GRAMMATICAL INFERENCE AND THEORIES OF LANGUAGE ACQUISITION

W.J.M. LEVELT

ABSTRACT

Some theorems from the theory of grammatical inference are used to clarify various issues in the now obsolete study of LAD as a model of human language acquisition. It is shown that large part of the psycholinguistic literature concerning the LAD model is based on two assumptions which find neither mathematical nor empirical support. Suggestions are made for a more promissing extension of the inference model.

Discussion paper for colloquium on "The role of grammar in inter-disciplinary linguistic research" to be held at the University of Bielefeld, December 1973.

Introduction

It is a fact in the history of science that certain issues which enjoy the active interest of the most outstanding minds during a certain period, become obsolete after some time without having been solved. One good example in psychology is the study of attention. This was a highly active topic around the turn of the century, but it disappeared from the scene between the two world wars, and it is only rather recently that interest in attention was revived by Broadbent and his coworkers. Another example is the study of inner versus outer form in language, which was a central issue in Wundt's days, fell into near complete disuse after, and finally got reintroduced as a brand new topic in Chomsky's theory of language.

The lesson to be learned is that obsolescence does not imply that adequate insight has been acquired, but only that the fad is over and the fashions have changed. It is often quite obscure how these changes come about, but it is not the aim of the precent article to discuss this issue. Instead, I will treat a topic which became obsolete only rather recently, but which is in my view a good example of an unsolved problem that deserves further study. In fact, I will try to add some additional steps to its solution.

The topic is that of grammatical inference, as related to the child's acquisition of language. The problem was introduced in a conference paper by Miller and Chomsky (1957), That paper was never published, and

meanwhile Miller lost all his copies (see Miller, 1967). But it seems that the problem posed went somewhat as follows: Given a language (natural or artificial) for which a (finite) grammar exists, could one conceive of a procedure for inferring the grammar from a finite set of observations? It was clear from the outset that, without further qualifications, this question could not be answered. And subsequently many qualifications were indeed made, as we shall see in a moment. It was also immediately obvious that an answer to the question could be highly relevant for the understanding of the child's acquisition of language. At a conference in 1960, Chomsky (1962) stated that relation as follows:

> ". . . ., we might attempt to construct a device of the kind
>
> (1) utterances of L → [ ] → formalized grammar of L
>
> This represents a function that maps a set of observed utterances into the formalized grammar of the language of which they are a sample. Given as input a sufficiently large and representative set of utterances of any language (English, Chinese, or whatever), the device (1) would provide as output a formalized grammar of this language. A description of this device would therefore represent a hypothesis about the innate intellectual equipment that a child brings to bear in language learning".

If such a "Language Learning Device", later rebaptized as "Language Acquisition Device" (LAD), could be conceived of, it could function as an ideal model for human language acquisition. As for any ideal model, the subsequent step should then be to compare the model with the

actual situation, i.e. the child's language acquisition, and see how the model has to be adapted in order to work in real time and to display the typical characteristics of the child's growing language. Chomsky (1965) denotes these two aspects of the problem by "adequacy-in-principle" and "feasibility" of LAD.

Within the psychology of languange the further study of grammatical inference went roughly as follows: First, there was the general assumption that some form of LAD could be adequate-in-principle for natural languages. Second, it was generally presupposed that the "adequate" form of LAD had to be of a strongly rationalistic (as opposed to empiricistic) character.

It was thought that this adequacy would not be dependent on semantic input. Semantic input would at most be a motivating factor (cf Chomsky, 1962, 1965). The device could in principle be of purely syntactic construction.

In short, it was generally assumed in the psycholinguistic literature that the question with respect to the adequacy-in-principle had been answered in the affirmative, and that LAD should be of an essentially syntactic and rationalistic brand.

These assumptions were sufficient to proceed to the feasibility-issue, and a new era of child language study was initiated. The writing of child grammars become the new fad, and enormous advances were made in a relatively short period (cf the studies by Bellugi, Braine, Brown, Ervin, Klima, McNeill, Slobin and many others). In accordance with the "principled" assumptions these grammars were given a rationalistic interpretation, and all sorts of "innate universals" popped up in the literature. Finally, the problem of grammatical inference in many respects fell into obsolescense in the psycholinguistic literature. The empirical work on language acquisition became largely independent of the assumptions on adequacy-in-principle from which it had started. It became

especially out-dated to think of a merely syntactic form of language acquisition. At the end of the sixties the attention turned to semantics and the old assumptions simply did not fit the new fashions. Also the notion "innate" became increasingly cryptic.(in accordance with developments in ethology, cf. Hinde, 1966), so that it became less attractive to defend a so-called rationalistic model of language acquisition.

