Schema Theorem in Language Acquisition: A Rags to Riches Story

Sean McLennan

Ling Lunch - Indiana University
December 4th, 2000

Poverty of Stimulus:

- One of the primary arguments that supports Universal Grammar (UG)
- **Basic premise:** Language is a productive system - i.e. we can both produce and understand utterances that we have never heard before. However, the amount of input data that we are exposed to during acquisition is not sufficient (in either quantity, nor quality) for a productive grammar to be inferred. (Miller and Chomsky, 1963).
- Inconsistent with current trends in other related fields: Psychology, Neuroscience, and Cognitive Science.
- Contradictory evidence from modeling: ex. Chalmers (1990), Elman (1995). It seems that statistically based algorithms (like neural nets) *can* learn tasks they were previously thought incapable of.

Genetic Algorithms:

- Three primary elements:
  1) a population of chromosomes (usually strings of ones and zeros - ‘1001101’)
  2) a fitness function (provides a qualitative judgement of ‘goodness’)
  3) mating and procreation
  ** for details see Mitchell (1996)*
- Each chromosome represents a single ‘solution’ in an enormous space of solutions
- Usually it is very easy to come up with a ‘solution’ rated 100% in only a few generations – i.e. GAs can efficiently search a ‘solution space’ with a relatively small number of instantiations
- How??

Schema Theorem:

- explanation for how GA’s can perform effective search; Holland (1975)
- introduces the notion of a ‘wildcard’ – * – into the chromosomes; provides a notion of ‘category’
  ex: **1**** is a category representation of **{100, 101, 110, 111 (1*1, 1*0, 11*, 10*)}**
- relationship is reversible:
  ex. **101** is an instantiation of **{***, 1**, *0*, **1, 10*, 1*1, *01, 101}**
- in general: each chromosome of length ***l*** represents $\mathbf{2}^{l}$ of a total of $\mathbf{3}^{l}$ possible schemata

Each instantiation implicitly represents all the categories it belongs to. When you judge the instantiation, you judge the categories as well – implicitly and simultaneously.

- GAs can act on the category information **implicit** in instantiations because:
  a) they exist in populations – judging multiple instantiations of the same schema is like **implicitly** calculating an average fitness for the schema
  b) there is **process** that acts on that information – i.e. mating and procreation is biased according to fitness

Although developed for a domain specific explanation, Schema Theorem is more profound and, if abstracted, is more generally applicable to other domains (like language acquisition).

In Language Acquisition:

Schema Theorem		GAs		Acquisition
representation	➢	chromosomes	➢	token memory
process	➢	selection	➢	learning
product	➢	solution	➢	high level categories

- **Token Memories:**
  - children are exquisitely sensitive to detailed sensory data and, at first, seem to lack the ability to 'parse' their environment. ex. Rovee-Collier (1991).
  - by token I just mean, some time - delineated perceptual memory passively acquired
  - importantly, *all* sensory information is simultaneously learned and correlated

- **Learning:**
  - in the most general Hebbian, correlative sense; on multiple levels

- **High Level Categories:**
  - phonological features; a single word; semantic features; transformations; word categories; i.e. anything treated as symbolic in current linguistic thought

Implications:

- "Poverty of Stimulus" is not a valid argument which places our current *view* of UG in question
- Language is not symbolic "all the way down"
- Features and categories are an emergent phenomenon
- Language and grammar in children and adults is not the same thing
- This ***does not*** preclude the idea that language is universally shaped by biology
- This ***does not*** preclude symbolic processing
- This ***does not*** preclude the validity of higher level descriptions
- It ***does*** change the focus of innate properties of language (UG, if you will)
- It ***does*** permit symbolic processing that can only be separated from low-level phenomena to a limited extent
- It ***does*** permit variability and language change
- It ***does*** imply that the tools of generativism are not useful to study all linguistic phenomena

Testing the Implications by modeling:

