Tommy Angelini
Physics 498BIO
Nigel Goldenfeld
12/18/01

Universal Grammar

Linguistics is an interesting field, in that it makes connections to a broad range of sciences, such as mathematics and the study of artificial intelligence, the social sciences (of course), and the study of complex systems in physics. I present work suggesting that a common view of language learning is flawed. The solution to this problem, according to many, is universal grammar. I introduce universal grammar in basic terms, and present work that studies universal grammar as a complex system in a formal and mathematically rigorous way.

Language Learnability

When children learn language they are seldom informed when they make grammatical errors, and when corrections are made, children often ignore them [1]. Children do eventually learn the language of their parents, however. This raises the question as to whether one can learn a given language by hearing a small number of sample sentences from that language. This is how human beings seem to do it. In developing artificial intelligence, the following model of language learnability was proposed [2]:

a. Strings from some language are repeatedly presented to a learner in either any arbitrary order or in some specific recursive way.
b. The learner takes a 'guess' by producing a string of text, drawing from the strings it has been given.
c. An informant may or may not correct the learner's error (by telling the learner whether or not the guess string is part of the language), depending on the model.

This process is iterated for a finite amount of time. If the guesses are consistently correct, the language was said to be *identified in the limit*. The most important finding, for the purposes of this essay, is that *none* of the relevant model languages were learnable in the absence of an informant, regardless of the method of string presentation. So, it seems that humans must have some mechanism very different from this model for learning languages.

Universal Grammar

The ways to arrange nouns, verbs, adjectives, independent and dependent clauses, subject, object, auxiliary, and so on, is really innumerable and by the time human beings begin to grasp language we have hardly heard enough to deduce a set of rules to follow when we speak. This was illustrated in the last section. The most popular solution to this problem is known as universal grammar.

Noam Chomsky's view of linguistics and of universal grammar is summarized as follows.

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by memory limitations, distractions, etc. … I have in mind certain biological properties, the most significant of these being properties that are genetically-determined and characteristic of the human species, which I will assume, for the purposes of this discussion to be genetically uniform, a further idealization. These properties determine the kinds of cognitive systems, language among them, that can develop in the human mind. In the case of language, I will use the term "universal grammar" to refer to these properties of human biological endowment [3].

Chomsky goes further to tell us what he *does not* mean by universal grammar. He is not interested in the characterization of language, as such, i.e. universal grammar is not the logically or conceptually necessary properties of language, but the biological properties humans happen to have, that determine the particular cognitive system called language.

Evolution of Universal Grammar

Now we look at language learning from the point of view of a physicist [4]. Consider that the universal grammar of human beings, U, allows for the learning of particular mental grammars (the grammar anybody happens to adopt and use), G_i, and that there are n different candidate grammars. Each grammar is a different set of rules for generating valid sentences. Let's say that the grammars overlap in some way: it may be possible to formulate a sentence in grammar G_i that could be also formulated (and have the same meaning) in grammar G_j. So we can call a_{ij} the probability a speaker using G_i formulates a sentence compatible with G_j. The probabilities are normalized as

$$\begin{gathered} 0 \le a_{ij} \le 1 \\ a_{ii} = 1. \end{gathered} \tag{1}$$

The payoff for successful communication between a particular pair of speaker-listeners is

$$F(G_i, G_j) = (a_{ij} + a_{ji})/2, \tag{2}$$

and the average payoff for each individual speaking grammar G_i is

$$f_i = \sum_j x_j F(G_i, G_j), \tag{3}$$

where x_j is the number frequency (i.e. the fraction) of individuals using grammar G_j. In reality, the 'payoff' is, of course, sex and hence more offspring.

Parents don't always successfully teach the right grammar to their children. The probability that a child with parents speaking G_i will learn G_j is Q_{ij}. With all of the aforementioned considerations taken into account, the time derivative of the fraction of people speaking G_i is

(4)

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j Q_{ij} - \phi x_i \, .$$

The first term just tells us that the rate of growth of a grammar depends on successful cross-grammar communication (by whatever means) mediated by mistakes in learning. In the second term, $\phi = \Sigma_i x_i f_i$ is the probability that a sentence said by one person is understood by another person, and is called average grammatical coherence. This term accounts for the fact that the more grammatically coherent the population, the slower the rate of change of x_i, and the less coherent, the quicker the change.

