chapter 1

Generative Grammar

Learning Objectives

After reading chapter 1, you should walk away having mastered the following ideas and skills:

1. Explain why language is a psychological property of humans.
2. Distinguish between prescriptive and descriptive rules.
3. Explain the scientific method as it applies to syntax.
4. Explain the differences between the kinds of data gathering, including corpora and linguistic judgments.
5. Explain the difference between competence and performance.
6. Explain the difference between i-language and e-language
7. Provide at least three arguments for Universal Grammar.
8. Explain the logical problem of language acquisition.
9. Distinguish between learning and acquisition.
10. Distinguish among observational, descriptive, and explanatory adequacy.

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" 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 as a complex machine. Think about it for a minute. You are reading this and understanding it, but you have no conscious knowledge of how you

Syntax: A Generative Introduction, Fourth Edition. Andrew Carnie.

Companion website: http://www.wiley.com/go/carnie

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, or the study of ***syntax*** and the people who study syntax are called ***syntacticians***.

There are many perspectives on studying linguistics. One could study language looking at languages across time, or one could study how language is used as a social too. But syntacticians typically take a different view. They look at language as a psychological or cognitive property of humans. That is, my mind contains certain principles that allow me to sit here and produce this set of letters, words and sentences, and you use similar principles that allow you to translate these squiggles back into coherent ideas and thoughts. At least I hope you can translate them back into coherent ideas!

There are several subsystems at work in when we use language. 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 pattern them appropriately. For example, speakers of English know that the made-up word *bluve* is a possible word of English, but the word *bnuck* is not. This is part of the science called ***phonology***. Then you take these groups of sounds and organize them into meaningful units (called morphemes) and words. For example, the word *dancer* is made up of two meaningful bits: *dance* and the suffix *-er*. The study of this level of language is called ***morphology***. Next you organize the words into phrases and sentences. One usage of the term ***syntax*** is the cover term for studies at this level of language. Finally, you take the sentences and phrases you hear and translate them into thoughts and ideas. This last step is what we refer to as the ***semantic*** level of language.

Syntax as a discipline studies the part of language knowledge that lies between words and the meaning of utterances: sentences. It is the level that mediates between sounds that someone produces (organized into words) and what they intend to say.

Perhaps one of the truly amazing aspects of the study of language is not the origins of the word *demerit,* or how to properly punctuate a quote inside parentheses, or how kids have, like, destroyed the English language, eh? Instead it's the question of how we subconsciously get from sounds and words to the meaning of sentences. This is the study of syntax.

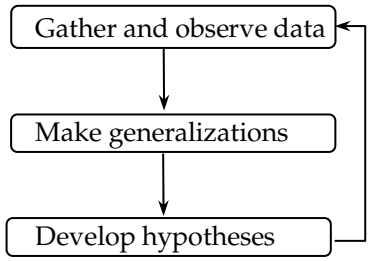
1. SYNTAX AS SCIENCE – THE SCIENTIFIC METHOD

For many people, the study of language properly belongs in the humanities. That is, the study of language is all about the beauty of its usage in fine (and not so fine) literature and its impact on human culture. However, there is no particular reason, other than tradition, 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[1].

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. But among scientists themselves, *science* typically refers to a particular methodology for study: the deductive scientific method. The scientific method dates back 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

[1] The notion that science should be the primary means of investigating linguistics has come under significant criticism by some members of indigenous communities (see for example, the opinions expressed by community members as reported in Czaykowska-Higgins 2009, the articles in Bischoff and Jany et al. 2019 and Rosborough, et al. 2017). The central idea is that by exclusively using a deductive scientific methodology in investigating their languages we are prioritizing a western European set of values on their traditions and cultures. When we as linguists work *with* native communities on their languages it is important that we acknowledge this perspective and think carefully about ways that our research can complement, connect to and support indigenous ways of knowing. We must also find ways in which we can be partners with these communities to help further their local agendas rather just observing their communities as objects of study and imposing our ideas upon them.

[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

about patterns in the data (e.g., in simple English declarative sentences, the subject precedes the verb). They then generate a hypothesis about these patterns and test the hypothesis against more syntactic data, and if necessary, go back and re-evaluate their hypotheses.

Hypotheses are only useful to the extent that they make ***predictions***. A hypothesis that makes no predictions (or worse yet, predicts everything) is useless from a scientific perspective. In particular, the hypothesis must be ***falsifiable***. That is, we must in principle be able to look for some data, which, if true, show that the hypothesis is wrong. This means that we are often looking for the cases where our hypotheses predict that a sentence will be grammatical (and it is not), or the cases where they predict that the sentence will be ungrammatical (contra to fact).

In syntax, hypotheses are called ***rules***, and the group of hypotheses that describe a language's syntax is called a ***grammar***. The term *grammar* can strike terror into the hearts of people. But you should note that there are two ways to go about writing grammatical rules. One is to tell people how they *should* speak (this is of course the domain of English teachers and copy-editors); we call these kinds of rules ***prescriptive rules*** (as they prescribe how people should speak according to some standard). Some examples of prescriptive rules include "never end a sentence with a preposition", "use *whom* not *who*" and "don't split infinitives". These rules tell us how we are supposed to use our language. The other approach is to write rules that describe how people *actually* speak, whether or not they are speaking "correctly". These are called ***descriptive rules***. Consider for a moment the approach we're taking in this book. Which of the two types (descriptive or prescriptive) is more scientific? Which kind of rule is more likely to give us insight into how the mind uses language? We are going to focus on descriptive rules. This doesn't mean that prescriptive rules aren't important (in fact, in the problem sets section of this chapter you are asked to critically examine the question of descriptive vs. prescriptive rules), but for our purposes descriptive rules are more important.

You now have enough information to answer General Problem Sets GPS1 & 2, as well as Challenge Problem Set CPS1 at the end of this chapter. For practice try Workbook Exercise WBE1 in chapter 1 of The Syntax Workbook, 2nd Edition*, an optional companion book to this text.*

Do Abstract Rules Really Exist?

As discussed in detail later in this chapter, the approach to grammar we are using here is supposed to be part cognitive psychology, so it's reasonable to ask whether formal rules *really* exist in the brain/minds of speakers. After all, a brain is a mass of neurons firing away, so how can abstract rules exist up there? Remember, however, that we are attempting to *model* language; we aren't trying to describe language exactly. This question confuses two disciplines: psychology and neurology. Psychology is concerned with the mind, which represents the output and the abstract organization of the brain.

---

working syntacticians this philosophical conundrum is often irrelevant, as we can just jump feet-first into both the hypothesis-forming and the data-analysis at the same time.

Neurology is concerned with the actual firing of the neurons and the physiology of the brain. Our approach doesn't try to be a theory of neurology. Instead it is a model of the psychology of language. Obviously, the rules per se don't exist in our brains, but they do model the external behavior of the mind. For more discussion of this issue, look at the readings in the further reading section of this chapter.

3.1 An Example of the Scientific Method as Applied to Syntax

Let's turn now to a real-world application of the scientific method to some language data. The following data concern the form of a specific kind of noun, called an ***anaphor*** (plural: ***anaphors***; the phenomenon is called ***anaphora***). These include the nouns that end with *-self* (e.g., *himself, herself, itself*). In chapter 5, we look at the distribution of anaphors in detail; here we'll only consider one superficial aspect of them. In the following sentences, as is standard in the syntactic literature, a sentence that isn't well-formed is marked with an ***asterisk*** (*) before it. For these sentences assume that *Bill* is male and *Sally* is female.

2) a) Bill kissed himself.
   b) *Bill kissed herself.
   c) Sally kissed herself.
   d) *Sally kissed himself.
   e) *Kiss himself.

Under the assumption that Bill is a cisgender male and Sally is a cisgender female, the ill-formed sentences in (2b and d) just look silly. It is obvious that Bill can't kiss herself, because Bill is male. There is a clear generalization about the distribution of anaphors here. In particular, the generalization we can draw about the sentences in (2) is that an anaphor must agree in ***grammatical gender*** with the noun it refers to (its ***antecedent***). So. in (2a & b) we see that the anaphor must agree in gender with *Bill*, its antecedent. The anaphor must take the masculine form *himself*. The situation in (2c & d) is the same; the anaphor must take the form *herself* so that it agrees in gender with the feminine *Sally*. Note further that a sentence like (2e) shows us that anaphors must have an antecedent. An anaphor without an antecedent is unacceptable. A plausible hypothesis (or rule) given the data in (2), then, is stated in (3):

3) An anaphor must (i) have an antecedent and (ii) agree in grammatical gender (masculine, feminine, or neuter) with that antecedent.