But were the initial questions really solved? The answer is no, at least not in the psycholinguistic literature. It was not solutions but dogma's that became obsolete. Much confusion, even in recent literature, can be attributed to the unclear state in which the adequacy-in-principle issue was left by psychologists.

In this article I will first turn to some results in computer science with respect to grammatical inference. To my knowledge, these results have never attracted any psychological attention, in spite of the fact that they contain at least partial answers to the question of adequacy-in-principle. Next, I will apply these results to the study of child language in order to clarify various confusions and to indicate empirical areas that, for theoretical reasons, deserve the special attention of research workers.

Some results in grammatical inference

It was stated above, that the original question of grammatical inference could not be answered without further qualifying the question. Such qualifications have been given in the computer science literature. I am especially referring to studies by Solomonoff (1958, 1964), and later work by Gold (1967) and by Horning (1969). The starting

point of this discussion will be a presentation of Gold's formulation of grammatical inference: it is the most explicit formulation available, whereas Horning's work can be summarized in the same terms as a stochastic generalization of Gold's model.

The answer to the question of adequacy-in-principle of LAD will depend on: (1) LAD's a-priori knowledge of the language it is going to learn. Let us denote this knowledge by LAD's hypothesis space. (2) The structure of LAD's observations, or input. Let us call this LAD's observation space. (3) The definition of what is meant by an "adequate grammar for L", to be derived by LAD.

The hypothesis space might be unrestricted: in that case LAD has to assume that L can be any recursively enumerable (type - 0) language. It can also be very narrow: LAD could know that L is one of a finite set of languages. Between these two extremes we have the whole Chomsky-hierarchy of languages: LAD can be constructed in such a way as to assume that L is regular, or context-free, or context-sensitive; or L is assumed to be a member of any other interesting subset of languages.

LAD's observation space could consist of a "representative set of utterances" in L, as in (1). It should then be defined what is meant by "representative". Gold makes the following distinctions. The input of LAD is a string of observations or instances. If L is the language for which a grammar is to be inferred, and x is a string of words from L's vocabulary, LAD can either obtain positive information, namely "x in L", or negative information, "x not in L". In the first case we speak of a positive instance, in the latter of a negative instance. A string of such instances is called an Information sequence. If all of the instances in the sequence are positive, we have a positive information sequence; if negative instances also occur, we have mixed information sequence.A complete information sequence is a mixed information sequence in which

all positive and negative instances are enumerated; such sequences are generally infinite in length. Such a sequence is also called an informant presentation; it is as if each possible string of words is presented to an informant whc provides the information "grammatical" or "ungrammatical". A complete positive information sequence is an enumeration of all positive instances; it is also called a text presentation, since it is as if LAD is reading a text containing all and only the grammatical strings of L. In his extension of Gold's treatment, Horning adds another possibility of information presentation, which he calls stochastic text presentation. If L is a probabilistic language, i.e. if a probability distribution has been defined over the sentences of L, a stochastic text presentation is any infinite sequence of positive instances, which is generated by randomly choosing from L, according to the probabilities of the sentences in L. Of course, there are many other possible ways of presenting information to LAD. Chomsky (1962) mentions the possibility of presenting "corrections", i.e. pairs of ungrammatical and grammatical strings. In the final section of this article another alternative will be presented, which could be called an "intelligent" text presentation.

What is meant, finally, by an adequate grammar for L? In Gold's definition any grammar which generates all and only the sentences of L is called adequate. It does not matter whether the grammar is clumsy or simple. Horning is much more restrictive on this point. In his case a grammar should be "optimal" according to certain criteria of simplicity. (In this place we cannot treat these or other technicalities in detail, the reader is referred to Levelt (1973/1974) or to the original publication)

Suppose we have a complete (text or informant) information sequence for a language of a given class (i.e. finite, context-free, etc.). The construction of LAD involves the development of an algorithm with the following characteristics:

1. Each time a new instance is presented to LAD. LAD outputs a grammar of the given class, which is consistent with all information received up to that moment (i.e. the grammar should generate all positive instances, and - for informant presentation - exclude all negative instances).
2. After a finite number of instances, the output remains constant: the grammar produced at each instance is the same as, or equivalent to the former output, and is moreover a grammar of L.

