Origins of language:  
A conspiracy theory  
Jeff Elman  
Models of Language Class  
Kenny Vaden

I. Introduction:  
language is puzzling

Alternative to language instinct:  
Connectionist perspective on language acquisition

Biology and development interact to produce lang.  
Certain forms of innateness are plausible, others not

Goals:  
1. Taxonomy of Innateness  
2. Two connectionist simulations  
3. How results fit into Conspiracy Theory

II. Ways to be innate

What does it mean to be innate?

Innateness would impose constraints on some level:

1. Representational constraints (brain-state)  
   Could genes pre-specify brain circuitry & connectivity?  
   Pinker’s favorite, but it is the most unlikely…

2. Architectures: unit, local, global constraints  
   Could genes specify some specific brain hardware for language?  
   Maybe connection pattern among modular systems is important…

3. Chronotopic constraints (timing)  
   Could genes affect endogenous and exogenous interactions?  
   Small changes in development could make language happen…

Is language a species-specific adaptation?

All humans acquire and use language, in all of its complexity,  
while no other animals can; similarities in languages and  
development patterns across cultures; (common patterns in  
language formation too).

Is language is an instinct, a genetically scripted for  
acquisition and use, as Pinker (1994) claims?

OR is Innateness an over-simplistic explanation?

Genome is required for all sorts of behaviors,  
Scant evidence of specifically genetic linguistic defects  
Consider complexity of gene function, versus Pinker’s claim
II. Ways to be innate
What does it mean to be innate?
Re-framed in connectionist terms: figure 1

<table>
<thead>
<tr>
<th>Source of connection</th>
<th>Examples of innateness</th>
<th>Examples in networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations</td>
<td>Non-epigenetic determinants</td>
<td>Weight on connections, development function, maturation factor, etc.</td>
</tr>
<tr>
<td>Architectures</td>
<td>Heritable packaging/density</td>
<td>Network architecture, modular organization</td>
</tr>
<tr>
<td>Timing</td>
<td>Sequential, local events, key events</td>
<td>Temporal gradients, allometric changes in bone length</td>
</tr>
</tbody>
</table>

III. Problem w. representational innateness

Representational Innateness is **wrong**

- **Numbers don't work:** Nematodes have identical connectivity, with only 959 neurons. No two humans share exact patterns; we have trillions of synapses.

- **Recycling:** Even for fruit flies, most genes play a role in multiple expressions and interact.

- **Specifying particular language microcircuits in human brains is too much of a burden for individual genes.**

IV. The importance of time

Architectural and timing constraints
In closed systems, **internal timing is critical**

ex 1: **temporal growth gradients:** allometric changes in bone length

ex 2: loss of interactions in tooth formation vs beaks

- Mouse EPD + Chick Mesenchyme: Tooth formation

“The ability of chick epithelium to participate in odontogenesis and to secrete enamel matrix proteins suggests that... an alteration in the behavior of cranial neural crest cells must have blocked the initiation of tooth development.” - Kollar & Fisher (1990)
V. The importance of starting small

Children must learn to communicate

Symbolic activity: function decoupled from content
Symbols have an arbitrary correspondence to referents
Structures have a complex relation to constituents

1. Theory of Semantics: meaning from sentence constituents
2. Theory of Syntax: what sort of structures are grammatical

Embedded information can then be successfully detangled:

The cat who the dogs chase runs toward me.

---

Gold’s Proof (1967)

Embedding cannot be learned inductively using positive input only. Violations should occur, then be corrected or identified.
Direct negative evidence is not present.

Critical knowledge about grammar must be innate.
Learning involves fine-tuning child’s UG to environment.

How specific is constraint to learn embedding?
Do constraints even need to be of a linguistic sort?

---

V. The importance of starting small

Artificial Language Time!

1. Grammatical categories: words belonged to different categories (eg. noun, verb, etc.)
2. Basic sentence structure: noun followed by a verb; transitive verbs followed by a second noun (eg. cat chased dog)
3. number agreement between subject noun & verb (eg. cat runs, dogs chase)
4. verb argument structure: some verbs transitive; others intransitive; others were optionally transitive (eg. lions eat vs. lions eat dogs)
5. relative clauses: nouns could be modified by a relative clause (eg., who the dogs chase); both subject relatives (girl who sees the boy) and object relatives (girl who the boys see) were possible.