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 wealthiest guy on the block.

h) Gary and Kevin ran themselves into exhaustion.
i) *Gary and Kevin ran himself into exhaustion.

Grammatical Gender vs. Sex vs. Personal Gender

Gender can be a politically charged and deeply personal issue for many people. In this chapter, I am talking about primarily about ***grammatical gender***. Grammatical gender is often confused with sex assigned at birth and with the gender identity/expression of the individual. This is because people often use grammatical gender to signal their sex or gender identity to others. But in the context that I'm using it here, it's a purely formal feature of words. In many languages grammatical gender, also called ***noun class***, has nothing to with actual sex or gender identity. For example, in Navajo grammatical gender is determined by shape, consistency and animacy and is quite distinct from their cultural understanding of gender identity. In other languages, grammatical gender does not need to correspond to gender expression – it can even be the opposite. In Modern Irish, for example, the word *cailín* 'girl' is masculine and the word *stail* 'stallion' is feminine.

Despite the objections of prescriptive language gurus, English has long used the pronoun *they* to refer to humans in a gender-neutral way. Recently this usage has been extended more regularly to people whose gender identity is non-binary. This new usage has some really interesting effects on the phenomenon of anaphora – in particular a new anaphor, *themself*, has been added to the grammatical system of many people, particularly younger speakers. General Problem Set GPS3 gives you a chance to explore the interplay of grammatical gender and personal gender with English anaphora and verb agreement.

Sentences (4a, b, & 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 & 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 about the actual structure of the sentence.

Now, look at the sentences in (4f–i) above; 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 correct predictions (6a), but it doesn't properly predict the acceptability 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 kissed myself.
   e) She kissed herself.

Even more revision to our hypothesis 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 addressee. ***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 11, 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 GPS3, WBE2, and CPS2 & CPS3

3.2 Sources of Data

If we are going to apply the scientific method to syntax, it is important to consider the sources of our data. One obvious source is in collections of either spoken or written texts. Such data are called ***corpora*** (singular: ***corpus***). There are many corpora available,

including some searchable through the internet. For languages without a literary tradition or languages spoken by a small group of people, it is often necessary for the linguist to go and gather data and compile a corpus in the field. In the early part of the last century, this was the primary occupation of linguists, and it is proudly carried on today by many researchers.

The linguist Heidi Harley reports in her blog[3] on an example of using search engines to do linguistic analysis on the huge corpus known as the web. Harley notes that to her ear, the expression *half full of something* sounds natural, but *half empty of something* does not. She does a comparison of *half empty* vs. *half full* and of *half empty of* vs. *half full of*. She finds that the ratio of *half full* to *half empty* without the *of* is roughly 1:1. The ratio of *half full of* to *half empty of* is approximately 149:1. This is a surprising difference. Harley was able to use the web to show that a fairly subtle difference in acceptability is reflected in the frequency with which the expressions are used.

But corpus searches aren't always adequate for finding out the information syntacticians need. For the most part corpora only contain grammatical sentences. Sometimes the most illuminating information is our knowledge that a certain sentence is ungrammatical (i.e., not a sentence of normal English), or that two similar sentences have very different meanings. Consider the pair of sentences in (9) as a starting point.

9) a) Marian blew the building up.
   b) Marian blew up the building.

Most native speakers of English will accept both of these sentences as acceptable sentences, with a preference for (9b). They also know that while the first sentence (9a) is unambiguous, the second one has two meanings (He destroyed the building using explosives vs. he blew really hard with his lungs up the stairwell). The second of these meanings is a bit silly, but it's a legitimate interpretation of the sentence.

Now contrast the sentences in (9) with the similar pair in (10). In these forms I've replaced "the building" with the pronoun "it":

10) a) Marian blew it up.
   b) Marian blew up it.

Here we find a different pattern of interpretation. (10a) is unambiguous just the way (9a) is, it refers to an act of explosion and cannot have an interpretation where Marian was blowing hard with her lungs up something. Sentence (10b), however, is a surprise. Unlike (9b), (10b) cannot have anything to do with explosives. It can only have the interpretation where Marian is blowing air up whatever "it" is. Recall that with (9) this "puff of air reading" was the silly or strange one. With a pronoun, however, it's the only available interpretation. This difference in interpretation would never be captured in a corpus, because the specific meanings of expressions and ambiguities are not indicated anywhere in the data source.

While corpora are unquestionably invaluable sources of data, they are only a partial representation of what goes on in the mind. More particularly, corpora often contain

[3] http://heideas.blogspot.com/2005/10/scalar-adjectives-with-arguments.html.

instances of only acceptable (or, more precisely, well-formed) sentences (sentences that sound "OK" to a native speaker). For example, the online *New York Times* contains very few ungrammatical sentences. Even corpora of naturalistic speech complete with the errors every speaker makes don't necessarily contain the data we need to test the falsifiable predictions of our hypotheses. So, corpora are just not enough: there is no way of knowing whether a corpus has *all* possible forms of grammatical sentences. In fact, as we will see in the next few chapters, due to the productive nature of language, a corpus could *never* contain all the grammatical forms of a language, nor could it even contain a representative sample. It also doesn't tell us about what sentences are ambiguous or what sentences are ungrammatical or strange. Those are really important sources of evidence for doing syntax. To really get at what we know about our languages we have to know what sentences are *not* well-formed. That is, in order to know the range of acceptable sentences of English, Italian or Igbo, we *first* have to know what are *not* acceptable sentences in English, Italian or Igbo. This kind of negative information is very rarely available in corpora, which mostly provide grammatical, or well-formed, sentences.

Consider the following sentence:

11) *Who do you wonder what bought?

For most speakers of English, this sentence borders on word salad – it is not a good sentence of English. How do you know that? Were you ever taught in school that you can't say sentences like (11)? Has anyone ever uttered this sentence in your presence before? I seriously doubt it. The fact that a sentence like (11) sounds strange, but similar sentences like (12a and b) *do* sound OK is not reflected anywhere in a corpus:

12) a) Who do you think bought the bread machine?
b) I wonder what Fiona bought.

Instead we have to rely on our knowledge of our native language (or on the knowledge of a native speaker consultant for languages that we don't speak natively). Notice that this is *not* conscious knowledge. I doubt there are many native speakers of English that could tell you why sentence (11) is terrible, but most can tell you that it is. This is subconscious knowledge. The trick is to get at and describe this subconscious knowledge.

The psychological experiment used to get this subconscious kind of knowledge is called the ***acceptability judgment task***. The judgment task involves asking a native speaker to read a sentence, and judge whether it is well-formed (i.e., grammatical), marginally well-formed, or ill-formed (ungrammatical).

There are actually several different kinds of acceptability judgments. Both of the following sentences are ill-formed, but for different reasons:

13) a) #The toothbrush is pregnant.
b) *Toothbrush the is blue.

Sentence (13a) sounds bizarre (cf. *the toothbrush is blue*) because we know that toothbrushes (except in the world of fantasy/science fiction or poetry or a dream) cannot be pregnant. The meaning of the sentence is strange, but the form of the sentence is okay. We call this ***semantic ill-formedness*** and mark the sentence with a #. By contrast, we can

glean the meaning of sentence (13b); 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, which is marked with an *. 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, but both kinds of judgment can help guide our analyses.

You now have enough information to do WBE3 & 4, GPS3 & 4, and CPS4–6.

Judgments as Science?

Many linguists refer to the acceptability judgment task as "drawing upon our native speaker intuitions". The word "intuition" here is slightly misleading. The last thing that pops into our heads when we hear the term "intuition" is science. Generative grammar has been severely criticized by many for relying on "unscientific" intuitions. But this is based primarily on a misunderstanding of the term. To the layperson, the term "intuition" brings to mind guesses and luck. This usage of the term is certainly standard. When a generative grammarian refers to "intuition", however, she is using the term to mean "tapping into our subconscious knowledge". The term "intuition" may have been badly chosen, but in this circumstance, it refers to a real psychological effect. Intuition (as an acceptability judgment) has an entirely scientific basis. It is replicable under strictly controlled experimental conditions (these conditions are rarely applied, but the validity of the task is well established). Other disciplines also use intuitions or judgment tasks. For example, within the study of vision, it has been determined that people can accurately judge differences in light intensity, drawing upon their subconscious knowledge (Bard et al. 1996). To avoid the negative associations with the term intuition, we will use the less loaded term ***judgment*** instead.

