

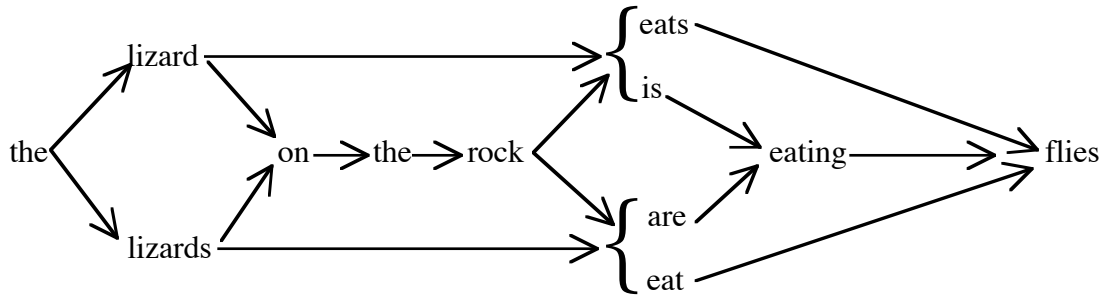
BASICS OF SYNTAX

syntax: “The way in which words are put together to form phrases and sentences.”
(American Heritage Dictionary)

Universal Grammar: The properties that must be part of the “language instinct” in order to account for the linguistic structures found in the world’s languages.

• What a model of syntax *cannot* be in Universal Grammar:

- (1) A “sentence dictionary”, i.e. a list of all the sentences of a language (why not?)
- (2) “Word salad”, i.e. words grouped randomly
- (3) A word-chain device (= a “finite state” or “Markov” model), i.e. a model whereby a sentence is produced one word at a time, with each successive word limiting the choice of what the next word might be (see *The Language Instinct*, pp. 81-90)



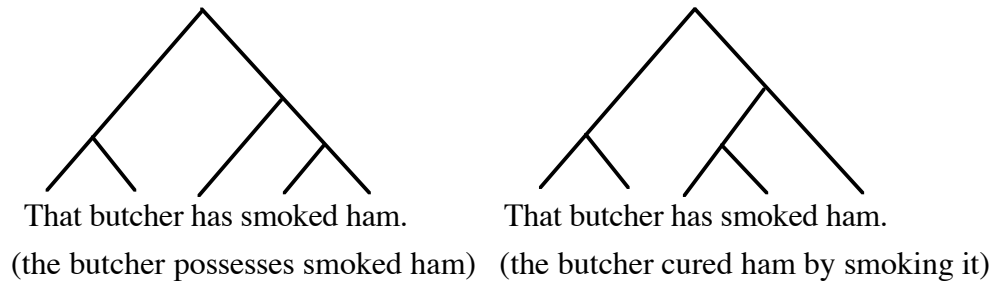
• What a model of syntax *must* account for in Universal Grammar:

(1) Word order

John showed the manager the shoplifter. ≠ John showed the shoplifter the manager.

Throw Momma a kiss from the train. *but* *Throw Momma from the train a kiss.

(2) Hierarchical structure



(3) Grammatical categories (*lexical* categories, i.e. “parts of speech”, and *phrasal* categories)

You should eat	$\left. \begin{array}{l} \underline{\text{fish.}} \\ \underline{\text{the fish.}} \\ \underline{\text{the fresh fish.}} \\ \underline{\text{fish from Japan.}} \end{array} \right\}$	<p>(<i>fish, the fish, the fresh fish, fish from Japan</i> are all Noun Phrases)</p>
You should eat	$\left. \begin{array}{l} \underline{\text{slowly.}} \\ \underline{\text{very slowly.}} \\ \underline{\text{much more slowly.}} \\ \underline{\text{slowly and carefully.}} \end{array} \right\}$	<p>(<i>slowly, very slowly, much more slowly, slowly and carefully</i> are all Adverbial Phrases)</p>

Compare the corresponding questions:

