Syntax

A Generative Introduction
Second Edition

Andrew Carnie

Blackwell Publishing

chapter 1

Generative Grammar

0. Preliminaries

Although we use it every day, and although we all have strong opinions about its proper form and appropriate use, we rarely stop to think about the wonder of language. So-called language "experts" like William Safire tell us about the misuse of *hopefully* or lecture us about the origins of the word *boondoggle*, but surprisingly, they never get at the true wonder of language: how it actually works. Think about it for a minute; you are reading this and understanding it but you have no conscious knowledge of how you are doing it. The study of this mystery is the science of linguistics. This book is about one aspect of how language works – how sentences are structured: ***syntax***.

Language is a psychological or cognitive property of humans. That is, there is some set of neurons in my head firing madly away that allows me to sit here and produce this set of letters, and there is some other set of neurons in your head firing away that allows you to translate these squiggles into coherent ideas and thoughts. There are several subsystems at work here. If you were listening to me speak, I would be producing sound waves with my vocal cords and articulating particular speech sounds with my tongue, lips, and vocal cords. On the other end of things you'd be hearing those sound waves and translating them into speech sounds using your auditory apparatus. The study of the acoustics and articulation of speech is called ***phonetics***. Once you've translated the waves of sound into mental representations of speech sounds, you analyze them into syllables and

like the rules of grammar you might have learned in school. These rules don't tell you how to properly punctuate a sentence or not to split an infinitive. Instead, they tell you the order in which to put your words (in English, for example, we put the subject of a sentence before its verb; this is the kind of information encoded in generative rules). These rules are thought to generate the sentences of a language, hence the name *generative* grammar. You can think of these rules as being like the command lines in a computer program. They tell you step by step how to put together words into a sentence. We'll look at precise examples of these rules in the next chapter. But before we can get into the nitty-gritty of sentence structure, let's look at some of the underlying assumptions of generative grammar.

Noam Chomsky

Avram Noam Chomsky was born on the 7th of December 1928, in Philadelphia. His father was a Hebrew grammarian and his mother a teacher. Chomsky got his Ph.D. from the University of Pennsylvania, where he studied linguistics under Zellig Harris. He took a position in machine translation and language teaching at the Massachusetts Institute of Technology. Eventually his ideas about the structure of language transformed the field of linguistics. Reviled by some and admired by others, Chomsky's ideas have laid the groundwork for the discipline of linguistics, and have been very influential in computer science, and philosophy.

Chomsky is also one of the leading intellectuals in the anarchist socialist movement. His political writings about the media and political injustice have profoundly influenced many. Chomsky is among the most quoted authors in the world (among the top ten and the only living person on the list).

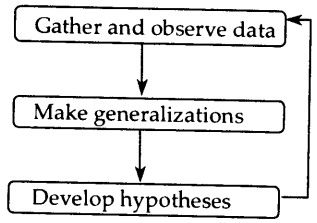
3. Syntax as Science – the Scientific Method

To many people the study of language properly belongs in the domain of the humanities. That is, the study of language is all about the beauty of its usage in fine (and not so fine) literature. However, there is no particular reason, other than our biases, that the study of language should be confined to a humanistic approach. It is also possible to approach the study of language from a scientific perspective; this is the domain of linguistics. People who study literature often accuse linguists of abstracting away from the richness of good prose and obscuring the beauty of language. Nothing could be further from the truth. Most linguists, including the present author,

enjoy nothing more than reading a finely crafted piece of fiction, and many linguists often study, as a sideline, the more humanistic aspects of language. This doesn't mean, however, that one can't appreciate and study the formal properties (or rules) of language and do it from a scientific perspective. The two approaches to language study are both valid; they complement each other; and neither takes away from the other.

Science is perhaps one of the most poorly defined words of the English language. We regularly talk of scientists as people who study bacteria, particle physics, and the formation of chemical compounds, but ask your average Joe or Jill on the street what science means, and you'll be hard pressed to get a decent definition. Science refers to a particular methodology for study: the scientific method. The scientific method dates backs to the ancient Greeks, such as Aristotle, Euclid, and Archimedes. The method involves observing some data, making some generalizations about patterns in the data, developing hypotheses that account for these generalizations, and testing the hypotheses against more data. Finally, the hypotheses are revised to account for any new data and then tested again. A flow chart showing the method is given in (1):

1)



In syntax, we apply this methodology to sentence structure. Syntacticians start[2] by observing data about the language they are studying, then they make generalizations about patterns in the data (e.g., in simple English declarative sentences, the subject precedes the verb). They then generate a hypothesis – preferably one that makes predictions – and test the hypothesis against more syntactic data, and if necessary go back and re-evaluate their hypotheses.