2. Syntax as a Cognitive Science

Cognitive science is a cover term for a group of disciplines that all have the same goal: describing and explaining human beings' ability to think (or more particularly, to think about abstract notions like subatomic particles, the possibility of life on other planets or even how many angels can fit on the head of a pin, etc.). One thing that distinguishes us from other animals, even relatively smart ones like chimps and elephants, is our ability to use productive, combinatory syntax. Language plays an important role in how we think about abstract notions, or, at the very least, it appears to be structured in such a way that it allows us to express abstract notions.[4] The discipline of linguistics is thus one of the important subdisciplines of cognitive science.[5] Sentences are how we get at expressing abstract thought processes, so the study of syntax is an important foundation stone for understanding how we communicate and interact with each other as humans.

[4] Whether language constrains what abstract things we can think about (this idea is called the Sapir–Whorf hypothesis) is a matter of great debate and one that lies outside the domain of syntax per se.
[5] Along with psychology, neuroscience, communication, philosophy, and computer science.

3. MODELS OF SYNTAX

One dominant theory of syntax that fits into the cognitive science mold is due to Noam Chomsky and his colleagues, starting in the mid 1950s and continuing to this day. This theory, which has had many different names through its development (Transformational Grammar (TG), Transformational Generative Grammar, Standard Theory, Extended Standard Theory, Government and Binding Theory (GB), Principles and Parameters approach (P&P) and Minimalism (MP)), is often given the blanket name ***generative grammar***. A number of alternate approaches to syntax have also branched off of this research program. These include Lexical-Functional Grammar (LFG) and Head-Driven Phrase Structure Grammar (HPSG). These approaches are also considered part of generative grammar; but we won't cover them extensively in this book. But I have included two additional chapters on these theories in the web resources for this book[6]. The particular version of generative grammar that we will mostly look at here is roughly the ***Principles and Parameters*** approach, although we will occasionally stray from this into the more recent version called ***Minimalism***.

The underlying thesis of generative grammar is that sentences are generated by a subconscious set of procedures (like computer programs). These procedures are part of our minds (or of our cognitive abilities if you prefer). The goal of syntactic theory is to model these procedures. In other words, we are trying to figure out what we subconsciously know about the syntax of our language.

4. COMPETENCE VS. PERFORMANCE

Consider sentences such as (14). Native speakers will have to read this sentence a couple of times to figure out what it means.

14) # Cotton shirts are made from comes from India.

This kind of sentence (called a ***garden path sentence***) is very hard to understand and process. In this example, the problem is that the intended reading has a noun, *cotton*, that is modified by a reduced relative clause: *(that) shirts are made from.* The linear sequence of *cotton* followed by *shirt* is ambiguous with the noun phrase *cotton shirts*. Note that this kind of relative structure is okay in other contexts; compare: *That material is the cotton shirts are made from.* Sentences like (14) get much easier to understand with really clear pauses (where … is meant to indicate a pause): *Cotton … shirts are made from … comes from India*. Or by insertion of a *that* which breaks up the potentially ambiguous *cotton shirts* sequence: *The cotton that shirts are made from comes from India*. What is critical about these garden path sentences is that, once one figures out what the intended meaning is, native speakers can identify them as acceptable sentences or at the very least as sentences that have structures that would otherwise be acceptable in them. The problem for us as

[6] http://www.wiley.com/go/carnie

linguists is that native speakers have a really hard time figuring out what the intended meaning for these sentences is on those first few passes!

A similar situation arises when we have really long sentences with complex syntactic relations. Look at (15). A first reading of this sentence will boggle your average speaker of English. But if you read it a couple of times, it becomes obvious what is intended. In fact, the sentence seems to be structured grammatically.

15) Who did Keisha say Monique claimed that Suzanne seems to have been likely to have kissed?

The reason this sentence is hard to understand is that the question word *who* is very far away from where it gets interpreted (as the object of *kiss*), and what lies in between those two points is quite a lot of sophisticated embeddings and structure. But once you get a chance to think about it, it gets better and better as a sentence. The most famous example of this kind of effect is called ***center embedding***. English speakers tolerate a small amount of stacking of relative clauses between subjects and verbs, so (16) – while a little clumsy – is still a good sentence for most speakers of English. We have some cheese, the kind that mice love, and it stinks. If you have trouble with this sentence put a big pause after *cheese* and before *stinks.*

16) Cheese mice love stinks.

But no pauses will fix a sentence in which we put another reduced relative right after *mice,* with the intended meaning that cheese which is loved by mice who are caught by cats is stinky:

17) #Cheese mice cats catch love stinks

This sentence is essentially uninterpretable for English speakers. Chomsky (1965) argued that the problem here is not one of the grammar (as English grammar allows reduced relative clauses after subjects and before verbs), but instead either a constraint on short-term memory[7] or a constraint on our mental ability to break apart sentences as we hear them. The English ***parsing*** system – that is the system that breaks down sentences into their bits for comprehension – has certain limits, and these limits are distinct from the limits on what it means to be "grammatical". Sentences (14), (15), and (16) are unacceptable to native speakers in a qualitatively different way than the ones in (13).

The distinction we've been looking at here is often known as the ***competence/performance*** distinction. When we speak or listen, we are performing the act of creating a piece of language output. This performance can be interrupted by all sorts of extraneous factors: we can be distracted or bored; we can cough or mumble our words; we can forget what we had previously heard; the noise of the bus driving past can blot out a crucial word. ***Performance*** refers to the kinds of language that are actually produced and heard. ***Competence***, by contrast, refers to what we *know* about our language; it is unimpeded by factors that might muddy the waters of performance. So, think about the

[7] The working memory hypothesis is suspicious because speakers of languages like Japanese and German can understand the similar sentences in their languages without problem.

really long complicated sentence in (15). The first time you read it, things like your memory and how complicated it was interfered with your ability to understand it. So the initial unacceptability of the sentence was due to a performance problem. But once you thought about it and stared at it a bit, you saw that it was actually a fairly standard grammatical sentence of English – just a really complicated one. When you did this, you were accessing your competence in (or knowledge of) English grammar. If syntax is part of cognitive science which is about what we know, then we should probably be most interested in *competence*.

This takes us to a new point. Listen carefully to someone speak (not lecture or read aloud, but someone really speaking in a conversation). You'll notice that they don't speak in grammatical sentences. They leave stuff off and they speak in fragments. They start and they stop the same sentence a couple of times. Everyone does this, even the most eloquent among us. So much of what you hear (or see in spoken language corpora) consists of actually "ungrammatical" forms. Nevertheless, if you're a native English speaker, you have the ability to judge if a sentence is acceptable or not. These two tasks, understanding spoken conversational language and being able to judge the well-formedness of a sentence, seem to actually be different skills corresponding roughly to performance and competence.

An analogy that might clarify these distinctions: imagine that you're a software engineer and you're writing a piece of computer code. First you run it on your own beautiful up-to-date computer and it behaves beautifully. The output of the computer code is one kind of performance of the underlying competence. Then you run it on your little sister's ancient PC. The program doesn't perform as you expect. It's really slow. It crashes. It causes the fan to run continuously and the processor to overheat. Now you go back and look at the code. There are no errors in the code. It meets all the requirements of the computer language. So, from the perspective of competence, your program is okay. The real problem here is not with your code, but with the machine you're running it on. The processor is too old, there isn't enough memory and you have a computer that tends to overheat. These are all performance problems.

What does this mean for the linguist using acceptability judgments as a tool for investigating syntax? It means that when using a judgment, you have to be really clear about what is causing the acceptability or unacceptability of the sentence. Is the sentence acceptable just because you have gleaned enough information from the conversational context (in which case we might consider it a performance effect)? If you hear a sentence that you judge as unacceptable, is it because someone was speaking too quickly and left out a word, or is it because the sentence really doesn't work as an English sentence at all? This distinction is very subtle, but it is one that syntacticians have to pay careful attention to as they do their work.

