituent type, though it is quite amazing to see how many variants of this type came up in this connection.

Leaving aside the question of the theoretical inadequacy of immediate constituent grammars for natural languages, the following fact has come to the fore during the last few years: If one wants to increase the degree of approximate practical adequacy of such grammars, one has to pay an enormous price for this, namely a proliferation of rules (partly, but not wholly, caused by a proliferation of syntactic catagories) of truly astronomic nature. The dialectics of the situation is distressing: the better the understanding of linguistic structure, and greater our mastery of the language — the larger the set of grammatical rules we need to describe the language, the heavier the preparatory work of writing the grammar, and the costlier the machine operations of storing and working with such a grammar.

It is very often said that our present computers are already good enough for the task of MT and will be more than sufficient in their next generation, but that the bottleneck lies mostly in our insufficient understanding of the workings of language. As soon as we know all of it, the problem will be licked. I shall not discuss here the extremely dubious character of this "knowing all of it," but only point out that the more we shall know about linguistic structure, the more complex the description of this structure will become, so long as we stick to immediate constituent grammars. It is known that in some cases transformational grammars are able to reduce the complexity of the description by orders of magnitude. Whether this holds in

addition to the training and learning, their brain is not a tabula rasa general purpose computer but a computer which, after all those hundreds of thousands of years of evolution mentioned before, is also special purpose structured in such a way that it possesses the unique faculté de langage which makes it so different from the brain of mice, monkeys, and machines. The fact that we know close to nothing about this structure does not turn the previous statement into a scholastic truism.

Years of most patient and skillful attempts at teaching monkeys to use language intelligently succeeded in nothing better than making them use four single words with understanding, and monkeys' brains are in many respects vastly superior to those of computers. True enough, computers can do many things better than monkeys or humans, computing for instance, but then we know the corresponding algorithms, and know how to feed them into the computer. In some cases we know algorithms which, when fed into the computer, will enable it to construct for itself computing algorithms out of other data and instructions that can be fed into it. But nothing of the kind is known with respect to linguistic abilities. So long as we are unable to wire or program computers so that their initial state will be similar to that of a newborn human infant, physically or at least functionally, let's forget about teaching computers to construct grammars.

Let me now turn to the first two items. What is the outlook for computers to master a natural language to approximately the same degree as does a native speaker of such a language? And by "mastering a language" I now mean, of course, only a mastery of its grammar, i.e. vocabulary, morphology, and syntax, to the exclusion of its semantics and pragmatics. Until recently, I think that most of us who dealt with MT at one time or another believed that not only was this aim attainable, but that it would not be so very difficult to attain it, for the practical purpose at hand. One realized that the mechanization of syntactic analysis, based on this mastery, would lead on occasion to multiple analyses whose final reduction to a unique analysis would then be relegated to the limbo of semantics, but did not tend to take this drawback very seriously. It seems that here, too, a more sober appraisal

which can be "trained" to perform certain tasks (such as pattern recognition) and indeed perform better after training than before, and though computers have been programmed to do certain things (such as playing checkers) and do these things better after a period of learning than before, it would be disastrous to extrapolate from these primitive exhibitions of artificial intelligence to something like translation. There just is no serious basis for such extrapolation. As to checkers, the definition of "legal move" is extremely simple and is, of course, given the computer in full. After a few years of work the inventor of the checker playing program [46] succeeded in formalizing a good set of strategies so that the training had nothing more to achieve than to introduce certain changes in the rank-ordering of these strategies. There never was any question of training the computer to discover the rules of checkers, or to expand an incomplete set of rules into a complete one, or to add new strategies to those given it beforehand. But some people do talk about letting computers discover rules of grammar or expand an incomplete set of such rules fed into it, by going over large texts and using "induction." But let me repeat, this talk is quite irresponsible and "induction " is nothing but a magic word in this connection. All attempts at formalizing what they believe to be inductive inference have completely failed, and inductive inference machines are pipe dreams even more than autonomous translation machines.

Now children do learn, as we all know, their native language up to an almost complete mastery of its grammar by the time they are four or five years old. But by the time they reach this age, they have heard (and spoken) surely no more than a few hundred thousand utterances in their native language (only a part of which are good textbook specimens of grammatical sentences). If they succeeded in mastering the grammar, apparently "by induction" from these utterances, why shouldn't a computer be able to do so? Even if we add the fact that these children were also told that so many word sequences were not grammatical sentences — whatever the form was by which they were given these pieces of instruction —,could not the same procedure be mirrored for computers? Well, the answer to these two questions can be nothing but an uncompromising No. The children are able to perform as splendidly as they do because, in