A language is called "learnable" if such an algorithm exists for it, for every complete information sequence. A class of languages is learnable if every language in it is learnable. In other words, there is a possible LAD for such classes of languages.

Gold studied this adequacy-in-principle question for various classes of languages and for the two presentation conditions(text and informant). The main results of this study are summarized in Table I (for the proofs I must refer the reader to the original publication).

Table I "Learnability" of various classes of languages by text and informant presentation.

Language Class	Text presentation	Informant presentation
Recursively enumerable		
Recursive		
Primitive recursive	-	+
Context-sensitive	-	+
Context-free	-	+
Regular	-	+
Finite	+	+

"+" refers to "learnable", "-" denotes "unlearnable"

The table shows that, under these definitions, only finite languages are learnable by means of text presentation. Informant presentation guarantees learnability up to the level of primitive recursive languages.

Where should we locate natural languages? This question has no definite answer. On the one hand it has been shown convincingly by transformational linguists that natural language is outside the class of context-sensitive languages [though this has never been proven). On the other hand, Peters and Ritchie (in press)have shown that the class of transformational grammars in Chomsky's Aspects of the theory of syntax (1965) generates all and only the recursively enumerable (or type-0) languages. This result holds as well for all transformational grammars which evolved from that book (including the generative semanticists' extension). However, this is an obvious overestimation. Natural languages are certainly more restrictive than recursively enumerable. They should at least be recursive (decidable), since humans are equally good at recognizing grammaticality as they are at recognizing ungrammaticality. (There are also various other reasons for assuming recursiveness of natural languages, see Levelt 1973/1974, Volume II). Furthermore, the distinction between recursive and primitive recursive is of no linguistic interest. Any recursive grammar of natural language is at the same time primitive recursive. This would mean that natural languages are "learnable" by informant presentation: one could in principle construct a language acquisition device that can infer a natural language in this way.

Before considering the relevance of these findings for a psychological theory of language acquisition, we must mention one further result. In his study. Horning analyzed learnability under stochastic text presentation. For this, he weakened the definition of learnability: the optimal grammar should only be reached asymptotically by LAD, i.e. any non-optimal grammar is to be rejected in finite time, but no time-limit is set on

inferring the optimal grammar. On the other hand, Horning strengthened the adequacy requirement: the inferred grammar should not only be a grammar for L, but also be the "best" grammar according to certain criteria. Homing's study Was limited to languages up to non-ambiguous context-free. His main result was that these languages are all "learnable" by stochastic text presentation.

Relevance for rationalistic and empiricistic models of language acquisition.

Let us first discuss how dogma's instead of theorems have guided the adequacy-in-principle issue in the psycholinguistic literature. Therefore, we should start at a distinction reintroduced by Chomsky (1965], namely that between rationalist and empiricist models of language acquisition:

"In short, empiricist speculation has characteristically assumed that only the procedures and mechanisms for the acquisition of Knowledge constitute an innate property of the mind, ... The form of knowledge, however, is otherwise quite free. On the other hand, rationalist speculation has assumed that the general form of a system of knowledge is fixed in advance as a disposition of the mind, and the function of experience is to cause this general schematic structure to be realized and more fully differentiated".

This distinction is easily cast in terms of the LAD-model, Chomsky and Miller C1963) mention two ways in which LAD might become a practical language learning device. One is to "uncorporate strong assumptions about the class of potential grammars that a natural language can have". This is quite clearly the rationalist point of view. LAD is assumed to have a very restrictive hypothesis space. The other is "to supply the language-learning device with some sort of heuristic principles that would enable it, given its input data and a range of possible grammars, to make a rapid

inferring the optimal grammar. On the other hand, Horning strengthened the adequacy requirement: the inferred grammar should not only be a grammar for L, but also be the "best" grammar according to certain criteria. Homing's study Was limited to languages up to non-ambiguous context-free. His main result was that these languages are all "learnable" by stochastic text presentation.

Relevance for rationalistic and empiricistic models of language acquisition.

