# Learnability

Geoffrey K. Pullum

Final version

**LEARNABILITY**. The mathematical theory of language learnability (also known as learnability theory, grammar induction, or grammatical inference) deals with idealized "learning procedures" for acquiring grammars on the basis of exposure to evidence about languages.

In one classic paradigm, presented in a seminal article by Mark Gold (1967), a learning procedure is taken to be an algorithm running an infinite loop on a never-ending stream of inputs. The inputs are grammatical strings chosen from a target language in a known class of languages. That language has to be identified by choosing a grammar for it from a known set of grammars. At each point in the process, any string in the language might be the next string that turns up (strings can turn up repeatedly). After each input the algorithm produces a guess at the grammar. Success in identifying a language consists in eventually guessing a grammar which is correct for the target language, and which is never subsequently abandoned in the face of additional input strings. A class of languages is called *identifiable in the limit from text* if and only if there exists some learning procedure which, given an input stream from any language in the class, always eventually succeeds in identifying that language.

For example, consider the class of all languages (i.e., sets of strings) over a vocabulary $V$ that consist of all possible strings over $V$ except for one. So if the strings over $V$ are $x_1, x_2, x_3, \ldots$, and the set of all strings over $V$ is $V^*$, the languages of this class are $L_1 = V^* - x_1$, $L_2 = V^* - x_2$, $L_3 = V^* - x_3$, and so on. This class is identifiable in the limit from text, because, as the reader can easily verify, the following procedure suffices to identify any member of it: after each input, guess that the target language is $V^* - \sigma$, where $\sigma$ is the alphabetically earliest string of the shortest length for which not all strings have yet been encountered.

Identifiability in the limit is a fragile property, however. In the case just cited, the addition of a single language to the class can make a new class that is not identifiable in the limit from text: if we add the language $V^*$, the above procedure fails: when $V^*$ is the target language, the procedure will keep guessing incorrectly forever. If there were any procedure that could identify this new class, it would have to guess $V^*$ after some input. Suppose that input is $x_k$. How can any procedure tell at that point that the real answer might not be $V^* - x_j$ for some $j \neq k$? If $V^* - x_j$ were actually the target language, $x_k$ would eventually turn up in the input (because no grammatical strings are eternally missing from the input sequence, and $x_k$ is not the lone ungrammatical string here). If $x_k$ triggers the incorrect guess $V^*$, and there will be no way to recover and guess $V^* - x_j$, because all future data will be compatible with both $V^*$ and $V^* - x_j$. The data is all positive, and no positive data can provide the information that the language currently guessed is a *proper superset* of the target language.

The very abstract and idealized view of language learning under consideration here is mainly used to prove that for certain sets of conditions there is *no* way acquisition from positive examples could take place. Notice, though, that the idealizing assumptions are not by any means unrelated to the situation of actual first language acquisition: the assumption of an unending input stream corresponds to the fact that normal learners may expect that people will continually talk in their presence throughout a learning period that has no set endpoint; the limitation of inputs to grammatical strings corresponds to the fact that learners typically get only evidence about what is grammatical, with no details concerning what is *not* grammatical; and the definition of success corresponds to the idea that it is never necessary to *know* that you have completed your language learning in order for you to be successful: it is sufficient to eventually arrive (without knowing it) at a point where you have nothing else to learn.

Gold proved that none of the standard classes of formal languages (e.g., the regular languages, the context-free languages, or the context-sensitive languages) are identifiable in the limit from text. In fact no class of languages is, provided that it contains all the finite languages and at least one infinite language (all over the same vocabulary).

Gold's results have been taken by both linguists and philosophers to constitute a powerful argument for the existence of innate knowledge of universal grammar that can assist in learning. But such an inference depends on many assumptions. At least six of them might reasonably be questioned. (1) The natural languages might be (unlike, for example, the class of all context-free languages) a text-learnable class; there are large and interesting classes of languages that do have that property. (2) It may be the case that learners receive considerable information about which strings are *not* grammatical — though perhaps indirectly rather than directly. (3) It is not clear that real language learners ever settle on a grammar at all. Feldman (1972) studied a conception of learning under which the learning procedure will eventually eliminate any incorrect grammar, and will be correct infinitely often, but does not necessarily always settle on a unique correct one and stay with it. The entire class of recursively enumerable languages (that is, the class of all languages that have any generative grammar at all) is a learnable class under this conception. But even Feldman's definition of success might be regarded as too stringent. Learners might continuously approach a correct grammar throughout life, adopting a succession of grammars that are each fallible and perhaps incorrect, but each closer to full correctness than the last. (4) Learners could *approximate* rather than exactly identify grammars. Wharton (1974) showed that approximate language learning, under a wide range of different definitions of approximation that allow for the learner to be right to within a certain tolerance but nevertheless wrong about the inclusion of certain strings in the language, is dramatically easier than identification in the limit — again, the entire class of recursively enumerable languages becomes learnable. (5) The idealized input to the learner need not be merely strings. Many investigators have considered the possibility that the learner in some way gains access to information about the structure of expressions as part of the input: for example, a learner is in a dramatically better position as regards identifying languages if the data are unlabeled tree structures, or strings paired with meanings, rather than just strings. (6) It is not necessarily true that language learning is a matter of exactly identifying some specific set of strings — that is, guessing a specific generative grammar. Matters are very different if grammars are taken to be sets of constraints partially characterizing linguistic structure but not necessarily defining a unique set of them.

