Preliminary Copy

ZTB-130

PROGRESS REPORT

RESEARCH IN INDUCTIVE INFERENCE FOR THE YEAR ENDING 31 MARCH 1959

R. J. Solomonoff

MAY 1959

CONTRACT AF49(638)-376

ZATOR COMPANY

140½ MOUNT AUBURN STREET, CAMBRIDGE 38, MASS.

ABSTRACT

PROGRESS REPORT

RESEARCH IN INDUCTIVE INFERENCE FOR THE YEAR ENDING 31 MARCH 1959

CONTRACT AF49(638)-376

R. J. Solomonoff

The concept of "language" has been generalized to include patterns of many extremely diverse types. This will make it possible to apply inductive inference methods that were originally devised for simple phrase structure languages to more complex pattern types, such as arithmetic problems, transformational languages, and translation between pairs of phrase structure languages and transformational languages. Work has been done on devising machines that will improve their own methods of operation. Work completed includes new methods for discovering the grammars of finite state languages and phrase structure languages, and methods for determining all possible phrase structures of an arbitrary sentence of a phrase structure language of known grammar.

Important work has been done on methods of discovering approximation languages. Routines of this type are not accurate all of the time but are useful for getting quick, approximate results.

Just about all the basic theoretical work has been done so that a machine to discover the grammars of phrase structure languages could be programmed. With what seems at present to be only a little more work a machine could be made to discover translation rules between certain simple language pairs, and learn to work several kinds of arithmetic problems. However, such a machine would be very slow in learning, would ask a very large number of questions, and would often be unable to learn a pattern if a simgle error were made in giving examples to it. Work is now progressing on a unified method of overcoming these difficulties.

TABLE OF CONTENTS

PROGRESS REPORT

RESEARCH IN INDUCTIVE INFERENCE
FOR THE YEAR ENDING 31 MARCH 1959

CONTRACT AF49(638)-376

R. J. Solomonoff

1. SUMMARY OF MAIN GOALS ACCOMPLISHED

Probably the most important single result of the year's work has been the generalization of the concept of "language" to include patterns of many extremely diverse types. This will make it possible to apply inductive inference methods that were originally devised for simple phrase structure languages to more complex pattern types. Some probable outgrowths of this generalization are: a machine that will learn to translate certain phrase structure languages in a suitable training situation; a machine that can learn to work arithmetic problems; a machine that will be able to discover the grammars of transformational languages; and a machine that can learn to translate between such languages.

Of much importance in the less immediate future is work that has been done on devising machines that will improve their own methods of operation.

Work that has been completed includes a new method for discovering the grammars of finite state languages and two methods for discovering the grammars of phrase structure languages. Also, two methods have been devised for determining all possible phrase structures of an arbitrary sentence of a phrase structure language of known grammar.

Important work has been done on approximation languages. This work eventually will make it possible to devise induction, mechanical translation and information retrieval routines that are not accurate all of the time, but are very useful for getting quick, approximate results, in cases in which the exact patterns are beyond the complexity of the language models being used.

In the early part of the year being reviewed much work was done in devising an adequate utility evaluation scheme for abstractions used in arithmetic learning. Although the system developed seemed adequate, this particular approach to inductive inference has been temporarily discontinued because the generalized language approach to inductive inference appears to be able to work arithmetic problems and accomplish all other goals of the old methods in a more unified manner. For example, we can now use the same machine for performing inductive inference on arithmetic, algebra, or linguistic problems. Some of the results of the earlier work on utility are still, however, used in one of the grammar discovery routines.

At the present time, just about all of the basic theoretical work has been done, which would enable me to program a machine to discover the grammars of phrase structure languages. With what seems at present to be only a little more work, I could make a machine to discover translation rules between certain simple language pairs, and learn to work several kinds of arithmetic problems. However, such a machine would be very slow in learning, would ask a very large number of questions, and would often be unable to learn a pattern if a single error were made in giving examples to it.

I am now working on a unified method of overcoming these difficulties, and will program a machine as soon as the difficulties have been adequately dealt with.

2. THE PRINCIPAL GOALS ACCOMPLISHED

2.1 The Generalization of the Concept of Languages, to Include Patterns of More Diverse Type.

This idea was first expressed in the paper "Linguistics, Mechanical

Translation, and Organic Cell Development".[1] My own impression is that this particular idea was the most important development in the year's work. Essentially, it makes it possible to view many, if not all, patterns of interest, as "languages" of a very general type.