Let us first discuss how dogma's instead of theorems have guided the adequacy-in-principle issue in the psycholinguistic literature. Therefore, we should start at a distinction reintroduced by Chomsky (1965], namely that between rationalist and empiricist models of language acquisition:

"In short, empiricist speculation has characteristically assumed that only the procedures and mechanisms for the acquisition of Knowledge constitute an innate property of the mind, ... The form of knowledge, however, is otherwise quite free. On the other hand, rationalist speculation has assumed that the general form of a system of knowledge is fixed in advance as a disposition of the mind, and the function of experience is to cause this general schematic structure to be realized and more fully differentiated".

This distinction is easily cast in terms of the LAD-model, Chomsky and Miller (1963) mention two ways in which LAD might become a practical language learning device. One is to "uncorporate strong assumptions about the class of potential grammars that a natural language can have". This is quite clearly the rationalist point of view. LAD is assumed to have a very restrictive hypothesis space. The other is "to supply the language-learning device with some sort of heuristic principles that would enable it, given its input data and a range of possible grammars, to make a rapid

selection of a few promissing alternatives, which could then be submitted to a process of evaluation, or that would enable it to evaluate certain characteristics of the grammar before others". According to the empiricist view, LAD's feasibility would mainly be due to the existence of strong heuristic procedures. Whether LAD will work in real time will depend on (1) the restriction of its hypothesis space, and (2) on the power of its heuristic procedures. This is a very useful distinction, and Chomsky and Miller do not hesitate to make their choice: "The proper division of labor between heuristic methods and specification of form remains to be decided, of course, but too much faith should not be put in the powers of induction, even when aided by intelligent heuristics, to discover the right grammar. After all, stupid people learn to talk, but even the brightest apes do not".

The care of this statement is not followed by Chomsky (1965). There we suddenly read: "In fact, the second question [the real-time issue] has rarely been raised in any serious way in connection with empiricist views ... since study of the first question [the adequacy-in-principle issue] has been sufficient to rule out whatever explicit proposals of an essentially empiricist character have emerged in modern discussions of language acquisition".

The facts are, however, that the adequacy-in-principle issue bad not been solved at all in 1965, let alone that any rationalist or empiricist solution had been obstained. Chomsky's statement is based on two assumptions or beliefs:

1. It is possible to construct a grammar-learning procedure for natural languages.
2. Such a procedure makes use of a very restricted hypotheses space (i.e. is rationalist in character).

These beliefs are, moreover, surrounded by an aura of scientific evidence in Chomsky's statement.

Many more such citations can be given. In spite of the recency of those citations, this first assumption became subject to erosion at the end of the sixties. A new interest in semantic aspects of language acquisition made the LAD-approach less appealing. "It is tragic to cut off from the domain of research the large field of cognitive relations which are found in early sentences (...) by assuming a priori that there are no interesting problems in their acquisition, Dogmatism without evidence is to say the least presumptuous", (Ervin-Tripp, 1971].

We have argued, however, that erosion or obsolescence of an issue is not, in general, a satisfactory solution. In our view the computer science results, presented in the preceding paragraph, can be helpful to bring the adequacy-in-principle issue nearer to solution, Under Gold's definition of learnability, we must conclude that natural languages are only learnable through informant presentation, not through text presentation. The empirical question, therefore, should be: are children essentially in a situation of informant presentation? The only psycho-linguist who studied this question explicitly (only put in different terms) is Braine (1971). His conclusion is that the child is essentially in a situation of text presentation. Some of the following arguments are taken from Braine's paper:

(a) Many children are in a situation were speech is never corrected, i.e. marked as ungrammatical (this would be essential for informant presentation).

lb) If ungrammatically is marked, it is strikingly ineffective (cf. Braine (1971), McNeill (1966), Brown et al (1970)) i.e. it cannot be considered as real input, since it is quite clearly rejected.

(c) One could consider adult non-reactions to ungrammatical utterances as forms of negative marking. However, if the child would work this way, it would become hopelessly confused since initially nearly all

of its utterances are ungrammatical (in the adult sense) and many of them do lead to adult reactions. These are, in general, understandable utterances, whether grammatical or not.

(d) Informant presentation requires presentation of negative instances. A study by Ervin-Tripp (1971) showed, however, that the language spoken to children is strikingly grammatical, while utterances not addressed to the child are generally denied by him, and therefore cannot be considered as real input.

