ON LEARNABILITY: A REPLY TO LASNIK AND CHOMSKY

Willem J.M. Levelt

Max-Planck-Projektgruppe für Psycholinguistik

Nijmegen, August 1979

ON LEARNABILITY: A REPLY TO LASNIK AND CHOMSKY

Linguists and psycholinguists generally agree that there is an important relation between learnability and linguistic theory. The fact that children normally acquire the language of their community to a substantial degree, and in a fairly short period of time, has long been recognized as posing severe restrictions on the class of biologically possible human languages, and therefore on linguistic theory. The issue has been raised by Chomsky (1962) in particular, and extensively discussed in Chomsky & Miller (1963), and Chomsky (1965).

In spite of the obvious importance of learnability consideration for linguistic theory, during the first decade since the initial publications, almost no substantial effort was made to formally specify mutual restrictions between acquisition devices and grammars. When I wrote my review of the learnability literature in 1972 as part of a larger text on formal grammars (English translation, Levelt 1974), the only published paper in which a systematic effort was made to prove learnability (resp. non-learnability) for different classes of languages, combined with different types of primary data, was by Gold (1967). This fundamental paper was, at the time, fully ignored by both linguists and psycholinguists, a regrettable state of affairs since the paper, whatever else it was worth, set new standards of precision and explicitness for the discussion of learnability.

One theorem in Gold's paper struck me as potentially important. Under his definition of learnability, and his definition of presentation modes of the language it could be shown that the class of recursively enumerable languages was not learnable. The potential importance of this result derived from the Peters and Ritchie proof (1973, then available to me in draft form) that generative grammars of the Aspects-type can generate all and only the recursively enumerable sets. Apart from reviewing the literature on learnability, I tried to link those two results and put forth for further consideration that, in order to guarantee the learnability of natural languages, it might be necessary to reduce the power of transformational grammars.

Part of the task would then be to study whether Gold's definition of learnability and primary data are realistic as a model for the language acquiring child. I went into great detail to show that this is not so (Vol III pp 148 - 156), and suggested what aspects of Gold's results could be maintained under more realistic assumptions (such as the requirement that the acquisition device is "error-proof", i.e. resistant

to occasional faulty information, that it can deal with ambiguity, and that the procedure is "feasible", i.e. executable within realistic time and space constraints).

Various alternatives were reviewed (stochastic models, models with "intelligent" text presentation, where the child is, successively presented with more and more inclusive subsets of the language, and semantic models where pairs of strings and meanings are presented to the acquisition device), and no firm conclusion was drawn:

> "For natural languages, ambiguity will continue to contribute a serious problem, but for the present we can only state that, with this type of presentation [i.e. "intelligent" text presentation[1]], learnability without additional semantic input cannot yet be excluded a priori. But even in that case it is not clear why semantic input in the model of language acquisition must necessarily occupy a secondary place. The acquisition mechanism might be able to work much more efficiently if it can dispose of such information at the same time. Finally, it is less a question of what is possible than of what is in fact the case" (p. 155).

In short, the text suggested a relation between power of the grammar and learnability of the language, and considered various forms this relation may take under different types of primary, or input data, and definitions of learnability. This approach is, I still think, basically correct, and others have developed it since 1972 much further than I have ever endeavoured to do (Wexler & Hamburger 1973, Wexler, Culicover & Hamburger 1975, and several other papers by the same authors[2]).

---

1) Recent work by Newport, Gleitman and Gleitman (1977) has convinced me that on empirical grounds at least, there is no reason to strive for an acquisition model which has "intelligent" text presentation as an essential feature. Motherese turns out not to be so simple as was generally claimed in the early seventies. Though "intelligent" ordering of the input material doesn't have any bearing on the adequacy-in-principle of a learning algorithm, it can be of relevance for the feasibility of an algorithm, as I pointed out (Vol. III, p. 154), i.e. intelligent presentation can affect the time and/or space constraints of a learning algorithm that is adequate-in-principle.

2) I have not seen the text of the forthcoming book by Wexler and Culicover, which presumably summarizes most of this work.