5. A CLARIFICATION ON THE WORD "LANGUAGE"

When I use the term *language*, most people immediately think of some particular language such as English, French, or KiSwahili. But this is not the way most syntacticians

use the term; when we talk about *language*, we are often really talking more about ***i-language*** (where the i- stands for "internal"), i.e., the internal psychological/cognitive tools i.e. our competence. Confusingly, we are often sloppy and also use the term l*anguage* to refer to ***e-languages*** (where the e stands for "external"). E-languages are the instantiations of the output or performance of an i-language. E-languages are what we commonly conceive of particular languages (such as French or English). In this book, we'll be using e-language as our primary data, but we'll be trying to come up with a model of i-language. The difference between e-languages and i-languages is very similar to the old distinction Ferdinand de Saussure made between langue (i-language) and parole (e-language)[8]. The linguist Seth Cable gave me a great analogy to help understand these notions. Imagine language is football. Competence is how deeply you understand the rules of the game. I-language is the rules of the game themselves. E-language is an actual game or match played and performance is how well you play in a that day.

To make things even more confusing linguistics sometimes use the term *language* to refer to the general *ability* of humans to acquire an i-language and to use that i-language to produce any (particular) e-language. Noam Chomsky calls this ability the ***Human Language Capacity (HLC)***. In our football analogy this would be how well you are equipped to play.

You now have enough information to answer WBE5, GPS5, and CPS7.

---

6. 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 yet 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 to develop a proper formalization of them. The way in 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.

6.1 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***. A lot of subconscious knowledge, like how to speak or the ability to visually

[8] For an accessible discussion of the notions of i-language/e-language, competence/performance and langue/parole, see Duffield (2018).

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 i-language seem to be built into our brains, or ***innate***.

You now have enough information to answer GPS6.

6.2 Innateness: Parts of i-Language as Instincts

If you think about the other types of knowledge that are subconscious, you'll see that many 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 that Noam Chomsky has made is that parts of i-language are also an instinct. That is many parts of i-language are built in, or ***innate***. Chomsky claims that much of i-language is an ability hard-wired into our brains.

Obviously, particular languages (e-languages) are not innate. It is never the case that a child of Slovak parents growing up in North America who has never been 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. But when we are talking about i-languages, there are very good reasons to believe, however, that a human facility (the Human Language Capacity) for language is innate. We call the innate parts of the HLC, ***Universal Grammar*** (or ***UG***).

6.3 The Logical Problem of Language Acquisition

What follows is a fairly technical proof of the idea that parts of our linguistic system are 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 6.4.

The argument in this section is that a productive system like the rules of Language probably could not be learned or acquired. Infinite systems are in principle, given certain assumptions, both unlearnable and unacquirable. Since we'll show that syntax is an infinite system, we shouldn't have been able to acquire 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:

Premise (i): Syntax is a productive, recursive and infinite system.
Premise (ii): Rule-governed infinite systems are unacquirable.
Conclusion: Therefore syntax is an unacquirable system. Since we have such a system, it follows that at least parts of syntax are innate.

There are parts of this argument that are very controversial. In the challenge problem sets at the end of this chapter you are invited to think very critically about the form of this proof. Challenge Problem Set CPS8 considers the possibility that premise (i) is false (but hopefully you will conclude that, despite the argument given in the problem set, the idea that Language is productive and infinite is correct). Premise (ii) is more dubious, and is the topic of Challenge Problem Set CPS9. Here, in the main body of the text, I will give you the classic versions of the support for these premises, without criticizing them. You are invited to be skeptical and critical of them when you do the Challenge Problem sets.

Let's start with premise (i): i-language is a productive system. That is, you can produce and understand sentences you have never heard before. For example, I can practically guarantee that you have never heard or read the following sentence:

18) The dancing chorus-line of elephants broke my television set.

The magic of syntax is that it can generate forms that have never been produced before. Another example of this productive quality lies in what is called ***recursion***. It is possible to utter a sentence like (19):

19) Rosie reads magazine articles.

It is also possible to put this sentence inside another sentence, like (20):

20) I think [Rosie reads magazine articles].

Similarly, you can put this larger sentence inside of another one:

21) Drew believes [I think [Rosie reads magazine articles]].

And, of course, you can put this bigger sentence inside of another one:

22) Dana doubts that [Drew believes [I think [Rosie reads magazine articles]]].

and so on, and so on ad infinitum. It is always possible to embed a sentence inside of a larger one. This means that i-language is a productive (probably infinite) system. There are no limits on what we can talk about. Other examples of the productivity of syntax can be seen in the fact that you can infinitely repeat adverbs (23) and you can infinitely add coordinated nouns to a noun phrase (24):

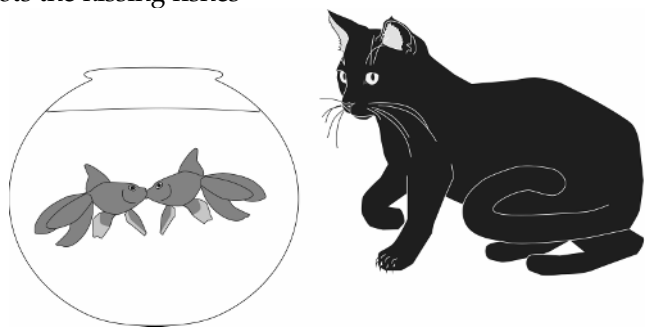
23) a) a very big peanut
    b) a very very big peanut
    c) a very very very big peanut
    d) a very very very very big peanut
    etc.

24) a) Dave left
    b) Dave and Alina left
    c) Dave, Dan, and Alina left
    d) Dave, Dan, Erin, and Alina left
    e) Dave, Dan, Erin, Jaime, and Alina left
    etc.

It follows that for every grammatical sentence of English, you can find a longer one (based on one of the rules of recursion, adverb repetition, or coordination). This means that the production of syntax is at least countably infinite. This premise is relatively uncontroversial (however, see the discussion in Challenge Problem Set CPS8).

Let's now turn to premise (ii), the idea that infinite systems are unlearnable. In order to make this more concrete, let us consider an algebraic treatment of a linguistic example. Imagine that the task of a child is to determine the rules by which her language is constructed. Further, let's simplify the task, and say a child simply has to match up situations in the real world with utterances she hears.[9] So upon hearing the utterance *the cat spots the kissing fishe*s, she identifies it with an appropriate situation in the context around her (as represented by the picture).

25) "the cat spots the kissing fishes" =



Her job, then, is to correctly match up the sentence with the situation.[10] 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 father making breakfast for her, etc.). This matching of situations with expressions is a kind of mathematical relation (or function) that *maps* sentences onto particular situations. 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 sentence she hears. The y is the number correctly associated with the situation.

[9] The task is actually several magnitudes more difficult than this, as the child has to work out the phonology, etc., too, but for argument's sake, let's stick with this simplified example.

[10] Note that this is the job of the child who is using Universal Grammar, not the job of UG itself.

26)

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 (26) is actually:

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

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, they will give some large number.

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 all the data, even if you were to hear 1 new 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 10 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 CPS8 & 9.

6.4 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 (28). A child might plausibly have heard sentences of these types (the underline represents the place where the question word *who* might start out – that is, as either the object or the subject of the verb *will question*):

28) a) Who do you think that Siobhan will question _____ first?
b) Who do you think Siobhan will question _____ first?
c) Who do you think _____ will question Seamus first?

The child has to draw a hypothesis about the distribution of the word *that* in English sentences. One conclusion consistent with these observed data is that the word *that* in English is optional. You can either have it or not. Unfortunately this conclusion is not accurate. Consider the fourth sentence in the paradigm in (28). This sentence is the same as (28c) but with a *that*:

d) *Who do you think that _____ will question Seamus first?

It appears as if *that* is only optional when the question word (*who* in this case) starts in object position (as in 28a and b). It is obligatorily absent when the question word starts in subject position (as in 28c and d) (don't worry about the details of this generalization). What is important to note is that *no one* has ever taught you that (28d) is ungrammatical. Nor could you have come to that conclusion on the basis of the data you've heard. The logical hypothesis on the basis of the data in (28a–c) predicts sentence (28d) to be grammatical. There is nothing in the input a child hears that would lead them to the conclusion that (28d) is ungrammatical, yet every English-speaking child knows it is. One solution to this conundrum is that we are born with the knowledge that sentences like (28d) are ungrammatical.[11] This kind of argument is often called the ***poverty of the stimulus*** argument for UG.

