The Richness of the Poverty of the Stimulus

Julie Anne Legate
University of Delaware
Charles Yang
Yale University

The Argument from the Poverty of the Stimulus in <u>LSLT</u>

"In <u>LSLT</u> the `psychological analogue' to the methodological problem of constructing linguistic theory is not discussed, but it lay in the immediate background of my own thinking."

"To raise the issue seemed to me, at the time, too audacious." Introduction

LSLT

"a grammar is justified by showing that it follows from a given abstract theory of linguistic structure."

"The abstract theory must have the property that for each language, the highest-valued grammar for that language meets the external criteria of adequacy" p65

LSLT

1. "A speaker of a language has observed a certain limited set of utterances in his language. On the basis of this finite linguistic experience he can produce an indefinite number of new utterances which are immediately acceptable to other members of his speech community. He can also distinguish a certain set of "grammatical" utterances, among utterances that he has never heard and might never produce." §1

LSLT

"It may be that along these lines we will be able to develop ... an explanation for the general process of projection by which speakers extend their limited linguistic experience to new and immediately acceptable forms." §110

<u>Syntactic Structures</u> (1957)

"we pose a condition of generality on grammars; we require that the grammar of a given language be constructed in accordance with a specific theory of linguistic structure in which terms such as `phoneme' and `phrase' are defined independently of any particular language." §6.1 p50

"a reasonable requirement, since we are interested not only in particular languages, but also in the general nature of Language." §2.1 p14

<u>Syntactic Structures</u> (1957)

"a grammar mirrors the behavior of the speaker who, on the basis of a finite and accidental experience with language, can produce or understand an indefinite number of new sentences." §2.2 p15

Review of <u>Verbal Behavior</u> (1959)

"in principle it may be possible to study the problem of determining what the built-in structure of an information-processing (hypothesis-forming) system must be to enable it to arrive at the grammar of a language from the available data in the available time." p58

Aspects
(1965)

"A consideration of the character of the grammar that is acquired, the degenerate quality and narrowly limited extent of the available data, the striking uniformity of the resulting grammars, and their independence of intelligence, motivation, and emotional state, over wide ranges of variation, leave little hope that much of the structure of the language can be learned by an organism initially uninformed as to its general character." p58

<u>Cartesian Linguistics</u>
(1966)

"seventeenth-century rationalism ... notes that knowledge arises on the basis of very scattered and inadequate data and that there are uniformities in what is learned that are in no way uniquely determined by the data itself. Consequently, these properties are attributed to the mind, as preconditions for experience." p65

Language and Mind
(1968)

"if we assume that the A-over-A principle is a part of the innate schematism that determines the form of knowledge of language, we can account for certain aspects of the knowledge of English possessed by speakers who obviously have not been trained and who have not even been presented with data bearing on the phenomena in question in any relevant way, as far as can be ascertained." p53

Structure Dependence in yes/no Questions

- **First mention of auxiliary fronting in yes/no questions requiring structure dependence: Aspects p55-56, developed in Language and Mind**

- **The subjects who will act as controls will be paid**

- **Will [the subjects who will act as controls] __ be paid?**

- ***Will the subjects who __ act as controls will be paid?**

Fodor 2001

"That's what Chomsky gets for offering an example that's easy to understand."

Three Challenges

- maybe children's grammar isn't that good

- maybe the data isn't that poor

- maybe empiricist learning isn't that hopeless

Objection from the Poverty of the Output

- **Error-free learning: intermediate states do not violate innate constraints**

  - **"No one ever makes mistakes to be corrected." (debate with Piaget re: SSC)**

  - **Rules and Representations 1980**

- **Objection: error-free learning must be established, not asserted**

To Err is Human?

- Crain & Nakayama (1987)

- 3-5 year olds

- Ask Jabba if [the boy who is watching Mickey Mouse] is happy.

- Is [the boy who is watching Mickey Mouse] __ happy?

- *Is [the boy who __ watching Mickey Mouse] is happy?

To Err is Human?