Ordinarily, a "language" is a subset of the set of all sequences of words, these words all being in a certain set, called the "terminal vocabulary" of the language. A "grammar" is a method of describing this subset of sequences of words. A grammar may consist of a set of rules by which words may be put together to form sentences, or a grammar may consist of a test to determine if a proposed sequence of words is in the language being defined. Other forms of grammars may exist.

In one generalization of the concept of language, we allow the "words" in our vocabulary to be objects of any sort, and sentences in these languages may be configurations in spaces of any number of dimensions. In conventional languages, two words α and β may be joined together in only two possible ways: they may form the phrase $\alpha\beta$, or the phrase $\beta\alpha$. In simple one-dimensional languages of this type, each word has but two "interfaces" or "poles" at which it may be joined to other words -- the junction being between the poles of the words being joined. In multidimensional languages, each word may have several poles by which it may join to other words. A simple example of a three-dimensional language is the set of all organic chemical compounds. Here the words correspond to chemical elements, and the junctions between words correspond to chemical bonds.

In "The Mechanization of Linguistic Learning",[2] it was shown that the set of pairs of equivalent sentences of two intertranslatable languages can be usefully regarded as a generalized language of this new type. This particular method of viewing languages makes it possible to regard the problem of learning to translate from one language into another as being identical to the problem of discovering the grammar of a language of the new, general type. Inasmuch as we now have routines for grammar discovery, we have,

ZTB 130

then, some methods by which machines can be programmed to learn to translate between certain kinds of language pairs. There are, however, limitations on these learning routines that will be discussed later.

Through this generalization of language, the concept of "phrase structure language" takes on new power -- and many languages that were formerly expressible only as transformational languages may be conveniently expressed as phrase structure languages of this new, more general kind. This is of much importance, since we have methods for discovering the grammars of phrase structure languages, but we could not, until now, do this at all for any essentially transformational languages.

We can generalize the concept of phrase structure language further, if we allow the grammar rules to describe combinations of operators and other abstractions. Using this more powerful formalism, it appears that we can express all transformational languages as special cases of generalized phrase structure languages. Further work must, however, be done to confirm this apparent result. A consequence of such a result may be that we will be able to use existing phrase structure grammar discovery routines to discover the grammars of all transformational languages, and perhaps be able to learn to translate between pairs of such languages. Much (if not all) of English, and several other ethnic languages, is representable by the transformational grammar formalism.

Another important result is that we can regard many problems in arithmetic learning as generalized phrase structure patterns. In particular, the problem of learning arithmetic addition in binary notation, with "carries" not explicitly written, can be accomplished easily with these new methods. This particular problem had been quite difficult to solve using the older methods that I had devised for arithmetic learning.

Using a somewhat modified arithmetic notation, and a kind of hierarchical organization of learning routines, it has been possible for the routines to learn arithmetic multiplication and division with "remainders".

Another kind of generalization of the language concept is afforded by continuous languages. An example is a language in which the sentences consist of a particular geometric figure and all of its projective transforms. Such languages are useful in making inductive inferences about directly observable phenomena in the physical universe.

These generalized languages are extremely important, and the next report to be written will describe them in some detail. It will show how the linguistic approach will go far toward providing a unified methodology for a very large set of inductive inference problems.

2.2 Investigation of the Problem of Devising an Inductive Inference Machine That Will Improve Itself

Suppose we have an inductive inference machine that is working on a certain set of problems, and we have a method of assigning a value, G, to the effectiveness with which this machine accomplishes its task. The value of G may involve the speed as well as the accuracy with which the machine makes inferences. We can look upon the description of such a machine as a set of symbols. We may then vary the set of symbols (i.e. the description of the machine), and, with the resultant new machine, a new value of G can be obtained. By varying the sets of symbols and obtaining the resultant G values for each such set, we obtain a set of sentences in a generalized language. Each sentence in this language consists of a set of symbols and its associated G value. It should be noted that this language is somewhat unusual, in that it contains both discrete symbols (which are used to describe machines) and continuous symbols (i.e., G values).

We may then employ a second inductive inference machine to look at the sentences of this language, determine an approximate grammar to fit those sentences, and use this grammar to extrapolate to patterns of higher expected G values than have ever been obtained before.

Several methods have been investigated by which an inductive inference machine may be used to choose sets of symbols so as to attempt to maximize the G values to be assigned to them. In particular, it is possible to make the functions of the second inductive inference machine be performed by the first machine – thus the machine will be helping to improve itself. The conditions under which this self-improvement can take place without unstable operation have been investigated to some extent. Also, languages which consist of both discrete and continuous symbols, have been examined.

