Part III

Learning structured representations
Hierarchical Bayesian models
Universal Grammar

Grammar

Phrase structure

Utterance

Speech signal

Hierarchical phrase structure grammars (e.g., CFG, HPSG, TAG)

\[
S \rightarrow NP \ VP
\]

\[
NP \rightarrow \text{Det} \ [\text{Adj}] \ \text{Noun} \ [\text{RelClause}]
\]

\[
\text{RelClause} \rightarrow \ [\text{Rel}] \ NP \ V
\]

\[
VP \rightarrow VP \ NP
\]

\[
VP \rightarrow \text{Verb}
\]

\[
S
\]

\[
NP
\]

\[
\text{Pronoun}
\]

\[
\text{Verb}
\]

\[
\text{Article}
\]

\[
\text{Noun}
\]

I shoot the wumpus
Outline

• Learning structured representations
  – grammars
  – logical theories

• Learning at multiple levels of abstraction
A historical divide

Structured Representations

Innate knowledge

vs

Unstructured Representations

Learning

(Chomsky, Pinker, Keil, ...)

(McClelland, Rumelhart, ...)
Representations

Causal networks

asbestos

lung cancer

coughing  chest pain

Grammars

\[ S \rightarrow NP \ VP \]
\[ NP \rightarrow Det \ [Adj] \ Noun \ [Rel\Clause] \]
\[ Rel\Clause \rightarrow [Rel] \ NP \ V \]
\[ VP \rightarrow VP \ NP \]
\[ VP \rightarrow Verb \]

Logical theories

\[ \forall x \ y \ Sibling(x, y) \leftarrow Sibling(y, x) \]
\[ \forall x \ y \ Ancestor(x, y) \leftarrow Parent(x, y) \]
Representations

Phonological rules

\[ [+syllabic] \rightarrow [+back] / [+syllabic] \]

Semantic networks

```
  ANIMAL
  /|
  / |
Has skin
Can move around
Eats
Breathes

  / |
  BIRD
Has wings
Can fly
Has feathers

  / |
  FISH
Has fins
Can swim
Has gills

  / |
  CANARY
Can sing
Is yellow

  / |
  OSTRICH
Has long thin legs
Is tall
Can't fly

  / |
  SHARK
Can bite
Is dangerous

  / |
  SALMON
Is pink
Is edible
Swims upstream to lay eggs
```
How to learn a $R$

- Search for $R$ that maximizes

$$P(R|\text{Data}) \propto P(\text{Data}, R) P(R)$$

- Prerequisites
  - Put a prior over a hypothesis space of Rs.
  - Decide how observable data are generated from an underlying $R$. 
How to learn a $R$

- Search for $R$ that maximizes

\[ P(R|\text{Data}) \propto P(\text{Data}, R) P(R) \]

- Prerequisites
  - Put a prior over a hypothesis space of Rs.
  - Decide how observable data are generated from an underlying $R$. 
Context free grammar

S → N VP  
VP → V  
N → “Alice”  
V → “scratched”  
VP → V N  
N → “Bob”  
V → “cheered”
Probabilistic context free grammar

1.0  
S → N VP  
   VP → V  
   VP → V N  
   N → “Alice”  
   V → “scratched”  
   N → “Bob”  
   V → “cheered”

S 1.0
  N 0.5  VP 0.6
   Alice V
   cheered

S 1.0
  N 0.5  VP 0.4
   Alice V
   scratched Bob

probability = 1.0 * 0.5 * 0.6  
= 0.3

probability = 1.0*0.5*0.4*0.5*0.5  
= 0.05
The learning problem

Grammar G:

\[
\begin{align*}
S & \rightarrow N \ VP \\
VP & \rightarrow V \\
N & \rightarrow \text{“Alice”} \\
VP & \rightarrow V N \\
N & \rightarrow \text{“Bob”} \\
V & \rightarrow \text{“scratched”} \\
V & \rightarrow \text{“cheered”}
\end{align*}
\]

Data D:

Alice scratched.  Alice cheered.
Bob scratched.   Bob cheered.
Alice scratched Alice. Alice cheered Alice.
Alice scratched Bob. Alice cheered Bob.
Bob scratched Alice. Bob cheered Alice.
Grammar learning

• Search for $G$ that maximizes

\[ P(G|\text{Data}) \propto P(\text{Data}|G)P(G) \]

• Prior: \[ P(G) \propto 2^{-\text{length}(G)} \]

• Likelihood: \[ P(\text{Data}|G) \]
  – assume that sentences in the data are independently generated from the grammar.