- MacWhinney (2004) -- learning is low-error not error-free

- e.g. Complex NP Constraint: Seth 38-42 months (Wilson & Peters <u>Language</u> 1988)

Seth Violates UG?

(Seth's running away at the store had been much discussed)
What did I get lost at [the __], Dad?

Is this a T?
No, that's a funny I.
What is this [a funny __], Dad?

What are you cookin' on [a hot __]?
Well, what AM I cookin' on a hot?
Stove!

Seth Speaks Slavic
(Or Perhaps Pama-Nyungan)

Iz kojeg grada je Petar sreo [djevojke __]
from which city is Peter met [girls __]
"Which city did Peter meet girls from?"

                                    (SC Stjepanovic 1998)

(Jangari mayi kanpa mardarni?)
(shanghai Q   aux   have)
("Do you have a shanghai?")

Yuwayi. Jirrama karna mardarni [jangari __]
yes   two   aux   have   [shanghai __]
"Yes. I have TWO shanghais!"

                                    (Warlpiri Laughren 1984)

To Err is Human

- Conclusions from such cases:

    - Limited to status of principle

    - No effect on broader claim of innate knowledge

    - Use of non-target but possible human grammars further evidence for innate knowledge

Objection from the Richness of the Stimulus

- Perhaps the child does have the evidence available to rule out wrong alternative hypotheses (Pullum & Scholz 2002)

- Will [the subjects who will act as controls] __ be paid?

- "Was the Argument that Made was Empirical?" Legate 1999

- "Empirical Reassessment of Stimulus Poverty Arguments" Legate & Yang 2002

Richness?

- **Adam corpus = 0 (20,372 sentences, 8,889 questions) (Legate 1999)**

- **Nina corpus = 0 (46,499 sentences, 20,651 questions) (Legate & Yang 2002)**

- **CHILDES = 1 (3 million sentences) (MacWhinney 2004)**

Existence vs Sufficency

- **If such evidence did exist, would it be sufficient to rule out incorrect hypotheses?**

- Situate within a comparative framework of language acquisition

- Could the child use this information given time and frequency of data?

- Comparable age of acquisition = comparable frequency of disambiguating evidence

Objection from the Richness of the Learner

- Piaget (1975/1980) "the `innate fixed nucleus' would retain all its properties of a `fixed nucleus' if it were not innate but constituted the `necessary' result of the constructions of sensorimotor intelligence" (p31)

- Putnam (1980) "Once it is granted that such multipurpose learning strategies exist, the claim that they <u>cannot</u> account for language learning becomes highly dubious" (p296)

Objection from the Richness of the Learner

- Relevant data not required--emerges from statistical properties of the input

- e.g. Lewis & Elman (2001), Reali & Christiansen (2003, 2004)

- Track transitional probabilities of adjacent words

- Differentiates between:

   (1) Is the little boy who is crying ___ hurt?

   (2) *Is the little boy who ___ crying is hurt?

Emergence?

- Does not learn movement

- Learns that "who is" is common in the input

   (1) Is the little boy who is crying ___ hurt?

   (2)*Is the little boy who ___ crying is hurt?

- Kam, Stoyneshka, Tornyova, Sakas, Fodor (2005)

- Differentiate who-rel vs who-wh and correct output is no longer selected (17% correct, 30% incorrect, 44% can't choose)

Ideas: Old and New I

- **Trends in linguistic sciences: computers, corpora, & babies, more powerful than "previously thought"**

- **What LSLT actually said …**

- **Distributional learning: "(perhaps more promising) way ... is through the analysis of clustering. We define the <u>distribution</u> of a word as the set of contexts of the corpus in which it occurs, and the distributional <u>distance</u> between two words..." (§34.5)**

- **Information theory: "This conception of syntactic analysis has an information-theoretic interpretation ... We have in fact defined the best analysis as the one that minimizes information per word in the generated language grammatical discourse" (§35.4)**

Ideas: Old and New II

- **Minimum Description Length (MDL) Principle: "it is interesting to note that any simplification along these lines is immediately reflected in the length of the grammar." (§26)**