More arguments can be given (see Levelt, 1973/1974), but these suffice to make Braine's conclusion quite likely. At the same time this empirical evidence shows that LAD can in principle not be a model of human language acquisition: natural languages are not "learnable" through text presentation. Out here we should be careful: we have to consider Gold's definition of "learnability". The definition is both too weak and too strong in comparison to human learning. Too weak, since "learnability" is concluded if any Turing Machine can be constructed that fits the requirements. However, contrary to Turing Machines, humans do not have infinite memory capacity. Therefore, text presentation would fail a fortiori with humans. Too strong, since the procedure should work for any complete information sequence, even the most "tricky" or misleading ones. This is certainly not a requirement to be put on children's language acquisition.

Much more realistic in this latter respect is Horning's "learnability": stochastic text reflects the probability distribution in the language, and Horning shows that his results are still valid if the text deviates rather strongly from this distribution. Also, Horning's procedure is rather "noise-proof". An occasional ungrammatical sentence among the text will not upset the procedure. (This seems to be true for humans as well at learning artificial languages - cf. an experiment by Braine (1971)).

Gold's procedure, however, would even be upset by a single instance of wrong information ("+" instead of "-" or inversily).

If we take Horning's definition, however, the question whether a procedure for natural language learning can be constructed in principle keeps to be open. We only know that non-ambiguous context-free languages are learnable, but natural languages are more complex.

This situation is less bothersome than it seems. If we make the (much too restinctive) assumption that natural languages are of the non-ambiguous context-free type then Gold's procedures breaks down with respect to the real time issue. Horning shows that a grammar as ALGOL 60, which is learnable in principle, will use a staggering amount of observations to be learnt, even if the cleverest heuristics are built into the learning procedure. The procedure is, therefore, completely unrealistic for natural languages, which are much more complex then ALGOL 60.

So it seems that no real time LAD of the sorts considered can be constructed. For the moment, we must conclude that, without further qualification, the first assumption has no feet to stand on. It has to be rejected or changed instead of being forgotten. One rather fargoing qualification will be given in the final section of this article.

Ad (2). The second assumption, the rationalist character of LAD, became completely contaminated with the first (LAD's constructibility).

At all places where one reads about LAD, one is informed that LAD should have a very restricted hypothesis space in order to "pick" the right grammar in a relatively short period of time. This a priori knowledge of the class of natural languages must be innate.

The brain-washing effect of contaminating LAD's feasibility with rationalist theory became so great that even the most independent researchers became confused. As we noted above Brain (1971) was the first to argue that children are essentially in a situation of text

presentation. From this, he quite correctly concluded that the LAD-model should be ruled out. He then proceeds to presenting his own model, called a discovery-procedures model, which in its simplest form takes text as input, and gives a grammar for L as output) the model has very little *a* priori knowledge of L. By definition, however, the model is a special case of LAD; it has a nearly unrestricted hypothesis space, and a set of powerful heuristics (which are worked out in some detail). It is simply an empiricist version of LAD. Now, as we have seen above, the text presentation argument proves the unfeasibility of any model of the LAD-family. This includes Braine's own discovery procedures model. I should add that Braine quickly extends his model in order to also allow semantic input, but his argument for introducing the discovery procedures model is not based on this extension.

In view of this complete contamination of the first and the second "belief", it should not be surprising that there is no agreement in the literature as to the empirical evidence that might be decisive for either the rationalist or the empiricist point of view. "Empiricists" never defended their case within the LAD-frame work; they did not believe in LAD's feasibility since they wrongly supposed it to be a rationalist model.

Let us, in spite of the mathematical evidence given above, for the moment assume that a syntactic learning procedure for natural languages is feasible. (In the last section we will shortly discuss some revised versions of LAD). How could we distinguish empiricist from rationalist models?

As we have seen, the distinction resides in the division of weight placed on restriction of hypothesis space versus power of intelligent heuristics.

In the first, rationalist, case the grammar can be inferred, even in the light of very little evidence, i.e. variations in the observation

space should have little effect on language acquisition. In view of transformational grammar, rationalists have made the following, more detailed assumption: the hypothesis space specifies the phrase structure part of the grammar, as well as some formal aspects of transformations. What is learnt, i.e. derived from the observation space, is mainly the language-specific transformational apparatus. Whatever the input language, it has been said, children will start using the a-priori base-grammar; they will start talking "deep structures" or "deep sentences". Whatever its position on this latter qualification, the rationalist view implies a language development which is relatively independent of the input corpus universals of language are mainly due to the limited hypothesis space (strong universals), and only to a minor degree to the structure of the heuristic procedures (weak universals). Since the hypothesis space is given from the start, early language should be largely universal accross languages and cultures.

