BASICS OF SYNTAX

Knowledge of language includes knowing how to speak and understand, but it also includes:

1) grammaticality judgements

- A word is a possible word in the language, or an impossible word .

  brick blick *bnick

- A sentence is grammatical or it is not grammatical.

  John went to the store
  **John to the store went*
  This is the report that I filed without reading.
  **This is the report that was filed without reading.*

2) Ambiguity 1: Structural

- What has four wheels and flies?

3) Ambiguity 2: Lexical

- Go ahead: Pick your nose

Grammar

- An internalized (mental) system of rules allows us to speak (or sign), understand, and make judgments about our language.

- Some rules are language-specific:
  SVO vs. SOV word order
  *bnick

- Other rules are universal:
  *Who did John and play the guitar?
  *This is the report that was filed without reading.

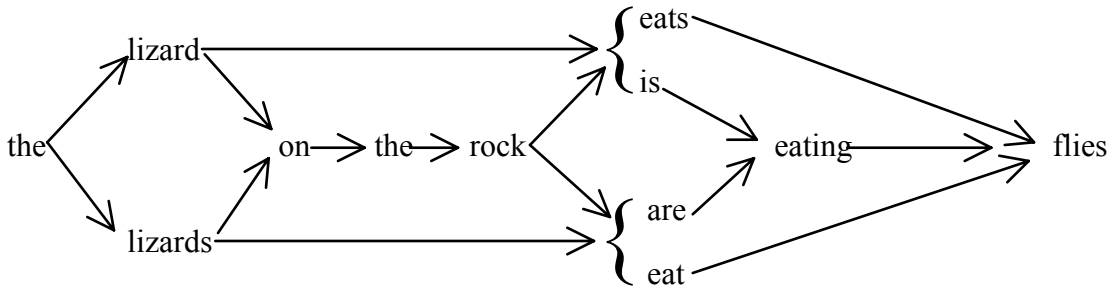
syntax: "The way in which words are put together to form phrases and sentences." (*American Heritage Dictionary*)

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?)- how do we know these? Memorized by hearing them? We can say sentences we've never heard before. Children learn lang by learning the which words to put together not by recording what comes before and after, but by which category follows other categories.

(2) 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)



Problems with Word Chains as a model of natural Language:

- Doesn't explain structural ambiguity
- Word Chains cannot handle long distance dependencies

• 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 { 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**)

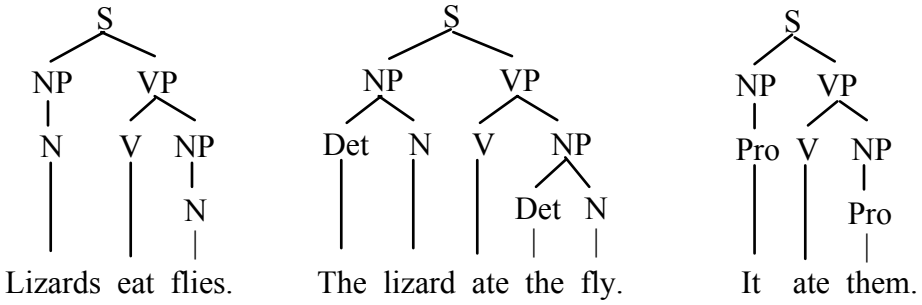
Compare the corresponding questions:

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

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

• 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.)



Terms

S, NP,V, flies, etc are nodes
Nodes are connected by branches
S dominates all other nodes
S immediately dominates NP1 and VP
NP1 and VP are sisters and daughters of S
S is the mother of NP1 and VP
Constituent

- How do we know what is a constituent?