[2] This is a bit of an oversimplification. We really have a "chicken and the egg" problem here. You can't know what data to study unless you have a hypothesis about what is important, and you can't have a hypothesis unless you have some basic understanding of the data. Fortunately, as working syntacticians this philosophical conundrum is often irrelevant, as we can just jump feet-first into both the hypothesis-forming and data-analysis at the same time.

The next step in the scientific method is to test this hypothesis against more data. Consider the additional data in (4):

4) a) The robot kissed itself.
   b) She knocked herself on the head with a zucchini.
   c) *She knocked himself on the head with a zucchini.
   d) The snake flattened itself against the rock.
   e) ?The snake flattened himself/herself against the rock.
   f) The Joneses think themselves the best family on the block.
   g) *The Joneses think himself the most wealthy guy on the block.
   h) Gary and Kevin ran themselves into exhaustion.
   i) *Gary and Kevin ran himself into exhaustion.

Sentences (4a, b, and c) are all consistent with our hypothesis that anaphors must agree in gender with their antecedents, which at least confirms that the hypothesis is on the right track. What about the data in (4d and e)? It appears as if any gender is compatible with the antecedent *the snake*. This appears, on the surface, to be a contradiction to our hypothesis. Think about these examples a little more closely, however. Whether sentence (4e) is well-formed or not depends upon your assumptions about the gender of the snake. If you assume (or know) the snake to be male, then *The snake flattened himself against the rock* is perfectly well-formed. But under the same assumption, the sentence *The snake flattened herself against the rock* seems very odd indeed, although it is fine if you assume the snake is female. So it appears as if this example also meets the generalization in (3); the vagueness about its well-formedness has to do with the fact that we are rarely sure what gender a snake is and not with the actual structure of the sentence.

Now, look at the sentences in (4f–i); note that the ill-formedness of (g) and (i) is not predicted by our generalization. In fact, our generalization predicts that sentence (4i) should be perfectly grammatical, since *himself* agrees in gender (masculine) with its antecedents *Gary* and *Kevin*. Yet there is clearly something wrong with this sentence. The hypothesis needs revision. It appears as if the anaphor must agree in gender and ***number*** with the antecedent. Number refers to the quantity of individuals involved in the sentence; English primarily distinguishes singular number from plural number. (5) reflects our revised hypothesis.

5) An anaphor must agree in gender and number with its antecedent.

If there is more than one person or object mentioned in the antecedent, then the anaphor must be plural (i.e., *themselves*).

Testing this against more data, we can see that this partially makes the right predictions (6a), but it doesn't properly predict the grammaticality of sentences (6b–e):

6) a) People from Tucson think very highly of themselves.
   b) *I gave yourself the bucket of ice cream.
   c) I gave myself the bucket of ice cream.
   d) *She hit myself with a hammer.
   e) She hit herself with a hammer.

Even more revision is in order. The phenomenon seen in (6b–e) revolves around a grammatical distinction called ***person***. Person refers to the perspective of the speaker with respect to the other participants in the speech act. First person refers to the speaker. Second person refers to the listener. Third person refers to people being discussed that aren't participating in the conversation. Here are the English pronouns associated with each person: (***Nominative*** refers to the ***case*** form the pronouns take when in subject position like *I* in "*I* love peanut butter;" ***accusative*** refers to the form they take when in object positions like *me* in "John loves *me*." We will look at case in much more detail in chapter 9, so don't worry if you don't understand it right now.)

7)

	Nominative		Accusative		Anaphoric	
	Singular	Plural	Singular	Plural	Singular	Plural
1	I	we	me	us	myself	ourselves
2	you	you	you	you	yourself	yourselves
3 masc	he	they	him	them	himself	themselves
3 fem	she		her		herself	
3 neut	it		it		itself	

As you can see from this chart, the form of the anaphor seems also to agree in person with its antecedent. So once again we revise our hypothesis (rule):

8) An anaphor must agree in person, gender and number with its antecedent.