In the empiricist version of LAD, it is the strong heuristics which, by scrutinizing the observation space, infer a grammar for L. In this case, language acquisition is highly dependent on characteristics of the observation space. Universals can arise either from commonalities in observation space between different languages (input universals), or from general aspects of the heuristic procedures (weak universals). Braine, the most explicit representative of the empiricist position, elaborates this point by assuming that the heuristics are designed to register frequently recurring input aspects at various levels of abstraction. It is, in fact, hard or impossible to conceive of an empiricist procedure which is essentially different from such a model. The empiricist prediction should then be soma form of frequency matching between observed language and early speech. This implies that for different linguistic environments early language will differ in so far as

frequent salient aspects of these languages differ. Another implication is that early language should be sensitive to training, i.e. changing frequences in the observation space should be reflected in an adaptation of frequences in the child's speech.

It should be clear from a comparison of these predictions that the existence of universals in itself does not discriminate between the two extremes. Both expect and can explain universal phenomena in language development. The rationalists have their strong and weak universals, the empiricists have weak and input universals. It is wrong, therefore, to construe the existence of weak universals as evidence for the rationalist position. Is it possible to empirically distinguish between strong universals and input universals? The only way to do this would be by showing that universals arise in the child's speech at a very early moment, i.e. when no substantial input has yet been given. Though it is quite clearly the case that there are indeed universal traits in early language of children, it is near to impossible to decide at the present time whether they are of the strong or the weak variety of universals. One sufficient condition for strong would be that the feature is part of the universal base. However, it is unknown what is and what is not part of "the" universal base. In fact, given the grammatical formalism of Aspects or related grammars, the universal base hypothesis is not even an empirical issue. All type-0 languages can be transformationally derived from an almost trivial base grammar (Peters 8 Ritchie, 1969). A strong (though probably not even sufficient) argument for weak universals is their simultaneous appearance in several non-linguistic forms of behavior. Though there is much talk about such universal aspects of information processing (cf. Bever, 1970 and Slobin, 1971) there is a near complete lack of experimentation to show that a particular form of information processing also exists for non-linguistic materials. Slobin's article which **was** especially written to comment on this link between early

universals and (perceptual) information processing does not give one example of an independent non-linguistic experiment. At this point, moreover, the discussion can get badly confused. Universal is, according to Slobin, that the child is able to register the order of information. This explains that even in early speech we find a reflection of the dominant word order of the native language. In other words, this "universal" explains that children's languages are universally different since their mother tongues are different. In sum, it seems to me that the existence of early universals cannot, for the time being, be used as an argument for the rationalist position. Moreover, I see no other means to establish their point by mean of direct evidence.

Of course, the rationalist position might be establishable by default. For this we have to show that empiricist predictions find nu support in the data. One explicitly empiricist prediction is frequency matching, as we have seen above. Early frequency matching cannot be explained from the rationalist model. It is again amazing to see that hardly any data have been collected to prove or disprove this crucial issue. The only evidence I could find is some data in Bowerman's (1971) study on the acquisition of Finnish. She gives the following table of frequences of the order of S, V and 0 in the mother's and child's speech. At the time the average sentence length of the child was 1.42 words.

	SV	VS	VO	OV	SVO	OSV	OVS	VSO	SOV
Mother	47	5	16	3	32	0	1	1	1
Child	44	4	4	1	7	1	0	0	1

It seems that at least this child takes over the dominant word order of the mother's speech. But much more evidence is necessary to show the general existence of frequency matching.

frequent salient aspects of these languages differ. Another implication is that early language should be sensitive to training, i.e. changing frequences in the observation space should be reflected in an adaptation of frequences in the child's speech.

