MICHAEL BARR

THE THEORY OF THEORIES AS A MODEL OF SYNTAX ACQUISITION*

We give a brief description of the mathematical theory known as the theory of equational theories in sufficient detail to show how it may serve as a model of a language acquisition theory. We see how, by restricting attention to a very special class of first order theories – the finitistic equational ones – we see how a machine could be programmed to produce an entire theory on the basis of a severely limited quantity of data.

Introduction

Much of the controversy surrounding Chomsky's linguistic theories is caused by his insistence on the necessity of innate capacities or ideas, a notion which seems to have a different meaning to every reader (and every writer). In fact it seems likely that what the critics find most objectionable may be notions (e.g. racial memory) that are entirely absent from and entirely unnecessary to Chomsky's theories.

To my mind, what must be supposed innate, in order to support these theories, is a strategy for organizing language data. I believe it might be useful to expose, by way of analogy, an example of a data-organizing strategy from one branch of mathematics. This strategy is sufficiently precise that it might readily be programmed into a computer. It takes as input data of a certain very restricted form and produces as output a theory (of an equally restricted form) that describes the data. New data will cause the theory to be modified (unless it is simply inconsistent with the previous theory. This is a technical condition which doesn't concern us.)

As far as language capacities are concerned there are three logical possibilities: that people are born already knowing their language; that they are born without language but with a special capacity for learning languages not used (or not much used) for other purposes; that language is learned by

* I would like to thank the National Research Council of Canada and the Ministère de l'Education du Québec for their support of this work.

0301-4428/78/0502-0013 $ 2.00

use of some general learning capacity in the same way as, say, mathematics is. The first alternative can obviously be rejected, although it is interesting to note that in the case of the analogous question for birdcalls, all three possibilities apparently occur. That is, some species are born with a species-specific call; some learn the location-specific call of the members of their own species nearby, but not that of other species; a few species are capable of imitating the calls of any species or even human speech.

Thus the question comes down to a choice between the second and third possibilities and I believe that the evidence against the third is overwhelming. By all evidence, ability to learn a language (*as a first language*) is present only in the first decade of life and then disappears forever. As the debacle with the "new math" has clearly demonstrated, the ability to learn math is not normally present to any great extent until adolescence and probably peaks during the 20's. I believe that this is the norm for any learning capacity – except that of language.

The strongest evidence remains, nonetheless, the disparity between the quantity and quality of the linguistic input and that of the output. In fact, if language were learned in the way supposed by the behavorists, the results I would expect are about what the primate language experiments have demonstrated (even allowing for the difference in innate intelligence): rather intensive training resulting in communicative behavior which is rather concrete and with *few or no grammatical complexities*. For it is the grammatical complexities and the ease with which they are learned that is the main point here.

There is nothing in the preceding inconcistent with either the mentalist or the mechanistic point of view. In fact I find the extreme positions of both points of view untenable and – given physical uncertainty – I find the less extreme positions indistinguishable.

It seems entirely reasonable to describe the syntax of a language as a theory of how that language operates. In the analogy I am going to pursue the analogue of sentences is to be mathematical structures of some kind (technically known is *algebras* or *models*) and the analogue of the syntax is to be the *theory* of which these models are structures.

Some fairly sophisticated (but not genuinely hard!) mathematics is required of the reader in order to pursue this analogy. Nearly forty years ago a new concept was introduced into mathematical thought, that of a *category*. Crudely speaking, a category is a class mathematical structures of some kind together with all structure-preserving mappings between them. Slowly, but with increasing speed over the past ten years, this notion has been supplanting set theory as the underlying conceptual basis of the working mathematician. (See [MacLane] and [Barr-Wells] for introductory accounts.) There is even a school of mathematicians, led by F. W. Lawvere; who would make it the theoretical basis as well. (See [Lawvere], [Johnstone] and the many references found there.) Since theory usually follows practice (eventually, anyway) it seems likely they will succeed. Thus the day may not be too distant that at least a smattering of category theory be a prerequisite to comprehension of modern mathematics.

