Linguistics - words & sentences

Anca Chereches, PhD candidate in linguistics

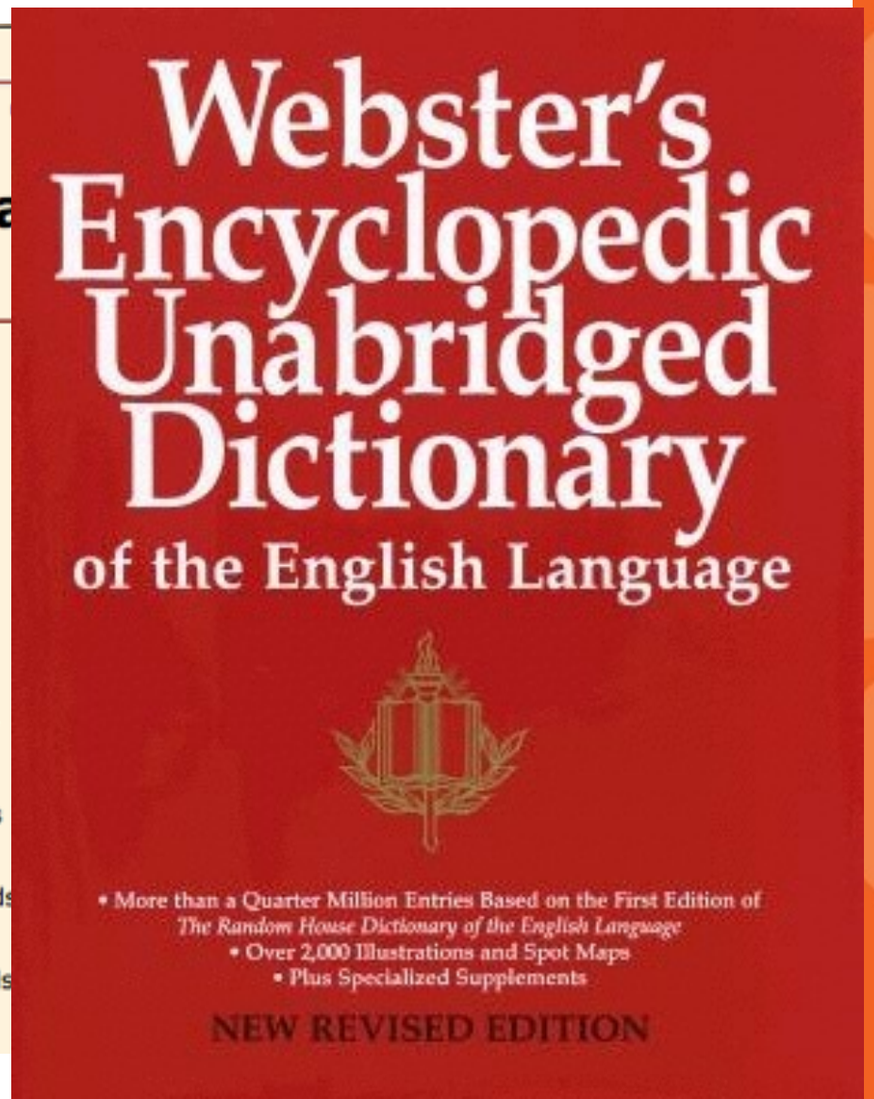
Guest lecture for COGST1101, Spring 2014

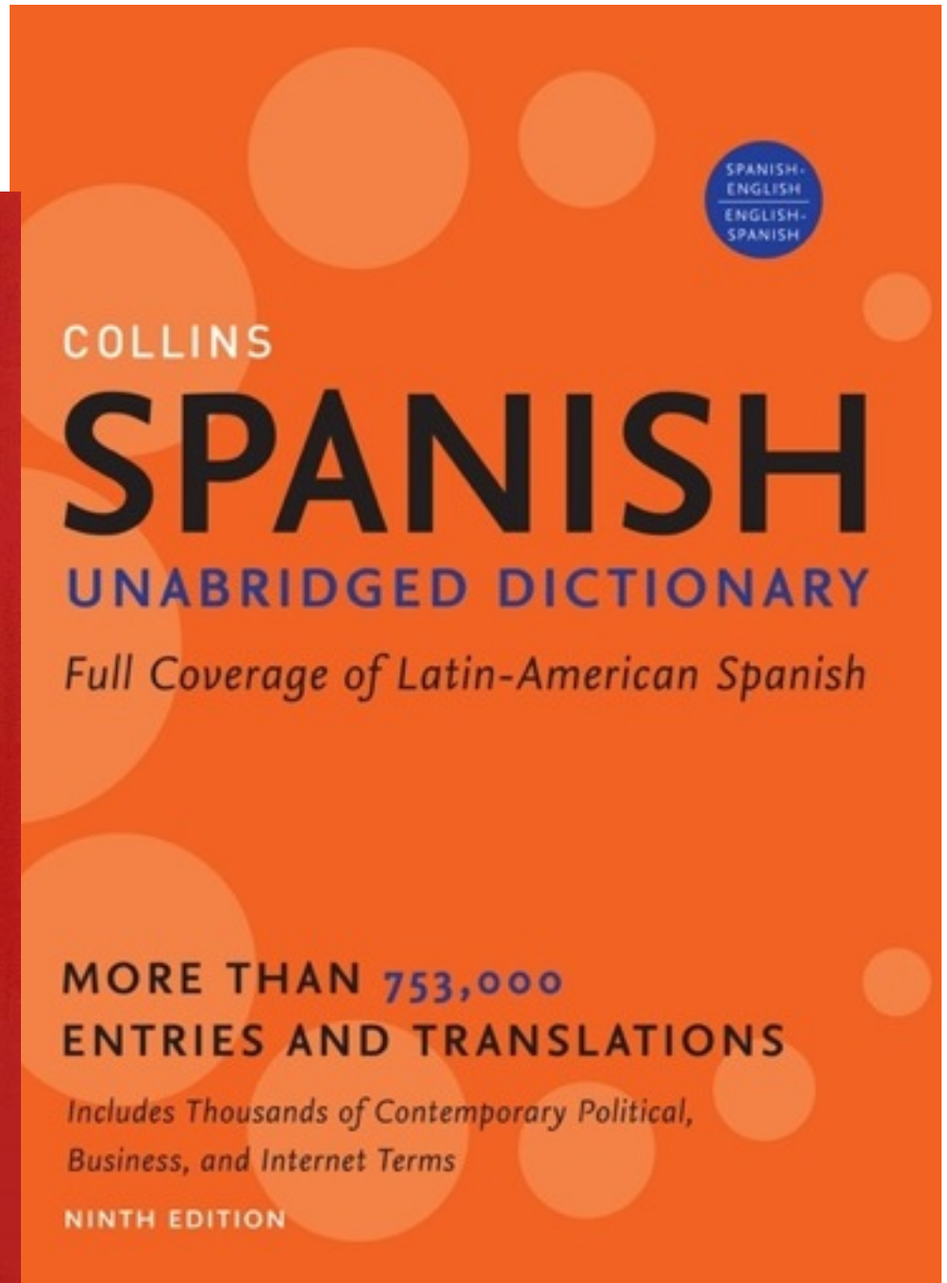
Language

- One of our most complex cognitive functions
- Uniquely human & inevitable
- Levels of analysis
  - Linguists - characterize linguistic knowledge
    - morphological & syntactic knowledge
  - Psycholinguists - determine the algorithms that implement this knowledge
    - mental lexicon (representations)
    - online sentence comprehension (modularity)
  - Neurolinguists - investigate neural mechanisms that realize these algorithms

Words

- How many words are in your native language?
- How many words do you know?





Words

- CLICKER QUIZ - How many words do you know?

A. Under 10,000.

B. Between 10,000 and 50,000.

C. Between 50,000 and 100,000.

D. Between 100,000 and 250,000.

E. Between 250,000 and 750,000.

Words

- Many estimates of adult vocabulary
    - between 40,000 and 150,000 words
    - recent estimates - around 40,000 for average American high school graduates
    - but 80,000 or more if we count names of people, places, idiomatic expressions etc.
- Oxford English Corpus
    - collection of texts, > 2 billion words
    - used by Oxford English Dictionary lexicographers

Vocab. size	% of OEC content	Example words
10	25%	the, of, and, to, that, have
100	50%	from, because, go, me, our, well, way
1000	75%	girl, win, decide, huge, difficult, series
7000	90%	tackle, peak, crude, purely, dude, modest
50,000	95%	saboteur, autocracy, calyx, conformist
>1,000,000	99%	laggardly, endobenthic, pomological

Words

- let's stick with 60,000 words. Should you be impressed?
    - animal vocalizations ("words")
        - vervet monkeys: ~ 5 alarm calls (for different kinds of predators), 4 grunts, 1 "intergroup wrr", …
        - chaffinch: ~ 8 basic calls + song (differs by sex & by season) (compare ~200 songs for nightingales)
        - chicken: ~18 basic calls + crowing (differs by sex & season)
        - bottlenose dolphins: ~ 40 whistles (not clear if they're distinct!)
    - animal language acquisition studies
        - chimps Washoe & Nim Chimpsky (!): ~ 200 ASL signs
        - Koko the gorilla: ~ 400 (handler claims ">1000 signs", but not evidence exists)
        - Kanzi the bonobo: ~ 400 lexigrams

