Models of Language Evolution
Session 10: Iterated Learning and the Evolution of Compositionality & Recursion

Michael Franke

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen
### Course Overview (definite?)

<table>
<thead>
<tr>
<th>date</th>
<th>content</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-4</td>
<td>MoLE: Aims &amp; Challenges</td>
</tr>
<tr>
<td>27-4</td>
<td>Evolutionary Game Theory 1: Statics</td>
</tr>
<tr>
<td>04-5</td>
<td>&gt; GT 2:: Macro-Dynamics</td>
</tr>
<tr>
<td>11-5</td>
<td>&gt; GT 3: Signaling Games (Gerhard Jäger)</td>
</tr>
<tr>
<td>18-5</td>
<td>&gt; GT 4: Micro-Dynamics &amp; Multi-Agent Systems</td>
</tr>
<tr>
<td>25-5</td>
<td>Evolution of Semantic Meaning</td>
</tr>
<tr>
<td>01-6</td>
<td>Semantic Meaning &amp; Conceptual Space</td>
</tr>
<tr>
<td>08-6</td>
<td>Evolution of Pragmatic Strategies (Roland Mühlenbernd)</td>
</tr>
<tr>
<td>15-6</td>
<td>PENTECOST — NO CLASS</td>
</tr>
<tr>
<td>22-6</td>
<td>Iterated Learning &amp; Compositionality</td>
</tr>
<tr>
<td>29-6</td>
<td>assignment of projects</td>
</tr>
<tr>
<td>06-7</td>
<td>work on student projects — NO CLASS</td>
</tr>
<tr>
<td>13-7</td>
<td>work on student projects — CONSULTATIONS, NO CLASS</td>
</tr>
<tr>
<td>20-7</td>
<td>PRESENTATIONS</td>
</tr>
</tbody>
</table>
Compositional Semantics

The meaning of a complex utterance depends systematically on the meaning of its parts and their way of combination.

(1)  a. John likes Mary.
    b. John abhors Mary.
    c. Mary likes John.
Recursive Syntax & Semantics

Complex expressions and meanings of type $x$ can be embedded in another expression to form type $x$.

(2)  
   a. John smokes.  
   b. Mary knows that John smokes.  
   c. Bill suspects that Mary knows that John smokes.  
   d. ...  

(3)  
   a. Hunde beißen.  
   b. Hunde, die Hunde beißen, beißen.  
   c. Hunde, die Hunde, die Hunde beißen, beißen, beißen.  
   d. ...
Syntactic Structure

Natural languages have seemingly idiosyncratic rules on what the “correct” way of forming a sentence and expressing a thought is.

(4) a. Hans raucht.
   b. Susanne weiß, dass Hans raucht.

(5) a. Hans raucht Pfeife.
   b. Susanne weiß, dass Hans Pfeife raucht.
   c. *Susanne weiß, dass Hans raucht Pfeife.

(6) a. Hans radelt viel, denn das ist gesund.
   b. Hans radelt viel, weil das gesund ist.
   c. *Hans radelt viel, weil das ist gesund.
The Innateness Hypothesis (Chomsky, 1965, and later)

Humans are biologically endowed with some knowledge of certain universal elements of the structure of human languages.

⇒ innate “language faculty” domain-specific?
⇒ specialized “language acquisition device” (LAD)?
⇒ shaped by biological evolution? (cf. Pinker and Bloom, 1990)
The Poverty of the Stimulus Argument (Chomsky, 1980)

consider a grammar $G$ of a language $L$

**P1:** children can rapidly and faithfully learn $G_L$

**P2:** data available during language acquisition underdetermines $G_L$

**P3:** adult competence matches $G_L$ also for unfamiliar expressions

---

**C:** some parts of grammatical competence must be innate
Argument for Biological Evolution

**P1:** what is innate is genetically encoded

**P2:** what is genetically encoded must have been shaped by biological evolution

**C:** the LAD is a product of biological evolution
The Iterated Learning Model (ILM) — Main Idea

- poverty of the stimulus ⇔ “learning bottleneck”
- grammatical competence passes the bottleneck repeatedly
- repeated “bottlenecking” shapes language, not vice versa
The Iterated Learning Model (ILM) — Main Idea (second shot)

- language learners have some domain-general learning capability including a (modest) capacity to generalize and extract patterns
- competent speakers have learned from learners . . .
  . . . who have learned from learners . . .
  . . . who have learned from learners . . .
  . . . who have learned from learners . . .
- **iterated learning can create structure which wasn’t there before**
  - given capability for generalization
  - given an appropriately sized bottleneck
Interdependencies in Language Evolution

Learning mechanisms determine cultural dynamics

Learning

Culture

Evolution

Genes shape learning mechanisms

Emergent structure affects fitness landscape

(from Kirby, 2007)
Evolution of Compositionality  

(Kirby and Hurford, 2002)

- 1 learner, 1 teacher
- teacher produces $n$ state-signal pairs
- learner acquires a language based on these
- (iterate:) learner becomes teacher for new learner
- learning model:
  - feed-forward neural network
  - backpropagation (supervised learning)
- production strategy: "obversion"
  - production optimizes based on individual comprehension
Learning Model: Feed-Forward Neural Network

- $8 \times 8 \times 8$ network for interpretation
- input: signal
  $i = \langle i_1, \ldots, i_8 \rangle \in \{0, 1\}^8$
- output: meaning
  $o = \langle o_1, \ldots, o_8 \rangle \in \{0, 1\}^8$
- initially arbitrary weights
Backpropagation