While self-improving inductive inference machines will, eventually, be of much importance, the work was temporarily discontinued since it is not at present a critical problem – i.e., we have other problems of more immediate importance that must be solved before our work can continue. However, the possibility of constructing a self-improving machine gives further motivation to work on simple pattern recognition routines. It appears at present that if an inductive inference machine is sufficiently complex, it can improve <u>very</u> rapidly in its ability to work problems through this self-improvement routine.

2.3 Two Methods for Discovering the Grammars of Phrase Structure Languages.

These methods are described in some detail in two papers: "The Mechanization of Linguistic Learning"[2] was given at the Second International Congress on Cybernetics, Namur, Belgium, September 3-10, 1958. "A New Method for Discovering the Grammars of Phrase Structure Languages"[3] is to be given at the International Conference on Information Processing, Paris, June 13-23, 1959.

While both methods do, indeed, make it possible to program a machine to discover the grammar of <u>any</u> language not employing context dependent substitution, they both require a "teacher" and must ask enormous numbers of questions. A "teacher," in the sense used, is a person or machine that will tell whether any sequence of symbols devised by the grammar discovering machine is an acceptable sentence in the language of interest. Of the two methods, the second[3] appears to be in a state of development that is closest to that needed for testing on a digital computer.

Each of these methods is also applicable to the discovery of phrase structure grammars of the generalized types that were referred to in Section 2.1. Since certain kinds of mechanical translation processes can be viewed as generalized phrase structure languages: (Solomonoff[3], Sec. 6) it is now possible to program a computer to learn to translate between a pair of such languages.

There are, however, several reasons why I have not attempted such programming. First, the languages types between which such translation can be done are fairly simple artificial languages (although parts of English and Russian form such a pair, and such a mechanical translation scheme would be at least as good, and probably much better, than the mechanical translation method demonstrated in 1954 by Georgetown University[4]). Second, the number of questions asked by the machine will be extremely large, so that a human being would be too slow to answer all of them. This means that we can only use truly artificial languages of known grammars, for which we are able to program a machine to act as "teacher." The method by which this may be done is discussed in Section 2.5. Third, if any mistakes are made in answering the machine's questions, it is very likely that the machine will be unable to arrive at any close approximation to the correct grammar. This practically rules out the possibility of using a human being as "teacher".

Methods by which these difficulties may be overcome, are discussed in Section 2.6

2.4 A New Method for the Discovery of the Grammars of Finite State Languages.

ZTB 1-30

This method was described in Solomonoff[1] Sections I.A and I.B. It is somewhat simpler than the method of Miller and Chomsky[5], and uses equivalence classes of phrases.

2.5 Two Methods for Finding All Possible Phrase Structures of Sentences in a Phrase Structure Language of Known Grammar.

This problem is of importance if we want to simulate the teacher (of Section 2.3) by a machine. It is also of importance in mechanical translation. The two routines have been studied in some detail.

One of them utilizes a kind of Boolean matrix multiplication. Ordinarily this would be quite time consuming, but by utilizing a special order on the IBM 704 or 709, we can perform the necessary analysis in about 10 seconds of computer time. By further refinement of the program, this can probably be reduced to 4 or 5 seconds. This routine is applicable to some fairly general types of phrase structure languages.

Another routine that has been analyzed is, in a certain sense, the fastest possible, but no estimates have been made for its expected speed. The routine is very economical of memory space, and may be simulated by hand, for many languages of interest.

For phrase structure languages of extremely general types, several other analysis routines have been devised – but none have been investigated in as much detail as the two methods referred to above.

2.6 Approximation Grammars.

Suppose we have a set of configurations of symbols, that conform to a complex pattern. One kind of approximate grammar would describe the set of configurations of symbols using less than, say, 20 grammar rules such that the rules "fit" the set of configurations of symbols in the best possible way – using some sort of stated criterion for "goodness of fit."

Such a grammar would be useful if we had two ethnic languages with fairly complex translation rules between them, and we wanted to find a set of approximate translation rules that were fairly simple. Simple translation rules of this sort would give rapid translations, using perhaps relatively little computer memory space.

Another application would be information retrieval – in this case, the assignment of descriptors to documents by means of machines. Here, we give the machine various documents along with the descriptor sets that have been assigned to those documents by humans. The relations between the descriptors, and the patterns of words in the documents are extremely complex, and humans are usually not completely consistent in applying such relations. As a result, it will almost always be necessary for a machine to look for approximate relations between the patterns of words in the documents and the descriptor sets. Again, using a suitable goodness-of-fit criterion, and a limited set of kinds of relations, a machine would be able to arrive at an optimum grammar describing the desired relations, and use this grammar to assign descriptor sets to new documents.