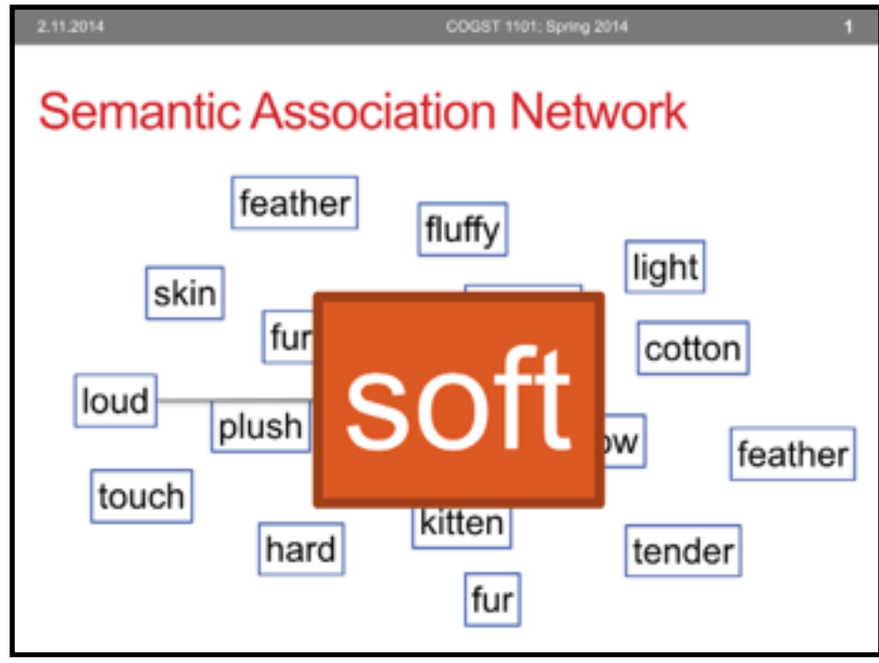
Words

- let's stick with 60,000 words. Should you be impressed?
    - human language acquisition
        - 60,000 words at 17 yrs old
        - start learning words at 1
            - => ~3750 words/yr, ~10 words/day, 1word/1.5hrs
- words in oral languages & signs in sign languages are **arbitrary** symbols
    - arbitrariness - a word's meaning is not predictable from its form
    - *dog* (Eng), *perro* (Spa), *inu* (Jap), *câine* (Rom) - there's nothing dog-like about these words
    - *bow-wow* (Eng), *guau guau* (Spa), *wan wan* (Jap), *ham ham* (Rom) - not as arbitrary, but still unpredictable, conventionalized, learned
        - compare to animal vocalizations, which are usually instinctual: a meow is a meow is a meow

- CLICKER POLL - Are the lexigrams that Kanzi the bonobo uses arbitrary symbols, in the same way that human words are?

A) Yes B) Mostly yes C) Mostly no D) No E) Who is Kanzi?

Words



- What does it mean to “know” a word?
- definitions? (**mental lexicon** is like a dictionary?)
  - do dictionary definitions cover all you need to know? (your experience with using a dictionary when learning a foreign language)
  - try defining “love”, “the”, and “nevertheless”
  - how are words organized? alphabetically vs…
    - by meaning: remember semantic association networks in the DRM False Memory paradigm
    - by sound: common slips of the tongue suggest a phonological network as well (*I hereby jeopardize you*; intended: *I hereby deputize you*)

Words

- What does it mean to “know” a word?
- many dimensions to knowing a word
    - spoken form, written form
    - frequency (*abode* vs *house*)
    - stylistic register (*request* vs *ask for, bucks* vs *dollars*)
    - grammatical behavior (noun, vb, etc.)
    - relationship to other words: **morphological [To be continued!]** (*sing - singer*), semantic (*warm - tepid - hot*)
    - referential (*aardvark*) & affective components (*puerile* vs *irresponsible, old* vs *ancient*)
    - collocational behavior: what other words does this word commonly occur with? (*take* vs *make a break*)

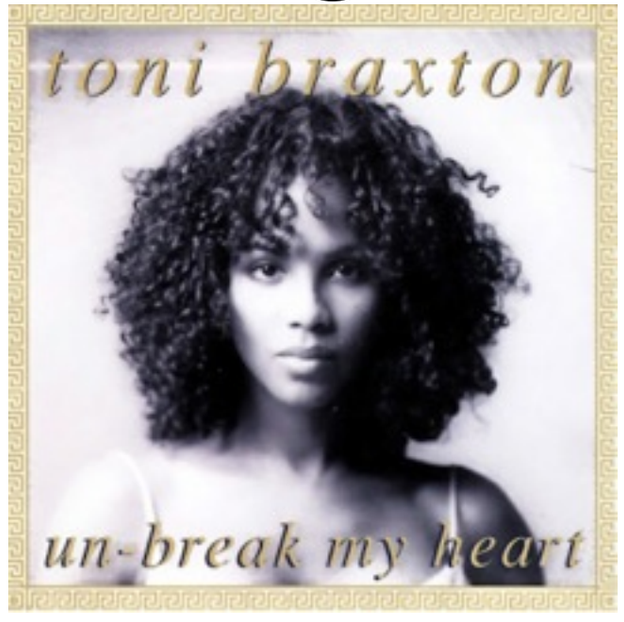
Words - morphological knowledge

- The structure of words
  - Compounds
  - Derivational affixes
  - Inflectional affixes
- Q: What does this suggest about the structure of the mental lexicon?
- Q: How many words are there in your language?

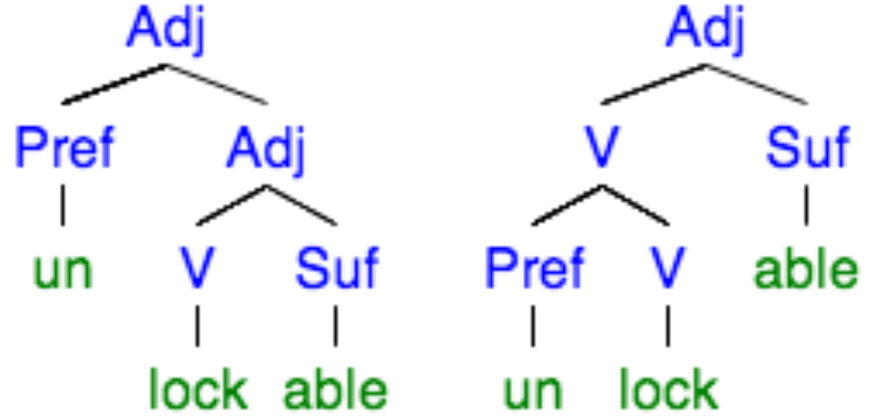
Words - morphological knowledge

- **Compounding** - a productive way to form new words in English
    - *black* + *bird* > *blackbird; black* + *board* > *blackboard*
    - Ger. *Rechtsschutzversicherungsgesellschaften* 'insurance companies providing legal protection'
    - all languages have compounds! (e.g. British Sign Language think + true > 'believe')
    - compounding is **productive**: new compounds formed every day (*crowdsourcing*, *Brangelina*)
- understanding/being able to infer a word's internal components
    - children are not as good as you are!
        - e.g. "it's called a birthday because you get presents and eat cake"
    - you're so eager to do this, you can go overboard
        - folk etymologies: *asparagus* > *sparrow grass*, *cater-corner* > *catty/kitty-corner* (after *cater* "four" became obsolete)
        - eggcorns: *acorn* > *eggcorn*, *respite* > *rest bite*
    - semantic organization of words in your mental lexicon

Words - morphological knowledge



- **Derivation** - another productive way to form new words in English
  - add an affix (*re*-, *un*-, -*able*, -*ify* etc.) to a stem => new meaning & sometimes a new part of speech
  - e.g. Adj + *ity* > N (*oddity*, **learnity*, **bossity*; **itty-odd*)
  - e.g. *re* + V > V (*replay*, **reodd*, **reboss*; **play-re*)
  - some affixes can attach to >1 type of stem
    - *un* + V > V (*undo*), *un* + Adj > Adj (*unaware*)
    - => **ambiguities**, eg *unlockable*
  - **productivity**:*retweet*,*defriend*



- Derivation to the extreme!
  - **Agglutinative languages**

Turkish derivation	
göz	eye
gözlük	glasses
gözlükçü	seller of glasses (oculist)
gözlükçülük	the occupation of oculists
gözlükçülükçü	a lobbyist for the oculist profession
gözlükçülükçülük	occupation of lobbying for the oculist profession

Words - morphological knowledge

- **Inflection**
  - add an affix (*-s*, *-ed* etc.) to a stem => express a grammatical category (tense, number, case, gender, etc.); doesn't form a new word or change the part of speech
  - English has lost most of its inflectional morphology

Person endings	Modern English		Old English	
I	steal	stole	stel**e**	stæl
you	steal	stole	stil**st**	stæl**e**
he/she/it	steal**s**	stole	stil**ð**	stæl
we/you/they	steal	stole	stel**aþ**	stǣl**on**

  - but what's left is very **productive**: *tweet* > *tweets, wug* > *wugs*
  - some languages have even less inflectional morphology!
    - e.g. in Standard Chinese, "zǒu" (走) = I/you/(s)he/it/we/they go/went
  - in some languages, inflection is prefixed rather than suffixed
    - e.g. in Swahili, *ni-na-soma* (I-present-read) 'I am reading'

Words - morphological knowledge

- Morphological knowledge - language at Marr's computational level of analysis
- (building blocks) morphemes = smallest units of meaning (roots, affixes)
- (rules of combination) stems can combine with each other and with affixes
    - e.g. affixes have preferences for:
        - certain types of stems (*oddity* / **learn-ity*, *learned* /* *student-ed*)
        - certain orders of combination
            - preffixes vs. suffixes (*unlock*/**lock-un*, *oddity*/**itty-odd*)
            - derivational vs. inflectional (*singers*/**sing-s-er*, *simplified*/**simpl-ed-fy*)
- => This raises a question at a different level of analysis: the algorithmic
    - How are morphologically complex words represented in the mental lexicon? Do all word forms get stored in long-term memory? What can be constructed online, in working memory?