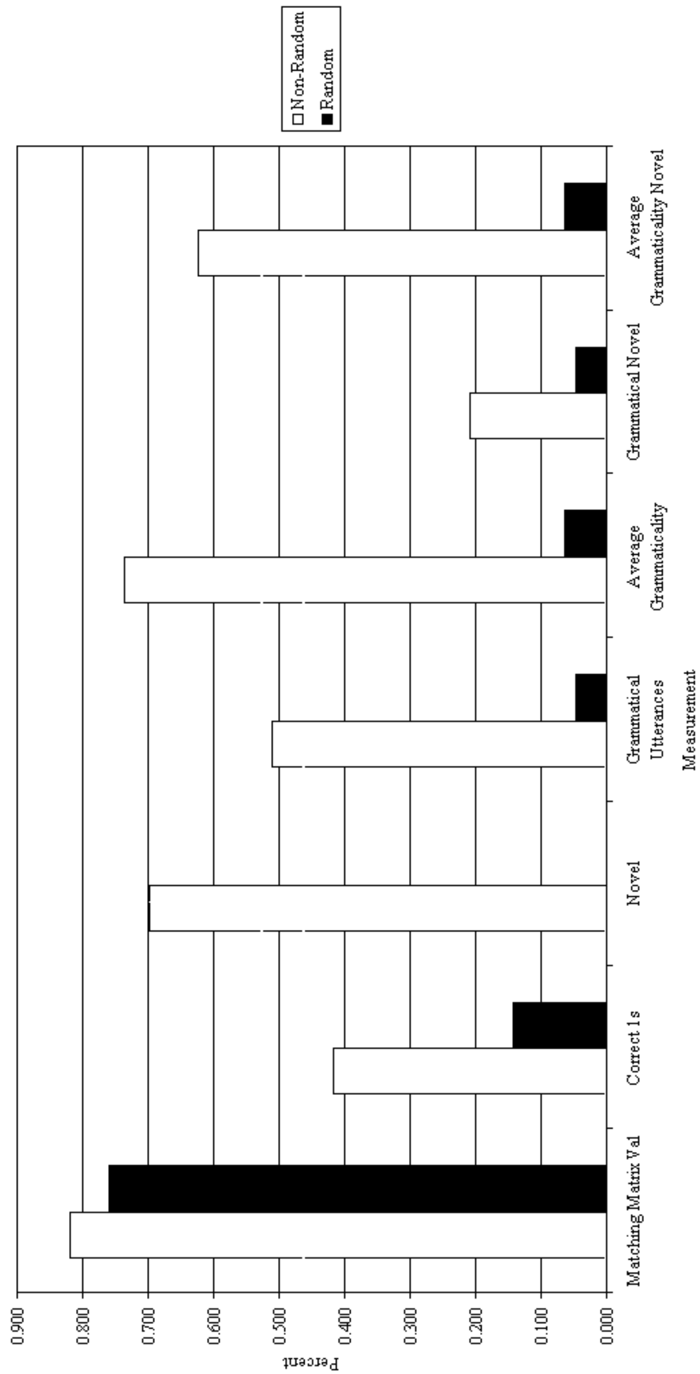
- **Question:**
  Can a data-driven approach give us:
  a) productivity?
  b) grammaticality?
  c) categories?

- **The model:**
  a) Intent is *demonstrative* – i.e. *nothing* in the system is intended to *directly* model a language phenomenon
  b) Input: utterances comprised of *articles*, *adjectives, nouns, and (intransitive) verbs* generated randomly using a simple Markov chain
  c) Learning: sensitive to only one dimension of information – what comes before / after entities (of any level) that have already been seen; system starts with no a prior information; inference based on similarity of *only* unanalyzable tokens (usually words).
  d) Output: new utterances generated with an inferred Markov chain

- **Results (based on 100 trials of 50 input / output utterances):**
  a) productivity?
  70% of output utterances are novel (i.e. not in input)
  b) grammaticality?
  51% of output utterances are 100% grammatical
  74% - average grammaticality
  21% of novel output utterances are 100% grammatical
  62% - average grammaticality of novel utterances
  c) categories?
  example of groupings made by a typical system:
  1) `{|slept| |slid|}`
  2) `{|smelled| |crashed|}`
  3) `{|window| |dog|}`
  4) `{|expensive| |dirty|}`
  5) `{|red| |deep| |empty|}`
  6) `{|dog| |car| |box| |rat| |window| |box|}`
  7) `{|opened| |crashed|}`
  8) `{|car| |box| |rat|}`

Comparison Ratings

(with random grammars of similar parameters - all significant)



References:

Chalmers, David. (1990). Syntactic Transformations on Distributed Representations. *Connection Science,* Vol. 2, Nos 1 & 2, 53-62.

Elman, Jeffery. (1995). Language as a dynamical system. In: Port et al.

Holland, John. (1975). *Adaption in Natural and Artificial Systems.* Ann Arbor, MI: University of Michigan Press.

Miller, G.A. and Noam Chomsky. (1963). Finitary models of language users. In: R.D. Luce, R.R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology, Vol. 2.* New York, NY: Wiley.

Mitchell, Melanie. (1996). *An Introduction to Genetic Algorithms.* Cambridge, MA: MIT Press.

Port, Robert and Timothy van Gelder. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.

Rovee-Collier, C. (1991). The "memory system" of prelinguistic infants. *Annals of the New York Academy of Sciences, 608,* 517-536.