Possibility (2) makes a huge difference: if inputs are strings paired with indications of whether they are grammatical or ungrammatical, then we have the learning condition that Gold (1967) calls *identification in the limit from an informant*. This is vastly easier. For example, the class of all context-sensitive languages is identifiable in the limit from an informant.

Note also that although the class of all context-free languages is not identifiable in the limit from text, if we choose a number $k$ and consider the class of all context-free languages generable by a context-free grammar with not more than $k$ rules, we find that every choice of $k$ defines a class that is identifiable in the limit from text. And the same is true for context-sensitive grammars.

The above considerations do not permit us to conclude from the facts of human language acquisition that there is some kind of innate language-specialized device in the brain of the human infant that facilitates language acquisition. The view that this is so — known as linguistic nativism — is widely held by linguists, but it cannot be said that mathematical research on learnability supports it.

In the work on learnability that Gold (1967) inspired it is common to adopt the mathematical convenience of equating sentences and grammars alike with numbers. A learner (or rather, a learning procedure) is then just a function from the natural numbers to the natural numbers. The mathematics on which the paradigm is based on recursive function theory. As such, the research in this field is applicable in principle to many other domains than natural language learning. It has applications in pattern recognition, molecular bioengineering, and the formal analysis of scientific investigation.

The development of statistical ways of modeling language learning, particularly by Leslie Valiant (1984), revolutionized learnability theory and led to a new period of growth in the subject during the 1980s and 1990s and the emergence of the concept of *probably approximately correct (PAC)* learning (see Haussler 1996 for an excellent review of the PAC literature to 1996). Statistical computational learning theory is now a major subfield of computer science, and topics have shifted to questions like the conditions under which a language can be learned within certain reasonable (e.g., polynomial) time bounds (i.e., limits on the number of computational steps allowed for learning). However, while biochemistry and molecular biology have been massively influenced by such work, relatively little of it so far has related to the learning of natural languages.

Learnability theory has not restricted itself to viewing languages as sets of strings. Some work that has been of particular interest within computer science has concerned the induction of grammars (actually, finite-state tree automata, which are weakly equivalent to context-free grammars) for infinite sets of trees from finite sets of presented trees. This might be regarded as a model for a learning situation in which a learner encounters a string of known or conjectured meaning and conjectures a structural description for it, basing the learning on the structural descriptions as well as the strings. This work has proved practically important in computational chemistry, in connection with the development of hypotheses about molecular correlates of properties of substances on the basis of a finite corpus of molecular structures of substances with known properties. Again, applications to the learning of natural languages have been relatively few.

A large number of more recent results in learnability theory are presented in the major overview text by Jain, Osherson, Royer, and Sharma (1999). Theorems are proved concerning the effects of many different limitations on learners, for example, requiring that the learner have only a finite memory for what has gone before, and on environments, for example, allowing the input to contain noise (i.e., a finite number of incorrect inputs included along with the correct ones).

Detailed discussion of the relation between learnability theory and the study of human language acquisition may be found in Wexler and Culicover (1980) and Pinker (1984). Kanazawa (1998) presents some interesting results on the learning of categorial grammars, and includes a review of results on languages generated by grammars with not more than some fixed number of rules.

**Bibliography**

Feldman, J. 1972. Some decidability results on grammatical inference and complexity. *Information and Control* 20.244–262.

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10.447–474.

Haussler, David. 1996. Probably Approximately Correct learning and decision-theoretic generalizations. In Paul Smolensky, ed., *Mathematical Perspectives on Neural Networks*, 651–718.

Jain, Sanjay; Osherson, Daniel N.; Royer, James S.; and Sharma, Arun. 1999. *Systems That Learn: An Introduction to Learning Theory*, second edition. MIT Press, Cambridge, Massachusetts.

Kanazawa, Makoto. 1998. *Learnable Classes of Categorial Grammars*. Stanford, CA: CSLI Publications.

Pinker, Steven. 1984. *Language Learnability and Language Development*. Harvard University Press, Cambridge, Massachusetts.

Valiant, Leslie. 1984. A theory of the learnable. *Communications of the ACM* 27 (11), 1134–1142.

Wexler, Kenneth, and Peter W. Culicover. 1980. *Formal Principles of Lan-*

*guage Acquisition*. Cambridge, Massachusetts: MIT Press.
Wharton, R. M. 1974. Approximate language identification. *Information and Control* 26.236–255.