Elman used these rules to generated a bunch of non-repeating artificial sentences in the form of orthogonal [0,1] sentence vectors (each entry is a random word in the artificial language).

ex. sentence: [00010000011000000000000000]

---

V. The importance of starting small

Can network learn embedding from positive evidence?

ex. sentence: [00010000011000000000000000]

Simple recurrent network used for prediction task
Rectangles: layer of nodes
Arrows: connections between layers
Broken lines: changing weights
Solid line: fixed weight = 1.

Network failed even to predict words from the training sets. Relative clause number agreement:
Ex. The boys who the girl sees see the dog.
V. The importance of starting small

Network learns embedding with increasingly complex sets
ex. sentence: [00010000001000000000000000]

FIVE EPOCHS: 10k Sentences Each
1) 100% Simple Sentences
2) 75:25 Ratio Simple:Complex sentences
3) 50:50 Ratio Simple:Complex sentences
4) 25:75 Ratio Simple:Complex sentences
5) 100% Complex sentences

Network predicts grammatical output for complex and simple sentences.
✓ Children start w producing simple sentences
  but their input is usually complex

V. The importance of starting small

A reasonable way to limit input complexity?
Sentences do not change, but child's internal abilities do change.
Children do not start with mature memory or perceptual systems.

Interaction of available resources and language learning
1. Limited but expanding capacity for context
2. Noisy input, but decrease noise over time

Result: Internal state-space converges on regularities more quickly

Unstructured: Structure in input structure
learning failed
Hidden unit activation

VI. Where cortical structure comes from

Functionally specialized language areas in cortex
First simulation demonstrates that developmental constraints could explain acquisition pattern instead of an innate LAD.

BUT how does specific functional organization arise across individuals, if the architecture is not pre-specified genetically?

Innate architecture or convergent development?

Development of memory capacity and perception can limit search space enough for learning to succeed.

Limitations instead of a specialized acquisition system that comes on-line
Less is more: critical periods as mature systems settling on unstructured solutions – too many possibilities.

Evidence from the kinds of error at different stages of maturity:

Adult L2 learners: incomplete control, frozen morphology
Young native learners: errors of omission
VI. Where cortical structure comes from
Convergence: Shrager & Johnson’s XOR Model

**Fun Biological Facts:**
Cortex begins with high pluripotentiality, and over time becomes a cascade of specialized, nearly modular filters.

*Arborization and thinning* begin in primary sensory and motor areas, and spread anteriorly through the cortex from ages 4-25.

Changes do not happen everywhere simultaneously - do functional specializations result from *trophic waves of plasticity*?

---

**Shrager & Johnson (1996) Model:**

- **30x30 Instant Cortex**
  - A, B: Send input to all neurons
  - Each neuron also sends output to nearby units.
  - Each neuron receives external input, plus neighbor-transformed input
  - Hebbian Learning Rule: strength grows among correlated-output units

*Units develop functions:*
- AND and OR, But not XOR

---

**Shrager & Johnson (1996) Model:**

- **Trophic Wave Component**
  - Leftward columns change first, while holding rightward columns static.
  - Then freeze left-most columns, and allow next column to learn.
  - Results in spatially distributed, functionally distinct areas.

*Primary change: lower-level detection AND, OR, not XOR*

*Subsequent areas: higher-level AND, OR, A=B, ¬[A and B], etc.*

---

**Spatial Trophic Plasticity Waves**

Shrager & Johnson found that first order functions present in fast traveling trophic waves, second order functions increase with slow trophic waves.

Rebotier & Elman replicated that result, and found XOR detectors using a spatial wave of trophic change.

Demonstrates that powerful non-linear relationships could be learned with plausible unsupervised Hebbian learning over time and space.
VII. Conclusion

Conspiracy Theory of Language

Language is unique to human species, and it takes a constrained set of possible forms. But this is not evidence for a radical evolutionary change, or a language gene, per se.

**Biological basis of language be can explained**

1. Nonlinear effects of small developmental changes on outcome
2. Conservative genome, nature of interactions within development

Phenotypic variation, even if a single gene leads to it – involves complex interplay of developmental and regulatory mechanisms

*Twists and tangles in communicative behaviors are allometric in nature*

*Account of the processes involved in language is the ultimate goal*