With this hypothesis, we have a straightforward statement of the distribution of this noun type, derived using the scientific method. In the problem sets below, and in chapter 5, you'll have an opportunity to revise the rule in (8) with even more data.

You now have enough information to try Challenge Problem Sets 1 & 2

Sentence (11a) sounds bizarre (cf. *the toothbrush is blue*) because we know that toothbrushes (except in the world of fantasy/science fiction or poetry) cannot be pregnant. The meaning of the sentence is strange, but the form is OK. We call this ***semantic ill-formedness*** and mark the sentence with a #. By contrast, we can glean the meaning of sentence (11b); it seems semantically reasonable (toothbrushes can be blue), but it is ill-formed from a structural point of view. That is, the determiner *the* is in the wrong place in the sentence. This is a ***syntactically ill-formed*** sentence. A native speaker of English will judge both these sentences as ill-formed, but for very different reasons. In this text, we will be concerned primarily with syntactic well-formedness.

You now have enough information to answer General Problem Set 2

4. Where do the Rules Come From?

In this chapter we've been talking about our subconscious knowledge of syntactic rules, but we haven't dealt with how we get this knowledge. This is sort of a side issue, but it may affect the shape of our theory. If we know how children acquire their rules, then we are in a better position for a proper formalization of them. The way by which children develop knowledge is an important question in cognitive science. The theory of generative grammar makes some very specific (and very surprising) claims about this.

4.1 Learning vs. Acquisition

One of the most common misconceptions about Language is the idea that children and adults "learn" languages. Recall that the basic kind of knowledge we are talking about here is subconscious knowledge. When producing a sentence you don't consciously think about where to put the subject, where to put the verb, etc. Your subconscious language faculty does that for you. Cognitive scientists make a distinction in how we get conscious and subconscious knowledge. Conscious knowledge (like the rules of algebra, syntactic theory, principles of organic chemistry, or how to take apart a carburetor) is ***learned***. Subconscious knowledge, like how to speak or the ability to visually identify discrete objects, is ***acquired***. In part, this explains why classes in the formal grammar of a foreign language often fail abysmally to train people to speak those languages. By contrast, being immersed in an environment where you can subconsciously acquire a language is much more effective. In this text we'll be primarily interested

in how people acquire the rules of their language. Not all rules of grammar are acquired, however. Some facts about Language seem to be built into our brains, or ***innate***.

You now have enough information to answer General Problem Set 3

4.2 Innateness: Language as an Instinct

If you think about the other types of knowledge that are subconscious, you'll see that many[3] of them (for example, the ability to walk) are built directly into our brains – they are instincts. No one had to teach you to walk (despite what your parents might think!). Kids start walking on their own. Walking is an instinct. Probably the most controversial claim of Noam Chomsky's is that Language is also an instinct. Many parts of Language are built in, or ***innate***. Much of Language is an ability hard-wired into our brains by our genes.

Obviously, particular languages are not innate. It isn't the case that a child of Slovak parents growing up in North America who is never spoken to in Slovak, grows up speaking Slovak. They'll speak English (or whatever other language is spoken around them). So on the surface it seems crazy to claim that Language is an instinct. There are very good reasons to believe, however, that a human facility for Language (perhaps in the form of a "Language organ" in the brain) is innate. We call this facility ***Universal Grammar*** (or ***UG***).

4.3 The Logical Problem of Language Acquisition

What follows is a fairly technical proof of the idea that Language is at least plausibly construed as an innate, in-built system. If you aren't interested in this proof (and the problems with it), then you can reasonably skip ahead to section 4.4.

The argument in this section is that a productive system like the rules of Language probably have not been learned or acquired. Infinite systems are in principle, given certain assumptions, both unlearnable and unacquirable. Since we all have such an infinite system in our heads, and we shouldn't have been able to aquire it. So it follows that it is built in. The argument presented here is based on an unpublished paper by Alec Marantz, but is based on an argument dating back to at least Chomsky (1965).

First here's a sketch of the proof, which takes the classical form of an argument by modus ponens:

[3] but not all!