Now we have defined the model, and we are interested in the stable solutions to equation 4. The first case we consider makes all grammars "equidistant" by setting $a_{ij} = a$. The solutions in which all $x_i = 1/n$ are called the *symmetric* solutions, and the solutions in which one grammar, G_i dominates are called the *asymmetric* solutions. When the number of candidate grammars, n, greatly exceeds $1/a$, the solutions are restricted to the conditions

$$\begin{aligned} q &> \frac{2\sqrt{a}}{1+\sqrt{a}} \equiv q_1 \; (\textit{asymmetric case}) \\ q &< 1 - \frac{1-a}{na} \equiv q_2 \; (\textit{symmetric case}) \end{aligned} \tag{5}$$

Since an asymmetric solution is required for grammatical coherence (the society must converge onto a single G_i), we see that $q > q_1$ is the necessary condition. So, we call q_1 the coherence threshold. One can solve the case for $n < 1/a$, and the solutions turn out to be $2n^{-1/2}$ and ½ for the asymmetric and symmetric cases, respectively.

If we attribute the number of, and closeness between, candidate grammars to the universal grammar, we see that the coherence threshold is a function of the universal grammar. Only a universal grammar that satisfies the coherence threshold can lead to grammatical communication.

We can apply this model to two specific cases: (1) the memoryless learner and (2) the batch learner (memorizes *everything* and compares new input to the batch). One can imagine that a human being is somewhere between these two, so the results from these two cases might serve as boundaries.

For the case of the memoryless learner, the teacher utters a sentence using grammar G_k, and the learner starts out assuming that his grammar G_i is the desired grammar. If the sentence is consistent with G_i, the learner maintains the hypothesis that G_i is the desired grammar. Otherwise the learner picks a new hypothesis for a candidate grammar G_j. Iterate this until all b possible sentences within the universal grammar have been said. The probability that the memoryless student has learned G_k is

$$q = 1 - \left[1 - \frac{1-a}{n}\right]^b . \tag{6}$$

Applying the coherence threshold condition, one can show that $b \sim n$. The batch learner, on the other hand, memorizes all b sentences and searches each grammar most consistent with *all* of the sentences. The probability that he picks G_k is

$$q = \left[1 - (1 - a^b)^n\right] / \left[na^b\right]. \tag{7}$$

Again, applying the coherence threshold condition, one can show that $b \sim log\ (n)$. So, this makes sense. The batch learner learns much more efficiently, and requires a 'smaller' universal grammar to satisfy the coherence condition.

One can take this model further, and add more realism. We can make some sets of grammars more likely to overlap than others, or we can make some more difficult to teach to one's offspring. Such variables would add realism to the models, and make their results more realistic too. Finally, one can compare competing universal grammars with different b's and n's. I will not discuss these here, but they can be found in ref [4].

Conclusion

To summarize, we have seen that traditional mathematical models of language learning fail to produce results similar to human learning, and that the idea of a universal grammar solves this problem. Additionally, we have seen the application of the concept universal grammar as a complex system, yielding quite nice results.

Finally, I see a few problems with this entire 'problem' in linguistics:

1. Gold's requirement for an informer can be satisfied without parental intervention: the child's own memory can easily serve as a quasi-informant, as long as the child can maintain focus on and recall repeated verbal references to whatever aspect of the grammar he is trying to learn. This relieves the need for a universal grammar.
2. Chomsky's universal grammar is a misnomer. Calling the properties of our biological endowment that allow us language a 'grammar' loads the dice. Rather, his universal grammar is actually *something* which allows for human beings to use one or many different grammars.
3. Even if universal grammar is a valid concept, there is no reason to believe that there should be a *quantized* set of candidate grammars, nor that a universal grammar must have n of these, nor that a universal grammar can be consistent with only b different sentences.

Although the idea of a universal grammar lends itself nicely to this population dynamics problem, I'm skeptical of whether we're actually referring to anything that exists at all.

Bibliography

1. McNeill, D. *Developmental Psycholinguistics* (1966).
2. Gold, E.M. *Information and Control* **10**, 447 (1967).
3. Chomsky, N. *Rules and Representations* (Columbia University Press, New York, 1980).
4. Nowak M.A., Komarova N.L., Niyogi, P. *Science* **291**, 114 (2001).