Words - morphological knowledge

How are morphologically complex words represented?

decompositional models — mixed models — full listing

- Argument 1 - lexical retrieval is fast
  - average speech rate - ~6 syllables/sec
  - lexical decision tasks (is this a word?) - RT ~200msec
- Argument 2 - storage space is limited
  - agglutinative languages (e.g. Turkish)
- Argument 3 - separate entries for stems & affixes
  - morphological priming
- Argument 4 - productivity

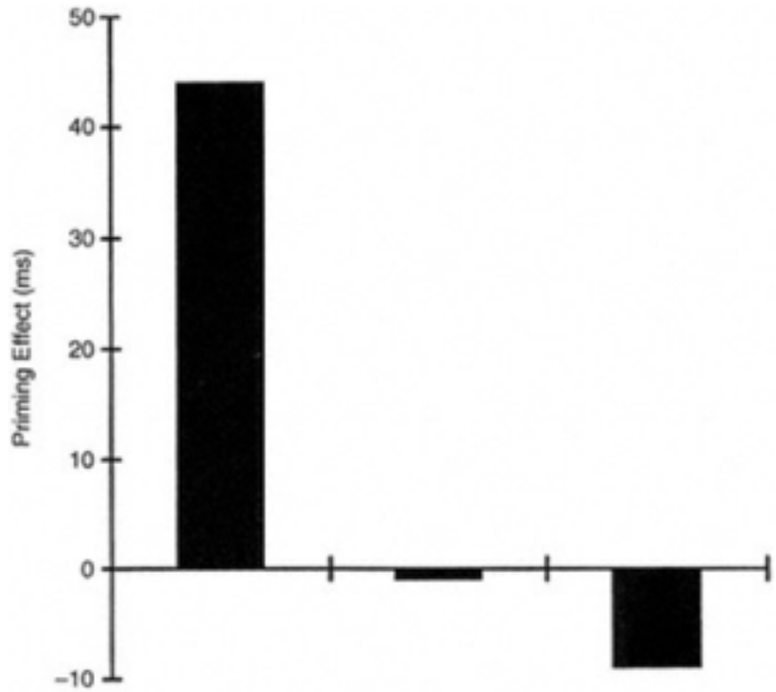


Fig. 4.1 Derivational morphology: Cross-modal priming between (i) morphologically and semantically related [+M +S] pairs, such as happiness/happy, (ii) morphologically but not semantically related [+M -S] pairs, such as apartment/apart, and (iii) unrelated [-M -S] pairs such as bulletin/bullet.

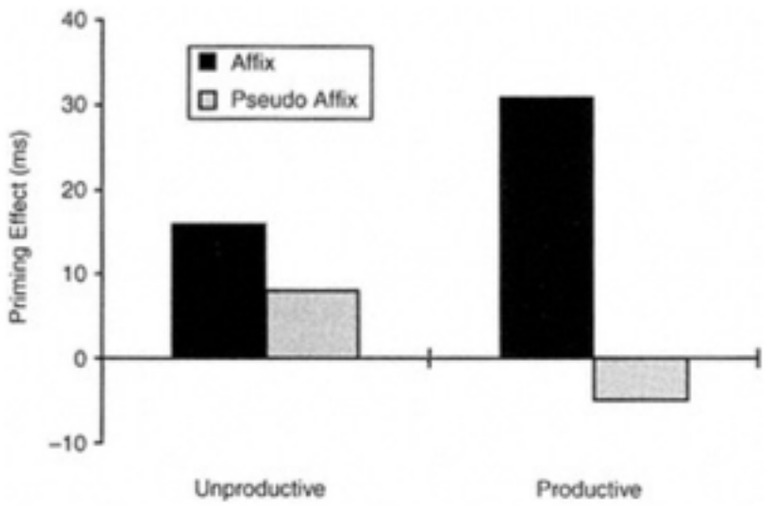


Fig. 4.4 Affix priming: Cross-modal priming between affixed primes (e.g. darkness) and affixed (toughness) and pseudo-affixed (harness) targets, broken down by the productivity of the affixes involved.

Figures from Marslen-Wilson 1999

Words - morphological knowledge

- Mini-review - words
  - arbitrariness
  - mental lexicon
  - e.g. morphological knowledge
    - building blocks (stems, affixes)
    - rules of combination
  - productivity
- Question: How many words are there in English / your native language?

Words

- CLICKER POLL: How many words are there in English / your language?

A. ~ 60,000

B. ~ 100,000

C. ~ 750,000

D. ~ 2 billion

E. Other.

Sentences

3-sign “sentences” by Nim Chimpsky		
Apple me eat	Eat me eat	More eat Nim
Banana Nim eat	Finish hug Nim	Nut Nim nut
Banana me eat	Give me eat	Play me Nim
Drink me Nim	Grape eat Nim	Tickle me Nim
Eat Nim eat	Hug me Nim	Tickle me eat
Eat Nim me	Me Nim eat	Yogurt Nim eat
Eat me Nim	Me more eat	

3/4-word sentences by a 2/3-yr-old child		
Mommy go store	I see manses	Tractor go floor
I want my cup	I want this	Give doggie paper
I camed here	More mac and cheese	Now put boots on
I no like it	I want green one	Mommy talking bout lady.
Where Daddy go?	Put truck window	See marching bear go?
I no want book	Adam put it box	
No Mommy do it	I ride horsie	

Sentences

- **Syntax** - knowledge of how to combine words in sentences
- Observation - basic word order
  - *Dog bites man.* vs *Man bites dog*.
  - English is an SVO language (subject-verb-object)
  - Compare with Japanese:

    Ann-ga bento-o kaimasu.

    Ann lunch buys.

    ‘Ann is buying a boxed lunch.’
- Compare with Malagasy:

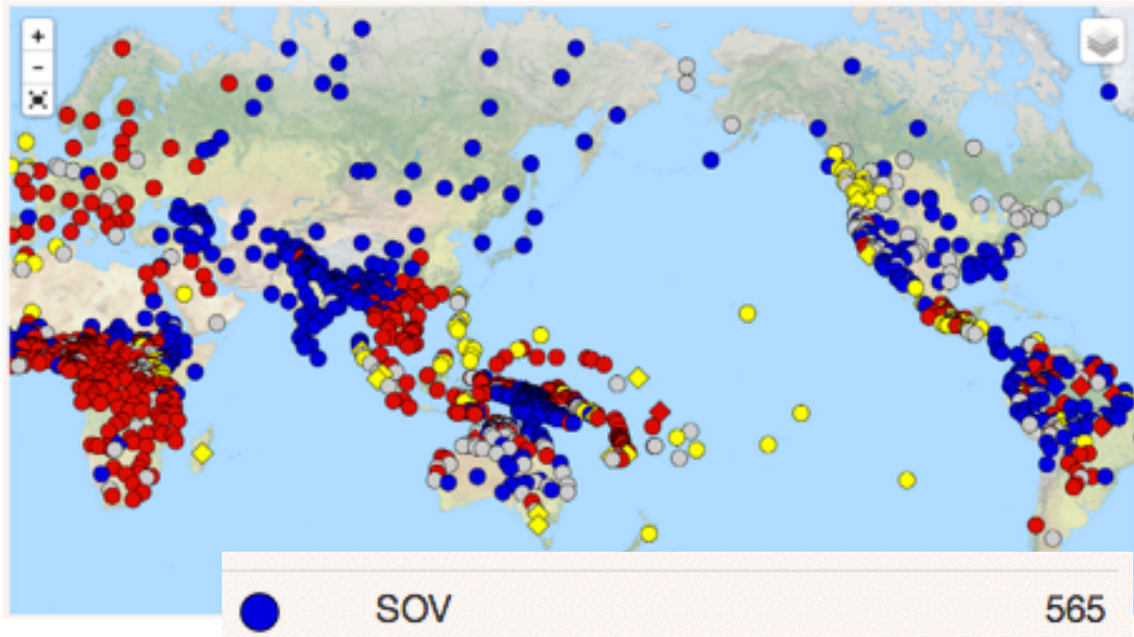
  Manasa lamba amin’ny savony ny lehilahy.

  washes clothes with the soap the man

  ‘The man washes clothes with the soap.’