Her job, then, is to correctly match up the sentences with the situation.[5] More crucially she has to make sure that she does *not* match it up with all the other possible alternatives, such as the things going on around her (like her older brother kicking the furniture, or her mother making her breakfast, etc.). This matching of situations with expressions is a kind of mathematical relation (or function) that *maps* sentences onto a particular situation. Another way of putting it is that she has to figure out the rule(s) that decode(s) the meaning of the sentences. It turns out that this task is, at least very difficult if not impossible.

Let's make this even more abstract to get at the mathematics of the situation. Assign each sentence some number. This number will represent the input to the rule. Similarly we will assign each situation a number. The function (or rule) modeling language acquisition maps from the set of sentence numbers to the set of situation numbers. Now let's assume that the child has the following set of inputs and correctly matched situations (perhaps explicitly pointed out to her by her parents). The x value represents the sentences she hears. The y is the number correctly associated with the situation.

20)

Sentence (input)	*Situation* (output)
x	y
1	1
2	2
3	3
4	4
5	5

Given this input, what do you suppose that the output where $x = 6$ will be?

6	?

Most people will jump to the conclusion that the output will be 6 as well. That is, they assume that the function (the rule) mapping between inputs and outputs is $x = y$. But what if I were to tell you that in the hypothetical situation I envision here, the correct answer is situation number 126. The rule that generated the table in (20) is actually:

21) $[(x-5)*(x-4)*(x-3)*(x-2)*(x-1)] + x = y$

With this rule, all inputs equal to or less than 5 will give an output equal to the input, but for all inputs greater than 5, will give some large number.

[5] Note that this is the job of the child who is using universal grammar, not the job of UG itself.

When you hypothesized the rule was $x = y$, you didn't have all the crucial information; you only had part of the data. This seems to mean that if you hear only the first five pieces of data in our table then you won't get the rule, but if you learn the sixth you will figure it out. Is this necessarily the case? Unfortunately not: Even if you add a sixth line, you have no way of being sure that you have the right function until you have heard *all* the possible inputs. The important information might be in the sixth line, but it might also be in the 7,902,821,123,765th sentence that you hear. You have no way of knowing for sure if you have heard all the relevant data until you have heard them all. In an infinite system you can't hear them all, even if you were to hear 1 sentence every 10 seconds for your entire life. If we assume the average person lives to be about 75 years old, if they heard one new sentence every ten seconds, ignoring leap years and assuming they never sleep, they'd have only heard about 39,420,000 sentences over their lifetime. This is a much smaller number than infinity. Despite this poverty of input, by the age of 5 most children are fairly confident with their use of complicated syntax. Productive systems are (possibly) unlearnable, because you never have enough input to be sure you have all the relevant facts. This is called ***the logical problem of language acquisition***.

Generative grammar gets around this logical puzzle by claiming that the child acquiring English, Irish, or Yoruba has some help: a flexible blueprint to use in constructing her knowledge of language called Universal Grammar. Universal Grammar restricts the number of possible functions that map between situations and utterances, thus making language learnable.

You now have enough information to try Challenge Problem Sets 3 & 4

4.4 Other Arguments for UG

The evidence for UG doesn't rely on the logical problem alone, however. There are many other arguments that support the hypothesis that at least a certain amount of language is built in.

An argument that is directly related to the logical problem of language acquisition discussed above has to do with the fact that we know things about the grammar of our language that we couldn't possibly have learned. Start with the data in (20). A child might plausibly have heard sentences of these types (the underline represents the place where the question word *who* plausibly starts out – that is either as the object or subject of the verb *will question*):

Statistical Probability or UG?

In looking at the logical problem of language acquisition you might be asking yourself "Ok, so maybe kids don't get all the data, but maybe they get enough to draw conclusions about what is the most likely structure of their grammar?" For example, we might conclude that a child learning English would observe the total absence of any sentences that have *that* followed by a trace (e.g., 22d), so after hearing some threshold of sentences they conclude that this sentence type is ungrammatical. This is a common objection to the hypothesis of UG. Unfortunately, this hypothesis can't explain why many sentence types that are extremely rare (to the point that they are probably never heard by children) are still judged as grammatical by the children. For example, English speakers rarely (if ever) produce sentences with seven embeddings (*John said that Mary thinks that Susan believes that Matt exclaimed that Marian claimed that Art said that Andrew wondered if Gwen had lost her pen*); yet speakers of English routinely agree these are acceptable. The actual speech of adult speakers is riddled with errors (due to all sorts of external factors: memory, slips of the tongue, tiredness, distraction, etc.). But children do not seem to assume that any of these errors, which they hear frequently, are part of the data that determines their grammars.