(1) Pronoun Substitution

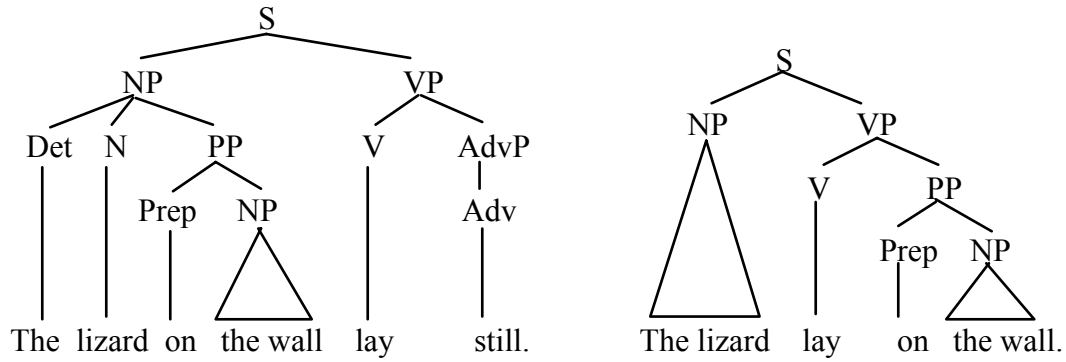
[The lizard] ate the fly.
It ate the fly.

(2) Question Word Substitution

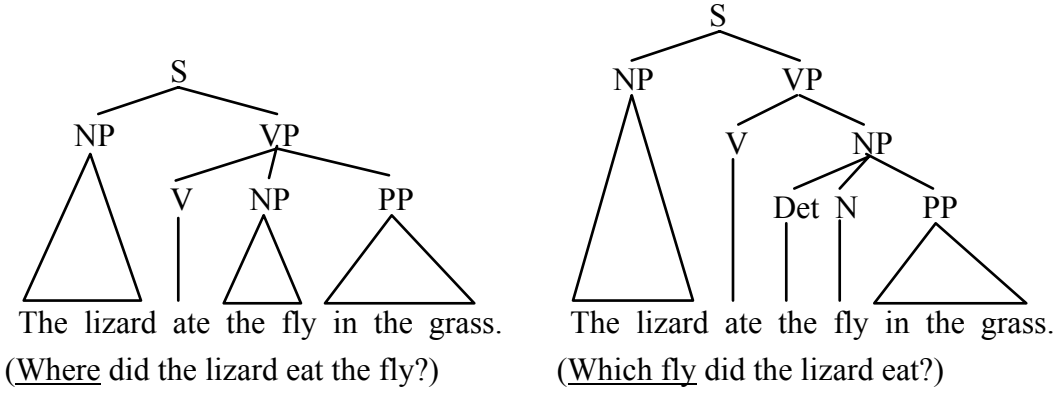
What ate the fly?
answer: the lizard

(3) *Do too, so do*

The spider ate the flies, and the lizard did too.
... and so did the lizard.



(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 → 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 → NP VP

NP → { (Det) N (PP) / Pro }

{ … } = 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

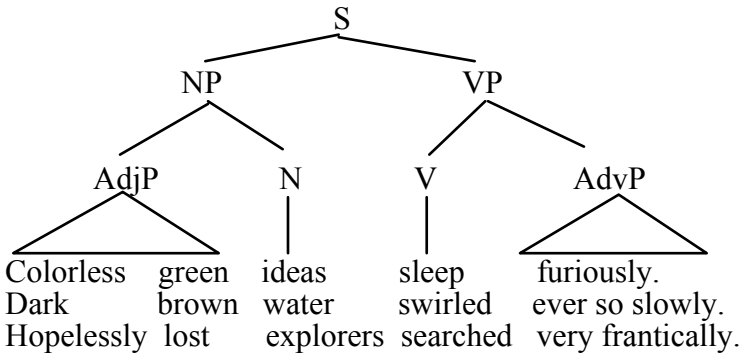
VP → V (NP) (AdvP) (PP)