	Proportion1	Proportion2	Examples
SOV	51.8	44.8	Latin
SVO	32.4	41.8	Mandarin
VSO	13.3	9.2	Irish
VOS	2.3	3.0	Fijian
OVS	—	1.2	Hixkaryana?
OSV	—	—	Warao?

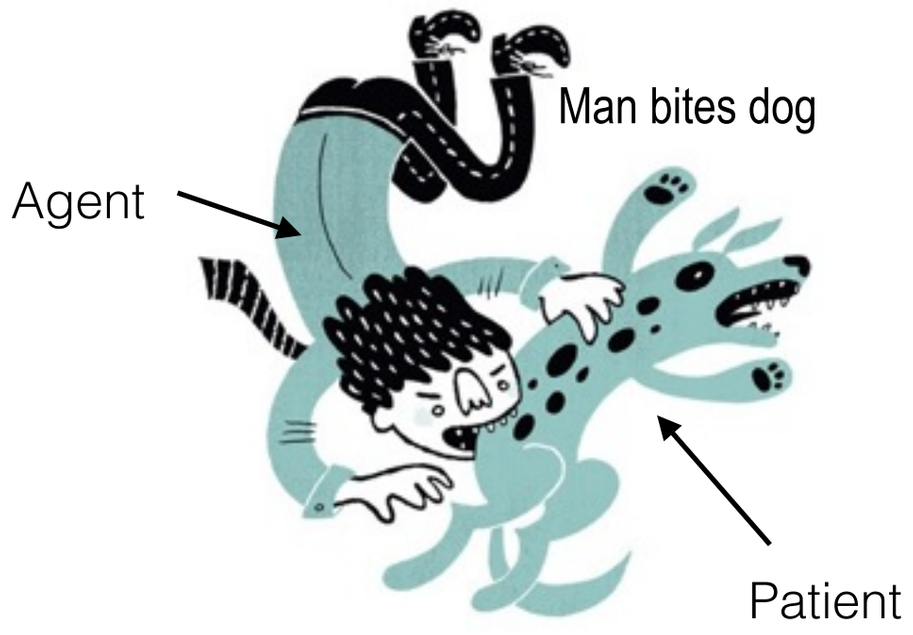
Proportion1 from Hawkins 1983; sample size 336
Proportion2 from Tomlin 1986; sample size 402
- total # of living languages - estimated at ~7,000



map from wals.info

Sentences

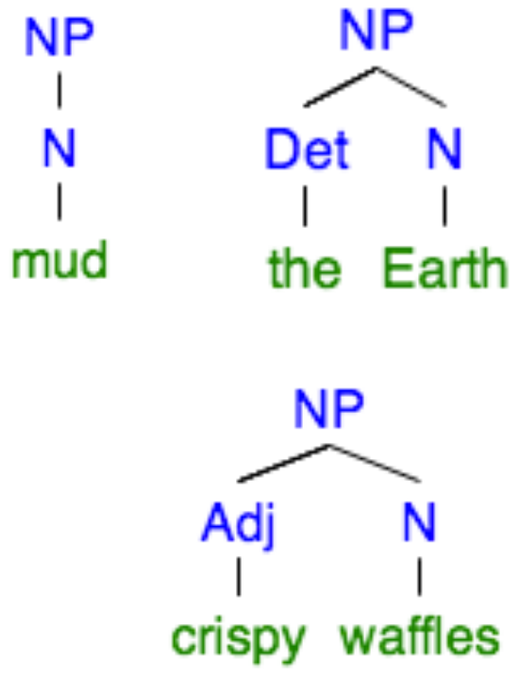
- Observation - basic word order
  - >95% of languages place subjects before objects —> Why?
  - information packaging: optimized for ease of processing
  - subjects - usually agents; given information -> easier to process
  - objects - usually patients; new information -> more effort to process
  - => addressee-oriented strategy



I read **an interesting article** last night. *It* was about **a man who bit a dog**. *He* said he didn't know what …

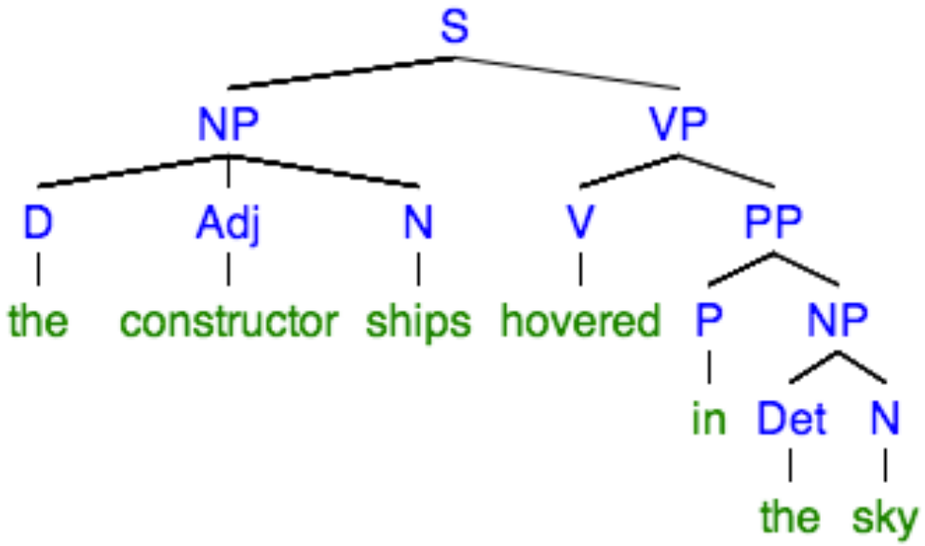
Sentences

- Observation 2 - constituents (groups of words that act as a unit)
    - *Sally read about {mud / the Earth / crispy waffles / the language of her parents / *crispy / *of / *quickly / *laugh}.*
    - *{mud / the Earth / a crispy waffle / the language of her parents / *crispy / *of / *quickly / *laugh} has a few defining features.*
    - chunks with the same distribution as a noun - Noun Phrases (NPs)
- what can an NP consist of? written as a **phrase structure rule**:
    - NP -> N (*mud, gold, salt*)
    - NP -> Determiner N (*the Earth, some people, a cat*)
    - NP -> Adj N (*crispy waffles, loud noises*)
- where did we find all of these NPs? after a preposition (*about*)
    - PP -> P NP (*about mud, about the Earth*, etc.)
- we've also seen PPs in another context (*of her parents*)
    - NP -> Det N PP (*the language of her parents*)



Sentences

- Observation 2 - constituents (groups of words that act as a unit)
    - *Sally {jumped / ate a pie / put a pie in the oven / *nice / *quickly / *towel}.*
    - chunks with the same distribution as a verb - Verb Phrases (VPs)
        - what can a VP consist of? more **phrase structure rules**:
            - VP -> V (*jumped, runs, boils*)
            - VP -> V NP (*ate a pie, ran an election campaign*)
            - VP -> V NP PP (*put a pie in the oven*)
- how do we form a sentence?
    - S -> NP VP (*Sally swam, The Vogon constructor ships hovered in the sky, Some people like vanilla*)



Sentences

- CLICKER POLL: What phrase structure rule do we need to add to our collection if we want to produce the sentence *Arthur thinks his friend is from Guildford*.

A. S -> NP VP NP VP

B. NP -> V NP

C. V -> V NP VP

D. VP -> V S

E. All we need is a towel.

Phrase structure grammar (so far)

N -> (Det) (Adj) N (PP)

