**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**

     That butcher has smoked ham.  
     (the butcher possesses smoked ham)  
     That butcher has smoked ham.  
     (the butcher cured ham by smoking it)
(3) Grammatical categories (lexical categories, i.e. “parts of speech”, and phrasal categories)

You should eat \{ fish, the fish, the fresh fish, fish from Japan \} (fish, the fish, the fresh fish, fish from Japan are all Noun Phrases)

You should eat \{ slowly, very slowly, much more slowly, slowly and carefully \} (slowly, very slowly, much more slowly, slowly and carefully are all Adverbial Phrases)

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

\[ S = \text{Sentence}, \ NP = \text{Noun Phrase}, \ VP = \text{Verb Phrase}, \ PP = \text{Prepositional Phrase}, \ N = \text{Noun}, \ V = \text{Verb}, \ Pro = \text{Pronoun}, \ Adv = \text{Adverb}, \ Det = \text{Determiner} \ (i.e. \ the, \ a, \ this, \ these, \ that, \ those, \ etc.) \]
The lizard on the wall lay still.

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

The lizard lay on the wall.

The lizard ate the fly in the grass.

(Where did the lizard eat the fly?)

(Which fly did the lizard eat?)

• 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!)

$S \rightarrow NP \quad VP$

$NP \rightarrow \{ (Det) \ N \ (PP) \} \quad \{ \ldots \} = \text{choose either the top line or the bottom line but not both}$

$(\ldots) = \text{the enclosed phrase is optional, i.e. it may be present or absent}$

$VP \rightarrow V \ (NP) \ (AdvP) \ (PP)$

$PP \rightarrow \text{Prep} \quad NP$

$AdvP \rightarrow \text{(Intens)} \quad \text{Adv} \quad \text{Intens = "intensifier" such as very, much more, etc.}$
• “Meaningful” vs. “grammatical”: The form of a sentence is independent of its meaning.

<table>
<thead>
<tr>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
</tr>
<tr>
<td>V</td>
</tr>
<tr>
<td>AdjP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Colorless</th>
<th>green</th>
<th>ideas</th>
<th>sleep</th>
<th>furiously.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dark</td>
<td>brown</td>
<td>water</td>
<td>swirled</td>
<td>ever so slowly.</td>
</tr>
<tr>
<td>Hopelessly lost</td>
<td>explorers</td>
<td>searched</td>
<td>very frantically.</td>
<td></td>
</tr>
</tbody>
</table>

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 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

Looks like a yawner
Phil. I got a feeling
we’re going to be re-
duced to writing about
the TV
coverage again.

Say...
That re-

minds me,

Rick.

We’ve just heard
that "Newsweek"
and "Time" are
gearing up for a
new round of
articles about the
press covering
itself.

You mean the newsweeklies
are actually covering the press
who are covering the press?

Right!

Sounds like
there could be a
story there.

Well, I thought
you might want
to check it out.

BY GARRY TRUDEAU

The newsweeklies cover the press who covers the press who …
The sentence below involves recursion by allowing an S to occur within a VP. See if you can figure out how to modify the VP rule above in order to account for this (potentially infinitely long) sentence.

I know John said he believed Mary claimed linguists have proved…

S_0  S_1  S_2  S_3  S_4

---

• Two major “ground plans” in syntactic systems
  (see notes on Discovering the Human Language, p. 21 of the APS book)

  (1) Relying on word order to convey meaning

  (2) Marking grammatical function on the words themselves (thus making order of the words less crucial for understanding the overall meaning)

  But most languages have a little of both

GROUND PLAN #1: LANGUAGES RELYING ON WORD ORDER

We may talk about languages in terms of the order for S(ubject), O(bject), and V(erb):

The lizard caught a fly.
S(ubject)  V(erb)  O(bject)

The possible orders of S, V, and O: SOV, SVO, VSO, VOS, OVS, OSV

With only a tiny number of exceptions, languages exhibit only the following orders:

SOV: Turkish, Japanese, Korean, Persian, Hindi, Georgian, Eskimo, etc.
SVO: English, French, Swahili, Hausa, Bole, Yoruba, Thai, Vietnamese, etc.
VSO: Tagalog, Irish, Berber, (Classical) Arabic, (Biblical) Hebrew, etc.

Of these, SOV is the most common! Note that of the 6 possible orders for S, O, and V, these are the three which put the subject before object in the sentence.

SOV (Korean): Tomapaymi phalilul capassta. ‘The lizard caught a fly.’
S  O  V

SVO (Bole): Gare ‘yuwu didi.'
S  V  O

VSO (Tagalog): Nakahuli ang butiki ng langaw
V  S  O

NOTE: These orderings refer to the basic order of “neutral” declarative sentences. Many languages use different order for questions, emphatic constructions, and other special meanings.
Correlations of ordering of other elements in “consistent” languages of the three main word order types:

<table>
<thead>
<tr>
<th></th>
<th>SOV</th>
<th>SVO, VSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. verb modifiers</td>
<td>precede verb</td>
<td>follow verb</td>
</tr>
<tr>
<td>2. postpositions or prepositions</td>
<td>postpositions</td>
<td>prepositions</td>
</tr>
<tr>
<td>3. possessive phrases</td>
<td>possessor precedes</td>
<td>possessor follows</td>
</tr>
<tr>
<td>4. determiner, adjective, numeral</td>
<td>precede noun</td>
<td>follow noun</td>
</tr>
<tr>
<td>5. relative clauses</td>
<td>precede noun</td>
<td>follow noun</td>
</tr>
<tr>
<td>6. order of NP elements</td>
<td>DET-NUM-ADJ-N</td>
<td>N-ADJ-NUM-DET</td>
</tr>
</tbody>
</table>

- 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).