It was mentioned above that analogue to a language is something which is called a theory or, more precisely, a finitistic equational theory. An equational theory (the finiteness condition will be introduced later) is a set of operations and equations between these operations in a formal sense. A *model* of such a theory is a set equipped with the operations and satisfying the equations. For example the theory of *monoids* has as operations a binary operation of multiplication and a nullary (or 0-ary) operation called a unit and is subject to associative and unitary identities. Both the operations and equations can be formalized (see appendix A) but it is easier to say what a model of the theory (i.e., a monoid) is. A monoid is a set M, together with a multiplication map, $x, y \to xy$ and a unit element $1 \in M$ such that for all $x, y, z \in M$, the following equations are satisfied:

$$(xy)z = x(yz)$$
$$1x = x1 = x.$$

There are many kinds of theories in mathematics. What make the kind of recognition device I am talking about feasible here is that it is not searching for an unknown example of an unknown kind of theory but rather it is searching from among all possible *equational* theories for a particular one of them. Moreover, as we shall see, it does not actually carry on a search in any sense but rather *constructs* the theory out of the data. In other words, the metatheory of equational theories is a data-organizing strategy of a particular kind.

The first section introduces the conceptual tool that is required to deal with this theory: categories (and functors and natural transformations). The second section describes theories. Section 3 introduces products and powers of sets, necessary for speaking of models in the fourth section. Section 5 finally describes how the theory recognition device actually functions and the sixth section contains a complete description of the analogy. There is a short appendix giving a formal description of an equational theory.

1. Categories

To simplify the exposition and to avoid unnecessary abstraction we define as a category here what is usually called a *concrete category*. By a *category* we mean the following data.

C1. A given class of sets $|\mathscr{C}|$ (called the objects of $\mathscr{C}$);
C2. For sets $C_1, C_2 \in |\mathscr{C}|$ a set (C_1, C_2) of functions from C_1 to C_2.

If $f: C_1 \to C_2$ belongs to (C_1, C_2) then we will say that f is a *mapping* of $\mathscr{C}$. These data are subject to two axioms.

C3. If $C \in |\mathscr{C}|$, then the function $i: C \to$ defined by $i(x) = x$

for all $x \in C$, belongs to $\mathscr{C}$. (This function is called the identity function or mapping on C).

C4. If $C_1, C_2, C_3 \in |\mathscr{C}|$ and
$f: C_1 \to C_2$, $g: C_2 \to C_3$

are mappings in $\mathscr{C}$ so is the function

$$g \circ f: C_1 \to C_3$$

defined by

$$g \circ f(x) = g(f(x))$$

for $x \in C_1$. (This is called the composite of f and g).

These axioms may be summarized by the slogans: "Identity functions are mappings," and "The composite of mappings is a mapping."

Here are some examples. Although the definition does not require this, we stick to categories whose objects are finite sets.

1. The category $\mathscr{P}$ of finite partially ordered sets (posets). A poset is a set equipped with a binary relation $\leq$ such that

(i) $x \leq y$ and $y \leq x$ if and only if $x = y$;
(ii) $x \leq y$ and $y \leq z$ implies $x \leq z$.

If P_1 and P_2 are posets, a function

$$f: P_1 \to P_2$$

is a mapping of $\mathscr{P}$ provided when $x, y \in P_1$,

$$x \leq y \text{ implies } f(x) \leq f(y).$$

2. The category $\mathscr{S}$ of finite inf semilattices. Suppose S is a partially ordered set with the property that for $x, y \in S$, there is a largest element, denoted $x \wedge y$, among all those z such that $z \leq x$ and $z \leq y$. In addition there is an element T such that $x \leq T$ for all $x \in S$. Then S is said to be an inf semilattice. A mapping $f: S_1 \rightarrow S_2$ is a function that satisfies:

$$f(T) = T$$
$$f(x \wedge y) = f(x) \wedge f(y).$$

It is an easy exercise to show that such an f is also a map of posets. (Hint: $x \leq y$ if and only if $x \wedge y = x$).