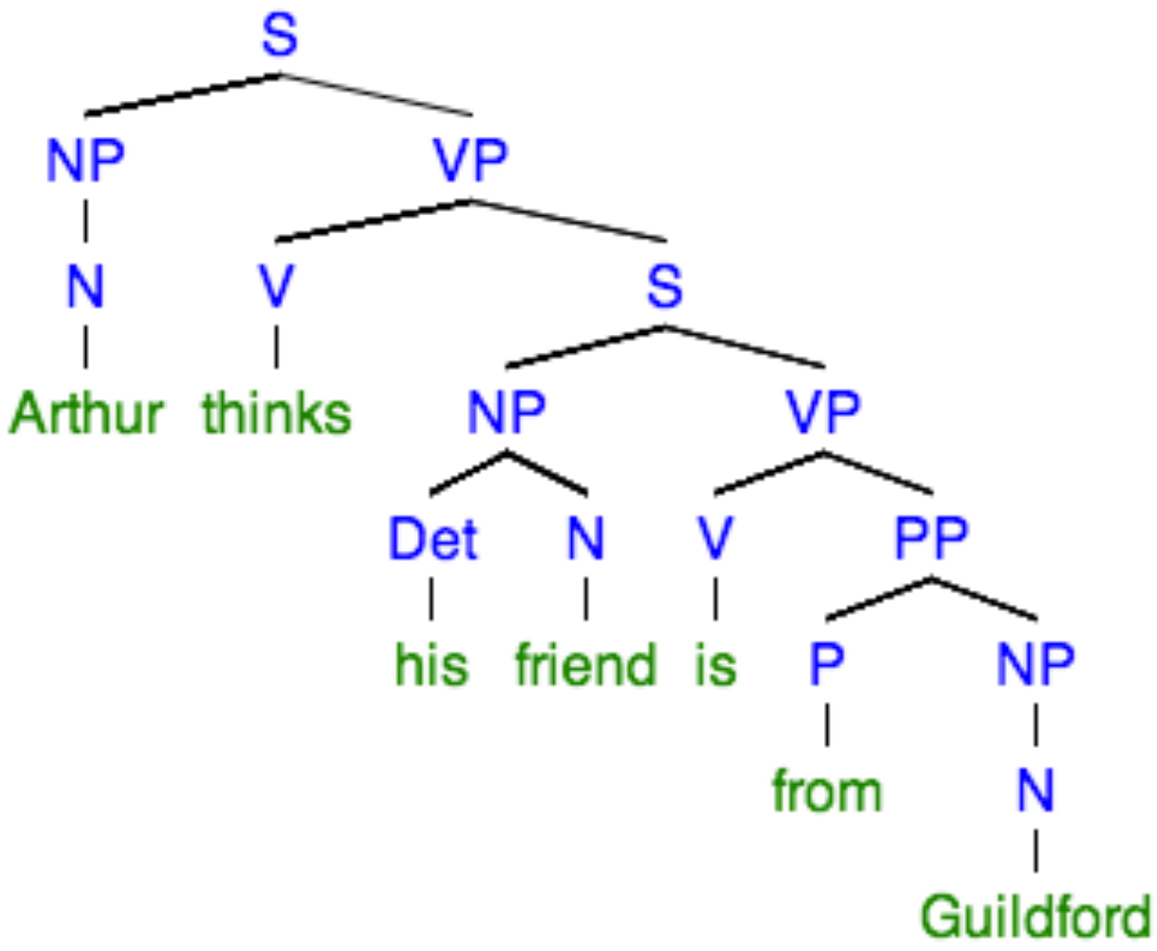
PP -> P NP

VP -> V (NP) (PP)

S -> NP VP

*Parentheses mark optionality.

Sentences



Phrase structure grammar (so far)

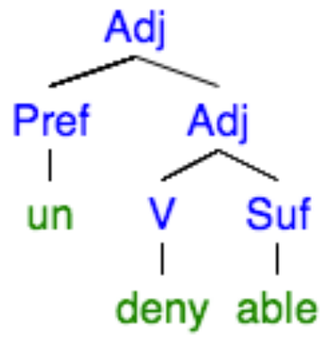
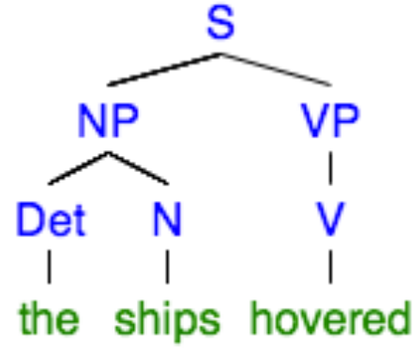
N -> (Det) (Adj) N (PP)

PP -> P NP

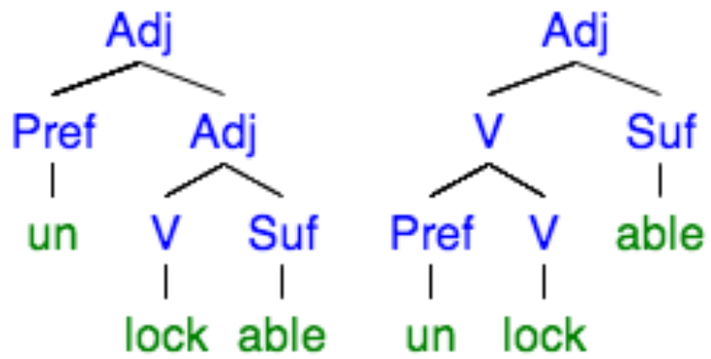
VP -> V (NP) (PP) **(S)**

S -> NP VP

Words & Sentences

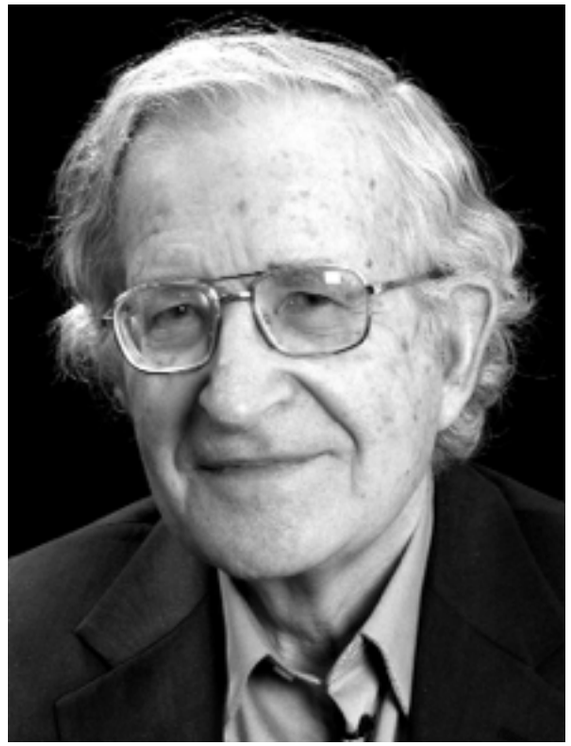
	Morphological knowledge (words)	Syntactic knowledge (sentences)
building blocks	morphemes (stems & affixes)	
rules of combination	Adj + *ity* > N *re* + V > V …	N -> (Det) (Adj) N (PP) VP -> V (NP) (PP) (S) P -> P NP S -> NP VP …
(we can represent these structures using) trees		

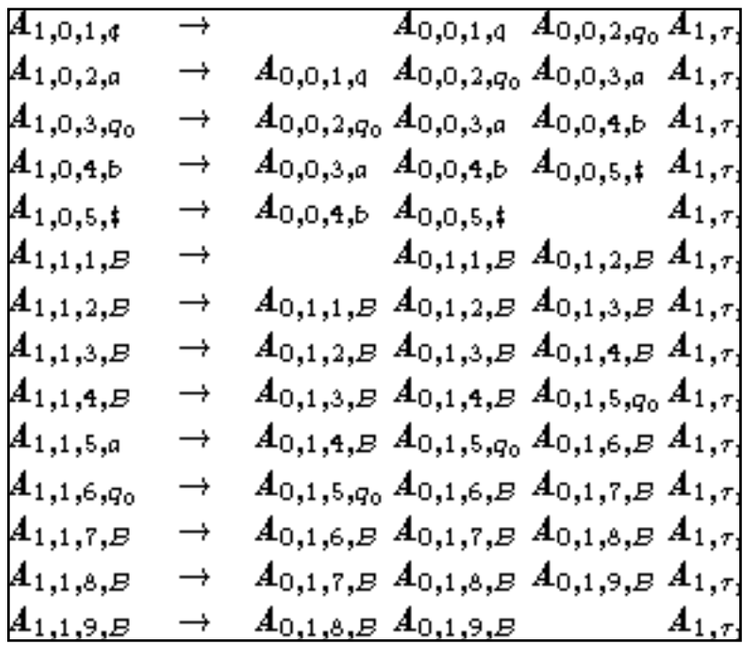
Words & Sentences

	Morphological knowledge (words)	Syntactic knowledge (sentences)
ambiguity		YES [to be continued]
productivity?	YES (*retweet, defriend, wugs*)	YES [**to be continued**]

Syntactic productivity

- Noam Chomsky, 1950s
    - introduced and popularized these grammars
        - in linguistics - mainly used in syntax
        - in computer science - mathematically precise cousins used in parsing applications & in theory of computing
- in cognitive science - one of the main contributors to the Cognitive Revolution
    - critique of behaviorist B.F. Skinner's book on language, *Verbal Behavior*
    - argued that language cannot be learned through conditioning

