1

Doing Science with Language
Introductory Concepts

1.1 What is Scientific Inquiry?

What differentiates a scientific inquiry from any other sort of inquiry or theorizing? One core feature of scientific inquiry is what we might term a **testable hypothesis**, one that makes predictions that we can test. A "testable" hypothesis is one that can potentially be falsified by data. Should data not match the predicted outcomes of such a hypothesis, then we might (1) reject the hypothesis in favor of a different one that makes better, more accurate predictions; (2) revise the hypothesis if the revision is straightforward; or, if there is no alternative hypothesis or obvious revision, (3) maintain the hypothesis but note the problem for future inquiry. In terms of getting at the truth of how something works, there is no great answer-book in the sky. The only tools that we have to discover the nature of things are hypothesis formation and testing. These form the basis of everything that we know about anything in the universe from a scientific standpoint.

Often, when people talk casually and dismissively about "theories" (e.g. "Oh, that's just a theory!"), they seem to mean raw speculation or wild and unsupported guesses. This isn't what scientists mean by the term. Let's consider a (**scientific**) **theory** to be an overarching framework of thought that embodies a collection of hypotheses – in the best case, ones that are borne out by data and thus have some empirical support. We can think of a theory as having what we might term **empirical weight** in direct proportion to the number of facts/amount of data that the hypotheses within that theory are successful in predicting/explaining. Some theory A can be considered "competitive" with another theory B if it can be shown that A has similar empirical weight to B, though the two theories might not explain all of the

Syntactic Analysis: *The Basics* Nicholas Sobin

same phenomena. But such comparisons are tricky. It may be that some theory has what appears to be a lot of empirical weight, but just can't explain certain nagging facts. Another theory comes along that can neatly predict/explain these nagging facts, though it may otherwise be incomplete and in need of some "filling out." It has sometimes turned out that the theory which could explain the "nagging facts" was in the end the right one.

Here's a quick example of the latter case. How would you answer the question, "Did the sun rise this morning?" Nearly everyone would say "Yes." The basis for this answer lies not in how the universe actually works, but in the visual impression that we have, and possibly in the medieval (and earlier) belief (based on such visual impressions) that the earth was the center of the universe and everything in the sky was going around it. That earth-centered ("Ptolemaic") theory of the universe was based on a vast multitude of observable facts – the sun, the moon, and every star appeared simply to be going around the earth. Further, no one on the earth had (or has) any direct sensation of the earth moving (rotating). However, there are five objects (the five visible planets) that didn't simply pass by in a linear fashion, but instead appeared to backtrack in their courses (something called "retrograde" motion). For those objects, the earth-centered view had no good explanation. But, if we count each of the smoothly progressing star paths as a "fact," then given the thousands of visible stars, it looks like the earth-centered view predicts the large mass of facts correctly, with only five exceptions – not too shabby. But completely wrong! Copernicus's work (and that of others) to explain the five exceptions put the sun, not the earth, in the center of the "solar system" (a new concept) with only the moon circling the earth, and this view has ultimately proved correct. Further, the sun never rises – the earth rotates.

This little story has two major points. First, hypotheses/theories based purely on visual impressions (doing science by looking out the window) might be quite wrong – you need experimentation and theorizing. And second, a theory that can make sense of the nagging facts, even though it doesn't match sensory experience or immediate intuition, may turn out to be the right one. So doing scientific inquiry isn't always easy, but it is essential to understanding how things in the world actually work.

1.2 The Science of Language – Linguistics

Linguistics is the scientific study of human languages and the human language capacity. Our understanding of how human languages are structured and learned is only advanced by hypothesis formation and testing. Human language is a strongly subconscious mental faculty. While all humans are able to acquire at least one language at an early age and are able to speak it and understand it almost effortlessly, they have no conscious access to it. Often

the "rules" that they firmly believe to hold in a language are wrong, even ridiculously off the mark, and are not followed by anyone speaking the language. Here's an example.

What is a pronoun? Many, maybe most, would say that it is a word that substitutes for a noun. Let's test that idea. Consider the noun *book* in (1):

(1) The red book is over there.

If you actually substitute a pronoun for the noun, you get (2), something that no English speaker would ever say:

(2) *The red it is over there.

What pronouns actually do is substitute for an abstract unit of structure called a *noun phrase*, a noun and all the words that immediately go with it. In (1), that would be *the red book*, and substituting *it* for *the red book* gives us the possible sentence (3):

(3) It is over there.