This latter work especially has led to an increased interest among linguists in the formal relations between grammars and learnability, a laudable development. One result of this renewed interest is a set of reactions to my original text, one by Lasnik (1978), and one by Chomsky (1979[3]). Both reactions challenge the relation between recursiveness and learnability, which formed the starting point (but not the conclusion, see above) of my text. Lasnik's paper seriously misrepresents my position for reasons that have not become clear to me: it only leads to confusion, not to clarification of any sort. The first part of the present paper will deal with these issues. Chomsky's lecture does not misrepresent my text, but his treatment of the learnability problem is, in my view, insufficient to show that recursiveness (and thus power) of the grammar and learnability are conceptually unrelated. I will try to show that, as far as his account of language learning goes, such a relation is still not excluded. This will be done in part two of this paper.

1. Reply to Lasnik

Lasnik cites me as follows:

> "A non-decidable language is unlearnable, even if the learner benefits from an informant. In short this means that there is no algorithm by which an (observationally) adequate grammar can be derived from a sequence of strings marked "grammatical" and "ungrammatical".

The citation is continued, and is then followed by the following remarks of his own:

> "In large part, these arguments hinge on a misunderstanding of the notion "linguistic competence". Clearly, it is the grammar of a language that is acquired by the child rather than the set of sentences that constitutes the language".

This is then embellished with a footnote which runs as follows:

> "Consider an analogue: mathematical ability is fairly widespread. Hence we might want to say that arithmetic can be acquired. Yet, as is well-known, the set of theorems of arithmetic is not recursive".

Then the text goes on to "clarify" as follows:

> "Suppose the theory of grammar allowed exactly one grammar, i.e., that human biology did

---

[3] In the following I cite from a preliminary draft of this text, with permission of the author.

> not admit of linguistic variation. Suppose further that a person acquires this grammar by exposure to one sentence of the generated language L. To use customary, though perhaps misleading, terminology, L is learnable under the stated conditions. Consider now the possibility that L is not a recursive set of sentences. This changes nothing in the example. Therefore, the logical possibility exists that a non-recursive language can be learned, contrary to the assertions of Putman and Levelt".

Lasnik's opening citation is a mutilation of my text at a highly essential point. The text says in fact this:

> "A non-decidable language is unlearnable, even if the learner benefits from an informant. For the precise meaning of "learnability" and "informant" we refer to the discussion in Volume I, Chapter 8, § 3. In short ..."

The referred to paragraph gives Gold's definition of learnability, and of text and informant presentation of the language. So, that is what I was talking about. If Lasnik had studied and understood that paragraph he would not have accused me of messing up the competence/performance distinction. Under Gold's definition of informant, resp. text presentation[4] the child is not learning a set of sentences. He guesses a grammar (or name) of the language at each instance (i.e. after each input string marked as within the language or without the language), consistent with the input so far. The language is called "identifiable in the limit", or "learnable" if, after some finite number of instances, the guesses become identical or equivalent. A class of languages is learnable if every language in the class in learnable. So, the child learns a grammar, not a list of sentences. If the language is infinite and learnable, a correct

---

4) This is not the place to outline the full technical definitions of Gold's text- and informant presentation. The reader is referred to Gold (1967), or to my text, Vol. I, p. 116. Here it suffices to say that text-presentation is any enumeration of the sentences of the language (i.e. only "positive instances"), whereas informant-presentation is any enumeration of both the sentences of the language, and of the complement of the language, the first type of instances marked as "positive", the latter type as "negative".

grammar will have been acquired after a finite number of instances have been presented. Gold, and I myself, for that matter, have managed to understand that an infinite list of sentences cannot be learned as a list.