Finally, there are a number of biological arguments in favor of UG. As noted above, Language seems to be both human-specific and pervasive across the species. All humans, unless they have some kind of physical impairment, seem to have Language as we know it. This points towards it being a genetically endowed instinct. Additionally, research from neurolinguistics seems to point towards certain parts of the brain being linked to specific linguistic functions.

With very few exceptions, most linguists believe that some Language is innate. What is of controversy is how much is innate and whether the innateness is specific to Language, or follows from more general innate cognitive functions. We leave these questions unanswered here.

You now have enough information to try General Problem Set 4

4.5 *Explaining Language Variation*

The evidence for UG seems to be very strong. However, we are still left with the annoying problem that languages differ from one another. This problem is what makes the study of syntax so interesting. It is also not an unsolvable

one. One way in which languages differ is in terms of the words used in the language. These clearly have to be learned or memorized. Other differences between languages (such as the fact that basic English word order is subject-verb-object (SVO), but the order of an Irish sentence is verb-subject-object (VSO) and the order of a Turkish sentence is subject-object-verb (SOV)) must also be acquired. The explanation for this kind of fact will be explored in chapter 6. Foreshadowing slightly, we'll claim there that differences in the grammars of languages can be boiled down to the setting of certain innate ***parameters*** (or switches) that select among possible variants. Language variation thus reduces to learning the correct set of words and selecting from a predetermined set of options.

Oversimplifying slightly, most languages put the order of elements in a sentence in one of the following word orders:

25) a) Subject Verb Object (SVO) (e.g., English)
b) Subject Object Verb (SOV) (e.g., Turkish)
c) Verb Subject Object (VSO) (e.g., Irish)

A few languages use:

d) Verb Object Subject (VOS) (e.g., Malagasy)

No (or almost no)[7] languages use

e) Object Subject Verb (OSV)
f) Object Verb Subject (OVS)

Let us imagine that part of UG is a parameter that determines the basic word order. Four of the options (SVO, SOV, VSO, and VOS) are innately available as possible settings. Two of the possible word orders are not part of UG. The child who is acquiring English is innately biased towards one of the common orders, when she hears a sentence like "Mommy loves Kirsten," if the child knows the meaning of each of the words, then she might hypothesize two possible word orders for English: SVO and OVS. None of the others are consistent with the data. The child thus rejects all the other hypotheses. OVS is not allowed, since it isn't one of the innately available forms. This leaves SVO, which is the correct order for English. So children acquiring English will choose to set the word order parameter at the innately available SVO setting.

In his excellent book *The Atoms of Language*, Mark Baker inventories a set of possible parameters of a language variation within the UG hypothesis.

[7] This is a matter of some debate. Derbyshire (1985) has claimed that the language Hixkaryana has object initial order.

7. SUMMARY

In this chapter, we've done very little syntax but talked a lot about the assumptions underlying the approach we're going to take to the study of sentence structure. The basic approach to syntax that we'll be using here is generative grammar; we've seen that this approach is scientific in that it uses the scientific method. It is descriptive and rule based. Further, it assumes that a certain amount of grammar is built in and the rest is acquired.

IDEAS, RULES, AND CONSTRAINTS INTRODUCED IN THIS CHAPTER

i) ***Syntax***: The level of linguistic organization that mediates between sounds and meaning, where words are organized into phrases and sentences.

ii) ***Language*** *(capital L)*: The psychological ability of humans to produce and understand a particular language. Also called the ***Human Language Capacity*** or ***i-Language***. This is the object of study in this book.

iii) ***language*** *(lower-case l)*: A language like English or French. These are the particular instances of the human Language. The data source we use to examine Language is language. Also called ***e-language***.

iv) ***Generative Grammar***: A theory of linguistics in which grammar is viewed as a cognitive faculty. Language is generated by a set of rules or procedures. The version of generative grammar we are looking at here is primarily the ***Principles and Parameters approach*** (P&P) touching occasionally on ***Minimalism.***

v) ***The Scientific Method***: Observe some data, make generalizations about that data, draw a hypothesis, test the hypothesis against more data.