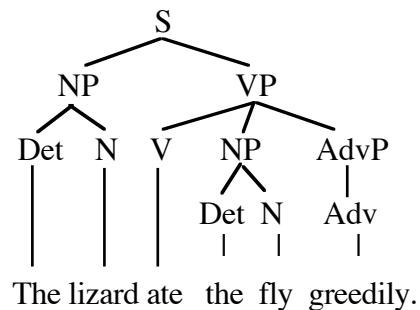
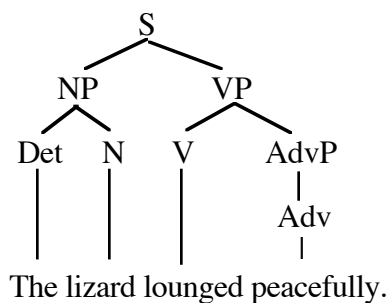
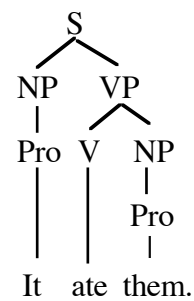
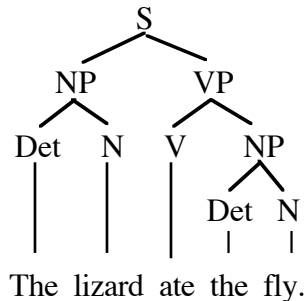
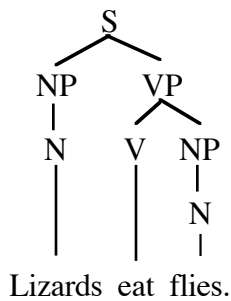
What should you eat? (*what* questions a **Noun Phrase**)

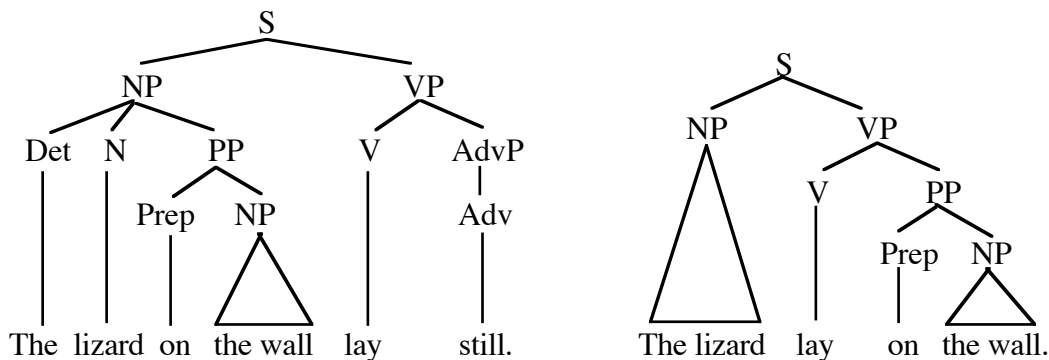
How should you eat? (*how* questions an **Adverbial Phrase**)

Words and/or phrases which can be substituted for each other in the same syntactic position belong to the same phrasal category.

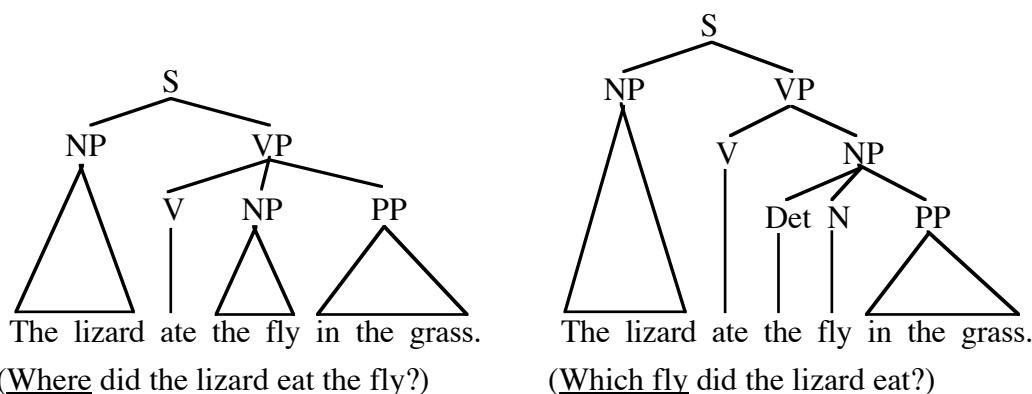
• Phrase Structure “trees” as a model for representing sentence structure

S = Sentence, **NP** = Noun Phrase, **VP** = Verb Phrase, **PP** = Prepositional Phrase, **N** = Noun, **V** = Verb, **Pro** = Pronoun, **Adv** = Adverb, **Det** = Determiner (i.e. *the, a, this, these, that, those*, etc.)





