Grammar & Learnability

Tom Wasow
Sym Sys 100
April 22, 2008

Language

- Universal across the species

- Acquired even w/o conscious instruction

- Unique to humans

- Hence, seen as key to human nature

Language vs. Animal Communication

Language	Animal Communication
Infinite combinations	Finite repertoire
About anything at all	About here and now
Express abstractions	No abstractions

How Is Language Acquired?

- Some imitation involved

- Comprehension goes beyond imitation

  - Novel utterances are understood

  - Parrots don't understand

- What makes it possible for humans? (but not for animals)

Quine on Language Acquisition

The operant act may be the random babbling of something like 'Mama' at some moment when, by coincidence, the mother's face is looming. The mother, being pleased at being named, rewards this random act, and so in the future the approach of the mother's face succeeds as a stimulus for further utterances of 'Mama'....

Quine, continued

We have imagined a child learning to send 'Mama', and also learning to parrot the word on hearing it, but we have not considered intelligent hearing. What would count as an intelligent response to the heard word 'Mama'...?...Perhaps something rather like this: the child hears 'Mama'...while sensing the mother in the periphery of his visual field, and then turns to the mother...

Quine, continued

Eventually the child becomes amenable to suggestion also in the initial utterance of new words. Mimicry... develops to the point where any new utterance from someone else becomes a direct stimulus for a duplicate. Once the child reaches this stage,...with little or no deliberate encouragement on the part of his elders, he proceeds to amass language hand over fist.

word&
object

WILLARD VAN ORMAN QUINE
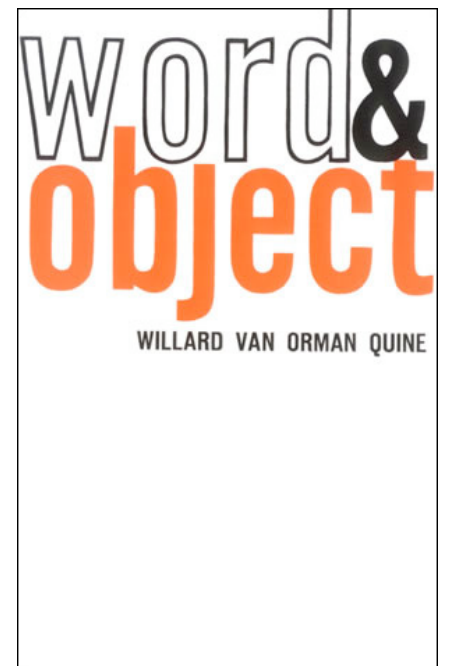
What's Wrong With This Story?

- Doesn't allow for language about things not here now

- Doesn't explain language about abstractions

- Doesn't explain novel utterances
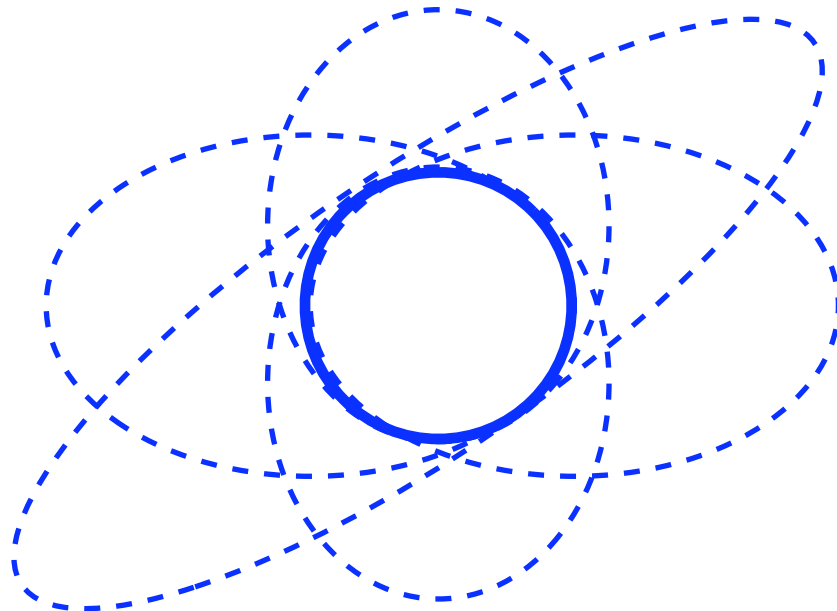
- Language acquisition is not just learning words

But People Believed It 50 Years Ago

...when Chomsky reviewed Skinner's *Verbal Behavior*

"Poverty of the Stimulus" Argument

- Versions differ in specificity

- Most general: just a case of the problem of induction: For any finite set of data, there are infinitely many logically possible generalizations consistent with it.