3. The category $\mathscr{L}$ of finite lattices. A semilattice L is called a lattice if, in addition, there is a smallest element F and for each $x, y \in L$ there is a smallest element, denoted $x \vee y$, among those w such that $x \leq w$, and $y \leq w$. A mapping of lattices is required to preserve both operations as well as T and F.

The reader will have perhaps noticed that every finite inf semilattice is actually a lattice. In fact, $x \vee y = z_1 \wedge z_2 \wedge \ldots \wedge z_n = z_1 \wedge (z_2 \wedge (\ldots \wedge (z_{n-1} \wedge z_n) \ldots)$ where $z_1, \ldots, z_n$ are the (finitely many) elementes z such that $x \leq z$ and $y \leq z$. Nonetheless, the *categories* $\mathscr{S}$ and $\mathscr{L}$ are different because in $\mathscr{S}$ the mappings are required only to preserve T and $\wedge$ while in $\mathscr{L}$, F and $\vee$ must be preserved as well. This illustrates an important feature about categories. The mappings are the significant things. In fact, at a higher level of abstraction, the objects may be omitted completely.

2. Finitistic theories

An *algebraic theory* **Th** consists of a sequence of sets $\Omega_0, \Omega_1, \Omega_2, \ldots$ called the operation sets. Elements of $\Omega_0, \Omega_1, \Omega_2, \Omega_3, \ldots, \Omega_n, \ldots$ are called nullary, unary, binary, ternary, n-ary operations, respectively. Nullary operations are also called constants for reasons that will become clearer. These sets have the following structure. The set Ω_n includes n distinguished operations named $\pi_{1n}, \ldots, \pi_{nn}$ called projection operations. Moreover if $\omega \in \Omega_n$ and $\omega_1, \ldots, \omega_n \in \Omega_m$ respectively there is a composite operation $\omega \circ (\omega_1, \ldots, \omega_n) \in \Omega_m$. There are certain equations between these operations but they are too complicated to state in this formulation. A complete description appears in the Appendix. Fortunately, the detailed description is not required here.

The special case $n = 0$ of the above requires a little explanation. It says that given $\omega \in \Omega_0$ and no other data, there is an operation $\omega \circ (\ ) \in \Omega_n$. In practice this means there are given functions $\Omega_0 \to \Omega_n$ for each n. It will simplify life if, for $\omega \in \Omega_0$, we simply let the same symbol ω stand for its image in each Ω_n.

An algebraic theory is called *finitistic* provided

(i) each set Ω_n is finite, and

(ii) there is some n such that for all $m > n$, every m-ary operation is built up using the $\circ$ composition above from projections and k-ary operations for $k \leq n$.

The significance of the second condition will be explained later.

In actual practice, you never describe all the operation sets of a theory but only a few generating operations and equations involving the generating operations, the projections and their composites. For example the theory of inf semilattices is generated by a single nullary operation ν (corresponding to T) and binary operation β (corresponding to $\wedge$). All other operations are gotten by using these two, projections and composites. For example, there is an operation $\beta \circ (\pi_{13}, \pi_{23}) \in \Omega_3$ which may be used to recursively construct still other operations. The equations assert that two apparently different operations are actually the same. In the present instance the theory of inf semilattices satisfies all relations entailed by

(2.1) $\beta \circ (\pi_{11}, \nu) = \beta \circ (\nu, \pi_{11}) = \pi_{11}$

(2.2) $\beta \circ (\pi_{22}, \pi_{12}) = \beta$

(2.3) $\beta \circ (\beta \circ (\pi_{13}, \pi_{23}), \pi_{33}) = \beta \circ (\pi_{13}, \beta \circ (\pi_{23}, \pi_{33}))$.

If all this seems mysterious, have faith. It will make much more sense in section 4 when we come to the actual interpretation of these identities in a model.