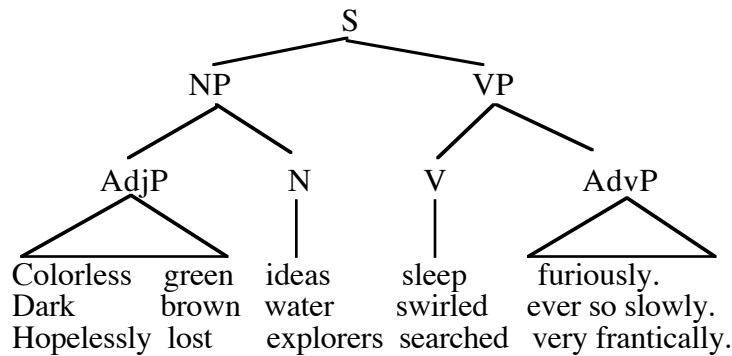
(A triangle indicates that the phrase has additional structure which is not spelled out.)



- **Phrase Structure rules:** a set of rules of the form $X \rightarrow Y Z$ (“grammatical category X is realized as grammatical category Y followed by grammatical category Z”); the little Phrase Structure grammar below “generates” all the trees above (and *infinitely* more!)

$S \rightarrow NP \quad VP$
 $NP \rightarrow \left\{ \begin{array}{l} (Det) \quad N \quad (PP) \\ \quad \quad \quad Pro \end{array} \right\}$ { ... } = choose either the top line or the bottom line but not both
 (...) = the enclosed phrase is optional, i.e. it may be present or absent
 $VP \rightarrow V \quad (NP) \quad (AdvP) \quad (PP)$
 $PP \rightarrow Prep \quad NP$
 $AdvP \rightarrow (Intens) \quad Adv$ Intens = “intensifier” such as *very*, *much more*, etc.

- “Meaningful” vs. “grammatical”: The form of a sentence is independent of its meaning.



Sentences can be ...

- Meaningful *and* grammatical. Most sentences we utter (we hope)!
 - Meaningless *but* grammatical. Colorless green ideas sleep furiously.
The lizard elapsed truth.
 - Meaningful *but* ungrammatical. Throw Momma from the train a kiss.
The balloon big and red.
 - Meaningless *and* ungrammatical. Ideas green colorless sleep furiously.
The lizard truth elapsed.
- The number of possible sentences is infinite, and there is no longest possible sentence—some ways that an infinite number of sentences can be created are

(1) Virtually incalculable number of word combinations

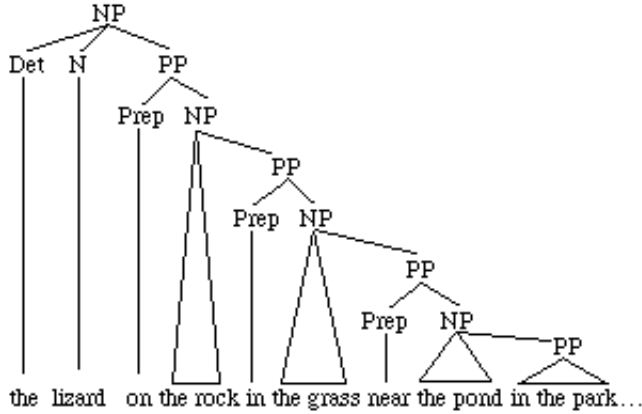
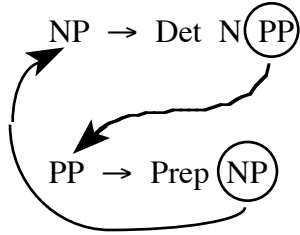
In the film, *Discovering the Human Language*, George Miller points out that in a sentence of just 10 words, if there were just 10 possible words to fill each of the ten slots in the sentence, the number of sentences would be 10^{10} .