"Poverty of the Stimulus" Argument

- Versions differ in specificity

- Most general: just a case of the problem of induction: For any finite set of data, there are infinitely many logically possible generalizations consistent with it.

- Linguistic experience is finite

- Languages are infinite

- How do people figure out which of the infinitely many possible languages they might be hearing is the right one?

Digression: Infinitude of Languages

- Why claim languages are infinite?

- Answer: No bound on length of sentences.

Some sentences go on and on and on and on and on...

This is the maiden all forlorn,
that milked the cow with the crumpled horn,
that tossed the dog,
that worried the cat,
that killed the rat,
that ate the malt
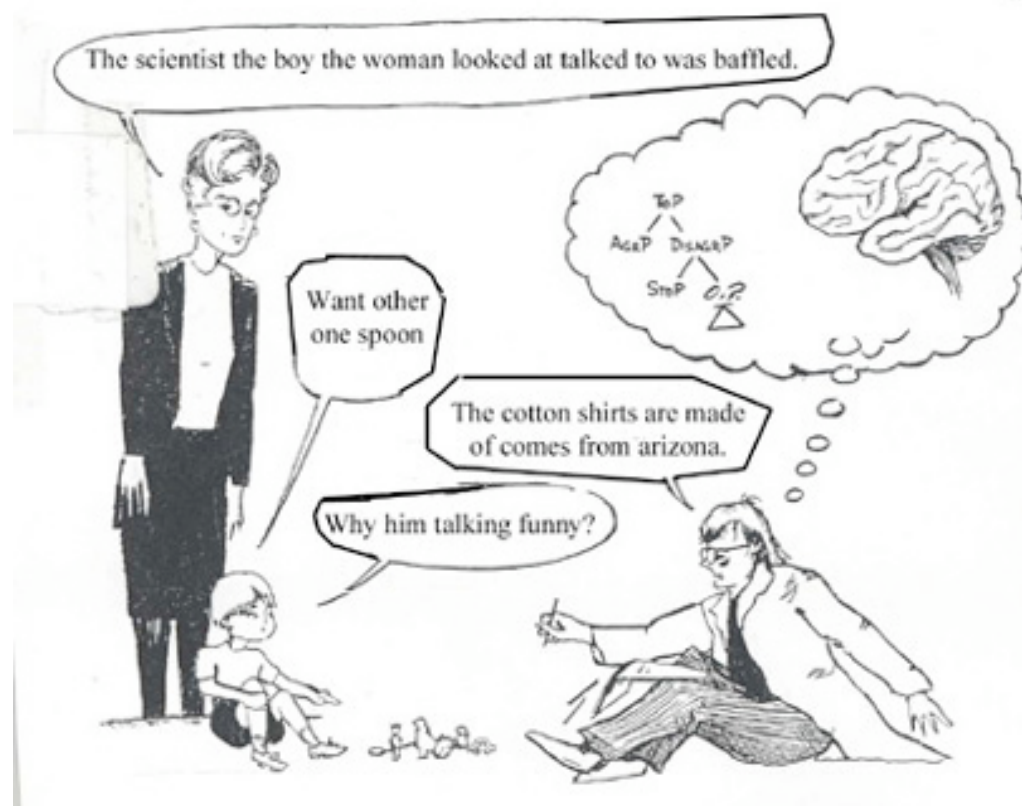*that lay in the house*
that Jack built.
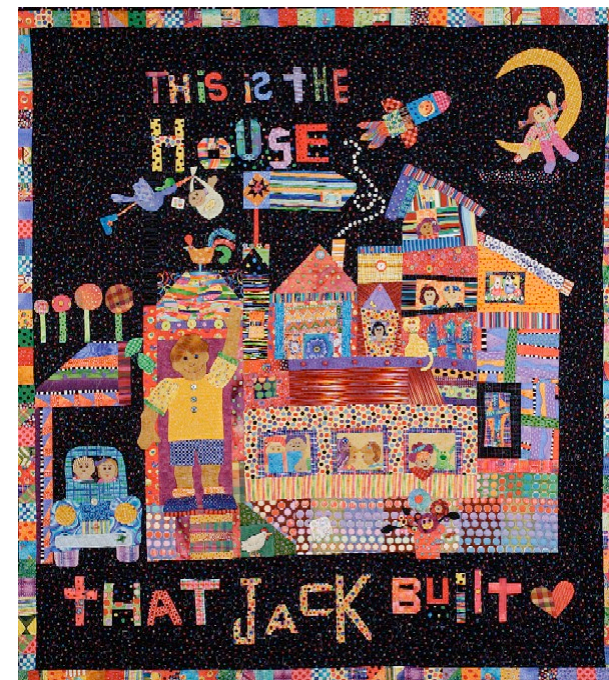
Chomsky's Solution to POS

