There seem to be two general approaches to the phenomenon of human language learning. One, which I might call *acquisition*, empirically addresses the observation that human children learn to speak and understand language. The other, which I might call *learnability*, theoretically addresses the question of whether, given certain assumptions about the natures of the learner and the language to be learned, such learning is actually possible. While acquisition is data-based, learnability is an entirely abstract and formal exercise motivated largely by computer science and theoretical questions of artificial intelligence.

Atkinson introduces Gold’s 1967 *language identification in the limit* model of learnability by first discussing its four-part framework adapted for language learning from cognitive science:

1. A definition of, or criterion for, learning: there exists some finite time in the learning process at which the learner will correctly identify the language he or she (or it, as the case may be, in a computational framework) is presented with;
2. A learner, or a mechanism by which conclusions about the identity of the language are considered;
3. A hypothesis space: a set of possible such conclusions, enumerated in some way;
4. Data available to the learner: the sentences, and perhaps non-sentences, of the language under consideration.

This model makes rather bold assumptions about time, memory, and language, notably:

1. (1) time is discrete and infinite, and the learner receives a datum at each moment; (2) a language is simply a set of sentences and can be described and identified as such; (3) the learning process is infinite and in it the learner encounters every sentence (or every sentence and every non-sentence) in the language; (4) the learner comes to the learning process equipped with infinite memory, a set of hypotheses (potentially infinite), and a default learning procedure: for any text presentation, at time \( t = n \), the learner is presented with a sentence \( s_1 \) from the unknown language; if \( s_1 \) belongs to the first language in the hypothesis space, \( L_n \), the learner hypothesizes that the language in question has identity \( L_n \) and waits for the next datum; if \( s_n \) does not belong to \( L_n \), the learner checks to see if it belongs to \( L_{n+1} \); if so, the learner hypothesizes \( L_{n+1} \) and waits for the next datum; if not, the learner moves on to \( L_{n+2} \) and repeats this process; after doing this for every piece of data in the sequence, if the learner’s hypothesis has reached a point of stabilization after which it does not change, we can say that the learner has successfully identified the language. These four assumptions, while not realistic, allow us to ponder some interesting questions about the theoretical possibility of language learning.

One of these questions asks whether there is a difference between learning from positive data (where the learner is presented only sentences that are in the language) and learning from both positive and negative data (where the learner is presented both sentences that are inside the language and sentences that are outside of it). Gold called these two types of data presentation *text* and *informant*, respectively. A legitimate text is
a data sequence in which every sentence of the language considered occurs at least once. A legitimate informant is a data sequence in which every sentence and every non-sentence of the language occurs, marked with its status as sentence or non-sentence, at least once.

To investigate this question, Atkinson works through several possible learning situations using two hypothetical classes of languages, \( L \) and \( L^* \), in what he says is “by now a standard way of illustrating the sort of result that can be obtained using Gold’s framework (p 13)”: 

Take the language class \( L \) to be a set of languages:

\[
L = \{ L_1, L_2, L_3, \ldots \},
\]

where

\[
L_1 = \{ a, aa, aaa, \ldots \}
\]
\[
L_2 = \{ aa, aaa, \ldots \}
\]
\[
L_3 = \{ aaa, \ldots \}
\]

etc.

We must note that \( L \) is an infinite class, and that its languages are enumerated here in a systematic way, i.e. the name of each language corresponds to the length of the shortest sentence it contains. We will now consider a hypothetical learning situation in which a learner is to be presented with some language from \( L \) of unknown identity and must correctly identify it from a text.

Suppose the language in question has identity \( L_n \). No matter what the value of \( n \), the concatenation of \( n \) ‘a’, the shortest sentence of the language, will occur at some point in the data sequence. If the learner’s hypothesis selection procedure sets \( i \) in the hypothesis \( L_i \) at the length of the shortest string presented so far, we can see that any language from the class \( L \) will be identified in the limit.

So \( L \) contains infinite languages only. Let’s see what happens when we look at a class that also contains finite languages. Take the class \( L^* \) to be a set of languages:

\[
L^* = \{ L_0, L_1, L_2, L_3, \ldots \},
\]

where

\[
L_0 = \{ a, aa, aaa, aaaa, \ldots \}
\]
\[
L_1 = \{ a \}
\]
\[
L_2 = \{ a, aa \}
\]
\[
L_3 = \{ a, aa, aaa \}
\]

etc.

All languages in \( L^* \) with the exception of \( L_0 \) are finite. And, like \( L \), they are systematically enumerated. We can try a hypothesis selection procedure like the one we used in identifying \( L \) languages, but instead of setting \( i \) in the hypothesis \( L_i \) to the length of the shortest string presented so far, we will set it to the length of the longest one. This will successfully identify all of the finite languages, but what of \( L_0 \)? Since \( L_0 \) has no longest member, at any point in the data sequence there will always be a string longer than any presented so far. Thus the learner’s hypothesis will never stabilize. What if we try a different hypothesis selection procedure, say, “set \( i \) in the hypothesis \( L_i \) at 0,” then let default hypothesis selection take over? If the target language is \( L_0 \), this procedure will
indeed successfully identify it. But, if the target language is one of the finite languages, there will be no datum that is not accepted by the hypothesis $L_0$, and the learner will never even consider any of the finite languages. So, while our first hypothesis selection procedure correctly identifies the all of the finite languages in $L^*$ but fails to identify the infinite one, our second procedure correctly identifies the infinite language but fails to identify all of the finite ones.