Now, clearly, we are not concerned with the learnability of individual languages, but of classes of languages. We want to know how the child selects one language from the finite or infinite class of possible natural languages. Since he should be able to do so whatever the language presented to him, all languages in the class should be learnable. Or to put it differently: given the foreknowledge of the child, i.e. the class of possible languages, an arbitrary language from the class will be learnable. This is what Gold's results are about, and which I summarize in table 8.2 (Vol. I) under the heading " 'Learnability' of languages of various classes according to text or informant presentation". There can be no misunderstanding whatsoever that that, and only that was meant when I wrote "A non-decidable language is unlearnable": if the foreknowledge is the class of non-decidable languages (for which there is a grammar, i.e. the class of recursively enummerable (r.e.) languages), then there is no algorithm guaranteeing learnability for an arbitrary language from the class. Chomsky, in his text, recognizes that this is what was meant, and that the statement is true under my (i.e. Gold's) technical definition of "unlearnable". It is amusing to see how Lasnik has to explain at great length that if the class contains just one language, it is learnable, even if r.e., since this trivially follows from the definition of learnability I used: If the class contains one language, all the names (or grammars) that the algorithm can produce are the same or equivalent from the very start. But after having developed this notion all by himself, Lasnik tells the reader that "Levelt's discussion of learnability indicates that he is unwilling to accept anything like the highly structured theory of grammar assumed by some linguists", which is probably true.

After this misrepresentation of my work, and after ignoring the Gold-definitions which I was using, it is interesting to read that

> "In one respect, Levelt misrepresents Gold. Gold nowhere claims, as Levelt does, that "it is impossible to 'learn' an infinite language only on the basis of text presentation". Furthermore the claim is patently false as Ken Wexler has pointed out to me".

So, at least Wexler read the paragraph, the reference to which was omitted by Lasnik. I do grant that taken in isolation that claim is false, but why does Lasnik, again, not acknowledge

the context, from which it appears that I am clearly dealing with classes of infinite languages, which might contain finite languages as well (just as all classes in the Chomsky hierarchy do contain finite languages). Under that interpretation the claim is correct.

As I mentioned above, at the time of writing my text the possible relevance of Gold's result was this: if it is the Aspects-formalism that constitutes the child's innate knowledge of natural languages, then, given the Peters and Ritchie results and Gold's technical definition of learnability, natural languages are unlearnable, since the class is r.e.

In my text I suggested two ways out of this dilemma: (i) to consider other technical definitions of learnability and presentation models; (ii) to reduce the class of grammars to decidable or even further, i.e. to something else than the class defined by the Aspects-formalism. Barring the acceptance of non-learnability of natural languages, these are the only ways out.

(i) As far as the definitions of learnability go, I have considered several alternatives in Volume III of the text, and others (esp. Wexler et al) have done a far more extensive and thorough job in this respect. Here, I would only like to add a remark on what, I think, should not be done, namely relating the acquisition of natural language to the acquisition of mathematics, as in done in Lasnik's footnote, cited above. This can only confuse matters. In the footnote Lasnik does not define the class of systems to be learned. As we have seen, if the class has cardinality 1, arithmetic, which is r.e., is trivially learnable, under Gold's definition of learnability. If the cardinality is greater than 1, nothing much can be said without precise specification of the class. But what is most important is that the presentation mode for learning arithmetic is essentially different from the presentation mode for learning a natural language. For arithmetic, at least part of the presentation consists of teaching the rules, i.e. inputting the grammar itself. This is totally different from natural language acquisition, where, as we know, explicit teaching of the rules is mostly not done, and certainly not required. So, surely, Gold's or similar presentation modes cannot be a model for learning arithmetic, but this is of no consequence for a model of language acquisition. Surprisingly, the same shortcut is made by Chomsky, who also argues that my (i.e. Gold's) definition of learnability must be the wrong one (fine!), under reference to arithmetic:

> "Note that arithmetic is learnable (though not in Levelt's technical sense of the word [this depends on the class, see above, W.L.]), though we might want to say that it is elicited rather than taught or learned" (Footnote 42)

I don't think this is so, at any rate the comparison to language learning is, as far as I can see, of no consequence for a correct definition of natural language learnability. Or, to cite Chomsky 1965, p. 56:

> "What one would expect, however, is that there should be a qualitative difference in the way in which an organism with a functional language-acquisition system will approach and deal with systems that are languagelike and others that are not".