- Much of knowledge of language is innate

- Humans have a species-specific, task-specific "mental organ" for language

- Its properties must manifest themselves in all languages
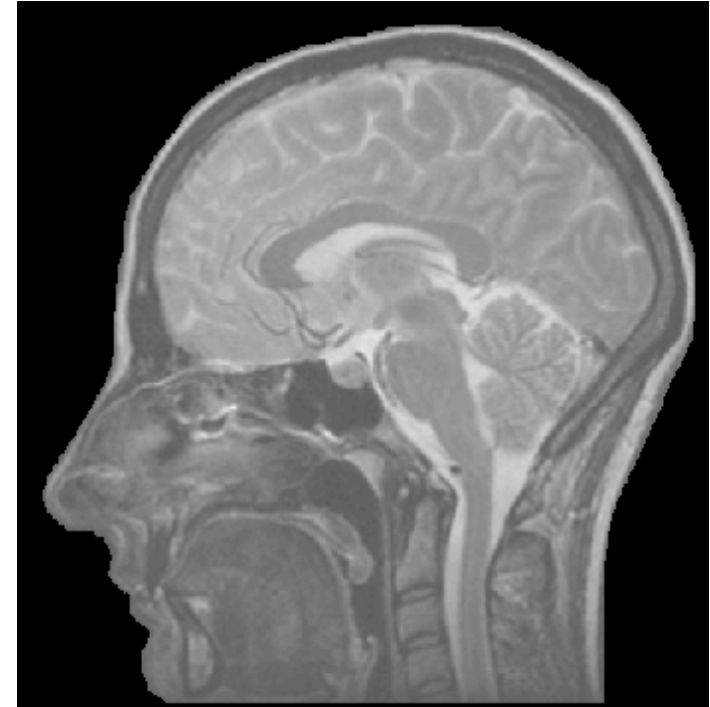
More Detailed Version of POS

- Languages have very complex structures
- Children learn them with little apparent effort
- The input to children is noisy
- Children get little if any explicit correction ("no negative evidence")
- Yet native speakers have robust intuitions separating well-formed from ill-formed sentences

The scientist the boy the woman looked at talked to was baffled.

Want other one spoon

The cotton shirts are made of comes from arizona.

Why him talking funny?

ToP
AgrP   DiskrP
StoP   O.?

The Premises: Complex Structure

- Learning a language involves not just vocabulary, but word order, case marking, agreement, etc., with a mix of very general facts and rather specific idiosyncrasies

- Learners beyond a certain age never get these entirely right

- Nobody has ever built a system that can correctly predict which strings of words will be judged acceptable sentences by native speakers of any natural language

Examples of Grammatical Complexity

- Verbs come at beginnings of clauses in Irish, at the ends of clauses in Japanese, and in the middle in English

- Some verbs in some languages require special case-marking, e.g. the Icelandic word for 'lack' requires an accusative subject, though subjects are usually nominative

