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

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

Course of early language development

<table>
<thead>
<tr>
<th>Vegetative sounds</th>
<th>0-6 weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooing</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Laughter</td>
<td>16 weeks</td>
</tr>
<tr>
<td>Vocal play</td>
<td>16 weeks – 6 months</td>
</tr>
<tr>
<td>Babbling</td>
<td>6-10 months</td>
</tr>
<tr>
<td>Single word utterances</td>
<td>10-18 months</td>
</tr>
<tr>
<td>Two-word utterances</td>
<td>18 months</td>
</tr>
<tr>
<td>Telegraphic speech</td>
<td>2 years</td>
</tr>
<tr>
<td>Full sentences</td>
<td>2 years 6 months</td>
</tr>
</tbody>
</table>

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)

The question

- How is this all learned . . . by age 5?
**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 **categorical**
- So can chinchillas, a type of South American rodent (Kuhl, 1981)

**Language experience affects phonological processing**

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

**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"

- 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
Comprehension vs. production

Syntax

How do children learn syntactic categories?

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: pardo (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.'

- 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
    Jack
  VP
    V
    put
    the car
    PP
      in the garage
```

Surface vs. Deep Structure

```
Surface
S
  NP
    the car
  VP
    V
    was put
    (trace)
    in the garage

Deep
S
  NP
    the car
  VP
    V
    was put
    in the garage
```

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 judgments concerning its distribution. How did that happen....?"

- See [http://www.princeton.edu/~browning/index.shtml](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
Structure connects with Phonetic Form of language (knowledge of speech sounds)

LF component
Structure connects with Logical Form of language (knowledge of meaning)

X-bar theory

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 computationally perfect solution 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 [to fail the exam]

In (4), as in (5), John is the understood subject of fail the exam. This fact is captured by deriving (6) 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, since 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 alternative 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 than 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 postulates that the case distinctions (nominal, accusative, etc.) morphologically manifested in languages like Latin and Russian are present on nominal expressions in all languages. The 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

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
- 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”

<table>
<thead>
<tr>
<th>Precedes</th>
<th>To</th>
<th>Be</th>
<th>Or</th>
<th>Not</th>
</tr>
</thead>
<tbody>
<tr>
<td>To</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Be</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Or</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Not</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Syntactic 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.


- The prediction task
- Model can learn long range dependencies

Trajectories through representational space

- (60-D projected onto two dimensions)

Back to the problem of exceptions

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

- 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

"Presence of overregularisation is obviously evidence of the child's discovery and initial overextension - of a rule!"

"Of course. They must memorise the exceptions!"

Past tense models (original)

Differences between 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:
  - Aphasia
  - Developmental disorders
  - Brain imaging
Dissociations

<table>
<thead>
<tr>
<th></th>
<th>Regular</th>
<th>Irregular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquired</td>
<td>Frontal lesions</td>
<td>Posterior lesions</td>
</tr>
<tr>
<td></td>
<td>Parkinson's D.</td>
<td>Alzheimer's D. (?)</td>
</tr>
<tr>
<td>Developmental</td>
<td>Specific Language impairment (?)</td>
<td>Williams syndrome (?)</td>
</tr>
</tbody>
</table>

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

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
  - "The power of human learning mechanisms you must not underestimate"
Possible reasons for language universals

1. Some universals may be part of the innate component of grammar
   - Why should all SVO languages put question words at the beginning of sentences, but all SOV languages put them at the end? Why should all 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