PP → Prep NP

AdvP → (Intens) 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

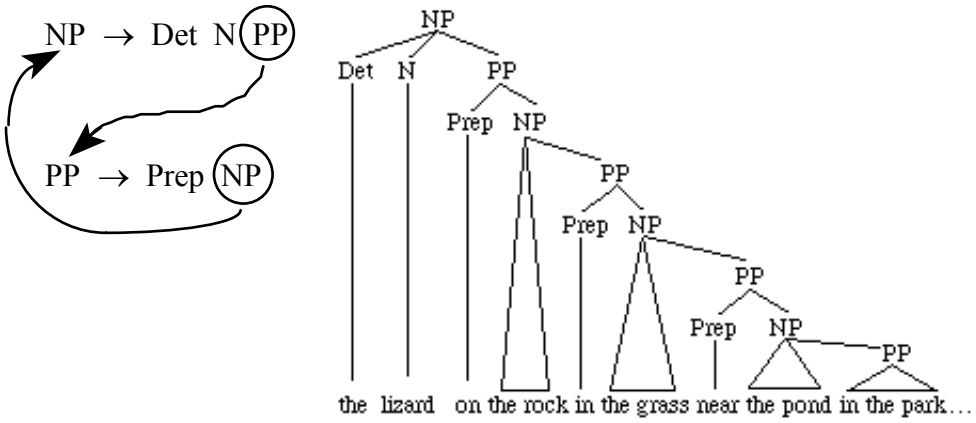
In the film, *Discovering the Human Language*, George Miller points out that in a sentence of just 10 words, if there were just 10 possible words to fill each of the ten slots in the sentence, the number of sentences would be 10^{10}.

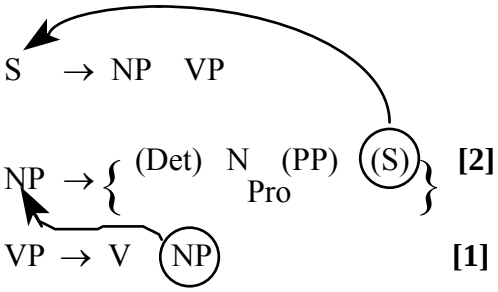
(2) Conjunction: any sentence can be extended by adding *and* or *or* followed by another sentence

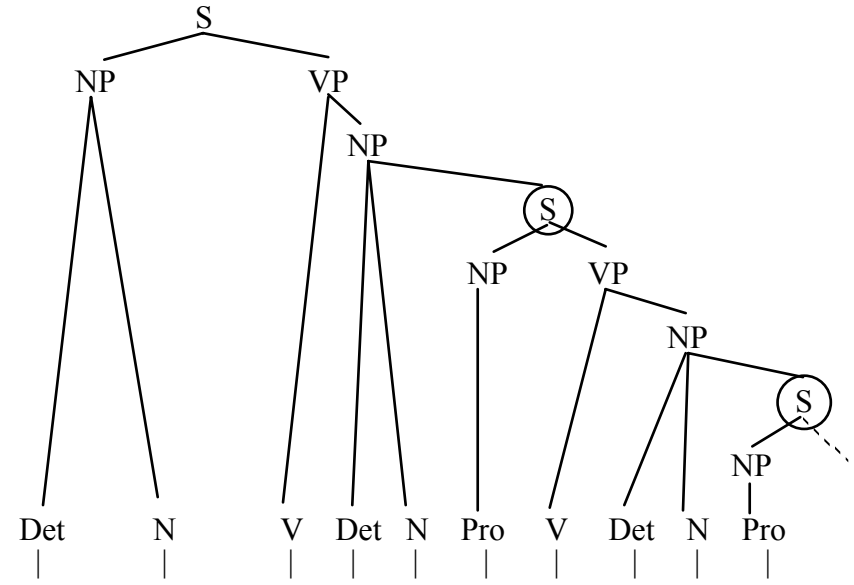
Washington was the first President *and* Jefferson followed him *and* …

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 …


NP → Det N (PP)
PP → Prep (NP)
NP
Det N PP
Prep NP
PP
Prep NP
PP
Prep NP
PP
the lizard on the rock in the grass near the pond in the park…


S → NP VP
NP → { (Det) N (PP) (S) / Pro } [2]
VP → V (NP) [1]


S
NP VP
NP
S
NP VP
NP
S
NP
Det N V Det N Pro V Det N Pro
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):	Tomapemi	p^halilɨl	capatt'a.	'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.

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):

Latin

	Subject form	**Object form**	**Genitive form**
'lizard' (masculine)	lacertus	lacertum	lacerti
'fly' (feminine)	mosca	moscam	moscae

Lacertus moscam cepit.
Moscam lacertus cepit.
Lacertus cepit moscam.
Moscam lacertus cepit.
Cepit lacertus moscam.
Cepit moscam lacertus.

'The lizard caught the fly.'

cauda lacerti = lacerti cauda 'lizard's tail'

Though word order does not play a strong role in languages such as Latin, 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.!?