Looking at this in terms of hypothesis testing, what we have just done is to compare two hypotheses, one stating that pronouns substitute for nouns, and another stating that pronouns substitute for NPs, and tested them. The first makes two incorrect predictions, namely that (2) is good and that (3) is impossible. The second hypothesis makes two correct predictions, namely that (2) is impossible and that (3) is good. So, we think the second hypothesis is correct, at least insofar as we can tell from such testing. And therefore we also think that there is such a thing as an NP, even though we can't directly observe it.

An important perspective here is that such testing is the only means that we have to discover things. We can't ever prove that something is ultimately right. The most we can say is that as far as we can tell from advancing and testing testable hypotheses, some particular vision of how things are (a theory) is the best that we can come up with. Those are the limitations of scientific inquiry.

1.3 The Cognitive Revolution

Alongside the goal of understanding how human language grammars are structured is the goal of explaining how human language grammars are **acquired** by children. Linguists use the term **acquisition** to refer to **subconscious learning**, the sort of learning that is characteristic of human language grammar. Young children hear only a relatively small number of sentences (relative to the infinitely many available ones) from which they

appear to "learn the language." No one tells them anything directly about the grammatical system underlying the language – in fact, linguists are still trying to discover what it consists of and how it works! Nonetheless, it is the grammatical system that a child must acquire in order to be able to speak the language, and, miraculously, that is exactly what each child manages to do. Though no one knows the details of how this is accomplished, there are some general ideas about how it works that are very likely on the right track. Here, we'll first consider a couple of what we might term "common sense" approaches to explaining language acquisition that don't actually pan out. Then we'll consider a more promising line of thought on the problem of acquisition.

1.3.1 Reinforcement

One early view of how language is acquired/learned is based on the stimulus–response model of learning. The idea is that positive and negative reinforcement would provide the "fundamental basis" for language "learning." I use the term "learning" here, since this would to some extent be the "conscious" learning of language. There are a number of problems with this view.

First, if the learner were to say a sentence like "Don't take some apples" and was told "No, that's wrong," the learner would not know what in particular was wrong, since the phrase "some apples" is perfectly correct in many circumstances. If the learner were told more specifically, "Don't say 'some'," then could that mean that the learner should never say it? Even if the learner were given a corrected form as in "Say 'Don't take *any* apples'," this still isn't sufficiently informative about why or when to say *any* (e.g. perhaps you're not supposed to say *some* with the word *apples*, or perhaps you're not supposed to say the word *some* after the word *take*, etc.). In sum, negative reinforcement and even corrections are much too ambiguous to drive language (grammar) learning/acquisition.

Second, studies show that when children are corrected, the corrections are mostly about the truth/accuracy of what the learner is saying and not about grammatical form. The amount of correction of sentence form that actually takes place is fairly minimal, far too minimal to be the basis for language learning/acquisition.

Third, studies show that children are not even capable of reproducing a "corrected" form before they are ready to do so, that is, before the grammar in their heads has developed to the point that that form is a possible production. Before this time, children famously persist in producing the **learner's forms** (utterance forms unique to children acquiring a language), regardless of the amount of correction that they are exposed to. It appears then that reinforcement is not and could not be the central device by which children acquire language.

1.3.2 Imitation

A somewhat popular "common sense" notion of how children acquire/learn a language is that they do it through imitation. This idea collapses rather quickly, however, on some of the same grounds as reinforcement does. The simple fact is that children at particular stages of language acquisition are not capable of imitating adult forms. Further, they persist in learner's forms (*I don't want some apples*) that are distinct from adult forms, and for which there is no model. Thus, imitation offers no explanation whatever of why learner's forms arise at all. It is also worth noting that imitation is nothing but the imitation of sentence forms; it is not at all clear how a learner would or could proceed from such imitation to acquiring the grammatical system, which is what the learner actually must acquire in order to be able to produce any of the infinitely many possible sentences in the language. So imitation also appears to fail as a plausible central device for language acquisition.

1.3.3 Innateness: Principles & Parameters

No one has been able to construct a plausible theory of how a child might use only the sentences that she or he hears to develop an explicit grammar of a language. The alternative is to think that perhaps human children are "hard-wired" to learn the language(s) in the surrounding environment. That is to say, the human brain may be genetically programmed for recognizing language input and knowing how to use it to construct the grammar for the language which that input exemplifies. Some general facts suggest that this is the correct approach.