- English expresses present time events with *be* plus a verb ending in *-ing*, rather than simple present: *I am speaking*, not *I speak*.
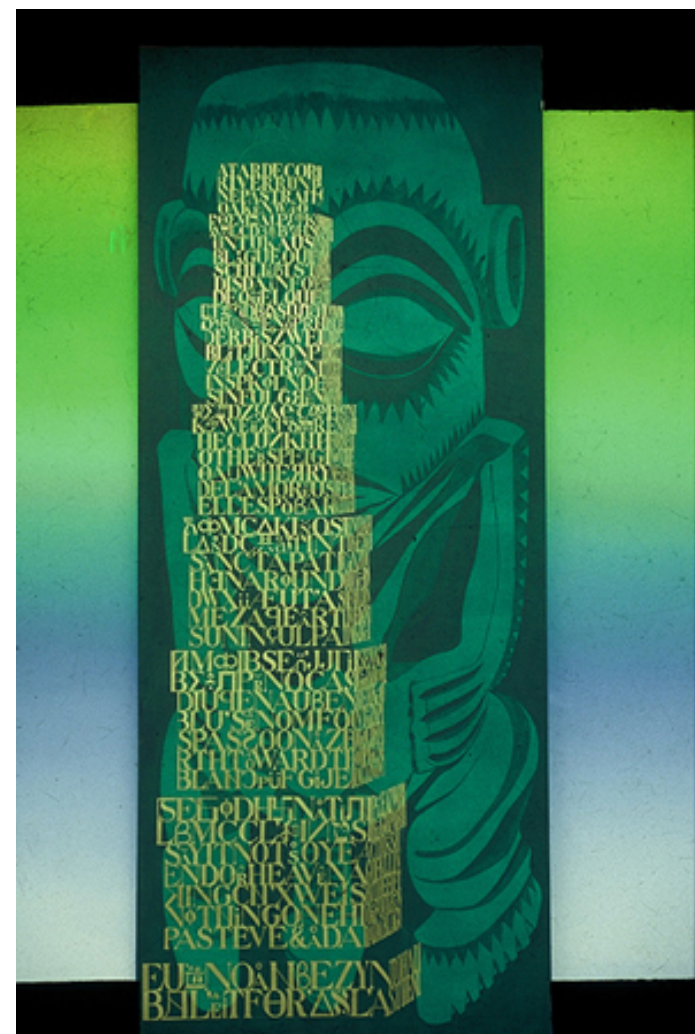
The Premises: Ease of Learning

- Children acquire productive language even before they begin formal schooling

- Other skills that look simpler (from an engineering perspective) require explicit instruction and years of practice, e.g. arithmetic, chess

- Anyone who fails to learn language is considered seriously impaired

The Premises: Noisy Input

- Natural speech is full of *uh*s, *um*s, repeated words, false starts, and unintentional errors, e.g.

  And, um, I, I, I think that what one thing that they were concerned probably was the fact it wasn't necessarily, you know, like the quantity of care but the quality of, care

The Premises: Noisy Input

- Natural speech is full of *uh*s, *um*s, repeated words, false starts, and unintentional errors, e.g.

  And, um, I, I, I think that what one thing that they were concerned [with] *probably was the fact* [that] *it wasn't necessarily, you know, like the quantity of care but the quality of, care*

The Premises: Noisy Input

- Natural speech is full of *uh*s, *um*s, repeated words, false starts, and unintentional errors, e.g.

  And, um, I, I, I think that what one thing that they were concerned [with] probably was the fact [that] it wasn't necessarily, you know, like the quantity of care but the quality of, care

- Some complex structures are extremely rare:

  What mail do you throw away___ without reading___?

The Premises: No Negative Evidence

- Parents reportedly correct for content much more than for structure

- When corrected, children (supposedly) ignore it

  Child: *Nobody don't like me.*

  Mother: *No, say, "Nobody likes me".*

  Child: *Nobody don't like me.* [Dialog repeated 8 times]

  Mother: *No, now listen carefully. Say, "Nobody likes me".*

  Child: *Oh! Nobody don't likes me.*

- Premise controversial: negative evidence might be implicit

The Premises: Robust Intuitions

- Examples from English

  Pat is likely to complain   vs.   **Pat is probable to complain*

  Which team do they think is strongest?   vs.
  *Which team do they think are strongest?

  She is never satisfied with herself.   vs.
  She is never satsified with her.   [*She≠her*]   vs.
  *She is bringing a friend with herself.   vs.
  She is bringing a friend with her.   [*She=her*]

The "Parade Case" of POS

- Consider:

  Pat is sleeping ~ Is Pat sleeping?

  Pat has slept ~ Has Pat slept?

  Pat could sleep ~ Could Pat sleep?

  Pat slept ~ Did Pat sleep?

- Generalization: To make an English yes-no question, move the auxiliary verb in the corresponding declarative to the front of the sentence; if there is none, use the appropriate form of *do* (with the main verb in base form)

The "Parade Case", continued

- What happens when there are two auxiliary verbs?

  Pat has been sleeping. ∼ *Has Pat been sleeping?*
  NOT **Been Pat has sleeping?*

- Suggests rule is to front the **first** auxiliary verb.

- But this runs afoul of

  The dog that was barking is sleeping ∼
  Is the dog that was barking____sleeping?
  NOT **Was the dog that____ barking is sleeping?*

- The rule must make reference to **structure**: front the first auxiliary verb **of the main clause**.

The "Parade Case", continued

- Chomsky claims people could go a lifetime w/o encountering the crucial data (examples like *Is the dog that was barking sleeping?*), though corpus searches cast doubt on this.