(Horning 1969; Stolcke 1994)
Experiment

- Data: 100 sentences

\[
\begin{align*}
S & \rightarrow NP \ VP \\
NP & \rightarrow \text{Det} \ N \\
VP & \rightarrow \text{Vt} \ NP \\
& \quad \rightarrow \text{Vc} \ PP \\
& \quad \rightarrow \text{Vi} \\
PP & \rightarrow P \ NP \\
\text{Det} & \rightarrow a \\
& \quad \rightarrow \text{the} \\
\text{Vt} & \rightarrow \text{touches} \\
& \quad \rightarrow \text{covers} \\
\text{Vc} & \rightarrow \text{is} \\
\text{Vi} & \rightarrow \text{rolls} \\
& \quad \rightarrow \text{bounces} \\
\text{N} & \rightarrow \text{circle} \\
& \quad \rightarrow \text{square} \\
& \quad \rightarrow \text{triangle} \\
\text{P} & \rightarrow \text{above} \\
& \quad \rightarrow \text{below}
\end{align*}
\]

the circle covers a square
a square is above the triangle
a circle bounces

(Stolcke, 1994)
Generating grammar:

S → NP VP
NP → Det N
VP → Vt NP
    → Vc PP
    → Vi
PP → P NP
Det → a
    → the
Vt → touches
    → covers
Vc → is
Vi → rolls
    → bounces
N → circle
    → square
    → triangle
P → above
    → below

Model solution:

S → NP VP
NP → Det N
VP → VI
    → X NP
X → VT
    → VC P
Det → a
    → the
Vt → touches
    → covers
Vc → is
Vi → rolls
    → bounces
N → circle
    → square
    → triangle
P → above
    → below
Predicate logic

• A compositional language

\[ \forall x \ y \ \text{Sibling}(x, y) \leftarrow \text{Sibling}(y, x) \]

For all x and y, if y is the sibling of x then x is the sibling of y

\[ \forall x \ y \ z \ \text{Ancestor}(x, z) \leftarrow \text{Ancestor}(x, y) \land \text{Ancestor}(y, z) \]

For all x, y and z, if x is the ancestor of y and y is the ancestor of z, then x is the ancestor of z.
Learning a kinship theory

Theory T:
\[ \forall x \, y \; \text{Sibling}(x, y) \leftarrow \text{Sibling}(y, x) \]
\[ \forall x \, y \, z \; \text{Ancestor}(x, z) \leftarrow \text{Ancestor}(x, y) \land \text{Ancestor}(y, z) \]
\[ \forall x \, y \; \text{Ancestor}(x, y) \leftarrow \text{Parent}(x, y) \]
\[ \forall x \, y \, z \; \text{Uncle}(x, z) \leftarrow \text{Brother}(x, y) \land \text{Parent}(y, z) \]

Data D:
\[
\text{Sibling}(\text{victoria}, \text{arthur}), \quad \text{Sibling}(\text{arthur}, \text{victoria}), \\
\text{Ancestor}(\text{chris}, \text{victoria}), \quad \text{Ancestor}(\text{chris}, \text{colin}), \\
\text{Parent}(\text{chris}, \text{victoria}), \quad \text{Parent}(\text{victoria}, \text{colin}), \\
\text{Uncle}(\text{arthur}, \text{colin}), \quad \text{Brother}(\text{arthur}, \text{victoria}) \quad \cdots
\]

(Hinton, Quinlan, ...)
Learning logical theories

• Search for T that maximizes

\[ P(T|\text{Data}) \propto P(\text{Data}|T)P(T) \]

• Prior: \( P(T) \propto 2^{-\text{length}(T)} \)