It should be clear from a comparison of these predictions that the existence of universals in itself does not discriminate between the two extremes. Both expect and can explain universal phenomena in language development. The rationalists have their strong and weak universals, the empiricists have weak and input universals. It is wrong, therefore, to construe the existence of weak universals as evidence for the rationalist position. Is it possible to empirically distinguish between strong universals and input universals? The only way to do this would be by showing that universals arise in the child's speech at a very early moment, i.e. when no substantial input has yet been given. Though it is quite clearly the case that there are indeed universal traits in early language of children, it is near to impossible to decide at the present time whether they are of the strong or the weak variety of universals. One sufficient condition for strong would be that the feature is part of the universal base. However, it is unknown what is and what is not part of "the" universal base. In fact, given the grammatical formalism of Aspects or related grammars, the universal base hypothesis is not even an empirical issue. All type-0 languages can be transformationally derived from an almost trivial base grammar (Peters & Ritchie, 1969). A strong (though probably not even sufficient) argument for weak universals is their simultaneous appearance in several non-linguistic forms of behavior. Though there is much talk about such universal aspects of Information processing (cf. Bever, 1970 and Slobin, 1971) there is a near complete lack of experimentation to show that a particular form of information processing also exists for non-linguistic materials. Slobin's article which was especially written to comment on this link between early

universals and (perceptual) information processing does not give one example of an independent non-linguistic experiment. At this point, moreover, the discussion can get badly confused. Universal is, according to Slobin, that the child is able to register the order of information. This explains that even in early speech we find a reflection of the dominant word order of the native language. In other words, this "universal" explains that children's languages are universally different since their mother tongues are different. In sum, it seems to me that the existence of early universals cannot, for the time being, be used as an argument for the rationalist position. Moreover, I see no other means to establish their point by mean of direct evidence.

Of course, the rationalist position might be establishable by default. For this we have to show that empiricist predictions find nu support in the data. One explicitly empiricist prediction is frequency matching, as we have seen above. Early frequency matching cannot be explained from the rationalist model. It is again amazing to see that hardly any data have been collected to prove or disprove this crucial issue. The only evidence I could find is some data in Bowerman's (1971) study on the acquisition of Finnish. She gives the following table of frequences of the order of S, V and O in the mother's and child's speech. At the time the average sentence length of the child was 1.42 words.

	SV	VS	VO	OV	SVO	OSV	OVS	VSO	SOV
Mother	47	5	16	3	32	0	1	1	1
Child	44	4	4	1	7	1	0	0	1

It seems that at least this child takes over the dominant word order of the mother's speech. But much more evidence is necessary to show the general existence of frequency matching.

With respect to the other empiricist prediction: trainability, there is again not much more available than anecdotical evidence.

Let us summarize this section by the following conclusion. For theoretical reason, any direct test of the rationalist model seems close to impossible. Indirect proof, by default of the empiricist model, has not been given since the empiricist predictions have hardly ever been put to test. The available evidence, moreover is probably more in agreement than in disagreement with the empiricist model. It seems, therefore, that not only the first assumption (adequacy-in-principle), but also the second (rationalist) assumption or "belief" connected to the IAD-model lacks support, both theoretically and empirically.

Extension of grammatical inference theory

Most children do learn their language, in spite of LAD's failure. Can the notion of a language learning device be expanded in such a way as to create new hopes that an adequate model might be attained?

In this final section we will consider a few lines along which extensions of the LAD-schema might be tried. What they have in common is that they, in first instance, concern LAD's observation space. We have namely observed that even very restricted hypothesis spaces and most powerful heuristic procedures do not guarantee LAD's feasibility. There is plenty reason, however to reconsider LAD's observations,

(a) Chomsky (1962) made the suggestion that LAD's input might consist of pairs of negative and positive instances, which he called "corrections". Nothing is known about the adequacy-in-principle for this type of presentation. However, the model is in any case quite unlikely, even less probable than a model with informant presentation. Corrections are more the exception than the rule in the child's linguistic world, as we have already discussed.

One could consider various sorts of semantic input. Though this is quite clearly a promising and moreover fashionable route to take, it rapidly leads to highly complicated problems in inference theory. The problem of adequacy-in-principle is worth considering only for the most simple forms of semantic input. In all other cases one might as well renounce completely the construction of a mechanical model for human language acquisition. Among the simplest models is one in which the observation space consists of paraphrastic equivalence classes. That is, a sentence is given, followed by one or a finite number of its paraphrases, a new sentence follows, etc. Though this does not seem to be very complex semantic information, there is no knowledge about its effectiveness. Moreover, just as the extension under (a) it does not seem to be a very realistic model.

