Language Acquisition

Michael Thomas
Birkbeck College

The power of syntax

had had had had had had had had had had had had

The power of syntax

had had had
had had had had had had had had had
had

The power of syntax

- John and Peter both took the English exam John where Peter had had had had had had had had had had had appeared in his answer then John would have received full marks

---

The power of syntax

- John and Peter both took the English exam. John, where Peter had had "had had had", had had "had had". Had "had had had" appeared in his answer, then John would have received full marks.

The power of syntax

The power of syntax

had had had had had had
had had had had had had
had

Lexicon
(grammatical class)

Context

Syntax

Prosody

John and Peter both took the English exam. John where Peter had had had had had had had had had had appeared in his answer than John would have received full marks.
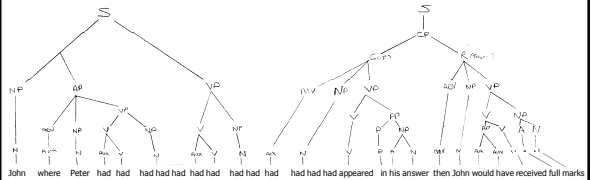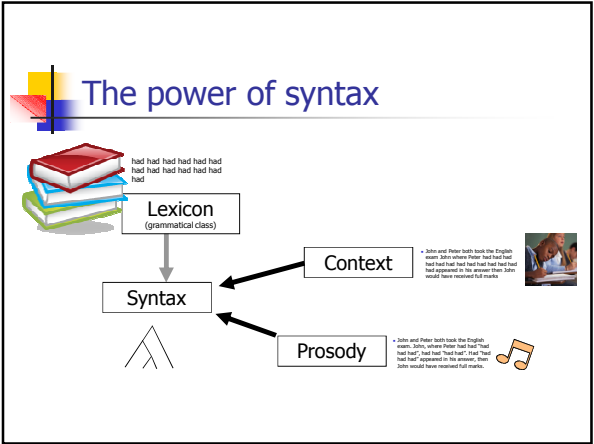
John and Peter both took the English exam. John, where Peter had had "had had had", had had "had had". Had "had had" appeared in his answer, then John would have received full marks.

The question

- How is this all learned . . . . *by age 5* ?

Language is innate, innit?

- Infants show precocious sensitivity to human speech sounds
- Word meanings can't be learnt without constraints on what labels refer to
- Syntax is uniquely human
- Adults don't teach children language
- Children don't simply imitate adults (certainly with respect to syntax)
- All normal children learn language when exposed to it in a normal language environment
- Language input seems too impoverished to explain fast acquisition of abstract information
- Children in impoverished linguistic environments (deaf parents, pidgin) re-invent syntax

---

Children's language errors

Rule-learning?

(1) Child: Doggie [pointing at a horse]
Adult: No, that's a horsie [stressed]/

(2) Adult: Say "Tur"
Child: Tur
Adult: Say "Tle"
Child: Tle
Adult: Say "Turtle"
Child: Kurka

Over-extension
Labelling

Resistance to correction

(3) Child: Mama isn't boy, he a girl.
Adult: That's right

Lack of correction (truth more important)

(4) Child: My teacher holded the rabbits and we patted them.
Adult: Did you say teacher held the baby rabbits:
Child: Yes.
Adult: What did you say she did?
Child: She holded the baby rabbits and we patted them.
Adult: Did you say she held them tightly?
Child: No, she holded them loosely.

(5) Adult: He's going out.
Child: He go out.
Adult: Adam, say what I say: Where can I put them?
Child: Where I can put them?

Course of early language development

Vegetative sounds	0-6 weeks
Cooing	6 weeks
Laughter	16 weeks
Vocal play	16 weeks – 6 months
Babbling	6-10 months
Single word utterances	10-18 months
Two-word utterances	18 months
Telegraphic speech	2 years
Full sentences	2 years 6 months

What has to be learned?

- Sources of knowledge required to use language:
  - **phonology** (the sounds words are made up from)
  - **semantics** (individual words and their meanings)
  - **syntax** (combinations of words)
  - **pragmatics** (how to use language in a social setting)

Phonology

Innate language perception?

