Gold’s Theorem and Language Learning

1 An abstraction of “learning” a language

A language, as we have seen, is abstractly represented as a set of strings. Each string represents a sentence. (We will use the word sentence interchangeably with string). The process of learning a language can be similarly abstracted: a language is learned when a learner (a student or child that does not already know the language) somehow acquires a means of judging, for any possible string, whether that string is in this language. The “means of judging” might be anything: a list of all correct sentences (obviously, this can only work for finite languages), a set description of the language plus a way of figuring out if a string is in that set, or an automaton of some kind that can respond with “accept” or “reject” for any string it is given. Whatever it is, we can think of it as a “grammar” that is built to handle the job of telling grammatical strings from ungrammatical ones.

Definitions: The language being learned is called the target language. A language is finite if it contains a finite number of strings, otherwise it is superfinite. (A superfinite language could be regular, context-free, or even more complex). A string (i.e., sentence) is called grammatical if it is an element of the target language.

In order to simplify enough to tackle the issue at all, assume that the learner is really good at building grammars based on the available evidence. How good could that be?

1.1 Positive and negative information

A lot depends on the kind of information available to the learner. Obviously the learner must encounter many strings from the target language, or no learning could take place. We assume that the learner is exposed to an unending sequence of strings from the target language, one by one, and that it (he/she) knows that the strings in this sequence are grammatical. This is positive information.

A student learning a foreign language may also try to construct sentences, and is told when they are ungrammatical. Knowledge that certain strings are not in the target language is called negative information.

1.2 The perfect learner

Suppose we assume that a learner has perfect memory, so that it remembers every grammatical string it has ever heard; and that it is really good at constructing grammars. Here is a learning algorithm based on positive evidence only:

---

This handout is adapted, with permission, from notes provided by David Powers, of the School of Informatics and Engineering, Flinders University, Adelaide, South Australia.
A learning algorithm: Every time you hear a sentence, check if your grammar can handle it. If it can, fine; otherwise, revise your grammar so it can handle both the new sentence and everything else you have heard in the past.

Definition: If a learner can eventually “learn” a grammar that accepts all sentences in the target language and rejects all that are not, then the learner has accomplished perfect learning.

2 Gold’s theorem

The question: If our only information is a series of sentences from the target language, which we hear one by one (positive information), is there some point at which the learner can be guaranteed to have learned a grammar perfectly?

Gold’s Theorem: Gold proves that perfect learning of a superfinite language is impossible, if the learner only has access to positive information.

This follows directly from the observation that after any finite time you have seen only a finite number of sentences, so you have a finite candidate language (which you may discard as not superfinite) as well as many superfinite extensions of it. E.g., recursive rules could insert extra words before or after any of the words. There is no basis for distinguishing between these. Any of these may generate sentences that are not in the language. Some of them will generate the right language—but we can’t tell which.

Providing negative information of the form yes/no to sentences composable from the lexicon will allow learning of regular, context-free and context sensitive languages (and other languages up to a level of complexity called primitive recursive). This kind of learning is just like learning with a teacher whom you can ask questions about things that you are unsure, (“Supervised learning”).

A theorem about negative information The family of candidate languages discussed above can be pruned by generating sentences and getting the teacher to say yes/no, until the wrong ones are all eliminated. “Primitive recursive” languages have the property that they can be listed systematically, so this is always possible. The learner could generate the sentences it wants answers for, or all possible sentences can be listed and their status (grammatical or ungrammatical) given to the learner. In either case, the bad grammars will be pruned after a finite time number of sentences.

3 Assumptions associated with Gold’s theorem

The above theorems depend on many assumptions about language and learning. Those who interpret the theorems make even more assumptions. All of these assumptions may or may not be true, and affect the applicability of Gold’s theorem to human language.

Note that we are not interested in assumptions like “the learner has perfect memory,” in this context: If it is impossible for a learner with perfect memory to learn a language, a learner with imperfect memory certainly could do no better!

1. Poverty of the stimulus: Only explicit judgements (corrections) are useful as negative information, and these are not available.
Linguistic studies found that parents (contrary to what they may believe) almost never correct their children’s use of syntax; parents are much more interested in the truthfulness and appropriateness of what their children say. Also, children who are corrected generally ignore corrections and persist in their (temporary) errors.

2. Natural languages are superfinite.
   **Objection:** The linguist Charles Frith showed that adjectives have a fixed order and are not recursive. Perhaps there really is only a finite number of possible sentences?

3. People have infinite heads: The type of automaton that goes with Context Free Languages needs an infinite stack.
   **Objection:** Our heads have finite (but very large) capacity, and without an infinite stack we can’t recognize higher than regular languages.

4. Sentences are the target of what’s being learned, and can involve arbitrarily many conjunctions (‘and’, ‘that’, ‘who’, participles, commas)

5. Conjunctions within sentences are intrinsically different from the punctuation/prosody that separates sentences. (Otherwise, what we learn as a “sentence” might end at a conjunction).

6. Nesting can be arbitrarily deep. In other words, no matter how many if/then’s, either/or’s, etc. we use, the sentence is grammatical if they are paired correctly.

7. Learners learn a language perfectly.
   **Objection:** If in fact we only learn our parent’s language approximately, then Gold’s theorem does not apply.

8. Probabilities do not play a role.
   **Objection:** Learning algorithms that include information about how frequently we see each kind of sentence are more powerful. Within two years of Gold’s paper, probabilistic proofs of superfinite learning appeared.

9. Semantics and pragmatics plays no role.
   **Objection:** The context of use, and feedback of a non-linguistic kind, do influence language learning.

3.1 Could evolution get around Gold’s theorem?

Chomsky’s conclusion from Gold’s theorem was that since children do somehow manage to learn language, we must all somehow be born knowing the hard stuff already.

But note that an adopted child from any part of the world can learn any other language, regardless of ancestry. So the specifics of any particular language cannot be genetically inherited; only the features that are common to all languages in the world can be in our common genetic inheritance. That’s the doctrine of **Universal Grammar**.

On the other hand, evolution is very slow, and genetic learning even slower.