Seidenberg (1997): Innate Biases ≠ Grammatical Knowledge

But what about learning more abstract things (like syntax) and language-independent things that are hard (or impossible) to observe?

... future work for connectionist models.

And innate knowledge?

“Innate capacities may take the form of biases or sensitivities toward particular types of information inherent in environmental events such as language, rather than a priori knowledge of grammar itself.”

“Brain organization therefore constrains how language is learned, but the principles that govern the acquisition, representation, and use of language are not specific to this type of knowledge.”


The Logic of Poverty of the Stimulus

1) Suppose there is some data.
2) Suppose there is an incorrect hypothesis compatible with the data.
3) Suppose children behave as if they never entertain the incorrect hypothesis.

Conclusion: Children possess innate knowledge ruling out the incorrect hypothesis from the hypothesis space considered.

Example case: Yes/No question auxiliary fronting (structure-dependent rules only)
Is Hoggle tis running away from Jareth?
Can someone who can solve the Labyrinth tis show someone who can’t how?


Child Input

Very frequent
Is Hoggle tis running away from Jareth?

Very infrequent, if ever
Can someone who can solve the Labyrinth tis show someone who can’t how?

Hypotheses for frequent data type
Structure-independent (linear)
Front first auxiliary, Front last auxiliary, ...

Structure-independent (hierarchical)
Front the first auxiliary following the first noun phrase, Front the first auxiliary preceding a verb phrase, ...

Structure-independent (creative)
Front the auxiliary closest to a noun, Front the auxiliary that is an odd-numbered position, ...


The Real Rule

Front the auxiliary following the subject noun phrase in the main clause.

But the unbiased child has to rule out all the other options, even ones that are simpler to compute. (For instance: front first auxiliary is much easier to compute.) We would expect to see errors of this type.

Is the dwarf who tis talking to Jareth is going to give Sarah the peach?

Real Children
But kids don’t seem to make this error (Crain & Nakayama, 1987).

Implication: They’ve already ruled out that hypothesis, even though they’re likely not seen much data (if any at all) incompatible with it. This is due to an innate bias to look for structure-dependent rules.


Pullum & Scholz 2002 (P&S)

Claim: But there is enough disconfirming data available to children. So this situation is not true - poverty of the stimulus does not hold here.

Assumption: Only trying to rule out the front first auxiliary hypothesis, not all the other ones, too. (This isn’t necessarily true, and the P&S argument is based on the idea that the hypothesis space contains many more potential hypotheses.)

What kind of data?

One kind of disconfirming data: yes/no questions with two auxiliaries, where first auxiliary is not fronted
“Is the dwarf who is talking to Jareth tis going to give Sarah the peach?” (rare)

Another kind: wh-questions with complex subject, where first auxiliary is not fronted
“How could anyone who has watched Labyrinth before tis not wince at this part?”
(how frequent?)

 But Existence of Data ≠ Sufficiency of Data

 We need to know if the amount of disconfirming (unambiguous) data is sufficient to learn the correct hypothesis by the time children seem to know it.

 How much data is enough?

 Gauging a threshold

 Suppose we have two learning problems: Problem 1 and Problem 2.
 Suppose both have only two hypotheses to choose from.
 Suppose the frequency of unambiguous data for Problem 1 is Frequency 1 and the frequency of unambiguous data for Problem 2 is Frequency 2.

 Idea: If children figure out Problem 1 and Problem 2 at the same time, and they’re learning from the data alone, we would predict that Frequency 1 and Frequency 2 should be about equal.


 Auxiliary-Fronting Threshold: Comparative

 Auxiliary-fronting: acquired by 3 years, 2 months (Crain & Nakayama 1987)
 Subject-drop: acquired by about 3 years (Vaillant 1991).

 Unambiguous data for subject-drop: 1.2% of the data

 Another problem learned by about 3 years: Verb-Second movement in German and Dutch (German: Outhwaite 1986, Yang 2000; Dutch: Lightfoot 1997, Yang 2000)

 Sarah must solve the labyrinth.

 German/Dutch: Sarah must the labyrinth solve.

 Unambiguous evidence for Verb-Second movement: 1.2% of the data

 Expectation: Auxiliary-fronting also needs 1.2% of the data to be unambiguous, in order for it to be learned by this age.


 So how much data is there really?

 Looking at the Nina corpus:
 46,499 sentences
 23,661 questions
 14 unambiguous data examples (all of wh-question type)
 Frequency of unambiguous data: 0.068% (much less than 1.2%)

 Looking at the Adam corpus:
 20,000 sentences
 8,189 questions
 4 unambiguous data examples (all of wh-question type)
 Frequency of unambiguous data: 0.046% (much less than 1.2%)

 Data is not frequent enough for children to learn by the time they do.
Baker (2001): Complex Systems

Navajo Code Talker Paradox

English must be very different from Navajo
Japanese could decode English, but
couldn’t decode Navajo (when they didn’t
know it was Navajo).

English must be similar enough to Navajo
English can be translated into Navajo and
back with no loss of meaning. Languages
are not just a product of the culture -
pastoral AZ lifestyle couldn’t have prepared
them for Pacific Island high tech warfare,
but translation was still possible.

Baker (2001): Complex Systems

HAL 9000 from 2001: A Space Odyssey (1968)
Perfect production and comprehension of
English.

1960s: Language not considered one of the
"hard"
problems of artificial intelligence.
Reality: Still not even close now to human-like
performance.

Contrast: Chess-playing. (Not about insufficient
computational power.)

Baker (2001): Complex Systems

Levels of Variation Between Languages

Word sense (vocabulary selection):
English “think”: think, know, wonder, suppose, assume, ...
Navajo “carry”: aah (solid round-ish object)
kaah (open container with contents)
Aw (flexible object)

Baker (2001): Complex Systems

Levels of Variation Between Languages

Sounds:
English “th”, “f”, “sh”, ...
Navajo “whispered l”, “nasalized a”, ...

Baker (2001): Complex Systems

Levels of Variation Between Languages

Prefix System:
English: invariant words
“Girl crying”: “I am crying”

Navajo: no invariant forms (ex: 100-200 prefixes for verb stems)
Af’ééd yicha. “Girl crying”
Yishcha. “I am crying” (y1 + sh + cha)
Nindáhwiishdlaad. “I am again plowing” (ni + nák + ho + hi + sh + l + di)
Baker (2001): Complex Systems

Similarities & Differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages.

Idea: A relatively small number of parameters yields a large number of different languages.

English

Japanese

Tagalog

French

Navajo

...

Baker (2001): Complex Systems

Similarities & Differences: Parameters

Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the environment determines what values these parameters should have.