- Sucking habituation technique with infants
- Infants as young as 1 month old can distinguish between two syllables that differ in only one distinctive phonological feature (ba-pa) (Eimas et al, 1971)
- Moreover, perception is <u>categorical</u>
- So can chinchillas, a type of South American rodent (Kuhl, 1981)

Language experience affects phonological processing

Newborn to 4 months	- Ability to discriminate own from non-native language - Preference for mother's voice - Preference for a story of song heard prenatally
6 months	- Some decline in ability to make non-native sound discriminations - Preference for pauses at clause boundaries in native language only
10 months	- No longer able to discriminate non-native phonemic difference (though brain knows the difference long after...) - Preference for words with common native stress pattern

---

Semantics

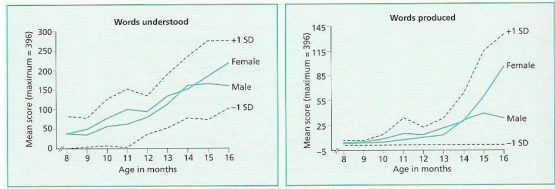
Early semantic development

- Clark & Clark (1977):
  - Mother pointed out and named a dog "bow-wow".
  - Child later applies "bow-wow" to dogs, but also to cats, cows, and horses

  - Mother says sternly to child: "Young man, you did that on purpose"
  - When asked later what "on purpose" means, child says: "It means you're looking at me"

Early semantic development

- Mapping problem
- Constraints:
  - Whole object assumptions
  - Taxonomic constraint
  - Mutual exclusivity assumption
  - Novel name – nameless category
- Joint attention
- Names used at basic level first (dog, not terrier or animal)
- Over-extension and under-extension in both comprehension and production
- Vocabulary explosion (slow up to 50-100 words, 18-24 months rapid acceleration)
- Comprehension-production asymmetry
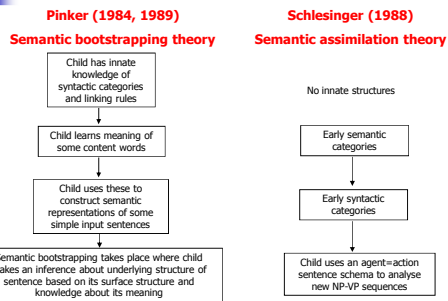- Girls better than boys

"DOG"

Comprehension vs. production (parental ratings)

Syntax

How do children learn syntactic categories?

Pinker (1984, 1989)
Semantic bootstrapping theory

- Child has innate knowledge of syntactic categories and linking rules
- Child learns meaning of some content words
- Child uses these to construct semantic representations of some simple input sentences
- Semantic bootstrapping takes place where child makes an inference about underlying structure of sentence based on its surface structure and knowledge about its meaning

Schlesinger (1988)
Semantic assimilation theory

- No innate structures
- Early semantic categories
- Early syntactic categories
- Child uses an agent=action sentence schema to analyse new NP-VP sequences

---

Syntactic categories = semantic categories? Problems…

- Not all verbs refer to actions, some refer to states
  - need, see
- Many adjective refer to states
  - hungry, nice
- If verbdom is based on semantic notion of actions and states, might expect mistakes with adjectives
  - "I hungries". Never found…

Chomskian approach to language acquisition

Language acquisition as setting the Parameters in Universal Principles

- Pro-drop

- In your language, can you drop the pronoun?
  - Italian: *parla* (speaks)   (see also Arabic)
  - English: *he speaks*   (see also French)
- Once pro-drop is set, generalises to other constructions
  - Italian: *cade la pioggia* (falls the rain)
  - English: *the rain falls*

- Pro-drop is a generalisation about how languages work

Exceptions to the rule

- Writing in your diary
  - 'Went to the shops.'