First, all children go through the same general "stages" of language acquisition, regardless of which particular language is being acquired. These stages include a **One-Word Stage**, in which the child creates only single-word expressions (though the intended meaning is more complex); a **Two-Word Stage**, in which two and only two words may be put into a sentence, regardless of the complexity of the intended meaning; and then a **Telegraphic Stage**, in which three or more content words (nouns, verbs, etc.) may be used to form longer sentences, but **function words** (e.g. *the*, *a*, *at*, *on*, *be*, etc.) are still largely absent. Children then proceed to develop use of the function words, along with developing more complex sentences. Though the stages and the reasons for them are not fully understood and are still the object of much research, the very presence of such cross-linguistically uniform stages strongly points to the presence of an innate program for language acquisition.

Second, children acquiring the grammar of a language produce normally irregular adult word forms (e.g. *swam*) as regular (e.g. *swimmed*), again without any external model for such regularization. This is one of many clear

indications that children are subconsciously forming rules and using them broadly. Children learning English show this behavior in regularizing irregular verb forms (such as *swimmed*) and irregular noun forms (so that *feet* becomes *foots*, and plural *fish* becomes *fish(i)es*).

Although it is clear that children are acquiring the rules of a grammar, until recently it was not clear how they were going about it. When linguists attempt to discover grammatical rules, both **positive data** (data about which word sequences are 'good' sentences in the language) and **negative data** (data about "bad" word sequences, ones that are not sentences in the language) are crucial. That is, linguists have to know what the speaker both can and cannot do in forming sentences. For example, the subconscious grammar of English allows you to say, *Who saw what?* and *What did Mary see?* but not * *What did who see?*. (The * marks a sentence that is not possible in the language.) Such positive and negative data are crucial to discovering the rules that regulate the appearance of *wh* words like *what* and *who*. (In fact, such data are needed in any science – chemistry, for instance, is founded on information about which elemental combinations are possible and which are not.) If a child were discovering the rules of the grammar like a linguist (or any scientist) does, then the child would also need both positive and negative data. However, research in child language acquisition has shown that the *primary data* (all of the language input that a child hears and uses to subconsciously construct the grammar of the surrounding language) contains almost exclusively positive data. (Recall what was said above about correction.) Key negative data as exemplified above are simply absent from the primary data, and the few negative data that are present are of questionable value. This indicates that the child could not be working at grammar construction like a linguist does. So on what basis does a child construct her or his grammar? How does a child do it with only positive data, something that no scientist could do?

A recent and very promising theory of how children form the grammar of a language is what is called the **Principles & Parameters model**. The idea is that while much information key to forming a grammar comes from the primary data, the framework for constructing a grammar is inborn/"hard-wired" into human cognition in the form of **Principles** and **Parameters**. We can think of Principles as inviolable rules that must be followed in the construction of a grammar for any language. These are an inherent part of the innate human language capacity, something that has been given the label **Universal Grammar(UG)**. Every human is born with it, and it is essential to allowing each human being as a child to discover the grammar of the surrounding language. Parameters form another part of UG. They are also rules of grammar formation that must be followed, but they contain an "open setting" whose value is determined by relevant items in the primary data. This conception of the nature of language acquisition is in its infancy and needs a lot of filling out – there has been and continues to be a lot of debate about what the actual

Principles and Parameters are. However, given that children only have positive primary data to work from in constructing a grammar, it is still the most promising general theory of language and language acquisition that has been proposed thus far.

We will not try to exemplify particular Principles or Parameters at this immediate point, since we haven't established enough of a linguistic analysis to make this exercise very meaningful or contentful. However, as we proceed through our development of a linguistic theory, we will note and consider various possible candidates.

In the chapters to follow, we'll try to accomplish three things. First, we'll introduce terms and concepts basic to the study of human language, and especially syntax; second, we'll explore the operational details of particular hypotheses/theories of syntactic structure; and third, we'll put some emphasis on argumentation and hypothesis testing – the sort of work that linguists actually carry out in order to advance linguistic theory.

Summary Points of This Chapter

- A "theory" in the empirical sciences is an overarching view of how some part of the world works based on successfully tested hypotheses.
- Linguistic theory is a theory of how the grammatical systems that produce human languages are built and how they are acquired.
- Linguistics, like all empirical sciences, employs hypothesis testing to advance linguistic theory.
- Language acquisition, the subconscious learning of the grammatical system of a language, may to a large extent be "hard-wired," as indicated in part by the existence of cross-linguistically uniform stages of acquisition.
- Universal Grammar, the "hard-wired" basis for any language-particular grammar, may involve completely fixed Principles, and partially fixed Parameters, the latter allowing for some of the differences in sentence structure that different languages exhibit.