Chapter 1: Generative Grammar

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[1] 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[2] 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.

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?

---

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

[2] 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).

GPS3. 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, indicate whether this unacceptability is

i) is due to a prescriptive or a descriptive judgment, and
ii) for all descriptive judgments indicate whether the ungrammaticality 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 grammar or look at Pinker's *The Language Instinct*.

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.

GPS4. 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.

GPS5. 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?

GPS6. 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.

GPS7. LEVELS OF ADEQUACY
[Application of Skills; Basic]
Below, you'll find the description of several different linguists' work. Attribute a level of adequacy to them (state whether the grammars they developed are observationally adequate, descriptively adequate, or explanatorily adequate). Explain *why* you assigned the level of adequacy that you did.

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

CHALLENGE PROBLEM SETS

CHALLENGE PROBLEM SET 1: PRESCRIPTIVISM
[Creative and Critical Thinking; Challenge]
The linguist Geoff Pullum reports[3] 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

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

linguists should probably ignore prescriptive rules if they're trying to model real human language.

CHALLENGE PROBLEM SET 2: 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.
c) Betsy loves herself in blue leather.
d) *Blue leather shows herself that Betsy is pretty.

Part 2: Now consider the following sentences:[4]

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

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

CHALLENGE PROBLEM SET 3: 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)?

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

CHALLENGE PROBLEM SET 4: 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.

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?

CHALLENGE PROBLEM SET 5: JUDGMENTS[5]
[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 brother a birthday present.
   ii. I gave a birthday present to my brother.
   iii. That horror movie almost gave my brother a heart attack.
   iv. That horror movie almost gave a heart attack to my brother.
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.

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

iii. Who is it obvious that Tasha likes?[6]
iv. Who is the fact that Tasha likes obvious?

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?

- The sentence is **ungrammatical** 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 *****.
- The sentence is **marginally grammatical**. One could imagine a native speaker saying this sentence, but it seems less than perfect syntactically, and should probably be marked with a **?** or **??**.
- 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 **%**.
- 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 **#**.

CHALLENGE PROBLEM SET 6: 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? Grammaticality judgments work for determining the competence underlying language; how might a cognitive scientist explore competence in other domains?

[6] The intended meaning for (iii) and (iv) is "Who is the person such that it is obvious that Tasha likes that person?" Or paraphrased another way: "It's obvious that Tasha likes somebody. Who is that somebody?"

CHALLENGE PROBLEM SET 7: 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 *E*. If a sentence *x* is an element of this set we write *E(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.

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

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*).

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

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):

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

Pullum and Scholz claim that the problem with this argument lies with the nature of the set E. 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 E, 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 E, 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 E 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.

CHALLENGE PROBLEM SET 8: 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.