<table>
<thead>
<tr>
<th></th>
<th>“Head Final” languages</th>
<th>“Head Initial” languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP</td>
<td>(S) O <strong>V</strong></td>
<td>(S) <strong>V</strong> O</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>V</strong> (S) O</td>
</tr>
<tr>
<td>PP</td>
<td>NP [PostP]</td>
<td>[PreP] NP</td>
</tr>
<tr>
<td>NP</td>
<td>Modifiers [N]</td>
<td>[N] Modifiers</td>
</tr>
</tbody>
</table>

SEE NEXT PAGE FOR EXAMPLES:
### EXAMPLES:

<table>
<thead>
<tr>
<th>Example</th>
<th>SOV</th>
<th>SVO</th>
<th>VSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2.</td>
<td>Tomapaymi pawi wi-ey nwwessta. pawi</td>
<td>Gare gandu ga ko gusho. gandu</td>
<td>Humiga ang butiki sa bato. ang butiki</td>
</tr>
<tr>
<td></td>
<td>\text{SOV:} Tomapaymi pawi wi-ey nwwessta. pawi</td>
<td>\text{SVO:} Gare gandu ga ko gusho. gandu</td>
<td>\text{VSO:} Humiga ang butiki sa bato. ang butiki</td>
</tr>
<tr>
<td></td>
<td>‘The lizard lay on the rock.’ (The phrase ‘on the rock’ shows both the position of a modifier of the verb—where did the lizard lie?—and whether the language uses a postposition or a preposition, shown in boxes.)</td>
<td>\text{‘The lizard lay on the rock.’}</td>
<td>\text{‘The lizard lay on the rock.’}</td>
</tr>
<tr>
<td>3.</td>
<td>tomapaym-uy kkoli</td>
<td>pata gare</td>
<td>buntot ng butiki</td>
</tr>
<tr>
<td></td>
<td>\text{SOV:} tomapaym-uy kkoli</td>
<td>\text{SVO:} pata gare</td>
<td>\text{VSO:} buntot ng butiki</td>
</tr>
<tr>
<td></td>
<td>‘lizard’s tail’</td>
<td>‘lizard’s tail’</td>
<td>‘lizard’s tail’</td>
</tr>
</tbody>
</table>
| 4.      | \begin{align*}
\{ & \text{ce tomapaym} \\
\text{khun tomapaym} \\
\text{twu tomapaym} \\
\} \\
\{ & \text{gare ama} \\
\text{gare-n siri} \\
\text{gare bolou} \\
\} \\
\{ & \text{butiki-ng iyon} \\
\text{butiki-ng malaki} \\
\text{but} \\
\text{dalawa-ng butiki} \\
\} \\
\end{align*} | \begin{align*}
\{ & \text{that lizard’} \\
\text{‘big lizard’} \\
\text{‘two lizards’} \\
\} \\
\end{align*} | \begin{align*}
\{ & \text{that lizard’} \\
\text{‘big lizard’} \\
\text{‘two lizards’} \\
\} \\
\end{align*} |
| 5.      | phali-lul cap-un tomapaym | gare la ’ywu didi ye | ang butiki-ng nakahuli ng langaw |
|         | \text{SOV:} phali-lul cap-un tomapaym | \text{SVO:} gare la ’ywu didi ye | \text{VSO:} ang butiki-ng nakahuli ng langaw |
|         | ‘the lizard that caught the fly’ | ‘the lizard that caught the fly’ | ‘the lizard that caught the fly’ |
| 6.      | i twu-mali khun tomapaym | gare-n asse bolou maine | itong dalawang malaking butiki |
|         | \text{SOV:} i twu-mali khun tomapaym | \text{SVO:} gare-n asse bolou maine | \text{VSO:} itong dalawang malaking butiki |
|         | ‘these two big lizards’ | ‘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**

<table>
<thead>
<tr>
<th>Subject form</th>
<th>Object form</th>
<th>Genitive form</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘lizard’ (masculine)</td>
<td>jáščeritsa</td>
<td>jáščeritsu</td>
</tr>
<tr>
<td>‘fly’ (feminine)</td>
<td>múxa</td>
<td>múxu</td>
</tr>
</tbody>
</table>

Jáščeritsa pojmála múxu.  
Múxu pojmála jáščeritsa.  
Pojmála jáščeritsa múxu.  
Pojmála múxu jáščeritsa.  
Jáščeritsa múxu pojmála.  
Múxu jáščeritsa pojmála.

‘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 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 → SPEC X, X → X 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.

   ![Word chain device diagram]

   a. Use this word chain device to create at least one GRAMMATICAL sentence of English.

   b. 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.

   c. 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**

   ![Cartoon image]

3. **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 Sheriff 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.

   Women who left their husbands to run-off with gigolos because their mothers-in-law wanted to see their sons in households with dishes that have flecks of food on them.

   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.

<table>
<thead>
<tr>
<th>Turkish</th>
<th>Hebrew</th>
</tr>
</thead>
<tbody>
<tr>
<td>kertenkele-nin</td>
<td>zanav shel</td>
</tr>
<tr>
<td>lizard-’s tail-its</td>
<td>tail of lizard</td>
</tr>
<tr>
<td>kuyruk-u</td>
<td>letaa gdol-a</td>
</tr>
<tr>
<td>iki</td>
<td>štev leta-ot</td>
</tr>
<tr>
<td>two lizard</td>
<td>two lizard-plural (fem.)</td>
</tr>
</tbody>
</table>

‘the lizard’s tail’
‘big lizard’
‘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?