Subject	Object	Verb	44%
Subject	Verb	Object	35%
Verb	Subject	Object	19%
Verb	Object	Subject	2%
Object	Verb	Subject	0%
Object	Subject	Verb	0%

Different word orders, as percentages of the world's languages

- Understanding Yoda does not blow our minds
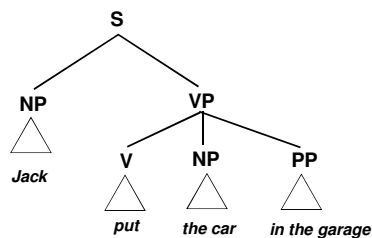 - 'Growing stronger the Dark Side is' (OSV)

- Rule-based theories struggle with exceptions
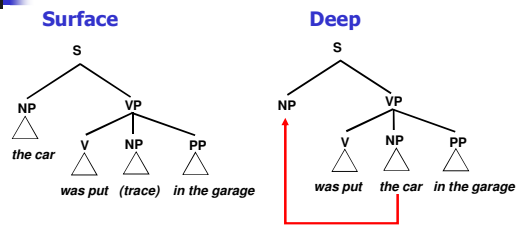

Intuitions about learnability

- "Most linguistically relevant properties are abstract, pertaining to phrase structure configurations, syntactic categories, grammatical relations … But these abstract properties are just the ones that the child cannot detect in the input prior to learning… The properties that the child can detect in the input – such as the serial positions and adjacency and co-occurrence relations among words – are in general linguistically irrelevant"

  (Pinker, 1984, pp. 49-50)

A phrase structure tree

```
              S
           /     \
         NP        VP
         /       /  |  \
       Jack     V   NP   PP
               put the car in the garage
```

Surface vs. Deep Structure

Surface

```
            S
         /     \
       NP        VP
       |       / | \
    the car   V  NP  PP
           was put (trace) in the garage
```

Deep

```
            S
         /     \
       NP        VP
              /  |  \
             V   NP   PP
          was put the car in the garage
```

- Surface structure captures relationships between sentence elements as they are directly expressed by the sentence
- A "trace" shows that the NP that serves as the object for put has been placed elsewhere
- The deep structure captures the relationships between sentence elements in terms of the general phrase structure rules, independent of whether the sentence is in the active or passive voice

---

Is the trace a linguistic invention or is it real?

- When is **contraction** permissible?
  - *Want to => wanna, You are => you're, I am => I'm*
- Claim – *not if it spans a trace*
  - I want to visit Fred
  - I wanna visit Fred
  - Who do you want to visit?
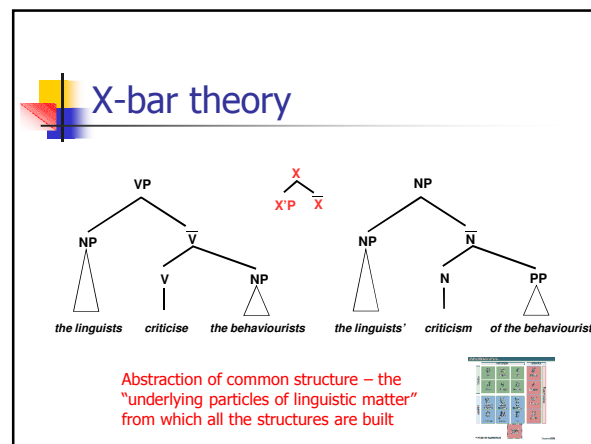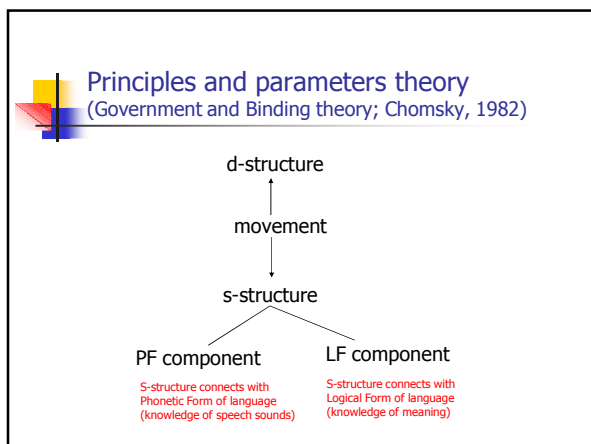 - Who do you wanna visit? (acceptable)

  - I want Jim to visit Fred
  - I wanna Jim visit Fred (not acceptable) – *no words allowed between want and to?*
  - Who do you want to visit Fred?
  - Who do you wanna visit Fred? (not acceptable)

The sort of things linguists say

- 'No one taught you that you couldn't say "Who do you wanna visit Fred?". In fact, if anyone ever said anything about *wanna* to you at all it was to tell you not to use it, period. Nevertheless, you have the ability to make subtle judgements concerning its distribution. How did that happen.....'