• Likelihood: \( P(\text{Data}|T) \)
  – assume that the data include all facts that are true according to T

(Conklin and Witten; Kemp et al 08; Katz et al 08)
Theory-learning in the lab

R(f,c) R(k,c) R(c,b)
R(f,l) R(k,l) R(c,l)
R(f,b) R(k,b) R(l,b)
R(k,h) R(c,h) R(b,h)

(cf Krueger 1979)
Theory-learning in the lab

Transitive: \( R(f,k). \ R(k,c). \ R(c,l). \ R(l,b). \ R(b,h). \)

\( R(X,Z) \leftarrow R(X,Y), \ R(Y,Z). \)
Learning time

Complexity

Theory length

(Kemp et al 08)
Conclusion: Part 1

• Bayesian models can combine structured representations with statistical inference.
Outline

• Learning structured representations
  – grammars
  – logical theories

• Learning at multiple levels of abstraction
Vision

(Han and Zhu, 2006)
Motor Control

symbolic representation of tasks e.g. goal

mid-level representation e.g. sequences of elements

low level dynamics e.g. elements of movements

(Wolpert et al., 2003)
Causal learning

- chemicals
- diseases
- symptoms

- asbestos
- lung cancer
- coughing
- chest pain

- mercury
- minamata disease
- muscle wasting

Patient 1: asbestos exposure, coughing, chest pain
Patient 2: mercury exposure, muscle wasting

(Kelley; Cheng; Waldmann)
Universal Grammar

Grammar

Phrase structure

Utterance

Speech signal

Hierarchical phrase structure grammars (e.g., CFG, HPSG, TAG)

\[ S \rightarrow NP \ VP \]

\[ NP \rightarrow \text{Det} \ [\text{Adj}] \ Noun \ [\text{RelClause}] \]

\[ \text{RelClause} \rightarrow [\text{Rel}] \ NP \ V \]

\[ VP \rightarrow VP \ NP \]

\[ VP \rightarrow \text{Verb} \]

P(grammar | UG)

P(phrase structure | grammar)

P(utterance | phrase structure)

P(speech | utterance)
A hierarchical Bayesian model specifies a joint distribution over all variables in the hierarchy:

\[
P(\{u_i\}, \{s_i\}, G \mid U)
\]

\[
= P(\{u_i\} \mid \{s_i\}) \ P(\{s_i\} \mid G) \ P(G \mid U)
\]
Top-down inferences

Universal Grammar
  ↓
Grammar
  ↓
Phrase structure
  ↓
Utterance

\[ \text{U} \quad \text{G} \]

\[ \begin{align*}
\text{S}_1 & \quad \text{S}_2 \\
\text{S}_3 & \quad \text{S}_4 \\
\text{S}_5 & \quad \text{S}_6 \\
\text{u}_1 & \quad \text{u}_2 \\
\text{u}_3 & \quad \text{u}_4 \\
\text{u}_5 & \quad \text{u}_6
\end{align*} \]

Infer \( \{s_i\} \) given \( \{u_i\}, G \):

\[ P( \{s_i\} | \{u_i\}, G) \propto P( \{u_i\} | \{s_i\} ) P( \{s_i\} | G) \]
Infer G given \( \{s_i\} \) and U:

\[
P(G| \{s_i\}, U) \propto P( \{s_i\} | G) P(G|U)
\]
Simultaneous learning at multiple levels

Universal Grammar
  ↓
Grammar
  ↓
Phrase structure
  ↓
Utterance

Infer $G$ and $\{s_i\}$ given $\{u_i\}$ and $U$:

$$P(G, \{s_i\} | \{u_i\}, U) \propto P(\{u_i\} | \{s_i\})P(\{s_i\} | G)P(G | U)$$
Word learning

Words in general
  ↓
Individual words
  ↓
Data

Whole-object bias
  Shape bias

car
monkey
duck
gavagai
A hierarchical Bayesian model

\[
θ ~ \text{Beta}(F_H, F_T)
\]

- Qualitative physical knowledge (symmetry) can influence estimates of continuous parameters \((F_H, F_T)\).
- Explains why 10 flips of 200 coins are better than 2000 flips of a single coin: more informative about...
Word Learning

“This is a dax.”