3. Products and powers of sets

Let X and Y be two sets. By $X \times Y$ we denote the set of all ordered pairs

$(x, y),\ x \in X,\ y \in Y.$

Two such pairs (x, y) and (x', y') are equal if and only if $x = x'$ and $y = y'$. We define two functions

$$p : X \times Y \to X, \; q : X \times Y \to Y$$

by

$$p(x, y) = x, \; q(x, y) = y$$

called the first and second coordinate projections, respectively. The set $X \times Y$ is called the *product* of X and Y with projections p and q.

If X, Y and Z are three sets, we have products $X \times (Y \times Z)$ and $(X \times Y) \times Z$ as well as a product $X \times Y \times Z$ which consists of all triplets

$$(x, y, z), \; x \in X, \; y \in Y, \; z \in Z.$$

These three product sets are distinct (i. e. their elements are not the same) but it is clear in an intuitive way all three are "essentially" the same. There is no particular difficulty in making this notion of "essentially the same" completely precise but only at the cost of considerable complication which I would like to avoid. Hence I will pretend that all three products are the same. Analogous observations hold for the products of more than three sets.

If $f : X \to X'$ and $g : Y \to Y'$ are functions, we let $f \times g$ denote the function $X \times Y \to X' \times Y'$ given by the formula

$$(f \times g)(x, y) = (f(x), g(y)).$$

Analogous formulas exist for three or more functions.

As a special case, we let X^n denote the product $X \times X \times \cdots \times X$ of n factors. It consists of all n-tuplets of elements

$$(x_1, \ldots, x_n)$$

taken from X. For an integer i, $1 \leq i \leq n$, we let $p_{in} : X^n \to X$ denote the projection on the i^{th} factor give by

$$p_{in}(x_1, \ldots, x_n) = x_i.$$

If there is any doubt about the set involved, we may write $p_{in}X$ instead. If $f : X \to Y$ write $f^n : X^n \to Y^n$ for the function $f \times f \times \cdots \times f$.

A very special case occurs when $n = 0$. In that case we define X^n to consists of one element. We write $X^n = 1$ and denote the single element of 1 by $*$. In this case there are no projections (or, rather, the number of them is 0).

4. Models of a theory

Let **Th** be a theory. By a model of a theory is meant a set X together with a function

$$\omega X : X^n \to X$$

for each $n \geq 0$ and each $\omega \in \Omega_n$. These are subject to the following conditions.

(i) $\pi_{in}X = p_{in} : X^n \to X$

(ii) If $\omega \in \Omega_n$, $\omega_1, \ldots, \omega_n \in \Omega_m$, then $\omega_0(\omega_1, \ldots, \omega_n)(x_1, \ldots, x_m)$ is $\omega(\omega_1(x_1, \ldots, x_m), \ldots, \omega_n(x_1, \ldots, x_m))$.

Note that a nullary operation ν determines a function $\nu X : 1 \to X$ in any model X. This is exactly the same thing as a fixed element of X. Such an element is called a constant and the operation ν is called a constant operation.

It should be understood that the data of a model consists of the set X together with all the functions ωX. We might write $M = (X, \{\omega X\})$ to denote a model and say that X is the *underlying set* of M. Usually we will write M interchangeably for the model as well as its underlying set.

If M_1 and M_2 are models (of the same theory **Th**) and

$$f : M_1 \to M_2$$

a function, we say that f is a mapping of models provided for every n-ary operation ω of **Th** we have that the functions

$$M_1^n \xrightarrow{f^n} M_2^n \xrightarrow{\omega M_2} M_2$$

and

$$M_1^n \xrightarrow{\omega M_1} M_1 \xrightarrow{f} M_2$$

are equal. We usually describe this situation by saying that the diagram

$$\begin{array}{ccc} M_1^n & \xrightarrow{f^n} & M_2^n \\ {\scriptstyle \omega M_1}\downarrow & & \downarrow{\scriptstyle \omega M_2} \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

commutes. We let Mod(**Th**) denote the category of models of **Th** and mappings of models.

In the very common situation that **Th** is presented by a few defining operations and equations between those operations it is necessary and sufficient that a model admit ωX only for those few operations subject, of course, to the condition that the relations be valid. I hope this will become clearer as we turn to the example of the theory of inf semilattices described in section 2.