- Crain & Nakayama (1987) had children convert declaratives to questions. They made errors, but never the error of adopting the structure-independent rule.

- Ambridge, Rowland, & Pine (2008) varied the experiment slightly and got what look like structure-independent errors.

Why the "Parade Case" Matters

- Chomsky claims the structure dependence of the yes-no question rule is innate knowledge
- The search space of possible generalizations from the data is narrowed by saying the rule must make reference to linguistic structure (phrases, not just word strings)
- He claims no language has any structure-independent rule
- Various critics have developed models of how yes-no questions could be learned w/o language-specific innate knowledge

Relevant Mathematical Result

- Much-cited paper by Gold (1967) developed a mathematical model of language learnability
- It ostensibly shows that interesting types of languages are not learnable from positive evidence alone, unless one assumes very strong constraints on possible languages (à la Chomsky)
- Gold's models are highly idealized, so relevance to real languages is debatable.
- I suspect most people who cite Gold haven't read the paper.

The Gold Paradigm

- A language is a set of finite strings of symbols, drawn from some fixed inventory (say, ascii)
- Sentences from a language are presented to a learner, which has to guess what language is being presented.
- The learner is an algorithm for guessing a language on the basis of the sentences presented.
- The sentences of the language can be presented in any order, so long as, at any point in the presentation, every sentence will be presented again at some future point. (That is, every sentence is presented infinitely often)

The Gold Paradigm, continued

- A learner "identifies a language in the limit" if, on every presentation of the sentences of that language, the learner picks the language being presented after a finite number of sentences and never changes its guess again

- Learnability is a property of **classes** of languages: if there is some algorithm that can identify-in-the-limit every language in the class, then the class is learnable.

The Famous Gold Theorem

- Any class of languages containing all finite languages and at least one infinite language is unlearnable.

- The proof relies on picking a presentation of the sentences of the infinite language in such a way as to fool the learner.

  - Since the learner must be able to identify each finite language in the limit, keep presenting the same sentences over and over until the learner guesses that these are the only sentences.

  - Then add a new sentence. The learner will need to revise its guesses.

  - In presenting the infinite language, this can be done infinitely often, so the infinite language won't be identifiable in the limit

Some Comments on Gold's Theorem

- All familiar classes of formal languages (e.g. finite-state, context-free,...) contain all finite languages and some infinite languages
- The proof will generalize to any class of languages that has an infinite sequence of languages, each of which is a proper subset of the next
- Gold's paper also had some positive learnability results, including one for presentation of both positive and negative evidence (i.e. strings marked as either in the language or not in the language)

Gold's Paradigm & Human Languages

- Some sentences in natural languages will never be uttered (because they're too long, or because their meanings are weird)

- We don't know what would happen to a child whose parents deliberately tried to fool it about what was in their language

- Children learning languages aren't just getting strings of words: sentences have **meanings**

- Why assume that the language children learn is **identical** to the language their parents speak?

Language Variation

- No two people speak exactly the same way

- Languages are constantly changing (even the language of a given individual)

- Children hear language from many other people, so the language they acquire won't be exactly the language they are exposed to

- So Gold's criterion of success -- picking out **exactly** the language being presented -- doesn't fit the real world of children

A More Realistic Approach

- Suppose the child's task is not to replicate the ambient language perfectly, but to acquire some approximation to it

- Given a metric of similarity between languages, success in learning a language can be defined as converging on a language suitably close to it

- On this criterion, many more classes of languages (including most familiar types of formal languages) are learnable

Another Alternative

- Suppose the learner assigns probabilities to strings of words (or to strings paired with possible meanings), instead of just guessing they're sentences or non-sentences

- The learning process involves adjusting probabilities, based on what the learner hears

- Success defined in terms of limiting fluctuations in the probability distribution

- Formalizations along these lines also escape the negative results of Gold's theorem

Status of POS Argument

- Chomsky and his followers continue to use it

- Almost every aspect of it is highly controversial

  - Does language learning require internalizing a grammar?

  - Is input really so noisy?

  - Do children get negative evidence?

  - Is there an end point to language acquisition?

What About Innateness Claims?