Finally, one could consider an "intelligent" mode of information presentation. As we have seen, Gold's inference procedure was designed to work for any complete presentation, whereas Horning's only required stochastic representativeness. But Horning makes a very interesting remark at the end of his dissertation. His conjecture is that the child might be presented with a very different information sequence: "he is confronted with a very limited subset, both in syntax and vocabulary, which is gradually (albeit haphazardly) expanded as his competence grows". This is an important suggestion. The initial hypothesis space of the child might be very restricted, even finite, whereas the language which is directed to him is roughly of the same limited form. Only after the child has acquired this very restricted grammar his observation space is slightly expanded, and a concomitant expansion of the hypothesis space follows. A new optimal grammar is developed, etc., until the adult competence has been acquired.

There is certainly some truth to Homing's suggestion. Ervin-Tripp (1971) reports on studies about the language of adults which is typically addressed to small children. That language appears to be strikingly simple: it consists of short sentences, contains few or no conjunctions, passives or relative clauses, but is full of imitations of the child's language. Also it contains many paraphrases and slight expansions of the child's utterances, in which, however, the syntactic form of the child's language is maintained to a high degree. Moreover, it turned out that mothers increase the complicity of their sentences with the age of the child. Ervin-Tripp remarks: "the input maintains a consistent relation to the child's interpretative skill". This is all in complete accordance with Horning's suggestion, and it would be highly interesting to study learnability for such forms of "intelligent" language presentation.

References

Bever, T.G., The influence of speech performance on linguistic structure. In: Advances in Psycholinguistics, G.B. Flores d'Arcais & W.J.M. Levelt, (eds.). Amsterdam: North-Holland, 1970.

Bowerman, M.F., Acquisition of Finnish. Ph.D.-dissertation. Harvard University, 1971.

Braine, M.D.S., On two models of the internalization of grammars. In: The ontogenesis of grammar. A theoretical symposium, D.I. Slobin, (ed.). New York: Academic Press, 1971.

Brown, R. et al., Psycholinguistics, selected papers. New York: The Free Press, 1970.

Chomsky, N., Explanatory models in linguistics. In: Logic, methodology, and philosophy of science: Proceedings of the 1960 International Congress, E. Nagel, P. Suppes & A. Tarski, (eds.). Stanford: Univ. Press, 1962.

Chomsky, N., Aspects of the theory of syntax. Cambridge, Mass.: MIT Press, 1965.

Chomsky, N. & Miller, G.A., Introduction to the formal analysis of natural languages. In: Handbook of mathematical psychology, R.D. Luce, R.R. Bush & E. Galanter, (eds.), New York: Wiley. 1963.

Deese, J., Psycholinguistics. Boston: Allyn and Bacon, 1970.

Ervin-Tripp, S., An overview of theories of grammatical development. In: The ontogenesis of grammar. A theoretical symposium, D.I. Slobin, (ed.). New York: Academic Press, 1971

Gold, E.M., Language identification in the limit. Information and Control, 10, 1967, 447-474.

Horning, J.J., A study of grammatical inference. Technical Report CS 139 Stanford Artificial Intelligence Project, Stanford: Computer Science Department, 1969.

Levelt, W.J.M., Formal grammars in linguistics and psycholinguistics. Vol. I. An introduction to the theory of formal languages and automata, Vol. II Applications in linguistic theory, Vol. III Psycholinguistic applications. Mouton, 1974. Also published in Dutch: Kluwer, 1973.

McNeill, D., Developmental psycholinguistics. In: The genesis of language, A psycholinguistic approach, F. Smith & G.A. Miller, (eds.). Cambridge, Mass.: MIT Press, 1966.

McNeill, D., The acquisition of language. New York: Harper & Row, 1970.

Miller, G.A., The psychology of communication. Seven essays. New York: Basic Books, 1967.

Miller, G.A. & Chomsky, N., Pattern conception. Paper for Conference on pattern detection, University of Michigan, 1957.

Peters, P.S. & Ritchie, R.W., A note on the universal base hypothesis. J. Linguistics, 5, 1969, 150-152.

Peters, P.S. & Ritchie, R.W., On the generative power of transformational grammars. In Press.

Slobin, D.I., Developmental psycholinguistics. In: A survey of linguistic science, W.O. Dingwall, (ed.). College Park: Linguistics Program Univ. of Maryland, 1971.

Solomonoff, R.J., The mechanization of linguistic learning. In: Proc. Second Intern. Congr. Cybernetics, Namur, 1958, 180-193.

Solomonoff, R.J., A formal theory of inductive inference. Part I & Part II. Information and Control, 7, 1964, 1-22, 224-254.