Most parents raising a toddler will swear up and down that they are teaching their child to speak and that they actively engage in instructing their child in the proper form of the language. The claim that overt instruction by parents plays any role in language

[11] The phenomenon in (28) is sometimes called the **that-*trace effect***. There is no disputing the fact that this phenomenon is not learnable. However, it is also a fact that it is not a universal property of all languages. For example, French and Irish don't seem to have the *that*-trace effect. Here is a challenge for those of you who like to do logic puzzles: If the *that*-trace effect is not learnable and thus must be biologically built in, how is it possible for a speaker of French or Irish to violate it? Think carefully about what kind of input a child might have to have in order to learn an "exception" to a built-in principle. This is a hard problem, but there is a solution. It may become clearer below when we discuss parameters.

development is easily falsified. The evidence from the experimental language acquisition literature is very clear: parents, despite their best intentions, do not, for the most part, correct ungrammatical utterances by their children. More generally, they correct the content rather than the form of their child's utterances (see for example the extensive discussion in Holzman 1997).

29) (from Marcus et al. 1992)
Adult: Where is that big piece of paper I gave you yesterday?
Child: Remember? I writed on it.
Adult: Oh that's right, don't you have any paper down here, buddy?

When a parent does try to correct a child's sentence structure, it is more often than not ignored by the child:

30) (from Pinker 1995: 281 – attributed to Martin Braine)
Child: Want other one spoon, Daddy.
Adult: You mean, you want the other spoon.
Child: Yes, I want other one spoon, please, Daddy.
Adult: Can you say "the other spoon"?
Child: Other … one … spoon.
Adult: Say "other".
Child: Other.
Adult: "Spoon".
Child: Spoon.
Adult: "Other … spoon".
Child: Other … spoon. Now give me other one spoon?

This humorous example is typical of parental attempts to "instruct" their children in language. When these attempts do occur, they fail. However, children still acquire language in the face of a complete lack of instruction. Perhaps one of the most convincing explanations for this is UG. In the problem set part of this chapter, you are asked to consider other possible explanations and evaluate which are the most convincing.

There are also typological arguments for the existence of an innate language faculty. All the languages of the world share certain properties (for example they *all* have subjects and predicates – other examples will be seen throughout the rest of this book). These properties are called ***universals*** of Language. If we assume UG, then the explanation for these language universals is straightforward – they exist because all speakers of human languages share the same basic innate materials for building their language's grammar. In addition to sharing many similar characteristics, recent research into Language acquisition has begun to show that there is a certain amount of consistency cross-linguistically in the way children acquire Language. For example, children seem to go through the same stages and make the same kinds of mistakes when acquiring their language, no matter what their cultural or linguistics background.

Derek Bickerton (1984) has noted the fact that creole languages[12] have a lot of features in common with one another, even when they come from very diverse places in the world and spring forth from unrelated languages. For example, they all have SVO order; they all lack non-specific indefinite articles; they all use modals or particles to indicate tense, mood, and aspect, and they have limited verbal inflection, and many other such similarities. Furthermore these properties are ones that are found in the speech of children of non-creole languages. Bickerton hypothesizes that these properties are a function of an innate ***language bioprogram***, an idea similar to Chomsky's Universal Grammar.

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 generative linguists believe that some i-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 GPS7 & 8 and CPS10.

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 perhaps 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., 28d), 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.). However, children do not seem to assume that any of these errors, which they hear frequently, are part of the data that determine their grammars.

---

[12] Creole languages are new languages that are formed when a generation of speakers starts using a trade language or pidgin as their first language and speak it natively in the home.

6.5 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.

The fact that an inborn system should allow variation won't be a surprise to any biologist. Think about the color of your eyes. Every sighted person has eyes. Having eyes is clearly an inborn property of being a human (or being a mammal). I doubt that anyone would object to that characterization. Nevertheless we see both widespread variation in eye color, size, and shape among humans, and widespread variation in form and position among various mammalian species. A closer analog to language might be bird song. In 1962, Marler and Tamura observed dialect variation among the songs of white-crowned sparrows. The ability and motivation for these birds to vocalize is widely assumed to be innate, but the particular song they sing is dependent upon the input they hear.

One way in which languages differ is in terms of the words used in the language. The different words of different languages clearly have to be learned or memorized and are not innate. Other differences between languages must also be acquired. For example the child learning English must determine that its word order is subject-verb-object (SVO), but the child learning Irish determines the order is verb-subject-object (VSO) and the Turkish child figures out subject-object-verb (SOV) order. The explanation for this kind of fact will be explored in more detail 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 elements in a sentence in one of the following word orders:

31) 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)[13] 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

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

"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 language variation within the UG hypothesis. This is an excellent and highly accessible treatment of parameters. I strongly recommend this book for further reading on how language variation is consistent with Universal Grammar.

You now have enough information to try GPS7, GPS8, and CPS 11.

---

7. Choosing among Theories about Syntax

There is one last preliminary we have to touch on before actually doing some real syntax. In this book we are going to posit many hypotheses. Some of these we'll keep, others we'll revise, and still others we'll reject. How do we know what is a good hypothesis and what is a bad one? Chomsky (1965) proposed that we can evaluate how good theories of syntax are using what are called the ***levels of adequacy***. Chomsky claimed that there are three stages that a grammar (the collection of descriptive rules that constitute your theory) can attain in terms of scientific adequacy.

If your theory only accounts for the data in a corpus (say a series of printed texts) and nothing more, it is said to be an ***observationally adequate grammar***. Needless to say, this isn't much use if we are trying to account for the cognition of an i-language. As we discussed above, it doesn't tell us the whole picture. We also need to know what kinds of sentences are unacceptable, or ill-formed. A theory that accounts for both corpora and native speaker judgments about well-formedness is called a ***descriptively adequate grammar***. On the surface this may seem to be all we need. Chomsky, however, has claimed that we can go one step better. He points out that a theory that also accounts for how children acquire their language is the best. He calls this an ***explanatorily adequate grammar***. The simple theory of parameters might get this label. Generative grammar strives towards explanatorily adequate grammars.

You now have enough information to try GPS9 and CPS12.

---

8. The Scientific Method and the Structure of this Textbook

Throughout this chapter I've emphasized the importance of the scientific method to the study of syntax. It's worth noting that we're not only going to apply this principle to small problems or specific rules, but we'll also apply it in a more global way. This principle is in part a guide to the way in which the rest of this book is structured.

In chapters 2–5 (the remainder of Part 1 of the book) we're going to develop an initial hypothesis about the way in which syntactic rules are formed. These are the phrase structure rules (PSRs). Chapters 2 and 3 examine the words these rules use, the form of the rules, and the structures they generate. Chapters 4 and 5 look at ways we can detail the structure of the trees formed by the PSRs.

In chapters 6–9 (Part 2 of the book), we examine some data that present problems for the simple grammar presented in Part 1. When faced withmore complicated data, *we revise our hypotheses*, and this is precisely what we do. We develop a special refined kind of PSR known as an X-bar rule. X-bar rules are still phrase structure rules, but they offer a more sophisticated way of looking at trees. This more sophisticated version also needs an additional constraint known as the "theta criterion", which is the focus of chapter 8.

In chapters 10–13 (Part 3) we consider even more data, and refine our hypothesis once again, this time adding a new rule type: the transformation (we retain X-bar, but enrich it with transformations). Part 4 of the book (chapters 14–18) refines these proposals even further.

With each step we build upon our initial hypothesis, just as the scientific method tells us to. I've been teaching with this proposal–then-revision method of theory construction for a couple of years now, and every now and then I hear the complaint from a student that we should just start with the final answer (i.e. the revised hypotheses found in the later chapters in the book). Why bother learning all this "other" "wrong" stuff? Why should we bother learning phrase structure rules? Why don't we just jump straight into X-bar theory? Well, in principle, I could have constructed a book like that, but then you, the student, wouldn't understand *why* things are the way they are in the latter chapters. The theory would appear to be unmotivated, and you wouldn't understand what the technology actually does. By proposing a simple hypothesis early on in the initial chapters, and then refining and revising it, building new ideas onto old ones, you not only get an understanding of the motivations for and inner workings of our theoretical premises, but you get practice in working like a real linguist. Professional linguists, like all scientists, work from a set of simple hypotheses and revise them in light of predictions made by the hypotheses. The earlier versions of the theory aren't "wrong" so much as they need refinement and revision. These early versions represent the foundations out of which the rest of the theory has been built. This is simply how science works.