vi) ***Falsifiable Prediction***: To prove that a hypothesis correct you have to look for the data that would prove it *wrong*. The prediction that might prove a hypothesis wrong is said to be falsifiable.

vii) ***Grammar***: Not what you learned in school. This is the set of rules that generate a language.

viii) ***Prescriptive Grammar***: The grammar rules as taught by so called "language experts." These rules, often inaccurate descriptively, prescribe how people should talk/write, rather than describe what they actually do.

ix) ***Descriptive Grammar***: A scientific grammar that describes, rather than prescribes, how people talk/write.

x) ***Anaphor***: A word that ends in *-self* or *-selves* (a better definition will be given in chapter 5).

xi) ***Antecedent***: The noun an anaphor refers to.

xii) ***Asterisk***: * used to mark syntactically ill-formed (unacceptable or ungrammatical) sentences. The hash mark, pound, or number sign (#) is used to mark semantically strange, but syntactically well-formed, sentences.

xiii) ***Gender (Grammatical)***: Masculine vs. Feminine vs. Neuter. Does not have to be identical to the actual sex of the referent. For example, a dog might be female, but we can refer to it with the neuter pronoun *it*. Similarly, boats don't have a sex, but are grammatically feminine.

xiv) ***Number***: The quantity of individuals or things described by a noun. English distinguishes singular (e.g., *a cat*) from plural (e.g., *the cats*). Other languages have more or less complicated number systems.

xv) ***Person***: The perspective of the participants in the conversation. The speaker or speakers (*I, me, we, us*) are called first person. The listener(s) (*you*), are called the second person. Anyone else (those not involved in the conversation) (*he, him, she, her, it, they, them*) is referred to as the third person.

xvi) ***Case***: The form a noun takes depending upon its position in the sentence. We discuss this more in chapter 10.

xvii) ***Nominative***: The form of a noun in subject position (*I, you, he, she, it, we, they*).

xviii) ***Accusative***: The form of a noun in object position (*me, you, him, her, it, us, them*).

xix) ***Corpus (pl. Corpora)***: A collection of real-world language data.

xx) ***Native Speaker Judgments (intuitions)***: Information about the subconscious knowledge of a language. This information is tapped by means of the grammaticality judgment task.

3. LEARNING VS. ACQUISITION
[Creative and Critical Thinking; Basic]
We have distinguished between learning and acquiring knowledge. Learning is conscious, acquisition is automatic and subconscious. (Note that acquired things are *not* necessarily innate. They are just subconsciously obtained.) Other than language are there other things we acquire? What other things do we learn? What about walking? Or reading? Or sexual identity? An important point in answering this question is to talk about what kind of evidence is necessary to distinguish between learning and acquisition.

4. UNIVERSALS
[Creative and Critical Thinking; Intermediate]
Pretend for a moment that you don't believe Chomsky and that you don't believe in the innateness of syntax (but only *pretend*!). How might you account for the existence of universals (see definition above) across languages?

5. INNATENESS
[Creative and Critical Thinking; Intermediate]
We argued that some amount of syntax is innate (inborn). Can you think of an argument that might be raised against innateness? (It doesn't have to be an argument that works, just a plausible one.) Alternately, could you come up with a hypothetical experiment that could *disprove* innateness? What would such an experiment have to show? Remember that cross-linguistic variation (differences between languages) is *not* an argument against innateness or UG, because UG contains parameters that allow minute variations.

6. LEVELS OF ADEQUACY
[Application of Skills; Basic]
Below, you'll find the description of several different linguists' work. Attribute a level of adequacy to them (state whether the grammars they developed are observationally adequate, descriptively adequate, or explanatorily adequate). Explain *why* you assigned the level of adequacy that you did.

a) Juan Martínez has been working with speakers of Chicano English in the barrios of Los Angeles. He has been looking both at corpora (rap music, recorded snatches of speech) and working with adult native speakers.
b) Fredrike Schwarz has been looking at the structure of sentences in eleventh-century Welsh poems. She has been working at the national archives of Wales in Cardiff.
c) Boris Dimitrov has been working with adults and corpora on the formation of questions in Rhodopian Bulgarian. He is also conducting a longitudinal study of some two-year-old children learning the language to test his hypotheses.