(2) Conjunction: any sentence can be extended by adding *and* or *or* followed by another sentence

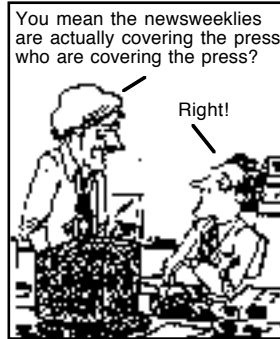
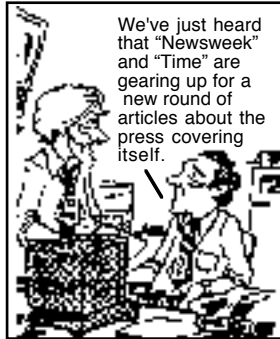
Washington was the first President *and* Jefferson followed him *and* ...

I can add one word *or* I can add two words *or* I can add three words *or* ...

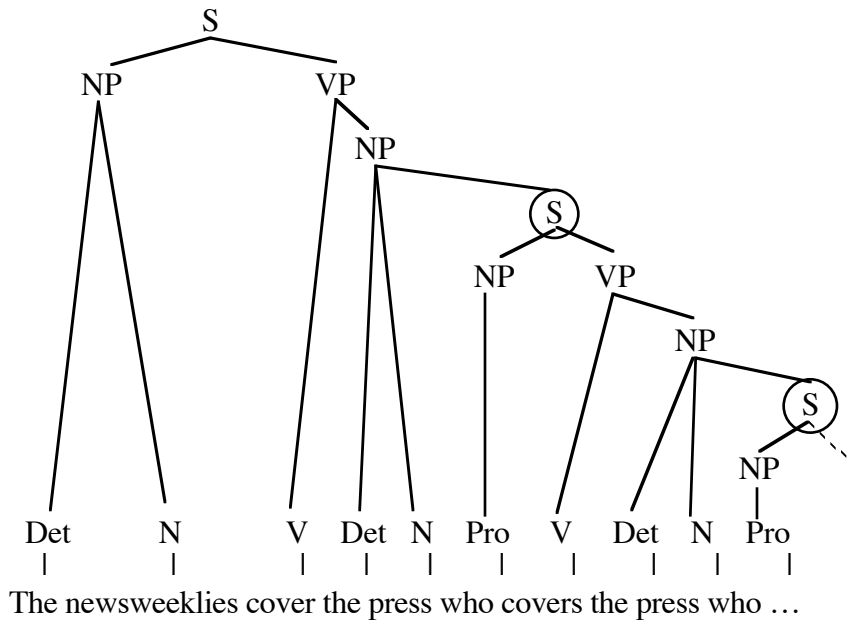
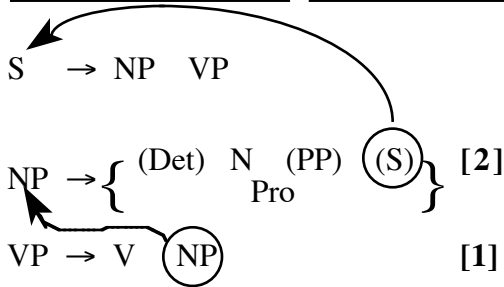
(3) **Recursion:** A syntactic category can contain a category of the same type which can contain a category of the same type which can contain a category of the same type which ...



DOONESBURY



BY GARRY TRUDEAU



Correlations of ordering of other elements in “consistent” languages of the three main word order types:

	SOV	SVO, VSO
1. verb modifiers	precede verb	follow verb
2. postpositions <i>or</i> prepositions	postpositions	prepositions
3. possessive phrases	possessor precedes	possessor follows
4. $\left. \begin{array}{l} \text{determiner} \\ \text{adjective} \\ \text{numeral} \end{array} \right\}$	precede noun	follow noun
5. relative clauses	precede noun	follow noun
6. order of NP elements	DET-NUM-ADJ-N	N-ADJ-NUM-DET

- Why do such correlations exist?

One answer: desire to be CONSISTENT in placement of the “**Head**” of a **Phrase** within its phrase

Head of a Phrase: The principal word of a phrase, i.e. the word from which a phrase type derives its name—the head of a VP is the Verb, the head of a NP is the Noun, the head of a PP is a Preposition or Postposition (depending on the type of PP a language has).