Another problem very closely related to that of devising approximate grammars is the problem of devising a grammar without being able to ask questions. Here the machine is given a large set of acceptable sentences in a given language, and asked to find a grammar that fits these sentences and is, in a sense, "most likely" – since there are an infinity of grammars that will fit any finite set of sentences "equally well".

The kind of grammar that is to be found is somewhat different from that which is ordinarily used. This kind of grammar assigns a relative probability to every conceivable sequence of terminal symbols. Sequences that are not sentences in the language are assigned relative probabilities of zero. For a phrase structure language, such a grammar could be described by giving different probabilities to the various possible legal substitutions for any nonterminal symbol.

ZTB 130 I have devised some formulae giving approximate a priori probabilities to all possible phrase structure grammars. Using this information, and

Bayes' Theorem, it is possible to define the grammar that is most likely to have given rise to the observed set of acceptable sentences. This, in a mathematical sense, solves the problem. However, it certainly does not tell us how to mechanize such a solution in any reasonable way – and I have been working on the problem of the physical realization of such a solution with existing equipment.

The solution to the problem of devising optimum grammars without using a teacher (i.e. without asking questions), would effectively deal with all of the objections listed in Section 2.3 to programming a machine to discover grammars using existing techniques.

2.7 Work on a General Method of Utility Evaluation.

The applications of the utility concept to inductive inference are discussed in Solomonoff[2], Section 2. An informal report[6] was written describing a fairly general system that seemed to work better than any other that had been tested. The system had been adapted, to some extent, to the problem of arithmetic learning, but it had not been adequately tested. Further work on utility evaluation has been temporarily discontinued. This is because the linguistic approach to inductive inference does not seem to need a complex utility evaluation system (though the grammar discovery method of Solomonoff[3] does use a rudimentary utility evaluation system) yet it is capable of arithmetic learning of great complexity, with much greater facility than my earlier methods which used a complex utility evaluation system.

3. SUMMARY OF THE PRESENT STATE OF THE WORK AND IMMEDIATE GOALS

At the present time, just about all of the basic theoretical work has been done which would enable me to program a machine to discover the grammars of phrase structure languages. With what seems at present to be only a little more work, I could make such a machine discover translation rules

ZTI
130

between certain simple language pairs, and learn to work several kinds of arithmetic problems. However, such a machine would be very slow in learning, would ask a very large number of questions and would often be unable to learn a pattern if a single error were made in giving examples to it.

In order to overcome these difficulties, I am working on a method by which the grammars of phrase structure languages may be discovered, without the use of a teacher, as is discussed in Section 2.6.

Also I am writing a paper that describes how the concept of language may be generalized, and shows how these generalized languages may be used to solve many different kinds of problems in inductive inference. In particular, I will show that many fairly complex pattern types can be viewed as generalized phrase structure languages, and that present methods of discovering the grammars of simple phrase structure languages are applicable to these more general phrase structure languages. Some types of mechanical translation, some (if not all) essentially transformational grammars, and many problems in arithmetic learning, will be shown to involve this type of generalized phrase structure language.

REFERENCES

1. R. J. Solomonoff, "Linguistics, Mechanical Translations and Organic Cell Development", Zator Company preliminary note V116, dated October 3, 1958 (to be published).

2. R. J. Solomonoff, "The Mechanization of Linguistic Learning", Zator Company ZTB No. 125; AFOSR-TN-59-246; ASTIA AD 212 226 (April 1959). Paper presented at the Second International Congress on Cybernetics, Association for Cybernetics, Namur, Belgium, September 3-10, 1958.

3. R. J. Solomonoff, "A New Method for Discovering the Grammars of Phrase Structure Languages", Zator Company ZTB No. 124; AFOSR TN 59-110; ASTIA AD 210 390 (April 1959). Paper to be presented at the International Conference on Information Processing, UNESCO, Paris, France, June 13-23, 1959.

4. Neil Macdonald, "Language Translation by Machine – A Report of the First Successful Trial", Computers and Automation, Vol. 3, No. 2 (February 1954).

5. G. A. Miller and N. Chomsky, "Pattern Conception", Air Force Cambridge Research Center Report No. AFCRC-TN-57-57; ASTIA AD 110 076 (August 7, 1957).

6. R. J. Solomonoff, "Utility Evaluation", Zator Company preliminary note V123, dated April 28, 1958 (to be published).