Se let X be a set together with functions $\nu X : 1 \to X$ and $\beta X : X \times X \to X$. We denote the (unique) value of νX by T and the value of $\beta X(x, y)$ by $x \wedge y$. Now consider equation (2.1). It says

$$\beta \circ (\pi_{11}, \nu) = \beta \circ (\nu, \pi_{11}) = \pi_{11}.$$

Now we have required that $\pi_{11}X = p_{11}$ and that function is defined by $p_{11}(x) = x$. The value of νX is the constant 1 so that the value of $\beta \circ (\pi_{11}, \nu)$ is defined to be

$$\beta \circ (\pi_{11}, \nu)(x) = \beta(x, \nu(x)) = \beta(x, T) = x \wedge T.$$

Similarly the value of $\beta \circ (\nu, \pi_{11})$ at x is $T \wedge x$. Hence the interpretation of (2.1) is

$$x \wedge T = T \wedge x = x.$$

Since $p_{22}(x, y) = y$ and $p_{12}(x, y) = x$, the interpretation of (2.2) in a model is that

$$y \wedge x = x \wedge y.$$

Finally (2.3) has the interpretation

$$x \wedge (y \wedge z) = (x \wedge y) \wedge z.$$

It is an easy exercise to show that these operations and equations serve to specify an inf semilattice. Again a function f between inf semilattices X and Y is a mapping if and only if

$$\begin{array}{ccc} 1 & \xrightarrow{\nu X} & X \\ \downarrow & & \downarrow f \\ 1 & \xrightarrow{\nu Y} & Y \end{array}$$

and

$$\begin{array}{ccc} X \times X & \xrightarrow{\beta X} & X \\ {\scriptstyle f \times f}\downarrow & & \downarrow{\scriptstyle f} \\ Y \times Y & \xrightarrow{\beta Y} & Y \end{array}$$

commute. That is if and only if

$$f(T) = T$$

and

$$f(x \wedge y) = f(x) \wedge f(y).$$

It is instructive to write out generating operations and equations for the theory of lattices. On the other hand, the category of posets is not the category of models of any algebraic theory.

5. The theory of a concrete category

We are now in a position to describe the procedure whereby given any concrete category we may construct an algebraic theory which best approximates it. In order to make it perfectly constructive, we will suppose that $\mathscr{C}$ is a category consisting of a finite number of finite sets and the mappings between them. At this juncture it will be necessary to clearly distinguish between the object C and its set of elements that we will denote UC. If $f : C_1 \to C_2$ is a mapping in, we let $Uf : UC_1 \to UC_2$ denote the function underlying f. It is easily seen that when $g : C_2 \to C_3$, $U(g \circ f) = Ug \circ Uf$. We let U^nC denote the set $(UC)^n$.

By a *natural transformation* ω from U to U is meant a collection of functions ωC, one for each $C \in |\mathscr{C}|$ such that

$$\omega C : UC \to UC$$

and such that whenever $f : C_1 \to C_2$ is a mapping in $\mathscr{C}$ the diagram

$$\begin{array}{ccc} UC_1 & \xrightarrow{\omega C_1} & UC_1 \\ {\scriptstyle Uf}\downarrow & & \downarrow{\scriptstyle Uf} \\ UC_2 & \xrightarrow{\omega C_2} & UC_2 \end{array}$$

commutes. Since there are only finitely many C and for each C only finitely many functions $UC \to UC$, it is clear that it is principly possible to compute the set Ω of all natural transformations $U \to U$. This proof of possibility leaves a lot to be desired in the way of feasibility but in actual practice many shortcuts present themselves. On sure example is that natural transformation π for which each πC is the identity function on UC. It is evident that the commutativity condition above is satisfied in that case.