	“Head Final” languages	“Head Initial” languages
VP	(S) O \boxed{V}	(S) \boxed{V} O \boxed{V} (S) O
PP	NP $\boxed{\text{PostP}}$	$\boxed{\text{PreP}}$ NP
NP	Modifiers \boxed{N}	\boxed{N} Modifiers

SEE NEXT PAGE FOR EXAMPLES:

EXAMPLES:

1, 2.	SOV: <u>Tomapaymi</u> <u>pawi</u> <u>wi-ey</u> <u>nwuwessta.</u> S rock on V SVO: <u>Gare</u> <u>gandu</u> <u>ga ko</u> <u>gusho.</u> S V at-on rock VSO: <u>Humiga</u> <u>ang</u> <u>butiki</u> <u>sa</u> <u>bato.</u> V S "on" rock	‘The lizard lay on the rock.’ (The phrase “on the rock” shows both the position of a modifier of the verb— <i>where</i> did the lizard lie?—and whether the language uses a postposition or a preposition, shown in boxes.)
3.	SOV: <u>tomapaym-uy</u> <u>kkoli</u> lizard-’s tail SVO: <u>pata</u> <u>gare</u> tail lizard VSO: <u>buntot</u> <u>ng</u> <u>butiki</u> tail “of” lizard	‘lizard’s tail’
4.	SOV: { <u>ce tomapaym</u> <u>khun tomapaym</u> <u>twu tomapaym</u> } SVO: { <u>gare ama</u> <u>gare-n siri</u> <u>gare bolou</u> } VSO: { <u>butiki-ng iyon</u> <u>butiki-ng malaki</u> <u>but dalawa-ng butiki</u> }	{ ‘that lizard’ ‘big lizard’ ‘two lizards’ }
5.	SOV: <u>phali-lul</u> <u>cap-un</u> <u>tomapaym</u> fly-obj. caught-“that” lizard SVO: <u>gare</u> <u>la</u> <u>’yuwu</u> <u>didi</u> <u>ye</u> lizard which caught fly the VSO: <u>ang</u> <u>butiki-ng</u> <u>nakahuli</u> <u>ng</u> <u>langaw</u> “the” lizard-“which” caught “the” fly	‘the lizard that caught the fly’
6.	SOV: <u>i</u> <u>twu-mali</u> <u>khun</u> <u>tomapaym</u> these two-classif. big lizard SVO: <u>gare-n</u> <u>asse</u> <u>bolou</u> <u>maine</u> lizard big two these VSO: <u>itong</u> <u>dalawang</u> <u>malaking</u> <u>butiki</u> these two big lizards	‘these two big lizards’

Note that English and Tagalog are not “consistent” in the order of elements in #6. That is, an SVO language (English) and a VSO language (Tagalog) would be expected to have the order NOUN-ADJ-NUM-DET, whereas both English and Tagalog have the opposite order, i.e. the one expected for SOV languages. In fact, few if any languages are entirely “consistent” in their order types. This is one of the facts which makes X-bar Theory complicated. Why?

GROUND PLAN #2: LANGUAGES RELYING ON MARKING ON WORDS THEMSELVES TO SHOW OVERALL MEANING

Latin, Russian, Sandawe, Warlpiri, etc.

We may talk about marking grammatical function, such as **S** (Subject), **DO** (Direct Object), **IO** (Indirect Object), **Gen** (Genitive or Possessor):

Russian

	Subject form	Object form	Genitive form
‘lizard’ (masculine)	jáščeritsa	jáščeritsu	jáščeritsy
‘fly’ (feminine)	múxa	múxu	múxi

Jáščeritsa poj mála múxu. Múxu poj mála jáščeritsa. Poj mála jáščeritsa múxu. Poj mála múxu jáščeritsa. Jáščeritsa múxu poj mála. Múxu jáščeritsa poj mála.	‘The lizard caught the fly.’
--	------------------------------

xvost jáščeritsy = jáščeritsy xvost ‘lizard’s tail’