- See http://www.princeton.edu/~browning/index.shtml for an example

Principles and parameters theory
(Government and Binding theory; Chomsky, 1982)

d-structure

↑

movement

↑

s-structure

PF component | LF component

S-structure connects with Phonetic Form of language (knowledge of speech sounds)

S-structure connects with Logical Form of language (knowledge of meaning)

X-bar theory

VP
- NP — the linguists
- V̄
  - V — criticise
  - NP — the behaviourists

X
- X'P
- X

NP
- NP — the linguists'
- N̄
  - N — criticism
  - PP — of the behaviourists

Abstraction of common structure – the "underlying particles of linguistic matter" from which all the structures are built

---

The Minimalist program in syntax in 1990s
(review article by Lasnik, 2002, TICS)

Chomsky put forward the audaciously Minimalist conjecture that the human language faculty might be a <u>computationally perfect solution</u> to the problem of relating sound and meaning, the minimal computational system given the boundary conditions provided by other modules of the mind...

The 'last resort' nature of syntactic movement
From its inception in the early 1990s, Minimalism has insisted on the 'last resort' nature of movement: Movement must happen for formal reason. The CASE FILTER, which was a central component of the GB system, was thought to provide one such driving force. A standard example involves 'subject raising'.

(4) John is certain [*t* to fail the exam]
(5) It is certain [that John will fail the exam]

In (4), as in (5), *John* is the understood subject of *fail the exam*. This fact is captured by deriving (4) from an underlying structure much like that of (5), except with an infinitival embedded sentence instead of a finite one:

(6) ___ is certain [John to fail the exam]

John in (6) is not in a position appropriate to any Case. By raising to the higher subject position, it can avoid a violation of the Case Filter, because the raised position is one where nominative case is licensed. But if the Case requirement of *John* provides the driving force for movement, the requirement will not be satisfied immediately upon the introduction of that nominal expression into the structure. Rather, satisfaction must wait until the next CYCLE, when a higher layer of structure is built, or, in fact, until an unlimited number of cycles later, as raising configurations can iteration:

(7) John seems [to be likely] [to fail the exam]]

Minimalism

…A Minimalist perspective favours an alterative where the driving force for movement can be satisfied immediately. Suppose that INFL has a feature that must be checked again the NP. Then as soon as that head has been introduced into the structure, it 'attracts' the NP or DP that will check its feature. Movement is then seen form the point of view of the target rather the moving item itself. The Case of the NP does get checked as a result of the movement, but that is simply a beneficial side effect of the satisfaction of the requirement of the attractor.

In an elegant metaphor, Uriagereka likens the attractor to a virus. Immediately upon its introduction into the body, it is dealt with (by the production of antibodies in the case of physical viruses, by movement to check the 'viral' feature in the syntactic case). The earlier Minimalist approach to the driving force of movement was called 'Greed' by Chomsky. This later one developed out of what I have called 'Enlightened Self Interest.'