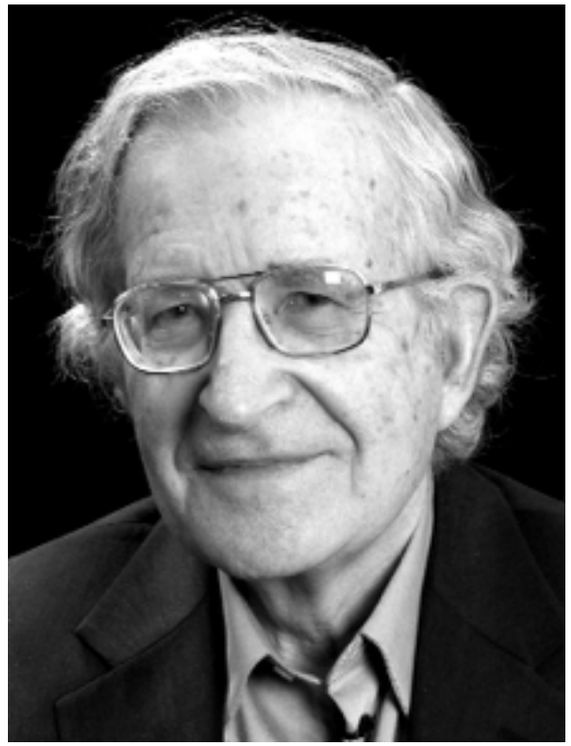




$A_{1,0,1,¢} \rightarrow A_{0,0,1,q} A_{0,0,2,q_0} A_{1,7}$
$A_{1,0,2,a} \rightarrow A_{0,0,1,q} A_{0,0,2,q_0} A_{0,0,3,a} A_{1,7}$
$A_{1,0,3,q_0} \rightarrow A_{0,0,2,q_0} A_{0,0,3,a} A_{0,0,4,b} A_{1,7}$
$A_{1,0,4,b} \rightarrow A_{0,0,3,a} A_{0,0,4,b} A_{0,0,5,\ddagger} A_{1,7}$
$A_{1,0,5,\ddagger} \rightarrow A_{0,0,4,b} A_{0,0,5,\ddagger} A_{1,7}$
$A_{1,1,1,B} \rightarrow A_{0,1,1,B} A_{0,1,2,B} A_{1,7}$
$A_{1,1,2,B} \rightarrow A_{0,1,1,B} A_{0,1,2,B} A_{0,1,3,B} A_{1,7}$
$A_{1,1,3,B} \rightarrow A_{0,1,2,B} A_{0,1,3,B} A_{0,1,4,B} A_{1,7}$
$A_{1,1,4,B} \rightarrow A_{0,1,3,B} A_{0,1,4,B} A_{0,1,5,q_0} A_{1,7}$
$A_{1,1,5,a} \rightarrow A_{0,1,4,B} A_{0,1,5,q_0} A_{0,1,6,B} A_{1,7}$
$A_{1,1,6,q_0} \rightarrow A_{0,1,5,q_0} A_{0,1,6,B} A_{0,1,7,B} A_{1,7}$
$A_{1,1,7,B} \rightarrow A_{0,1,6,B} A_{0,1,7,B} A_{0,1,8,B} A_{1,7}$
$A_{1,1,8,B} \rightarrow A_{0,1,7,B} A_{0,1,8,B} A_{0,1,9,B} A_{1,7}$
$A_{1,1,9,B} \rightarrow A_{0,1,8,B} A_{0,1,9,B} A_{1,7}$

Context-free grammar

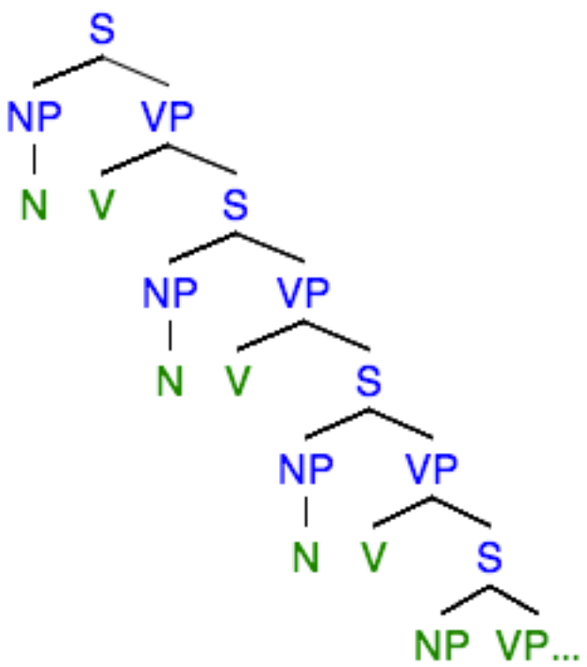
Syntactic productivity



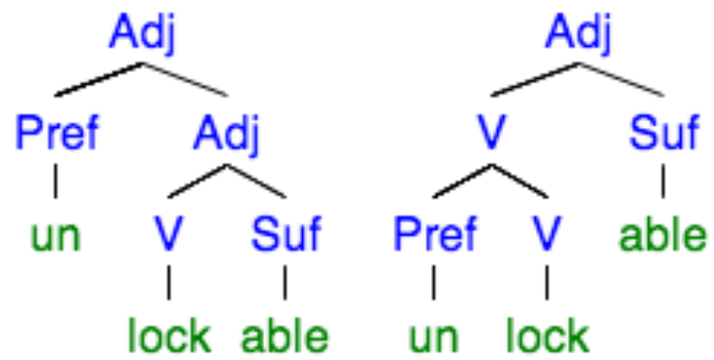
- Noam Chomsky, 1950s
  - anti-behaviorist: argued that language cannot be learned through conditioning
  - 1 of arguments: **productivity**
  - there is an infinite number of possible sentences in any human language

I think (that) Sally knows (that) Kanzi discovered (that) Nim signed that …

- how? recursion in the grammar
  - **S** -> NP VP; VP -> V **S**
  - **NP** -> Det N PP; PP -> P **NP**
- we could not have learned these sentences through conditioning; we create them on-the-fly, using our linguistic building blocks & rules of combination

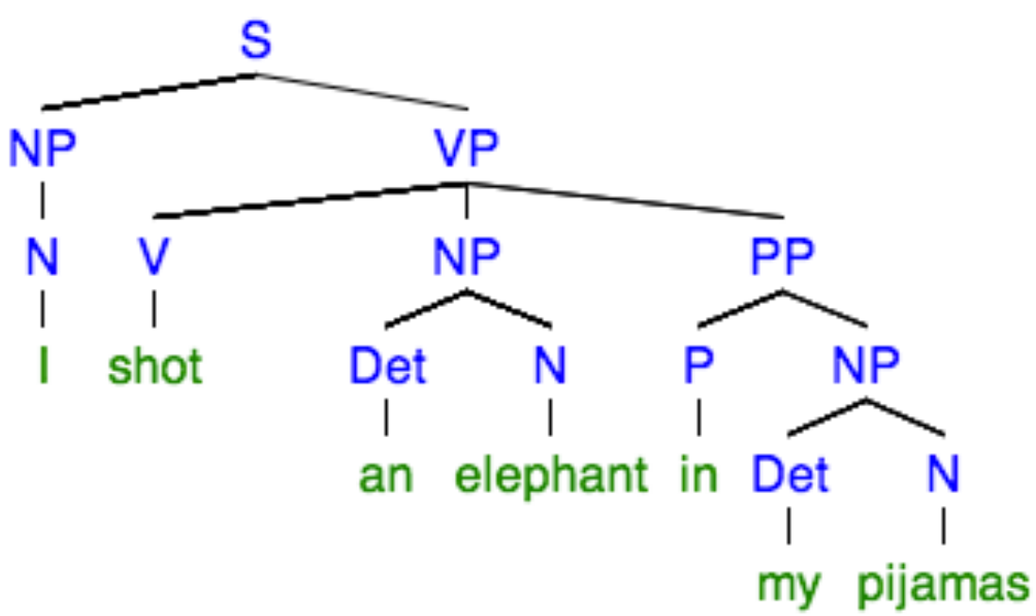


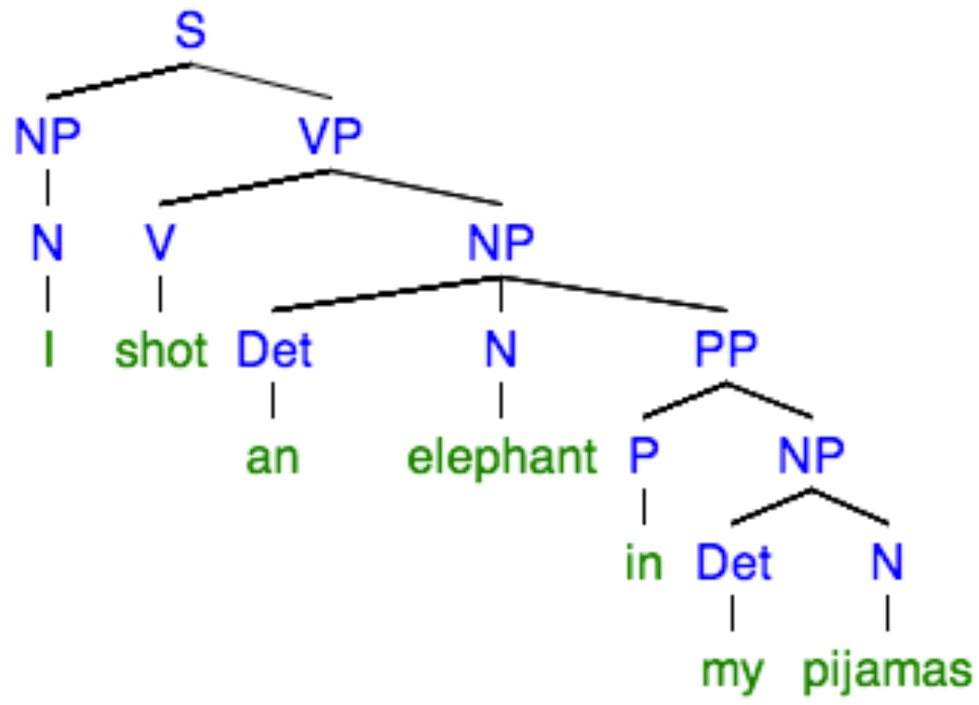
Words & Sentences

	Morphological knowledge (words)	Syntactic knowledge (sentences)
ambiguity		YES [**to be continued**]
productivity?	YES (*retweet, defriend, wugs*)	YES (recursion)