- **Probability in grammar: "though we have strong reasons for a nonstatistical conception of the form of grammar, it might turn out to be the case that statistical considerations are relevant to establishing, e.g., the absolute, nonstatistical distinctions between G and $\bar{G}$ ... (fn. Note that there is no question being raised here as to the legitimacy of a probabilistic study of language". (§36.3)**

- **many other revived, or rediscovered ...**

Ideas: Old and New III

- **Statistical Learning: "Investigation of his [Z. Harris] data seems to indicate that word boundaries can be placed much more effectively than morpheme boundaries by this method." (§45, fn)**

  - **8-month-old infants can do this (Saffran, Aslin, & Newport 1996)**

  - **"Learning Rediscovered … it flies in the face of received wisdom... learning is much more powerful than previously believed, and arguments about the innateness of language and other forms of cognition need to take that undeniable fact into account" (Bates & Elman 1996)**

- **§45, fn (Cont.): "The problem is whether this can be done on the basis of a corpus of a reasonable size ..."**

- **Psychological plausibility: "a stronger theory, i.e., of converting this evaluation procedure into a practical discovery procedure" (§38)—-not just corpus analysis**
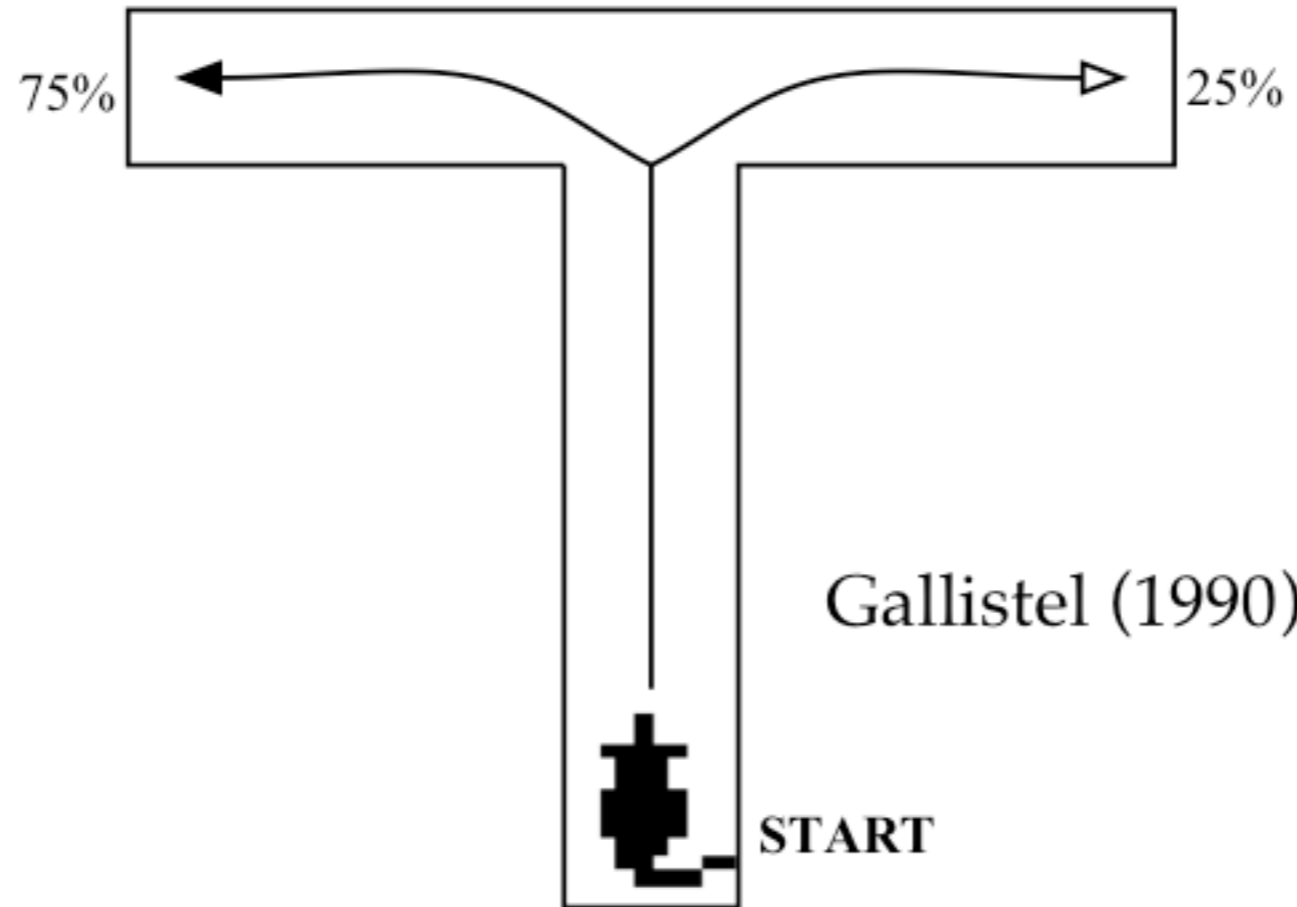
"Whether This Can Be Done"