“Show me the dax.”

• 24 month olds show a shape bias
• 20 month olds do not

(Landau, Smith & Gleitman)
Is the shape bias learned?

- Smith et al (2002) trained 17-month-olds on labels for 4 artificial categories:

- After 8 weeks of training 19-month-olds show the shape bias:

“wib”
“lug”
“zup”
“div”

“This is a dax.”
“Show me the dax.”
Learning about feature variability

(cf. Goodman)
Learning about feature variability

(cf. Goodman)
A hierarchical model

Meta-constraints

↓

Bags in general

↓

Bag proportions

↓

Data

↓

Color varies across bags but not much within bags

mostly red

mostly yellow

mostly brown

mostly green

... mostly blue?
A hierarchical Bayesian model

Meta-constraints

Bags in general

Bag proportions

Data

\[ \alpha = 0.1 \]

\[ \beta = [0.4, 0.4, 0.2] \]

Within-bag variability

Data

[1,0,0] [0,1,0] [1,0,0] [0,1,0] \ldots [0.0,0,1]

\[ [6,0,0] [0,6,0] [6,0,0] [0,6,0] \ldots [0,0,1] \]
A hierarchical Bayesian model
Shape of the Beta prior

$$F_H = 0.5, F_T = 0.5 \quad F_H = 0.5, F_T = 2$$

$$F_H = 2, F_T = 0.5 \quad F_H = 2, F_T = 2$$
A hierarchical Bayesian model

Meta-constraints

Bags in general

Bag proportions

Data

\[
p(\{y^i\}, \{\theta^i\}, \alpha, \beta | \lambda)
\]
A hierarchical Bayesian model

Meta-constraints → Bags in general → Bag proportions → Data

$\alpha \sim \text{Exponential}(\lambda)$
$\beta \sim \text{Dirichlet}(1)$
$\theta^i \sim \text{Dirichlet}(\alpha, \beta)$
$y^i \sim \text{Multinomial}(\theta^i)$

$p(\{\theta^i\}, \alpha, \beta|\{y^i\}, \lambda)$
Learning about feature variability

Meta-constraints

Categories in general

Individual categories

Data

\( \alpha, \beta \)

\( \theta^1 \)

\( \theta^2 \)

\( \theta^3 \)

\( \theta^4 \)

\( \theta^5 \)
<table>
<thead>
<tr>
<th>Category</th>
<th>“wib”</th>
<th>“lug”</th>
<th>“zup”</th>
<th>“div”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>1 1 2 2 3 3 4 4</td>
<td>1 1 2 2 3 3 4 4</td>
<td>5 5 6 6</td>
<td>5 5 6 6</td>
</tr>
<tr>
<td>Texture</td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
<td>9 10 9 10</td>
<td>9 10 9 10</td>
</tr>
<tr>
<td>Color</td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8</td>
<td>9 10 10 9</td>
<td>9 10 10 9</td>
</tr>
<tr>
<td>Size</td>
<td>1 2 1 2 1 2 1 2</td>
<td>1 2 1 2 1 2 1 2</td>
<td>1 1 1 1</td>
<td>1 1 1 1</td>
</tr>
</tbody>
</table>

“dax”
Model predictions

Choice probability

"Show me the dax:"
Where do priors come from?