- training items \( \langle i, o \rangle \) are presented
- network computes its output \( o' \) for given \( i \)
- error \( \delta = o - o' \) is propagated back through all layers
- weights are adjusted accordingly

Obverter Strategy

- feed-forward net only defines interpretation strategy
- production as best choice given the speaker’s own interpretation:
  - suppose teacher wants to express meaning \( o \in \{0, 1\}^8 \)
  - she then chooses an \( i_c \in \{0, 1\}^8 \) that triggers network output \( o' \in [0, 1]^8 \) if \( i_c \) maximizes confidence:

\[
i_c = \arg \max_{i \in \{0, 1\}^8} C(o|i)
\]

defined as:

\[
C(o|i) = \prod_{k=1}^{8} C(o_k|o'_k)
\]

\[
C(o_k|o'_k) = \begin{cases} 
    o'_k & \text{if } o_k = 1 \\
    1 - o'_k & \text{if } o_k = 0 
\end{cases}
\]
Results (20 Trainings Items)

dotted: difference teacher-learner language
solid: proportion of meaning space covered

from Kirby and Hurford (2002)
Results (2000 Trainings Items)

- dotted: difference teacher-learner language
- solid: proportion of meaning space covered

from Kirby and Hurford (2002)
Results

(50 Trainings Items)

dotted: difference teacher-learner language
solid: proportion of meaning space covered

from Kirby and Hurford (2002)
Compositionality

- compositionality arises for medium-sized bottlenecks, e.g.:

\[
\begin{align*}
  o_1 &= 1 \quad \leftrightarrow \quad i_3 = 0 \\
  o_2 &= 1 \quad \leftrightarrow \quad i_5 = 0 \\
  o_3 &= 1 \quad \leftrightarrow \quad i_6 = 0 \\
  o_4 &= 1 \quad \leftrightarrow \quad i_1 = 0 \\
  o_5 &= 1 \quad \leftrightarrow \quad i_4 = 1 \\
  o_6 &= 1 \quad \leftrightarrow \quad i_8 = 1 \\
  o_7 &= 1 \quad \leftrightarrow \quad i_2 = 0 \\
  o_8 &= 1 \quad \leftrightarrow \quad i_7 = 1 
\end{align*}
\]
Summary: Compositionality from Iterated Learning

- iterated learning creates compositionality . . .
  - if bottleneck size is appropriate given size of meaning and signal spaces
  - by generalizing over sparse training data
  - by informed innovation (where necessary)

- other learning mechanisms possible:
  - other kinds of neural networks (e.g. Smith et al., 2003)
  - finite state transducers (e.g. Brighton, 2002)
Evolution of Recursive Structure  

(Kirby and Hurford, 2002)

- ILM mainly as before

- state space: small logical language with
  - individual constants $C$ ("John", "Mary", dots)
  - 2-placed predicates $P$ ("loves(·,·)", …)
  - propositional attitude predicates $Q$ ("thinks(·,·)"")
  - $|C| = |P| = |Q| = 5$
  - language: $S = P(c,c) \mid Q(c,S)$

- signal space: finite strings from alphabet $\Sigma = \{a,b,c,\ldots,z\}$
Representation of Production Competence  (≈ Teacher Behavior)

- **definite clause grammar**, e.g.:

  \[ S/P(c_1,c_2) \rightarrow N/c_1 \quad V/P \quad N/c_2 \]

- \[ V/\text{love} \rightarrow g \]

- \[ N/\text{John} \rightarrow ff \]

- \[ N/\text{Mary} \rightarrow h \]

- \[ S/\text{love}(\text{Mary},\text{Mary}) \rightarrow lkjaa \]

- **informed innovation**: if no rule available for coding a given meaning, then . . .
  - choose the most similar meaning that you can express
  - make the smallest ("lowest") change to the parse tree to express new meaning
Grammar Induction  

- input: string-meaning pairs  
  e.g.: observe \( \langle ffgh, \text{love(John, Mary)} \rangle \)

- add holistic rule  
  e.g.: add rule \( S / \text{love(John, Mary)} \rightarrow ffgh \)

- if possible, **merge** or **chunk** rules
Chunk

• suppose we have rules:
  \[ S/\text{love}(\text{John}, \text{Mary}) \rightarrow abc \]
  \[ S/\text{love}(\text{John}, \text{Sue}) \rightarrow abd \]

• replace by least general rule that subsumes both:
  \[ S/\text{love}(\text{John}, c_1) \rightarrow ab \ N/c_1 \]

• add appropriate rule for invented category \( N \):
  \[ N/\text{Mary} \rightarrow c \]
  \[ N/\text{Sue} \rightarrow d \]
Merge (example)

- suppose we have rules:
  \[ N/Mary \rightarrow g \]
  \[ M/Mary \rightarrow g \]
- then drop, e.g., the \( M \)-based rule
- replace non-terminal \( M \) with \( N \) throughout
Results

from Kirby and Hurford (2002)
Reflection

How good and account of emerging structure is the ILM?
Homework

• read my project proposals:
  • http://www.sfs.uni-tuebingen.de/~mfranke/MoLE2011/material/project_proposals.pdf

• find out more about your favorites

• if none of my proposals suits you, prepare a proposal yourself to present in class next week

Background Reading Material for this Session

Please Join Our Experiment!

- 8 Euro for ca. 40min
- make appointment: versucheb1@sfb833.uni-tuebingen.de

“Genau eine Schere ist mit keiner ihrer blauen Formen verbunden.”

Wahr □

Falsch □
References