Essential terminology

- **CASE**
  Case theory posits that the case distinctions (nominative, accusative, etc.) morphologically manifested in languages like Latin and Russian are present on nominal expressions in all languages. This more abstract notion of case is called CASE

- **CASE FILTER**
  The requirement that in the course of a derivation, a nominal expression must eventually pass through or wind up in a position appropriate to its case. Such a position is called a position where the case is licensed

- **INFL**
  The head containing tense information (e.g., past versus present) and agreement information (person, number, gender). Takes VO as its complement to form an I(nfl)P (a clause)

- **CYCLE**
  Under cyclicity, a domain of application of transformations and/or the sequence of transformations that applies in that domain

- **NP / VP / DP / PP**
  Noun / verb / determiner / preposition Phrase

The evidence: Grammaticality judgements of adult speakers

Edelman & Christiansen (2003). "How seriously should we take Minimalist syntax?" TICS

The point

- Some very complex abstract mechanisms have been postulated purely on the basis of adult grammaticality judgements
  - "Does this sentence sound grammatical to you (yes/no)?"

- A huge edifice built on tenuous foundations?

Tricky questions

- Why does language acquisition take so long and show so many errors if it is mostly "pre-programmed"?
  - Perhaps certain information needs to be learned before innate grammatical can become relevant (e.g., phonology, lexicon)
  - Perhaps it's delayed maturation of innate language structures, like, you know, puberty or something
- What about bilingual kids learning two languages with *different* parameter settings?
  - Stop asking awkward questions . . . you're letting data get in the way of a beautiful theory

---

Are grammar and vocabulary truly independent? Sentence length vs. complexity

---

Do young children really have productive syntax?

Tomasello (1992): Verb island theory of early syntactic development

- "I documented virtually all of my English-speaking daughter's earliest verbs and linguistic constructions from 15-24 months of age"

  - 162 verbs used
  - ~50% in only **one** construction type
  - > 2/3 in only one or two types
  - construction types, e.g.
    - Draw car, Draw tree [Draw X]
    - Draw on paper [Draw on X]
    - Mommy draw [X draw]
    - Draw with pencil [Draw with X]

Verb island theory (cont.)

- Unevenness in how verbs used (even those close in meaning).
  - E.g., 'cut' only used in [Cut + X] while simultaneously 'draw' used in many constructions
- Unevenness of syntactic marking across verbs
  - E.g., 'eat with spoon' vs. 'hit spoon'
- Morphological marking on verbs also uneven: most unmarked, some past, some future, only 2% both past and future
  - I drew it
  - I will draw it
- Within verb's development, great continuity and gradual expansion with small addition or modification (e.g., marking of tense, or addition of new participant)

- Conclusion: Early child language not very productive and rather conservative

I can't imagine how language could be learned

It must be innate

Taking the input seriously

- What information is actually available in child directed speech?

- What can be extracted from it by simple learning mechanisms?

Co-occurrence statistics

Corpus: "To be or not to be"

Follows

	To	Be	Or	Not
To		2		
Be			1	
Or				1
Not	1			

Precedes

Syntactic information available in Child Directed Speech

Pronouns, Pronouns + Aux, Aux, Aux + Negation (49)
WH-, WH- + Aux, Pronoun + Aux (53)
Verb (105)
Verb (92)
Verb, Present Part. (50)
Determiner, Possessive Pronoun (29)
Conjunction, Interjection, Proper Noun (91)
Proper Noun (19)
Preposition (33)
Noun (317)
Adjective (92)
Proper Noun (10)

Syntactic/semantic information available in Child Directed Speech

---

Connectionism

- Use simply learning devices to evaluate how much can be learn from input

- Connectionist networks seem easily implementable in the brain (unlike rule-based computer programs)

Learning syntax

- Long range dependencies require representation of underlying syntactic structure.
- Can't just learn them from word co-occurrences

THE BOYS WHO LIVED IN THE LITTLE HOUSE AT THE END OF THE ROAD BOUGHT SOME FLOWERS FOR THEIR MOTHER

Simplified language for model

RECIPE

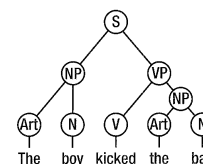S (sentence) => NP (noun phrase) + VP (verb phrase)

NP => Art (article) + N (noun)
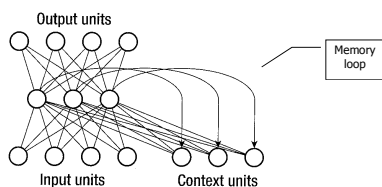
VP => V (verb) + NP

N => boy, house, flowers, ball

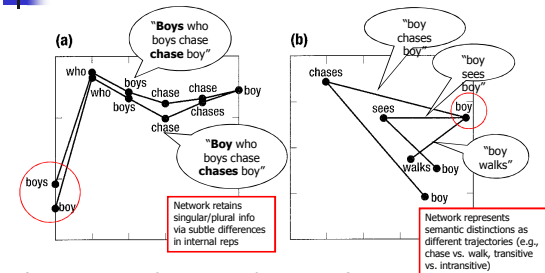V => lived, bought, kicked

Art => the, etc.