Syntactic ambiguities

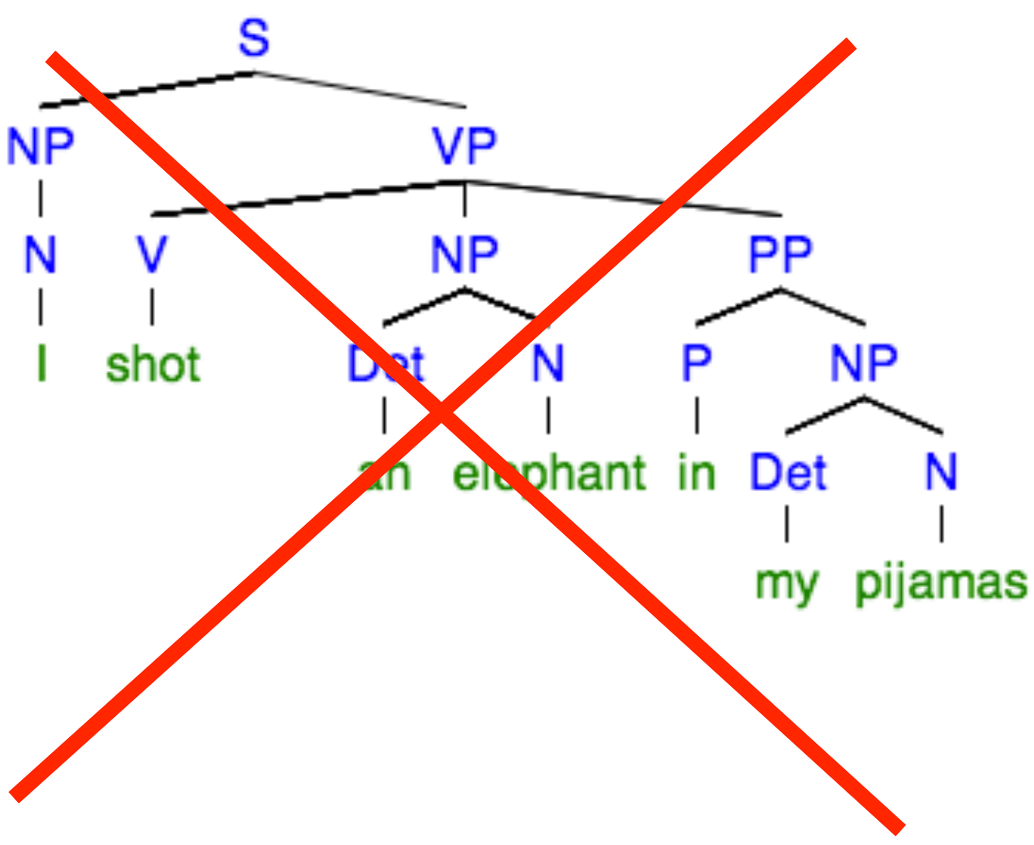
- *Animal Crackers*, Marx Brothers, 1930
  - "One morning, I shot an elephant in my pajamas."

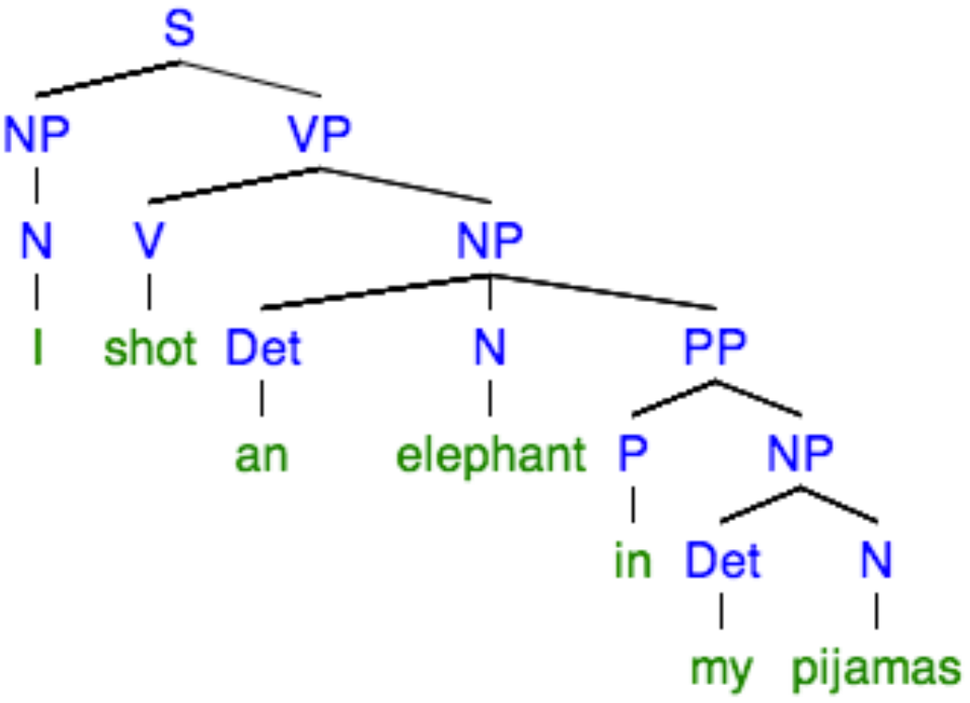




Syntactic ambiguities

- (*Animal Crackers*, Marx Brothers, 1930) Global ambiguity
  - "One morning, I shot an elephant in my pajamas."





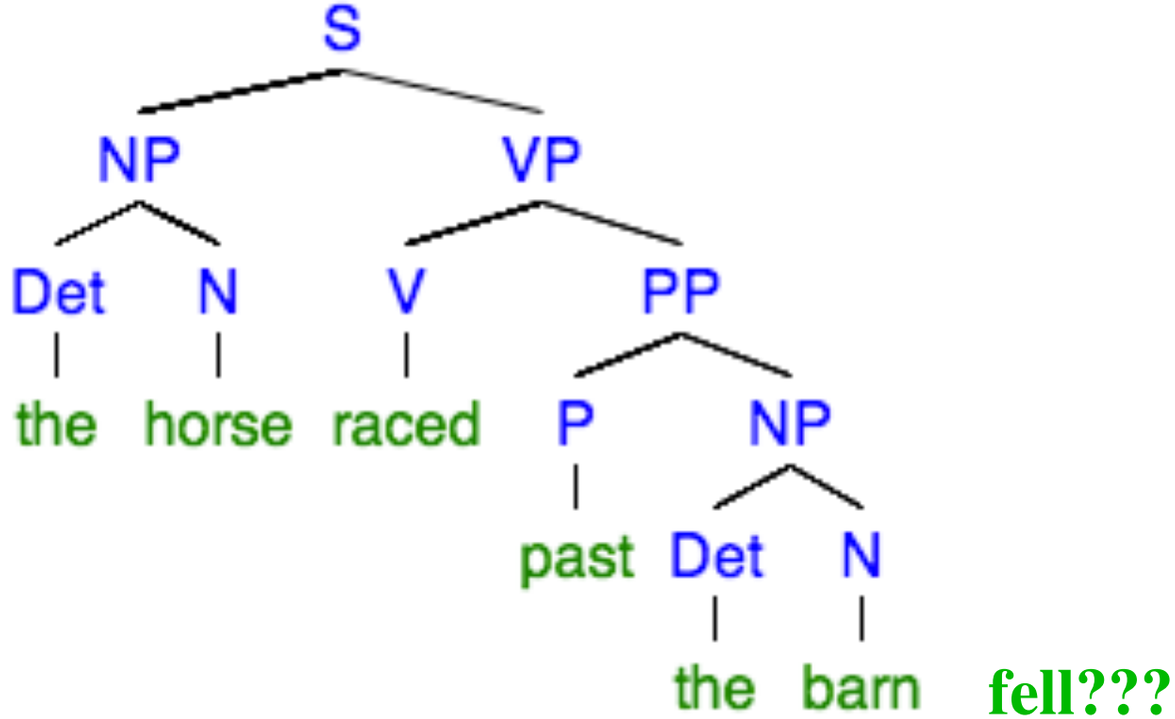
  - "How he got in my pajamas, I don't know."