So much for the hypothesis selection procedures governed by this enumeration of the hypothesis space. What about other enumerations of $L^*$? We shall divide all the possible enumerations of $L^*$ into two classes: (1) enumerations in which $L_0$ precedes some finite language $L_n$, and (2) enumerations in which every finite language precedes $L_0$. In the case of (1) with target language $L_n$, when $L_0$ is reached, it will be selected, and since $L_n \subset L_0$, there is no datum from $L_n$ that can cause the learner to switch hypotheses. In the case of (2) with target language $L_0$, there is no finite time at which the learner will stabilize on the correct hypothesis, since there is an infinite number of finite languages to hypothesize before reaching $L_0$. Thus, regardless of the enumeration, $L^*$ is not identifiable from text.

Now that we have seen what happens when the learner is presented with positive instances of the languages from $L$ and $L^*$, let’s examine what happens when the learner is presented with both positive and negative instances. Will the introduction of negative data change whether $L^*$ is identifiable?

Let’s modify our previous hypothesis selection procedure to take advantage of negative data. We shall set $i$ in the hypothesis $L_i$ initially at 0; if $l$ is the length of the shortest non-sentence presented so far, we can successfully identify the language in question by setting $i$ in the hypothesis $L_i$ to $(l - 1)$. In the case that $L_0$ is the language in question, and since there is no negative data in an informant data presentation ($L_0$ contains every possible combination of ‘a’), the learner’s initial hypothesis will be maintained and $L_0$ will be successfully identified. In the case that some finite language $L_n$ is the language in question, any legitimate data sequence will eventually present the learner with the shortest non-sentence in $L_n$, and the learner’s hypothesis at that moment will switch to $L_n$ and remain there. Thus, while not identifiable from text, all languages in the class $L^*$ are identifiable from an informant! The introduction of negative data dramatically affects learnability.

Let’s see what generalizations we can make from these results. First of all, what is the relationship between $L$ and $L^*$? In Chomsky’s (1956) model of formal language types, commonly called the Chomsky hierarchy, the class of languages generated by unrestricted rewrite systems (grammars consisting of rules that rewrite any string of symbols as any other string of symbols; i.e. $abAxBcb \rightarrow XaaYb$) is a superset of the class of languages generated by context-sensitive grammars (grammars consisting of rules that rewrite one non-terminal symbol as any string of symbols in a given context; i.e. $ab[X]cd \rightarrow ab[xy]cd$), which is a superset of the class of languages generated by context-free grammars (grammars consisting of rules that rewrite one nonterminal symbol as any string of symbols in any context; i.e. $X \rightarrow xy$ or $X \rightarrow aAbB$), which is a superset of the class

---

1 I am adopting here the convention of naming nonterminal symbols by upper-case letters and terminals by lower-case letters.
of languages generated by *finite state grammars* (grammars consisting of rules that rewrite one nonterminal as any string of symbols containing at most one nonterminal symbol; i.e. $X \rightarrow aA$, $X \rightarrow abYcy$, and $X \rightarrow ab$). We shall call these language classes $L_0$, $L_1$, $L_2$, and $L_3$ respectively. If a the class $L_0$ is not identifiable in the limit in a particular model, then neither are $L_1$, $L_2$, or $L_3$: “if there is no effective strategy for searching through a set of hypotheses, this situation will not be improved by extending the set” (p 20).

Where among these classes do $L$, $L^*$ and the natural languages fit? To attempt to place the natural languages in Chomsky’s hierarchy is a task far too extensive for our time and purposes, but to place $L$ and $L^*$ is simple, and revealing. To begin, we can easily see that all of the finite languages can be generated by grammars that have one rule for each sentence: to generate the language $L_3 = \{a, aa, aaa\}$ from $L^*$, we need only the rules \{S$\rightarrow$a, S$\rightarrow$aa, S$\rightarrow$aaa\}; for a language $L_n$, we need only $n$ such rules. The infinite language $L_0 = \{a, aa, aaa, aaaa, …\}$ from $L^*$ can be generated by a finite grammar with a recursive rule: \{S$\rightarrow$aS, S$\rightarrow$a\}. All of these rules rewrite one nonterminal symbol as a string of symbols containing at most one nonterminal. Thus, $L$ and $L^*$ are subsets of $L_3$, and we arrive at the interesting conclusion that *none* of the language classes in the Chomsky hierarchy are identifiable in the limit.

So if English or any other natural language can be generated by some finite set of rewrite rules, then it follows that the entire class of natural languages is *not* identifiable in the limit; not learnable! The question now is whether any of the natural languages is, in fact, generable by a finite grammar, and whether our definition of learnable is plausible in the case of human language acquisition. Given the vastness and centrality of these questions to linguistics, I shall not attempt to further explore them here.