9. Conclusion

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) ***Linguistics***: The scientific study of language.

ii) ***Syntax*** (the field): The scientific study of sentence structure

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

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

v) ***Falsifiable Prediction***: To prove that a hypothesis is 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.

vi) ***Grammar***: Not what you learned in school. This is the set of mental rules or procedures that generate a sentence.

vii) ***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.

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

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

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

xi) ***Asterisk (*)***: The mark 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.

xii) ***Grammatical Gender***: Masculine vs. Feminine vs. Neuter. Does not have to be identical to the actual sex or gender identity 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.

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

xiv) ***Person***: The perspective of the participants in the conversation. The speaker or speakers (*I, me, we, us*) are called the ***first person***. The addressee(s) (*you*) is 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***.

xv) ***Case***: The form a noun takes depending upon its position in the sentence. We discuss this more in chapter 11.

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

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

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

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

xx) ***Semantic Judgment***: A judgment about the meaning of a sentence, often relying on our knowledge of the context in which the sentence was uttered.

xxi) ***Syntactic Judgment***: A judgment about the form or structure of a sentence.

xxii) ***Garden Path Sentence:*** A sentence with a strong ambiguity in structure that makes it hard to understand.

xxiii) ***Center Embedding:*** A sentence in which a relative clause consisting of a subject and a verb is placed between the main clause subject and verb. E.g., *The house [Bill built] leans to the left.*

xxiv) ***Parsing:*** The mental tools a listener uses to process and understand a sentence.

xxv) ***Competence***: What you know about your language.

xxvi) ***Performance***: The real-world behaviors that are a consequence of what you know about your language.

xxvii) ***i-language***: This is the cognitive structure underlying your ability to speak a language. The *i-* stands for "internal". This is the primary object of study in this book.

xxviii) ***e-language***: The outward expression of a particular language like English, French or Mandarin. The *e-* stands for "external". These are the particular instances of the human ability to speak an i-language. The data sources we use to examine i-language are e-languages.

xxix) ***Human Language Capacity (HLC)***. The general ability to have an i-language and to express an e-language.

xxx) ***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), and we will be touching occasionally on ***Minimalism,*** a more recent approach**.**

xxxi) ***Learning***: The gathering of conscious knowledge (like linguistics or chemistry).

xxxii) ***Acquisition***: The gathering of subconscious information (like language).

xxxiii) ***Innate***: Hard-wired or built-in, an instinct.

xxxiv) ***Recursion***: The ability to embed structures iteratively inside one another. Allows us to produce sentences we've never heard before.

xxxv) ***Universal Grammar (UG)***: The innate (or instinctual) part of each language's grammar.

xxxvi) ***The Logical Problem of Language Acquisition***: The proof that an infinite system like human language cannot be learned on the basis of observed data – an argument for UG.

xxxvii) ***Poverty of the stimulus***: The idea that we know things about our language that we could not have possibly learned – an argument for UG.

xxxviii) ***Universal***: A property found in all the languages of the world.

xxxix) ***Bioprogram Hypothesis:*** The idea that creole languages share similar features because of an innate basic setting for language.

xl) ***Observationally Adequate Grammar***: A grammar that accounts for observed real-world data (such as corpora).

xli) ***Descriptively Adequate Grammar***: A grammar that accounts for observed real-world data and native speaker judgments.

xlii) ***Explanatorily Adequate Grammar***: A grammar that accounts for observed real-world data and native speaker judgments and offers an explanation for the facts of language acquisition.

FURTHER READING: Baker (2001b), Barsky (1997), Bickerton (1984), Chomsky (1965), Duffield (2018), Ghomeshi (2010), Jackendoff (1993), Sampson (1997), Uriagereka (1998)

GENERAL PROBLEM SETS

GPS1. PRESCRIPTIVE RULES

[Creative and Critical Thinking; Basic]

In the text above, we claimed that descriptive rules are the primary focus of syntactic theory. This doesn't mean that prescriptive rules don't have their uses. What are these uses? Why do societies have prescriptive rules?

GPS2. OBLIGATORY SPLIT INFINITIVES

[Creative and Critical Thinking, Analysis; Intermediate]

The linguist Arnold Zwicky has observed[14] that the prescription not to split infinitives can result in utterly ungrammatical sentences. The adverb *soon* can be reasonably placed before the infinitive (a) or after it (b) and, for most native speakers of English, also in the split infinitive (c):

a) I expect soon to see the results.
b) I expect to see the results soon.
c) I expect to soon see the results.

Zwicky notes that certain modifiers like *more than* or *already* when used with a verb like *to double,* obligatorily appear in a split infinitive construction (g). Putting them anywhere else results in the ungrammatical[15] sentences (d–f):

d) *I expect *more than* [to double] my profits.
e) *I expect [to double] *more than* my profits.
f) *I expect [to double] my profits *more than*.
g) I expect [to *more than* double] my profits.

[14] http://itre.cis.upenn.edu/~myl/languagelog/archives/000901.html.

[15] To be entirely accurate, (d) and (e) aren't wholly ill-formed; they just can't mean what (g) does. (d) can mean "I expect something else too, not just to double my profits" and (e) can mean "I expect to double something else too, not just my profits." The * marks of ungrammaticality are for the intended reading identical to that of (g).

Explain in your own words what this tells us about the validity of prescriptive rules such as "Don't split infinitives". Given these facts, how much stock should linguists put in prescriptive rules if they are following the scientific method?

GPS3: NON-BINARY PRONOUNS AND ANAPHORS[16]
[Data Analysis, Critical Thinking; Advanced]

BACKGROUND: In the chapter above, we discussed how anaphors must agree in person, number and gender with the noun they refer to. But we didn't do a very deep investigation of what we mean by "gender". Let's consider the following commonly accepted distinction that social scientists use: *Sex* refers to the biological characteristics of an individual[17] and *gender* refers to a social construct that can correlate with sex, but it doesn't have to. In many cultures, gender is typically defined by how individuals identify themselves. In other countries, often those who have more socially conservative perspectives, the society classes people into genders based on their outward appearance. Either way, gender can be distinct from the sex assigned at birth. People whose gender does not align with their biological sex assigned at birth are often known as *transgender* and those people whose gender corresponds to their biological sex are called *cisgender*. Somewhat confusingly, in English, we use the same terms to describe sex assigned at birth and gender: "male" and "female". Needless to say, conflation of these terms has led to a lot of conflict and misunderstanding.[18] Of particular interest for our question about pronouns, there are also people for whom the traditional two-way male/female gender distinction is not appropriate. They identify as having gender characteristics outside the traditional male and female distinction. These individuals are often called *non-binary*[19]. Below we'll be talking about the use of pronouns and anaphors for non-binary people. Let us refer to this notion of gender, which is tightly tied to identity as ***personal gender.***

To complicate matters further let's add in another definition: ***grammatical gender***. Many languages have two or more "genders" or "noun classes" that they use to classify all the nouns in their language. We use different terms to describe grammatical gender from personal gender. Rather than using "male" and "female", we use "masculine",

---

[16] Many thanks to my Facebook posse for very helpful feedback on this problem set, including Emily Bender, Claire Bowern, Elizabeth Cowper, Joe Dupris, Megan Figueroa, Andrew Garrett, Jason Merchant, Tel Monks, and Dana Sussman.

[17] There are of course people who have the physical characteristics of both sexes. This can be either by choice or due to their genetics. There are also people who were assigned an inaccurate or arbitrary sex at birth.

[18] The Wikipedia page on gender (https://en.wikipedia.org/wiki/Gender; accessed Jan 12, 2018) has a fairly nuanced description of these distinctions as well as links to major academic sources on the distinction.

[19] There are other related identities – including gender non-conforming, gender-queer, and gender-fluid. This isn't a book about gender, so I'm going to pretend that these all fit neatly under the term "non-binary", with the recognition that this is a gross oversimplification that doesn't recognize the complexity and nuance that many people have about their gender identity. Please forgive the simplification.

"feminine" and "neuter". Often these genders have no obvious correlation to sexual characteristics. For example, in French, *maison* 'house' is feminine and is used with the feminine article *la*, by contrast *camion* 'truck' is masculine and used with the masculine article *le*. In Modern Irish the word *cailín* 'girl' is masculine and the word for 'stallion', *stail*, is feminine! Clearly there is no actual sex or personal gender underlying this categorization. English doesn't express gender on most nouns, but the pronoun system shows what we might think of as grammatical gender in the distinction between *he, she* and *it* pronouns.