- Gambell & Yang (2003): computational modeling using CHILDES data

- Experiments vs. reality

- Statistical learning does not scale up

  - not enough "big" words.

- Is it a "practical discovery procedure"?

  - P(A->B)=P(AB)/P(A), every time hear A, must change P(A->B) for all B's (thousands in practice)

- Statistical learning can work well when constrained by what appears to be innate knowledge of how prosody marks linguistically significant structures (Gambell & Yang 2005): <u>another POS argument</u>

The Richness of Data

- Too much data, and too powerful a statistical learner, are both <u>very bad things!</u> (Goodman, Quine)

- Statistical learning is no Universal Acid: the learner pays attention to certain statistical correlations but not others (Nespor et al. 2002, Newport & Aslin 2004, Newport et al. 2004)

- Language acquisition as <u>innately guided learning</u>: "The animal is innately equipped to recognize when it should learn, what cues it should attend to, how to store the new information and how to refer to it in the future" (Gould & Marler 1987)

- Learning demonstrates innateness: Auxiliary Inversion, Word segmentation, ...

Yalies vs. Rats

75%   25%

Gallistel (1990)

START

Probabilities and Parameters

- Probability matching behavior (Bush & Mosteller 1951, Herrnstein 1961); cf. Labov (1994)

- Yang (2002): parameter setting as probability matching (cf. Jakobson 1968, Mehler 1974, Piatelli-Palmirini 1989, Clark 1993, Werker & Tees 1999, Roeper 2000, Kroch 2001, Crain & Pietroski 2002)

  - Head Initial/Final: domain-specific space

  - Quantitative correlations between development and frequency of specific linguistic data in the input

- Bloom (2001): no WAD but domain-specific syntactic and semantic constraints remain indispensable (Lasnik 1989, Gleitman 1990, Pinker 1994 etc.)

- Recall statistical learning in word segmentation.

Representations vs Mechanisms

A nativism of domain specific information needn't, of course, be <u>incompatible</u> with a nativism of domain specific acquisition mechanisms ... But I want to emphasize that, given his understanding of POSAs [Poverty of Stimulus Arguments], Chomsky can with perfect coherence claim that innate, domain specific PAs [Propositional Attitudes] mediate language acquisition, while remaining entirely agnostic about the domain specificity of language acquisition <u>mechanisms</u>. (Fodor 2001: p107-8)

Language Design

- **<u>Three Factors in Language Design</u>** (2005):

  - "Genetic endowment ... which interprets part of the environment as linguistic experience"

  - "Experience, which leads to variation, within a fairly narrow range ..."

  - "Principles not specific to language ... (a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints ... including principles of efficient computation"

- <u>LSLT</u> (1955): "At the present stage of our knowledge we must surely keep an open mind ... " (§36)

Thank you Noam!