Elman (1991, 1993)


- The prediction task
- Model can learn long range dependencies

Trajectories through representational space

(a)

"**Boys** who boys chase **chase** boy"

"**Boy** who boys chase **chases** boy"

Network retains singular/plural info via subtle differences in internal reps

(b)

"boy chases boy"

"boy sees boy"

"boy walks"

Network represents semantic distinctions as different trajectories (e.g., chase vs. walk, transitive vs. intransitive)

- (60-D projected onto two dimensions)

- Structure of the internal representations when the network is as good as it can get at the prediction task
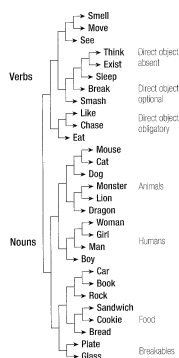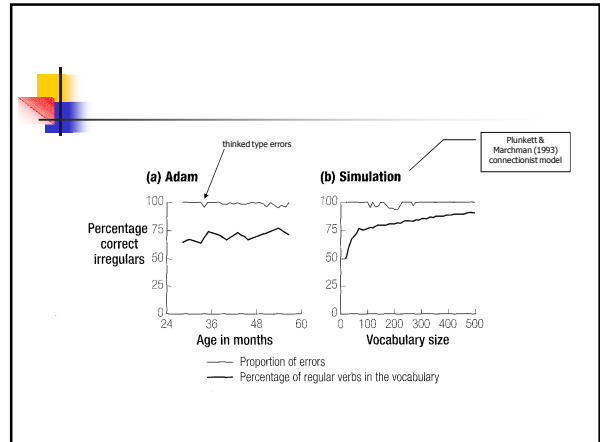
Back to the problem of exceptions

- If a child is learning grammatical rules, how do they deal with exceptions?

TALK -> TALKED

"THINKED"

English past tense

- Regular: **TALK - TALKED**
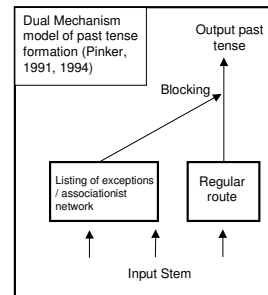- Irregular: **THINK - THOUGHT, HIT - HIT**
- Rule: **WUG - WUGGED**

---

thinked type errors

Plunkett & Marchman (1993) connectionist model

(a) Adam **(b) Simulation**

Percentage correct irregulars

Age in months

Vocabulary size

— Proportion of errors
— Percentage of regular verbs in the vocabulary

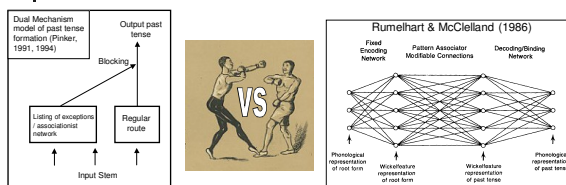# My teacher holded the baby rabbits and we patted them

"Presence of over-regularisation is obviously evidence of the child's discovery – and initial over-extension – of a rule!"

"Of course, they must memorise the exceptions."