More generally by a natural transformation $\omega : U^n \to U$ is meant a collection of functions ωC, one for each $C \in \mathscr{C}$ such that

$$\omega C : U^n C \to UC$$

and for each mapping $f : C_1 \to C_2$

$$\begin{array}{ccc} U^n C_1 & \xrightarrow{wC_1} & UC_1 \\ {\scriptstyle U^n f}\downarrow & & \downarrow{\scriptstyle Uf} \\ U^n C_2 & \xrightarrow{wC_2} & UC_2 \end{array}$$

Here $U^n f = (Uf)^n$. When $n = 0$, the sets on the left have one element and then $U^n f$ is the only function there is. When $n > 0$, there is the natural transformation for each $1 \le i \le n$, $\pi_{in}, : U^n \to U$ defined by

$$\pi_{in} C = p_{in},$$

the projection on the i^{th} coordinate. The required commutation is trivial. Equally easy is the proof that if $\omega : U^n \to U$ and $\omega_1, \ldots, \omega_n : U^m \to U$ are natural so is $\omega \circ (\omega_1, \ldots, \omega_n) : U^m \to U$ defined by $\omega \circ (\omega_1, \ldots, \omega_n)(x_1, \ldots, x_m)$

$$= \omega(\omega_1(x_1, \ldots, x_m), \ldots, \omega_n(x_1, \ldots, x_m)),$$

$(x_1, \ldots, x_m) \in U^m C$ (i.e., $x_1, \ldots, x_m \in UC$). The upshot is that if we let Ω_n be the set of all natural transformations $U^n \to U$, then the sets $\Omega_0, \Omega_1 (= \Omega), \Omega_2, \ldots$ form an equational theory that we call $\mathbf{Th}(\mathscr{C})$. (Technically it is the theory of U but I have defined things in such a way that U is part of the structure of $\mathscr{C}$. This is not the usual way for abstract categories but is appropriate for concrete ones.)

We make several observations here, without proof. Their proofs would take us far afield and are, in any case, found in standard sources, are trivial or they are mere practical observations.

1. For each fixed n, the computation of the finite set Ω_n is possible in principle. In practice it may be unfeasible, but usually is not.

2. It looks like you have to consider infinitely many n. In fact there is an n_0 depending only on $\mathscr{C}$ such that all operations are generated by those of arity $\leq n_0$. Take for example, n_0 to be the product of the number of elements of the finitely many objects of $\mathscr{C}$. That looks fearfully large but I personally have never seen anything worse than ternary operations arise in a serious mathematical discourse. (I have also seen infinitary operations but that is not possible with the finitistic theories considered here.) It would be perfectly feasible to replace the problem of finding the best algebraic theory that approximates $\mathscr{C}$ with that of finding the best theory whose operations were generated by those arity ≤ 2 or ≤ 3.

3. The objects of $\mathscr{C}$ are models of $\mathbf{Th}(\mathscr{C})$. The value of an ω on C is ωC. The mappings of $\mathscr{C}$ are mappings of models. The best way to describe $\mathbf{Th}(\mathscr{C})$ is to say that it has the largest class of operations for which both of the above statements are true and the largest set of equations for which the first one is.

4. This means that the category $\mathscr{C}$ is embedded in the category of models of $\mathbf{Th}(\mathscr{C})$. If C_1 and C_2 belong to $\mathscr{C}$ it may happen that some function $f : C_1 \to C_2$ is a mapping of models without being a mapping in $\mathscr{C}$.

6. The analogy

In this section we recapitulate the capacities of the theory generating device. After each mathematical term I add parenthetically the linguistic term it is analogous to.

Let **Th** be a theory (language – or perhaps better, the syntax of a language) known to be determined by a finite number of finite models (sentences). Let M be the category of finite models (corpus of all grammatical sentences). Imagine the computer (child) to be presented, not with all of M (M is infinite) but with a sequence of models (sentences) of M, one at a time and at random. Then the computer's (child's) input data (knowledge) is a sequence of subcategories (subsets of the corpus) $\mathscr{C}_1 \to \mathscr{C}_2 \to \mathscr{C}_3 \to \ldots$ of M each embedded in the next by adding one model in M or one mapping of such models (one sentence). Then the computer (child) is capable, by calculating natural transformations (???), of computing a sequence of theories and mappings

$$\mathbf{Th}(\mathscr{C}_1) \leftarrow \mathbf{Th}(\mathscr{C}_2) \leftarrow \mathbf{Th}(\mathscr{C}_3) \leftarrow \ldots$$