- Chomsky's claim that humans have a highly specialized "language organ" might be true, even if the POS is unsound

- Other arguments put forward for it

  - Species specificity

  - Language universals

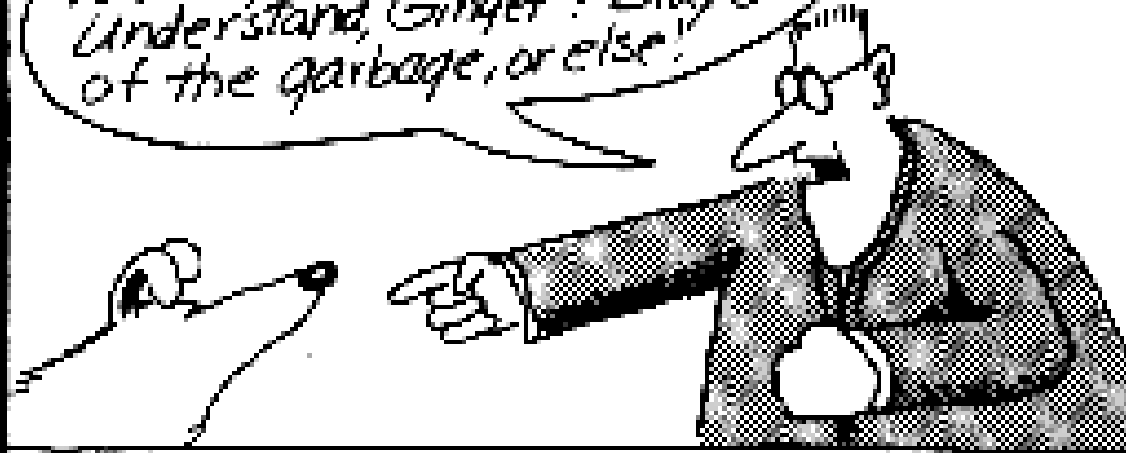 - Critical age for learning

- All controversial

Species Specificity

- Even highly intelligent non-human animals can't learn (much of) human languages

- Suggests specialized mechanism unique to humans

- Frequent claims to the contrary largely unconvincing
  - Focus on number of words
  - Little attention to syntax or range of meanings
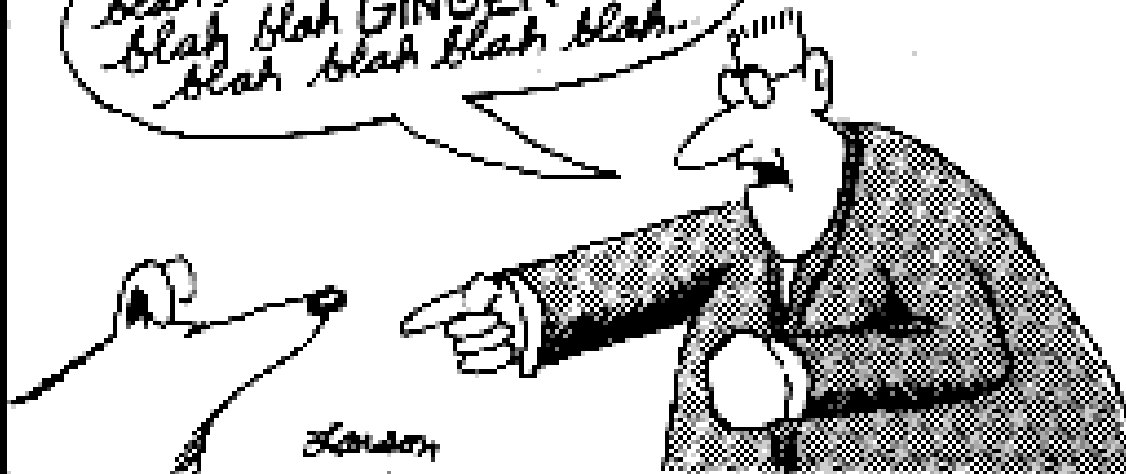
- Possible exception: the bonobo Kanzi

Language Universals

- Innate language organ entails language universals

- Many proposed in recent decades

- Most are statistical tendencies (not exceptionless) -- e.g. few languages have a dominant word order with objects before subjects

- Many reflect cognitive strategies not specific to language, e.g. long phrases tend to follow short phrases

Critical Age

- Most language learning takes place before about age 13
- People learning a language after that age never attain full mastery
- Includes people deprived of language in early years (notably, many deaf people)

- Many other skills are learned better later, suggesting a specialized ability, lost around puberty

Concluding Remarks

- Nobody doubts that innate abilities are crucial to language acquisition

- Chomsky's position is controversial because it posits a mechanism that is task-specific and species-specific

- His central argument, POS, doesn't hold up very well under close scrutiny

- Other arguments are also controversial

- In short, the jury is still out.