Dual Mechanism model of past tense formation (Pinker, 1991, 1994)

Output past tense

Blocking

Listing of exceptions / associationist network

Regular route

Input Stem

Past tense models (original)

Dual Mechanism model of past tense formation (Pinker, 1991, 1994)

Output past tense

Blocking

Listing of exceptions / associationist network

Regular route

Input Stem

VS

Rumelhart & McClelland (1986)

Fixed Encoding Network | Pattern Associator Modifiable Connections | Decoding/Binding Network

Phonological representation of root form | Wickelfeature representation of root form | Wickelfeature representation of past tense | Phonological representation of past tense


Evidence for dual mechanisms

- Differences between regulars and irregulars in:
  - Frequency effects
  - Deverbalisation/denominalisation
    - (e.g., flew = past tense of fly, fly-ball = noun, turn back to verb, past tense now regularise = "the pitcher flied it")
- Dissociations in:
  - Aphasia
  - Developmental disorders
  - Brain imaging

Dissociations

	Regular	Irregular
Acquired	Frontal lesions Parkinson's D.	Posterior lesions Alzheimer's D. (?)
Developmental	Specific Language Impairment (?)	Williams syndrome (?)

(?) = data still controversial

Past tense models (state of the art)
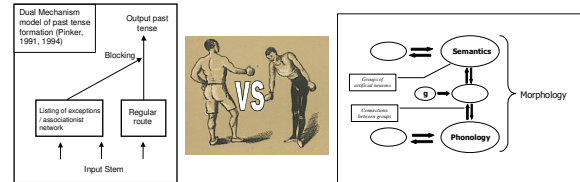
- Multiple phonological rules
- Associative network stores both regular and irregular forms based on frequency and can generalise

Both models now use duality to explain dissociations. No current evidence definitely requires rules (Lavric et al., 2001)

- Distinction between phonology and word-specific information
- Multiple inflection all paradigms (nouns, verbs)

Conclusions

Only one of these two guys has developed language

Conclusions

- Language universals need to be explained
  - Why are all human languages similar in some ways?
  - What scope do they have to differ?

- The power of human learning mechanisms cannot be estimated from an armchair (statistical learning in infants)
  - Or rather, 'the power of human learning mechanisms you must not underestimate'

Conclusions

- The question of the initial constraints to be built into a language learning system must be resolved through modelling

- And at some point, psycholinguistics may want to pay attention to how the brain processes language, and what we know about genetics and brain development

- "Innate / learned" distinction now irrelevant => It is about specifying the process. How detailed is the innate contribution compared to the final structure of adult language?

Reading

- Harley, T. A. (2001). *The psychology of language: From data to theory.* Psychology Press. (esp. ch4)
- Whitney, P. (1998). *The psychology of language.* Houghton Mifflin. (esp. ch2 & ch10)
- Lasnik, H. (2002). The minimalist program in syntax. *Trends in Cognitive Sciences, 6(10)* , 432-437.
- Palmer-Brown, D., Tepper, J. A., & Powell, H. M. (2002). Connectionist natural language parsing. *Trends in Cognitive Sciences, 6(10),* 437-442.
- Edelman, S. & Christiansen, M. (2003). How seriously should we take Minimalist syntax? *Trends in Cognitive Sciences, 7(2),* 60-61.
- Phillips, C. & Lasnik, H. (2003). Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Sciences, 7(2), 61-62.*
- Tomasello, M. (2000). Acquiring syntax is not what you think. In D. V. M. Bishop & L. B. Leonard (eds.) Speech and language impairments in children. Psychology Press.
- Lavric, A., Pizzagalli, D., Forstmeier, S., & Rippon, G., (2001). Mapping dissociations in verb morphology. *Trends in Cognitive Sciences, 5,* 301-308.

Possible reasons for language universals

- 1.Some universals may be part of the innate component of grammar
    - Why should <u>all</u> SVO languages put question words at the beginning of sentences, but <u>all</u> SOV languages put them at the end? Why should <u>all</u> SVO languages put prepositions before nouns but all SOV put postpositions after the noun – systematic patterns need to be explained
- 2. Some universals might be part of an innate component of cognition
- 3. Constraints on syntactic processing make some word orders easier to process than others. Language evolves so that they are easy to understand
    - No one likes a passive
- 4. Universals might result from strong features of the *human environment* that are imposed on us from birth and make their presence felt in all languages

Not yet clear which one is right, all lack sufficient detail