Number (singular vs. plural) also seems to have a grammatical usage distinct from the reference of the word. There are a set of nouns that are grammatically plural but clearly refer to singular entities. These words are known as *pluralia tantum*. For example, *pants* and *scissors* have no singular form and they show up with the plural form of verbs (*the pants <u>are</u> in the drawer* vs. **The pant(s) <u>is</u> in the drawer*.) Despite taking plural agreement, they can refer to singular elements. So, number, like gender, can both be personal and grammatical.

PART 1*: Non-binary they and subject number agreement.*

Many non-binary people (as well as some other transgender people and some cisgender people) are now choosing to refer to themselves with pronouns other than the binary male/female contrasts. It is common for many non-binary people to identify the pronouns *they*, *them*, and *theirs* as the correct ones to use in reference to them.

Consider a person whose name is Chris. They are non-binary and they find that the pronouns that correctly identify their gender are *they*, *them* and *theirs*. Consider the following sentences, which all refer to singular Chris.

a) Chris is leaving.
b) *Chris are leaving.
c) *They is leaving.
d) They are leaving.

Are is the form of the verb *to be* used with 3rd person plurals and *is* is the form used with 3rd person singulars. Based on the sentences in (a-d), figure out which of the following determines the form of the verb: biological sex and number, personal gender and number, or grammatical gender and number. How can you tell?

PART 2*: Gender in anaphors*

Now let's consider the form of anaphors. Again, assume in this section that Chris is non-binary (i.e. identifies as neither male nor female in personal gender) and they use the *they/them/their* pronouns. Consider the following sentences[20]:

e) Chris is introducing *himself.
f) Chris is introducing *herself.
g) Chris is introducing themselves/themself.
h) They are introducing *himself.

[20] For more on the phenomena discussed here, see Bjorkman (2017) and Konnelly and Cowper (2016).

i) They are introducing *herself.
j) They are introducing themselves/themself.

Ignore for the moment the issue of whether it should be *themselves* or *themself*, for the moment assume either is ok. What determines the form of the first part of the anaphor (the *them/him/her* part)? Is it grammatical gender (cf. part 1 above) or is it personal gender? How can you tell? Is this the same as what determines subject agreement?

***PART 3**: Number in anaphors*
There are at least two speech varieties concerning the *self/selves* part of the anaphor when the singular *they/them/their* set is used. Some speakers will tend to use *themselves*, others will tend to use *themself*. In the following sentences, *they*, *themselves* and *themself* are both meant to have a singular referent: non-binary Chris.

k) *Dialect 1:* Chris is introducing themselves to the president.
l) *Dialect 2:* Chris is introducing themself to the president.
m) *Dialect 1:* They are introducing themselves to the president.
n) *Dialect 2:* They are introducing themself to the president.

What determines the number expressed on the anaphor (*themselves* vs. *themself*) in each of the dialects? Is it personal number (where non-binary *they* is singular) or is it grammatical number (where non-binary *they* is plural)? How can you tell?

***PART 4**: Revising our hypothesis.*
In the text above we proposed the following hypothesis about the form of anaphors:

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

Propose a revision to our hypothesis about what determines the form of anaphors that takes into account your observations in parts 2 and 3.

GPS4. JUDGMENTS
[Application of Skills; Intermediate]

All of the following sentences have been claimed to be ungrammatical or unacceptable by someone at some time. For each sentence,

i) indicate whether this unacceptability is due to a prescriptive or a descriptive judgment, and
ii) for all descriptive judgments indicate whether the unacceptability has to do with syntax or semantics (or both).

One- or two-word answers are appropriate. If you are not a native speaker of English, enlist the help of someone who is. If you are not familiar with the *prescriptive* rules of English grammar, you may want to consult a writing guide or English.

a) Who did you see in Las Vegas?
b) You are taller than me.
c) My red is refrigerator.

d) Who do you think that saw Bill?
e) Hopefully, we'll make it through the winter without snow.
f) My friends wanted to quickly leave the party.
g) Bunnies carrots eat.
h) John's sister is not his sibling.

GPS5: PERFORMANCE VS. COMPETENCE
[Application of Skills (Basic)]

For each of the scenarios described below indicate whether the phenomena being described are best thought of as exhibiting the traits of performance or competence or a mix of the two.

a) Joe-Ellen is talking to her father about lending him some money, she starts to speak and says "Dad you need to get a new job, one where your boss …" and then she gets a text from her best friend telling her about the latest chapter in the on-going drama between two of their mutual friends. She reads the text which ends with "those guys", and she then continues in her conversation with her dad saying "respect you", using the agreement inflection that would be triggered by "those guys" rather than "your boss".

b) Josh has been reading a novel from the regency period and comes across a sentence that sounds very convoluted and odd to him. So he thinks about the sentence to determine if this sentence is acceptable to him or not.

c) I'm lecturing to my introductory syntax class about phrase structure and distracted by cat sitting on the windowsill of the classroom. So, I stop mid-sentence and I go off on a tangent about my cats and how they're fantastic. I never really come back to what I was saying before and my students are confused and uncertain about the topic I was lecturing about.

d) It's been a terrible flu season, and Raini has lost her voice to laryngitis. Mike is over bringing her some soup. Raini is trying to tell him that she wants him to go to the pharmacy to get some medicine. But he can't really hear her, and all he hears is "Go". He's terribly insulted because he was just trying to help, so he leaves and refuses to talk to her for a month.

e) Anita is doing field work on Basque word order. She's collected a bunch of stories and recorded some Basque language TV, but now she's working with a local native speaker. She's written down a bunch of sentences that she thinks might be okay and asks the speaker "Do these sentences sound ok to you?"

GPS6. 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.

GPS7. 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?

GPS8. 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 variation within the set of possibilities allowed for in UG.

GPS9. 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 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!

CPS1. PRESCRIPTIVISM

[Creative and Critical Thinking; Challenge]

The linguist Geoff Pullum reports[21] that he heard Alex Chadwick say the sentence below on the National Public Radio Show "Day to Day". This sentence has an interesting example of a split infinitive in it:

> But still, the policy of the Army at that time was not to send – was specifically to **not** send – women into combat roles.

Here, Mr. Chadwick corrects himself from not splitting an infinitive (*was not to send*) to a form where the word *not* appears between *to* and *send*, thus creating a classic violation of this prescriptive rule. One might wonder why he would correct the sentence in the wrong direction. Pullum observes that the two versions mean quite different things. *The policy was not to send women into combat* means that it was not the policy to send women into combat (i.e. negating the existence of such a policy). The sentence with the split infinitive by contrast, means that there was a policy and it was that they didn't send women into combat. It's a subtle but important distinction in the discussion. Note that putting the *not* after *send* would have rendered the sentence utterly unintelligible. With this background in mind, provide an argument that linguists should probably ignore prescriptive rules if they're trying to model real human language.

CPS2. ANAPHORA

[Creative and Critical Thinking, 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.

[21] http://itre.cis.upenn.edu/~myl/languagelog/archives/002180.html.

c) Betsy loves herself in blue leather.
d) *Blue leather shows herself that Betsy is pretty.

Part 2: Now consider the following sentences:[22]

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?

CPS3. 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!

Part 1: Are all anaphors allowed in sentences like (a)? Which ones are allowed there, and which ones aren't?

Part 2: Where is the antecedent for *yourself*? Is this a counterexample to our rule? Why is this rule an exception? It is easy to add a stipulation to our rule; but we'd rather have an explanatory rule. What is special about the sentence in (a)?

CPS4. CONSTRUCT AN EXPERIMENT
[Creative and Critical Thinking; Challenge]

Linguists have observed that when the subject of a sentence is close to the verb, the verb will invariably agree with that subject.

a) She is dancing.
b) They are dancing.
c) The man is dancing.
d) The men are dancing.

But under certain circumstances this tight verb–subject agreement relation is weakened (sentence taken from Bock and Miller 1991).

e) The readiness of our conventional forces are at an all-time low.

The subject of the sentence *readiness* is singular but the verb seems to agree with the plural *forces*. The predicted form is:

f) The readiness of our conventional forces is at an all-time low.

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

One hypothesis about this is that the intervening noun (*forces)* blocks the agreement with the actual subject noun *readiness*.