CHALLENGE PROBLEM SETS

Challenge Problem Sets are special exercises that either challenge the presentation of the main text or offer significant enrichment. Students are encouraged to complete the other problem sets before trying the Challenge Sets. Challenge Sets can vary in level from interesting puzzles to downright impossible conundrums. Try your best!

CHALLENGE PROBLEM SET 1: ANAPHORA
[Creative and Critical Thinking and Data Analysis; Challenge]
In this chapter, as an example of the scientific method, we looked at the distribution of anaphora (nouns like *himself*, *herself*, etc.). We came to the following conclusion about their distribution:

> An anaphor must agree in person, gender, and number with its antecedent.

However, there is much more to say about the distribution of these nouns (in fact, chapter 5 of this book is entirely devoted to the question).

Part 1: Consider the data below. Can you make an addition to the above statement that explains the distribution of anaphors and antecedents in the very limited data below?

a) Geordi sang to himself.
b) *Himself sang to Geordi.
c) Betsy loves herself in blue leather.
d) *Blue leather shows herself that Betsy is pretty.

Part 2: Now consider the following sentences:[8]

e) Everyone should be able to defend himself/herself/themselves.
f) I hope nobody will hurt themselves/himself/?herself.

Do these sentences obey your revised generalization? Why or why not? Is there something special about the antecedents that forces an exception here, or can you modify your generalization to fit these cases?

CHALLENGE PROBLEM SET 2: YOURSELF
[Creative and Critical Thinking; Challenge]
In the main body of the text we claimed that all anaphors need an antecedent. Consider the following acceptable sentence. This kind of sentence is called an "imperative" and is used to give orders.

a) Don't hit yourself!

[8] Thanks to Ahmad Lotfi for suggesting this part of the question.

While given the extreme view in section 4.3 is logically true, consider the following alternative possibilities:

a) We as humans have some kind of "cut off mechanism" that stops considering new data after we've heard some threshold number of examples. If we don't hear the crucial example after some period of time we simply assume it doesn't exist. Rules simply can't exist that require access to sentence types so rare that you don't hear them before the cut off point.

b) We are purely statistical engines. Rare sentences types are simply ignored as "statistical noise." We consider only those sentences that are frequent in the input when constructing our rules.

c) Child-directed speech (motherese) is specially designed to give you precisely the kinds of data you need to construct your rule system. The child listens for very specific "triggers" or "cues" in the parental input in order to determine the rules.

Question 1: To what extent are (a), (b) or (c) compatible with the hypothesis of Universal Grammar. If (a), (b) or (c) turned out to be true, would this mean that there was no innate grammar? Explain your answer.

Question 2: How might you experimentally or observationally distinguish between (a), (b), (c) and the infinite input hypothesis of 4.3? What kinds of evidence would you need to tell them apart?

Question 3: When people speak, they make errors. (They switch words around, they mispronounce things, they use the wrong word, they stop mid-sentence without completing what they are saying etc.) Nevertheless children seem to be able to ignore these errors and still come up with the right set of rules. Is this fact compatible with any of the infinite hypothesis, (a), (b), or (c)?

CHALLENGE PROBLEM SET 5: LEARNING PARAMETERS: PRO DROP
[Critical Thinking, Data Analysis; Challenge]
Background: Among the Indo-European languages there are two large groups of languages that pattern differently with respect to whether they require a pronoun (like he, she, it) in the subject position, or whether such pronouns can be "dropped". For example, in both English and French, pronouns are required. Sentences without them are usually ungrammatical:

a) He left
b) *Left
c) Il est parti (French)
   he is gone
   "he left"
d) *est parti (French)

In languages, such as Spanish and Italian, however, such pronouns are routinely omitted (1s = first person, singular):

e) Io telephono (Italian)
   I called.1s
   "I called (phoned)"
f) telephono
   called.1s
   "Called"

Question 1: Now imagine that you are a small child learning a language. What kind of data would you need to know in order to tell if your language was "pro drop" or not? (Hint. Does the English child hear sentences both with and without subjects? Does the Italian child? Are they listening for sentences with subjects or without them?

Question 2: Assume that one of the two possible settings for this parameter (either your language is pro-drop or it is not) is the "default" setting. This default setting is the version of the parameter one gets if one doesn't hear the right kind of input. Which of the two possibilities is the default?

Question 3: English has imperative constructions such as:

g) Leave now!

Why doesn't the English child assume on the basis of such sentences that English is pro-drop?