What else has Lasnik to remark about presentation models? After having repeated my conclusion that r.e. sets cannot be learned by informant presentation, Lasnik remarks:

> "In passing, I note that this conclusion indicates that Levelt has the wrong notion of language learning. There is no evidence that informant presentation plays any role in actual acquisition. See Wexler and Culicover (In press, Chapter 2) for a survey of the literature on this question".

If Lasnik had read my survey of the literature (up to early 1972) in the same text he is dealing with (Volume III, pages 150-156), he would have found my conclusion that:

> "..the child, at best, is in a situation of text presentation, not of informant presentation. There are indeed strong arguments to support this, and we shall mention a number of them". (p. 149)

Then follows a review of the relevant literature. Why does Lasnik construct a fake opponent, and give him my name, instead of reading the text he endeavours to attack? This festival goes on and on, but it does not help Lasnik to come with clear and explicit statements of his own as to realistic technical definitions of learnability and presentation mode.

My present position with respect to presentation mode is not much different from what I stated in the 1974 text. I think that any realistic model of human language acquisition requires a presentation mode with at least the following properties. (1) The presentation should be one of positive instances, i.e., sentences of the language. (2) Each, or a substantial part of the positive instances should be paired with structural information, as suggested in Aspects (p. 32), but it should be essentially semantic in character (for more details see Levelt, 1975). (3) The presentation should be allowed to contain erroneous information (i.e. ungrammatical strings unmarked as such, or erroneous pairings of strings and structural information). Clearly, this

requires the acquisition algorithm to be "fool-proof", or self-correcting to a certain degree. These requirements are no more than beliefs on my part. The real work would be to implement them on an acquisition algorithm, and to show that such an algorithm is both possible and feasible (i.e. within the time and space constraints of the child).

(ii) Let us now turn to the second way out of the dilemma, reduction of the class of grammars. There are two seemingly different ways to realize this: the "intensional" and the "extensional" way. The intensional way consists of defining "possible grammar" in such a way that the class is small. Going from r.e. languages to decidable to context-free would be such a step. Such moves have been taken by various linguists in the past decade (for a recent example, see Gazdar, 1979), but at the time of writing my text only Joshi's (1972) grammars were especially constructed to be decidable. This is not the route taken by Lasnik or by Chomsky. Both are quite willing to accept that natural languages are r.e., or not even that. This latter move (made by Chomsky) would mean that there are no grammars for natural languages, and Chomsky writes that:

> "Given the epiphenomenal nature of the notion "language", this would not be particularly disturbing discovery. It would mean that the real systems that are mentally represented do not happen to specify recursively enumerable languages".

If this turns out to be so, special provisions have to be made in the theory to guarantee learnability of natural languages (if that is what one wants to guarantee). The provisions suggested by Chomsky (the "extensional way") are insufficient for that purpose. But before turning to that, I would like to treat the reader to one more misrepresentation of my work by Lasnik. One of the arguments I give to consider natural languages as recursive is that people's intuitions on the grammaticality of strings are not strikingly different from their intuitions on the ungrammaticality of strings:

> "Native speakers will in general be as capable of judging that a sentence belongs to their language as of judging that that is not the case".

It should be noticed that "as capable" does not mean "capable", but just <u>equally</u> capable. I go on to say that there are many unclear cases, but that I find it "more elegant" to ascribe these to psychological than to linguistic factors (here I refer to an extensive earlier discussion in the same book, (pp. 6-7).

Now, if we do so, and therefore assume that the ideal native speaker has a recognition device for

the sentences of his language (which makes the language r.e.), then the mentioned equality requires us to assume at the same time that the ideal native speaker has a recognition device for the complement of the language. But if the complement is r.e. as well, the language itself is recursive. Such was the argument. The other alternative, clearly, is to ascribe unclear cases to linguistic factors, and I continue (cited by Lasnik):

> "If on the ground of this objection [the existence of unclear cases, W.L.] we drop the recursive enumerability of the complement of the language (the ungrammatical strings), on the ground of the same objection we must also drop the recursive enumerability, and therefore the type-0 character, of the language itself".