Syntactic ambiguities

- Temporary ambiguity
  - *The horse raced past the barn fell.*

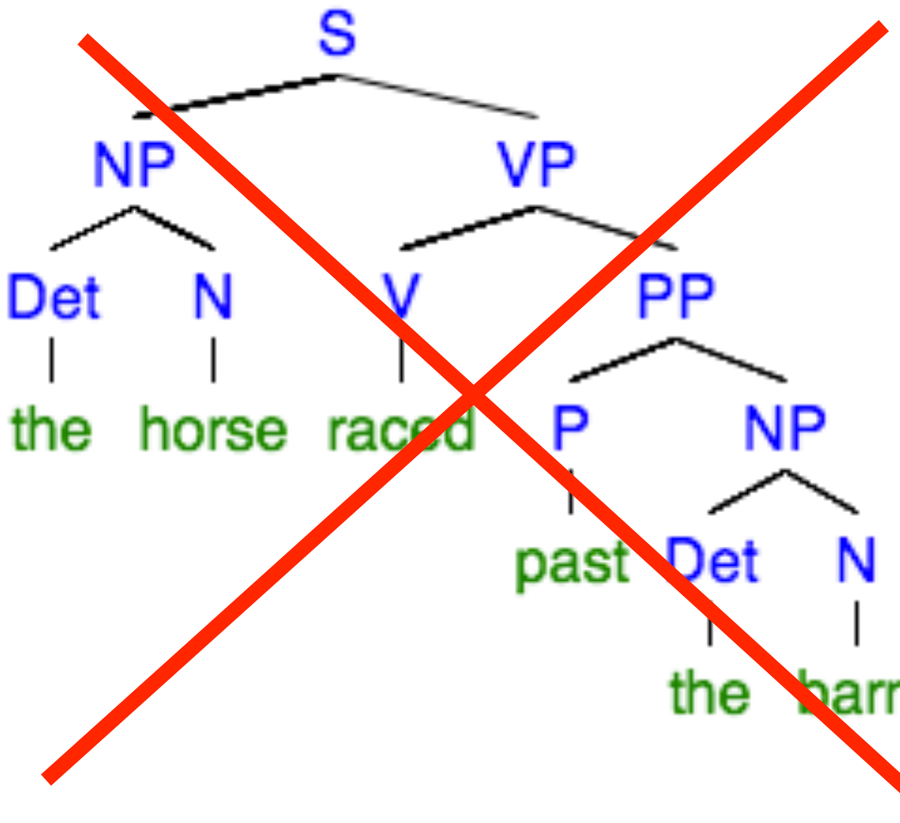
Syntactic ambiguities

- Temporary ambiguity
  - *The horse raced past the barn fell.*

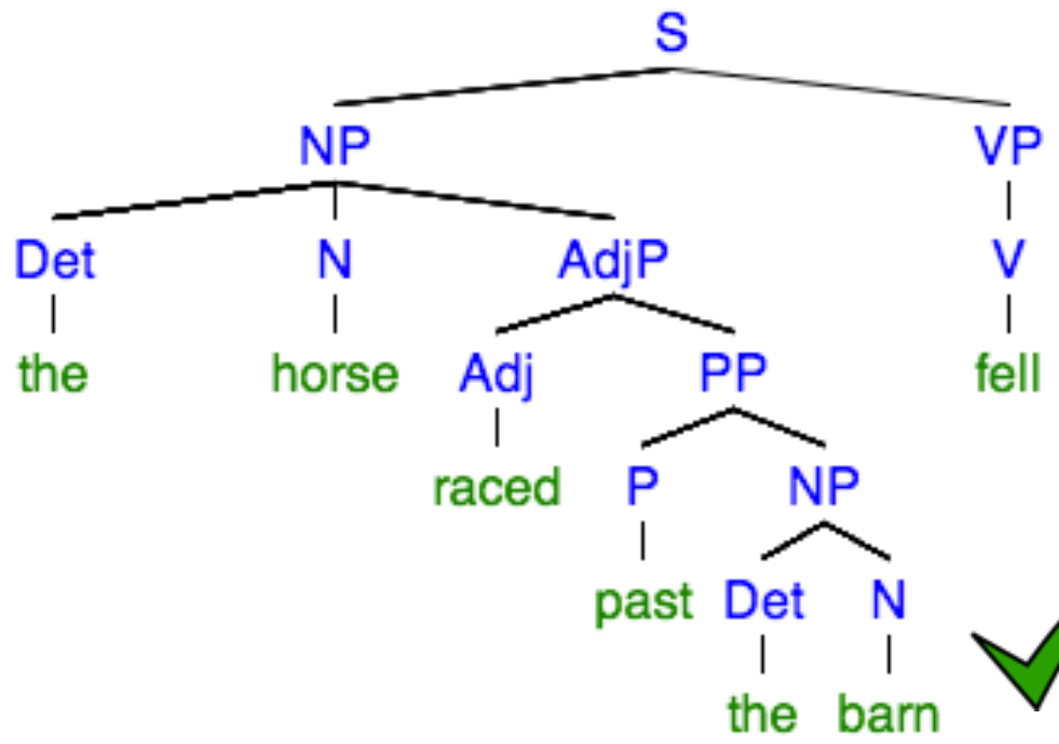


Syntactic ambiguities

- Temporary ambiguity
  - *The horse raced past the barn fell.*



fell???



- *The horse **that was** raced past the barn fell.* -> much easier, not ambiguous, but just as grammatical!

Syntactic ambiguities

- Temporary ambiguity - “garden path” sentences
  - *While Mary was mending a sock …*
  - incremental parsing!
    - VP -> V NP
      - *While [Mary]$_{NP}$ [was mending a sock]$_{VP}$ …*
    - *VP -> V*
      - *While [Mary]$_{NP}$ [was mending]$_{VP}$ [a sock]$_{NP}$ …*

Syntactic ambiguities

- Temporary ambiguity - “garden path” sentences
    - *While Mary was mending a sock …*
    - incremental parsing!
    - VP -> V NP
        - *While [Mary]$_{NP}$ [was mending a sock]$_{VP}$ …fell on the floor :(*
        - *…… Bill called :)*
- *VP -> V*
    - *While [Mary]$_{NP}$ [was mending]$_{VP}$ [a sock]$_{NP}$ … fell on the floor. :)*
    - *…… Bill called. :(*

Syntactic ambiguities

- how does the parser decide which rule to use?

- option 1 - choose a rule based on some **syntactic heuristic**, go with that option until you're done / it fails (modular! syntax only)

    - e.g. heuristic: "Right association" - If possible, attach new items into the clause or phrase postulated most recently.

        - here, attach *a sock* to the VP of *mended*

        - if *fell on the floor* follows, parser fails & reanalyzes => slowing down

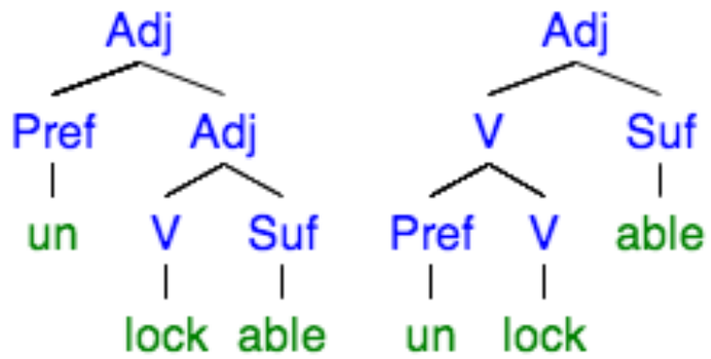
- option 2 - activate **all possible analyses to varying degrees**, depending on how much support they receive from various sources of information (syntax, morphology, semantics, intonation, general plausibility etc => not modular)

    - e.g. *The spy saw the cop with the revolver*. <— alternative interpretation is semantically odd, so does not get highly activated

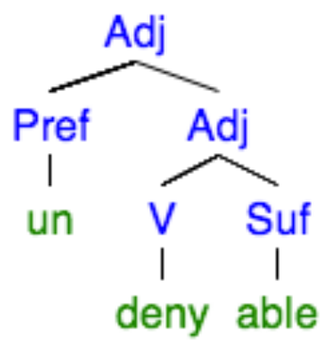
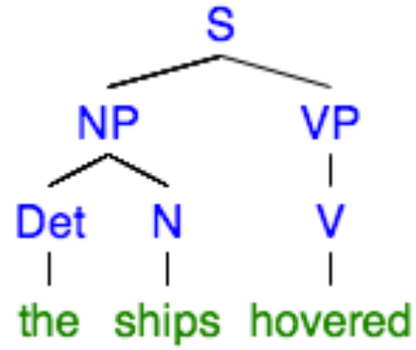
    - competing analyses => slowing down

- middle road: syntactic preferences exist (as in option 1), but they interact with other aspects of the speech stream

Words & Sentences

	Morphological knowledge (words)	Syntactic knowledge (sentences)
ambiguity		YES (garden path sentences)
productivity?	YES (*retweet, defriend, wugs*)	YES (recursion)

Words & Sentences

	Morphological knowledge (words)	Syntactic knowledge (sentences)
building blocks	morphemes (stems & affixes)	parts of speech
rules of combination	Adj + *ity* > N *re* + V > V …	N -> (Det) (Adj) N (PP) VP -> V (NP) (PP) (S) P -> P NP S -> NP VP …
(we can represent these structures using) trees		

Language

- One of our most complex cognitive functions
- Uniquely human & inevitable
- Levels of analysis
  - Linguists - characterize linguistic knowledge
    - morphological & syntactic knowledge
  - Psycholinguists - determine the algorithms that implement this knowledge
    - mental lexicon (representations)
    - online sentence comprehension (modularity)
  - Neurolinguists - investigate neural mechanisms that realize these algorithms