Construct an experiment that would test this hypothesis. What kind of data would you need to confirm or deny this hypothesis? How would you gather these data?

CPS5. OFF WE GO[23]

[Critical thinking; application of skills; Challenge]

Consider the expressions *off we go* and *in you go.* There seems to be some limits on which prepositions and verbs can be used in this construction.

Considering only two classes of verbs: Stative verbs like *sit, sleep, and live* and motion verbs like *go, dance, run*. Use the scientific method construct a hypothesis about kinds of verbs can appear in this construction and which cannot. Test your hypothesis with other verbs.

CPS6. JUDGMENTS[24]

[Data Analysis and Application of Skills; Challenge]

Consider the following sentences:

a) i. The students met to discuss the project.
   ii. The student met to discuss the project.
   iii. The class met to discuss the project.

b) i. Zeke cooked and ate the chili.
   ii. Zeke ate and cooked the chili.

c) i. He put the clothes.
   ii. He put in the washing machine.
   iii. He put the clothes in the washing machine.
   iv. He put in the washing machine the clothes.

d) i. I gave my sister a birthday present.
   ii. I gave a birthday present to my sister.
   iii. That horror movie almost gave my sister a heart attack.
   iv. That horror movie almost gave a heart attack to my sister.

e) Where do you guys live at?

f) i. It is obvious to everybody that Tasha likes Misha.
   ii. The fact that Tasha likes Misha is obvious to everybody.
   iii. Who is it obvious that Tasha likes?[25]
   iv. Who is the fact that Tasha likes obvious?

---

[23] This problem set was inspired by a discussion on a Facebook post by Gary Thoms.

[24] This problem set is thanks to Matt Pearson.

[25] The intended meaning for (iii) and (iv) is "Who is the person such that it is obvious that Tasha likes that person?" or "It's obvious that Tasha likes somebody. Who is that somebody?"

Some of these sentences would be judged acceptable by all (or nearly all) speakers of English, while other sentences would be judged unacceptable by at least some speakers. Find at least five native English speakers and elicit an acceptability judgment for each of these sentences (present the sentences to your speakers orally, rather than having them read them off the page). Give the results of your elicitation in the form of a table. Discuss how your consultants' reactions compare with your own native speaker judgments. If a sentence is judged unacceptable by most or all speakers, what do you think is the source of the unacceptability? Choose from the options listed below, and briefly explain and justify each choice. Are there any sentences for which it is difficult to determine the reason for the unacceptability, and if so, why?

1) The sentence is **unacceptable** in the linguistic sense: It would not be produced by a fully competent native speaker of English under any context, and is unlikely to be uttered except as a performance error. It should be marked with a *****.

2) The sentence is **marginally acceptable**. One could imagine a native speaker saying this sentence, but it seems less than perfect syntactically, and should probably be marked with a **?** or **??**.

3) The sentence is fully grammatical in the linguistic sense, but only in *some* varieties of English. It is likely to be treated as 'incorrect' or 'poor style' by some speakers because it belongs to a **stigmatized variety** (an informal or colloquial register, or a non-standard dialect), and is not part of formal written English. We might choose to indicate this with a **%**.

4) The sentence is syntactically well-formed, but **semantically anomalous**: It cannot be assigned a coherent interpretation based on the (normal) meanings of its component words, and should be marked with a **#**.

CPS7. COMPETENCE VS. PERFORMANCE
[Creative and Critical Thinking; Extra Challenge]

Performance refers to a set of behaviors; competence refers to the knowledge that underlies that behavior. We've talked about it for language, but can you think about other cognitive systems or behaviors where we might see examples of this distinction? What are they? Acceptability judgments work for determining the competence underlying language; how might a cognitive scientist explore competence in other domains?

CPS8. IS LANGUAGE REALLY INFINITE?
[Creative and Critical Thinking; Extra Challenge]

> *[**Note to instructors**: this question requires some background in either formal logic or mathematical proofs.]*

In the text, it was claimed that because language is recursive, it follows that it is infinite. (This was premise (i) of the discussion in section 4.3.) The idea is straightforward and at least intuitively correct: if you have some well-formed sentence, and you have a rule that

can embed it inside another structure, then you can also take this new structure and embed it inside another and so on and so on. Intuitively this leads to an infinitely large number of possible sentences. Pullum and Scholz (2005) have claimed that one formal version of this intuitive idea is either circular or a contradiction.

Here is the structure of the traditional argument (paraphrased and simplified from the version in Pullum and Scholz). This proof is cast in such a way that the way we count the number of sentences is by comparing the number of words in the sentence. If for *any* (extremely high) number of words, we can find a longer sentence, then we know the set is infinite. First some terminology:

Terminology: call the set of well-formed sentences *S*. If a sentence *x* is an element of this set we write *S(x)*.

Terminology: let us refer to the length of a sentence by counting the number of words in it. The number of words in a sentence is expressed by the variable *n.* There is a special measurement operation (function) which counts the number of words. This is called μ. If the sentence called *x* has 4 words in it then we say μ(x) = 4.

Next the formal argument:

Premise 1: There is at least one well-formed sentence that has more than zero words in it.

$$\exists x[S(x) \,\&\, \mu(x) > 0]$$

Premise 2: There is an operation in the PSRs such that any sentence may be embedded in another with more words in it. That means for any sentence in the language, there is another longer sentence. (If some expression has the length *n*, then some other well-formed sentence has a size greater than *n*).

$$\forall n\,[\exists x[S(x) \,\&\, \mu(x) = n]] \rightarrow [\exists y[S(y) \,\&\, \mu(y) > n]]$$

Conclusion: Therefore for every positive integer *n*, there are well-formed sentences with a length longer than *n* (i.e., the set of well-formed English expressions is at least countably infinite):

$$\therefore \forall n\,[\exists y[S(y) \,\&\, \mu(y) > n]]$$

Pullum and Scholz claim that the problem with this argument lies with the nature of the set S. Sets come of two kinds: there are finite sets which have a fixed number of elements (e.g. the set {a, b, c, d} has 4 and exactly 4 members). There are also infinite sets, which have an endless possible number of members (e.g., the set {a, b, c, … } has an infinite number of elements).

Question 1: Assume that S, the set of well-formed sentences, is finite. This is a contradiction of one of the two premises given above. Which one? Why is it a contradiction?

Question 2: Assume that S, the set of well-formed sentences, is infinite. This leads to a circularity in the argument. What is the circularity (i.e., why is the proof circular)?

Question 3: If the logical argument is either contradictory or circular what does that make of our claim that the number of sentences possible in a language is infinite? Is it totally wrong? What does the proof given immediately above really prove?

Question 4: Given that S can be neither a finite nor an infinite set, is there any way we might recast the premises, terminology, or conclusion in order not to have a circular argument and at the same time capture the intuitive insight of the claim? Explain how we might do this or why it's impossible. Try to be creative. There is no "right" answer to this question. Hint: one might try a proof that proves that a subset of the sentences of English is infinite (and by definition the entire set of sentences in English is infinite) or one might try a proof by contradiction.

Important notes:

1) Your answers can be given in English prose; you do not need to give a formal mathematical answer.
2) Do not try to look up the answer in the papers cited above. That's just cheating! Try to work out the answers for yourself.

CPS9. ARE INFINITE SYSTEMS REALLY UNLEARNABLE?

[Creative and Critical Thinking; Challenge]

In section 4.3, you saw the claim that if language is an infinite system then it must be unlearnable. In this problem set, you should aim a critical eye at the premise that infinite systems can't be learned on the basis of the data you hear.

While 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 sentence 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 alternative hypotheses: (a), (b), or (c)?

CPS10. INNATENESS AND PRESCRIPTIVISM?
[Creative and Critical Thinking; Challenge]

Start with the assumption that i-language is an instinct. How is this an argument against using prescriptive rules?

CPS11. 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 telefono.
I call.1S
"I call (phone)."
f) Telefono. (Italian)
call.1S
"I call."

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?

CPS12. Explanatory Adequacy
[Creative and Critical Thinking; Challenge]

In the text above, we said that in order for a hypothesis to be explanatory it had to account for how the system came to be, e.g. it had to account for how children acquire the system. Discuss the following question: Does a syntactician actually have to conduct experiments with children for the hypothesis to be explanatory? Or does it suffice that their hypothesis simply make predictions about how children learn a language, for that hypothesis to be explanatory?