So, if we take it to be a linguistic fact that the complement of the language cannot be recognized, the consequent of the above mentioned equality must be to assume that the language itself cannot be recognized, i.e. is not r.e. and thus has no grammar. Chomsky, apparently, prefers to ascribe the unclear cases to linguistic factors:

> "Levelt argues that it is "more elegant" to ascribe the failure of intuition to "psychological circumstances". I do not see why this is the case; rather, the question of fact remains open".

and he accepts the possibility that natural languages are not r.e. (see above). Lasnik does not do so, but only accuses me of having "the implications reversed":

> "I would not wish to conclude that, say, English is not recursive from the observation that English speakers do not always have sharp intuitions about their language".

If "some linguists" would only wish to understand what they read.

Let us now turn to the "extensional" way of restricting the class. This is, as far as I can see, the proviso Chomsky suggests for guaranteeing learnability if one is not willing to restrict the power of the grammar itself.

2. Reply to Chomsky

Let us start with the essential citation:

> "It might be that universal grammar provides exactly 83 possible grammars, all of which generate non-recursive sets, and that the task of the learner is to select among them, which might easily be possible from finite data, depending on the character of these sets".

Chomsky then goes on to suggest that even if the set is large, these grammars could be arrayed in a hierarchy of accessibility, so that the learner would be quickly led to the correct grammar.

So, even if the grammars are powerful, by restricting the extension of the class[5], and/or providing it with some sort of evaluation metric, a feasible model of acquisition would be constructable. In this way, Chomsky writes, "recursiveness and learnability (in any empirically significant sense of this notion) are unrelated conceptually".

To analyze whether this holds, let us consider the case where universal grammar provides the child not with 83, but with just 2 possible grammars, G1 and G2, specifying languages L1 and L2, respectively (i.e. we assume that L1 and L2 are r. e. - the argument only becomes stronger if that assumption is not made). Both grammars, moreover, have an evaluation value, e(G); we assume e(G1) > e (G2). We can now distinguish the following cases:

(1) $L1 \cap L2 = \emptyset$

(2) $L1 \cap L2 \neq \emptyset$, but $L1 \not\subset L2$, $L2 \not\subset L1$

(3a) $L1 \subset L2$,

(3b) $L2 \subset L1$

Let us consider these in turn:

(1) $L1 \cap L2 = \emptyset$. The two languages are completely disjoint. Presentation of just one positive instance, s, from either L1 or L2 will suffice to select the correct grammar. This is, at least, the case if the evaluation measure is not used to check L1 before L2 : if the correct grammar is G2, checking L1 will never stop. So the algorithm should ignore the evaluation measure and check the grammar in a not strictly ordered way, but otherwise learnability is guaranteed. However, this case is highly unrealistic: natural languages do overlap. The notion of universal grammar derives in part from the striking correspondences between languages.

(2) $L1 \cap L2 \neq \emptyset$, but $L1 \not\subset L2$, $L2 \not\subset L1$. The two languages overlap without strict inclusion. A positive instance s can now either be an element of just L1, of just L2, or of both L1 and L2, but the child has no a priori way of knowing whichever is the case. Let us assume the learner considers L1 first because it has the most highly valued grammar. If $s \in L1$, the learner will recognize this, but it is insufficient basis for selecting G1, since it could be that $s \in L1 \cap L2$. The next step should thus be to check whether $s \in L2$. If it is, it will be recognized

---

[5] Since the extension of the class is, ultimately, restricted by the properties of universal grammar, the distinction between intensional and extensional restriction is only an illusory one, as I indicated above.

and the conclusion is that s ∈ L1 ∩ L2, i.e. the instance does not help in selecting the grammar. If it is not, however, the learner will never recognize that fact, since the complement of an r.e. language cannot be recognized (in the technical sense, see Levelt, Vol. I, p. 113), the learner will keep waiting ad infinitum. Notice that this would be possible if L2 is decidable (and mutatis mutandis if LI is decidable). The situation is, in fact, independent of the evaluation measure: the order in which the grammars are checked is irrelevant, they can as well be checked simultaneously. If the evaluation measure would lead the child to select the highest valued grammar in his position of indecision he will select the wrong one if in fact the lower valued grammar is correct.