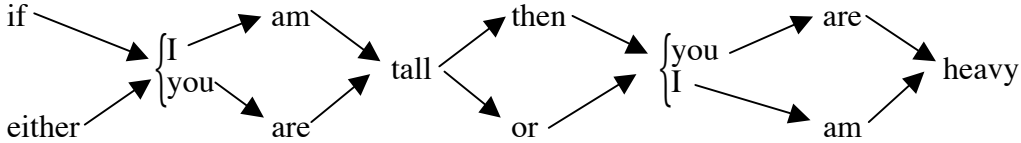
Though word order does not play a strong role in languages such as Russian, one can think of the word endings as forming a *hierarchical mental network* of grammatical relations like those diagrammed by phrase structure trees in languages with more rigid word order.

• **Conclusion:** All languages share a design (Universal Grammar) whereby actor, action, object of action, modifiers of various kinds, etc. combine according to systematic rule in hierarchical relationships. No other animal communication system has such a design, nor do any other complex communication systems used by humans. What would notions like “subject of action”, “modifier”, etc. mean in mathematics, computer programming languages, music, etc.!?

A NOTE ON \bar{X} (“X-BAR”) THEORY: Pinker in *The Language Instinct*, pp. 97-105, refers to “X-bar Theory” in his discussion of syntax in Universal Grammar. The idea of X-bar Theory is that all phrase types have similar internal structures. The phrase types include VP, NP, PP, AdvP, AdjP, and even Sentence, which, in X-Bar Theory is CP for “Complementizer Phrase”. Because of the similar structures, we can in effect reduce all Phrase Structure rules to “meta-rules” of the form $XP \rightarrow \text{SPEC } \bar{X}, \bar{X} \rightarrow X \text{ XP}$, in which any of the categories V, N, P, Adv, Adj, C(omplementizer) can be substituted for X. The way this theory is implemented to account for any particular language is complex and technical, going far beyond what can be covered in an introductory course (the UCLA Linguistics Department has a three quarter GRADUATE level sequence which covers just the basics!), and I don’t find Pinker’s discussion of this theory very enlightening. For our purposes it will suffice to understand (1) that a set of rather simple and general rules such as those presented in class can account for an infinite range of sentences and (2) that a model of this type would be needed for every human language. This is the basis of X-bar Theory as well, but X-bar Theory attempts to take the concept to a higher order of abstraction.

Discussion Questions on Syntax

1. **Word chain devices:** A model for the structure of sentences might be a device where you choose a word, which then leads to a limited set of choices for the next word and so on until you end up with a string of words that makes a sentence. Below is a word chain device that would yield a number of English sentences. The bracketed items mean that you can choose either the top one or the bottom one.



- Use this word chain device to create at least one GRAMMATICAL sentence of English.
 - Using the words in this word chain device, make up at least one UNGRAMMATICAL sentence of English that the structure of this device would PREVENT you from creating.
 - Use this word chain device to create at least one UNGRAMMATICAL sentence of English. Explain why your sentence demonstrates the shortcoming of a word chain device as a model for sentence construction in natural human languages.
2. **Hierarchical structure:** Below are some headlines (in capitals), some phrases, and a cartoon. In each case there is an ambiguity that involves grouping the words in different ways. For each item, draw two tree structures showing the different word groupings. For this question, don't try to label the trees with NP, VP, etc.

ENRAGED COW INJURES FARMER WITH AX
 KILLER SENTENCED TO DIE FOR 2ND TIME
 the design has big squares and circles
 they said she would go yesterday



- Grammatical categories:** Below are some headlines and phrases where there is an ambiguity that involves interpreting a word as belonging to either of two grammatical categories. Identify the word and the two categories that it allows.

JUVENILE COURT TO TRY SHOOTING DEFENDANT
 smoking grass can be nauseating
 the horse looked very fast

On the next page are some further examples that involve a combination of grouping and grammatical category ambiguity. Analyze the ambiguities as in the above examples.