This sequence may or may not, after a finite number of steps, ever reach the whole theory **Th**. If it does, it stabilizes (i. e. does not change from then on). That is, no additional data can change it. On the other hand it might not attain **Th**. If **Th′** is the theory actually attained there is a mapping $\mathbf{Th} \to \mathbf{Th}'$ and we may discuss its possible failure to be a $1-1$ correspondence. It may fail in two ways, each of which has a potential linguistic analogue. First, there may be an operation (syntactic rule) valid in all the models (sentences) which happens to have been sampled but not in all of M. Second, there may be two operations in **Th** (rules of syntax) that are equal (fall together) on all the sampled data. Here is an example from English morphology: It is certainly possible for a native speaker of English to have simply never sampled the past participle of "cleave". I can't think of any syntactic example of this phenomenon. That is perhaps not surprising as rarely used rules and rarely used distinctions will evidently tend to disappear.

At any rate, the attained theory **Th′** is determined exclusively by the data, not by the order in which it is input or any other irrelevancy. Its failure to be the original theory **Th** can only be the result of the lack of sufficient data. In fact, **Th′** is the best, in a very strong sense, theory that describes the data received. The construction of **Th′** instead of **Th** may be more usefully viewed as analogous to a restructuring of the grammar rather than as a failure of the machine to operate properly. When new data are received the machine will continue to construct better and better approximations to **Th**. This needn't ever cease although as a practical matter adults probably treat new syntactic information as irrelevant (or incorrect) variation, a behavior not available to a machine.

Appendix: A complete description of a theory

The only really efficient description of what a theory is is in terms of abstract categories. I will suppose henceforth that this notion is familiar to the reader.

Let **Th** consist of sets $\Omega_0, \Omega_1, \ldots$, have projection operators $\pi_{in}\, 1 \leq i \leq n$, in Ω_n and a composition law as described in section 2. We attempt to construct a category as follows. The objects of the category are formal symbols [0], [1], . . ., [n], . . . where n is a positive integer. We define the homsets by

$$\mathbf{Th}([m], [n]) = \Omega_m^n.$$

That is, a mapping $[m] \to [n]$ is to consist of an n-tuple

$$(\omega_1, \ldots, \omega_n)$$

of n-ary operations. Among the maps $[n] \to [n]$ is the map

$$(\pi_{1n}, \ldots, \pi_{nn}).$$

If

$$\omega = (\omega_1, \ldots, \omega_n) : [m] \to [n]$$

and

$$\tau = (\tau_1, \ldots, \tau_k) : [n] \to [k],$$

The composite $\tau \circ \omega$ is defined as $(\tau_1 \circ (\omega_1, \ldots, \omega_n), \ldots, \tau_k \circ (\omega_1, \ldots, \omega_n))$ which is an element of Ω^k_m. This defines a law of composition. Then the requirements that **Th** be a category is summarized by the requirement that the above law of composition determines the structure of a category for which the identity map of [n] is

$$(\pi_{1n}, \ldots, \pi_{1n}).$$

It is customary to denote this category by **Th**. The category has finite products. In fact, it follows essentially by definition that [n] is the n^{th} power of [1]. Also [0] is the terminal object. Then it is quite straightforward to show that the category of models of **Th** can be identified as exactly the category of set-valued product-preserving functors on **Th**. Mappings of models are exactly natural transformations between such functors.

REFERENCES

Barr, M., and C. Wells, Category Theory (to appear).

Johnstone, P. T. (1977), Topos Theory. Academic Press.

Lawvere, F. W. (1976), Variable quantities and variable structures in topoi. pp. 101–131 in: Heller, A. and M. Tierney (Eds.), Algebra, Topology and Category Theory: a collection of papers in honor of Samuel Eilenberg, Academic Press.

MacLane, S. (1971), Categories for the Working Mathematician. Springer Verlag.