(3a) L1 ⊂ L2. The higher valued grammar generates a language which is strictly included in the language generated by the lower valued grammar. I don't know how realistic this is but let us consider it for completeness' sake.

Assume s ∈ L1. This will be recognized. The learner, moreover knows a priori that s ∈ L2, so this information is to no avail for selecting the grammar. Assume s ∈ L2. This will be trivially recognized, and the next step should be to check whether s ∈ L1. If it is, it will be recognized, and we are back in the first situation. If it is not, recognition is not guaranteed, since the complement of L1 is not r.e. Again, this would be possible if L1 were decidable, and again, the evaluation measure can be of no further assistance. So, if only positive instances are presented, there will be no guaranteed selection procedure.

(3b) L2 ⊂ L1. The evaluation measure can be of no avail here either, and the case reduces to (3a).

In conclusion, if only positive instances are presented, there can be no decision procedure for selecting between two grammars for r.e. languages, except if we are willing to assume that the languages are totally disjoint. It is simple to show that under informant presentation (i.e. including a complete enumeration of the negative instances) such decision procedures do exist, but as has been discussed above, that does not seem to be realistic as a model for the child's acquisition situation. An algorithm does exist, however, for text-presentation if LI and L2 are decidable.

So it appears that reducing the cardinality of the class, or using some evaluation measure, are in themselves insufficient to guarantee learnability as long as the grammars generate r.e. languages: the child will never know how to exclude a grammar. Or in other words: as far as these proposals go, the connection between power of the grammar and learnability which I suggested in my original text, still holds. If one wants to maintain unrestricted power for grammars of natural languages, the only way out is to build the learning algorithm in such a way that it relies essentially on the input of structural information. This should be done anyhow (see my remarks cited earlier), but Chomsky's lecture-text does not contain any suggestions along these lines.

References

Chomsky, N. (1962) "Explaining models' in linguistics", in E. Nagel, P. Suppes, and A.Tarski, eds., Logic, Methodoloqy and Philosophy of Science, Stanford University Press, Stanford.

Chomsky, N. (1965) Aspects of the theory of syntax, MIT Press, Cambridge, Massachusetts.

Chomsky, N. (1979) Rules and Representations, Columbia University Press, New York (in press).

Chomsky, N. and G.A. Miller (1963) "Introduction to the formal analysis of natural languages", in R.D. Luce, R.R. Bush and E. Galanter, eds., Handbook of Mathematical Psychology, Wiley, New York.

Gazdar, G. (1979) "English as a context-free language", University of Sussex (mimeo).

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

Joshi, A.K. (1972) "A class of transformational grammars", in M. Gross, M. Halle and M.P. Schützenberger, eds., Formal Language Analysis, Mouton, The Hague.

Lasnik, R. (1978) "A note on learnability and theory construction in linguistics", University of California, Irvine (mimeo).

Levelt, W.J.M. (1974) Formal Grammars in Linguistics and Psycholinguistics, Vol. I, II, III, Mouton, The Hague.

Levelt, W.J.M. (1975) What became of LAD?, Peter de Ridder Press, the Hague.

Newport, E.H., H. Gleitman and L.Gleitman (1977) "Mother, I'd rather do it myself: some effects and non-effects of maternal speech style", in C.E. Snow and C. A. Ferguson, *eds.,* Talking to Children: Language Input and Acquisition, Cambridge University Press, Cambridge, Mass.

Peters, P.S. and R.W. Ritchie (1973) "On the generative power of transformational grammars", Information Sciences, 6, 49-83

Wexler, K. and H. Hamburger (1973) "On the insufficiency of surface data for the learning of transformational languages*, in K.J.J. Hintikka, J.M.E. Moravcsik and P. Suppes, eds., Approaches to Natural Language, Reidel, Dordrecht.

Wexler, K., P. Culicover and H. Hamburger C (1975) “Learning theoretic foundations of linguistic universals”, Theoretical Linguistics, 2, 215-253.

Wexler, K. and P. Culicover (1979) Formal Principles of Language Acquisition, MIT Press, Cambridge, Mass. (in press).