SQUAD HELPS DOG BITE VICTIM
 TEACHER STRIKES IDLE KIDS
 SERBIAN FORCES FLYING UNAUTHORIZED FLIGHTS (from the *Bruin* about flights over Bosnia after the UN peace forces went in)
 REQUEST TO BAR STUDENT DENIED (from the *Bruin* about an attempt to keep a student out of the UC Davis medical school)
 BLOCK HEADS BACK ON CAMPAIGN TRAIL (from the *LA Times* referring to the late Sherman Block running for Sherrif again)

4. Using the little grammar on page 27, draw LABELED trees, i.e. trees including NP, VP, etc. for the following sentences:

- a. It flies erratically.
- b. He won the race through the hills easily.
- c. The runner from Kenya beat the Tanzanian by inches.
- d. Guards guard guards of guards of guards.

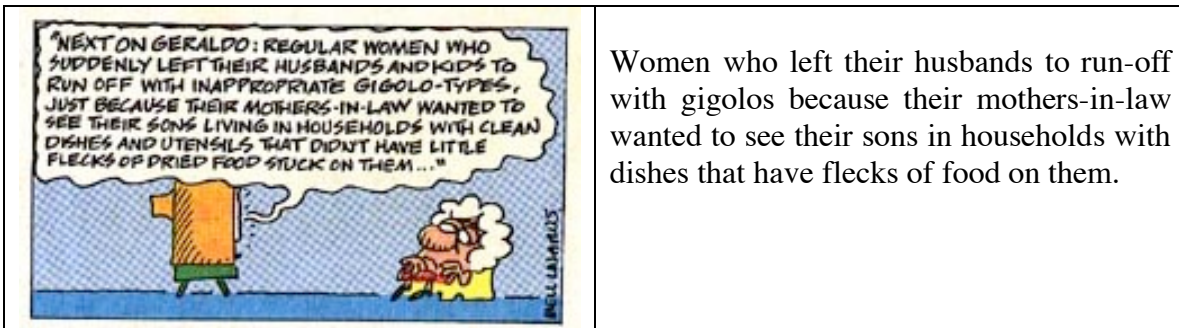
5. The following sentence is meaningful. Using the term *grammatical* to mean “conforming to the rules of a grammar”, it is also grammatical with respect to the little grammar on page 27:

Sisters are siblings of their brothers.

- a. Make a change in the sentence such that it is *meaningful but ungrammatical*.
- b. Make a change in the sentence such that it is *meaningless but grammatical*.
- c. Make a change in the sentence such that it is *meaningless and ungrammatical*.

6. Using the little grammar on page 27, create a sentence with a tree structure that involves at least two cases of recursion.

7. Here is a panel from a “Momma” cartoon and a simplified version of the sentence in the cartoon.



Make the following three modifications to the grammar on page 27, then figure out how the recursive properties of the grammar can account for the right-hand sentence, which could potentially go on infinitely. (Treat the hyphenated items as single words.)

- a. Add “S” to the NP rule, as in the grammar under the “Doonesbury” cartoon on p. 29.
- b. Add “(to VP)” to the end of the VP rule.
- c. Add “(because S)” to the end of the AdvP rule.

8. Here are sentences from two languages from the Middle East:

Turkish: kertenkele bir sinek yakala-di 'the lizard caught a fly'
 lizard a fly catch-past

Hebrew: ha-letaa taps-a zvuv 'the lizard caught the fly'
 the-lizard caught-it(fem.) fly

a. Using our typological categorizations of SOV, VSO, etc., what *types* do these languages fall into?

b. Here are the phrases meaning 'on (the) rock' in the two languages. Which phrase goes with which language? How do you know?

al ha-sela kaya-da
 on the-rock rock-on

c. Here are some further phrases in the two languages. State whether these are the orders you expect for the phrase types and why you have those expectations.

Turkish	Hebrew	
<u>kertenkele-nin</u> <u>kuyruk-u</u> lizard-'s tail-its	<u>zanav</u> <u>shel</u> <u>letaa</u> tail of lizard	'the lizard's tail'
<u>büyük</u> <u>kertenkele</u> big lizard	<u>letaa</u> <u>gdol-a</u> lizard big-fem.	'big lizard'
<u>iki</u> <u>kertenkele</u> two lizard	<u>štey</u> <u>leta-ot</u> two lizard-plural (fem.)	'two lizards'

9. In a language you know other than English, think about the order of words of phrases like those in 8a-c. What is the order for each type of phrase? Is the language CONSISTENT in the way it orders HEADS and MODIFIERS in the various phrase types?