Meta-constraints

Categories in general

Individual categories

Data

$\theta^1$  $\theta^2$  $\theta^3$  $\theta^4$  $\theta^5$

$\alpha, \beta$
### Knowledge representation

#### Mendeleev’s Periodic Table of 1869

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ti</td>
<td>Zr</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>Nb</td>
<td>Ta</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>94</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td>Cr</td>
<td>Mo</td>
<td>W</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>96</td>
<td>186</td>
</tr>
<tr>
<td></td>
<td>Mn</td>
<td>Rh</td>
<td>Pt</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>104.4</td>
<td>197.4</td>
</tr>
<tr>
<td></td>
<td>Fe</td>
<td>Ru</td>
<td>Ir</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>104.4</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td>Ni, Co</td>
<td>Pd</td>
<td>Os</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td>106.6</td>
<td>199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>Mg</td>
<td>Zn</td>
</tr>
<tr>
<td>9.4</td>
<td>24</td>
<td>65.2</td>
<td>112</td>
</tr>
<tr>
<td>B</td>
<td>11</td>
<td>Al</td>
<td>?</td>
</tr>
<tr>
<td>27.4</td>
<td></td>
<td>68</td>
<td>116</td>
</tr>
<tr>
<td>C</td>
<td>12</td>
<td>Si</td>
<td>?</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>70</td>
<td>118</td>
</tr>
<tr>
<td>N</td>
<td>14</td>
<td>P</td>
<td>As</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>75</td>
<td>122</td>
</tr>
<tr>
<td>O</td>
<td>16</td>
<td>S</td>
<td>Se</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>79.4</td>
<td>128</td>
</tr>
<tr>
<td>F</td>
<td>19</td>
<td>Cl</td>
<td>Br</td>
</tr>
<tr>
<td>35.5</td>
<td></td>
<td>80</td>
<td>127</td>
</tr>
<tr>
<td>Li</td>
<td>7</td>
<td>Na</td>
<td>K</td>
</tr>
<tr>
<td>19</td>
<td>23</td>
<td>39</td>
<td>85.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ca</td>
<td>Sr</td>
<td>Ba</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>87.6</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>Ce</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Er?</td>
<td>La</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td></td>
<td>94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yb?</td>
<td>Dy</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td></td>
<td>95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>In</td>
<td>Th</td>
<td></td>
</tr>
<tr>
<td>75.6?</td>
<td></td>
<td>118?</td>
<td></td>
</tr>
</tbody>
</table>
The discovery of structural form

BIOLOGY
- mouse
- squirrel
- chimp
- gorilla

POLITICS
- Ginsburg
- Scalia
- Stevens
- Thomas

COLOR

FRIENDSHIP

CHEMISTRY
Children discover structural form

• Children may discover that
  – Social networks are often organized into cliques
  – The months form a cycle
  – “Heavier than” is transitive
  – Category labels can be organized into hierarchies
A hierarchical Bayesian model

Meta-constraints
  ↓
  Form
  ↓
  Structure
  ↓
  Data

M
  ↓
  Tree
  ↓
  mouse
  ↓
  squirrel
  ↓
  chimp
  ↓
  gorilla

<table>
<thead>
<tr>
<th></th>
<th>whiskers</th>
<th>hands</th>
<th>tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>mouse</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>squirrel</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>chimp</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>gorilla</td>
<td>○</td>
<td>●</td>
<td>○</td>
</tr>
</tbody>
</table>
A hierarchical Bayesian model

Meta-constraints

M

F: form

S: structure

D: data

Tree

mouse
squirrel
chimp
gorilla

Meta-constraints

M

P(S, F | D, n) \propto P(D | S) P(S | F, n) P(F)
Structural forms

Partition  Order  Chain  Ring

Hierarchy  Tree  Grid  Cylinder
$P(S|F,n)$: Generating structures

- Each structure is weighted by the number of nodes it contains:

$$P(S|F) \propto \begin{cases} 0 & \text{if } S \text{ inconsistent with } F \\ \theta(1 - \theta)|S| & \text{otherwise} \end{cases}$$

where $|S|$ is the number of nodes in $S$
P(S|F, n): Generating structures from forms

- Simpler forms are preferred

\[ P(S|F, n) \]: Generating structures from forms

All possible graph structures S

Chain

Grid
A hierarchical Bayesian model

Meta-constraints

F: form

S: structure

D: data

Tree

Mouse

Squirrel

Chimp

Gorilla

Meta-constraints

F: form

S: structure

D: data

Tree

Mouse

Squirrel

Chimp

Gorilla

P(S, F | D, n) \propto P(D | S) P(S | F, n) P(F)
p(D|S): Generating feature data

- Intuition: features should be smooth over graph S

Relatively smooth

Not smooth
$p(D|S)$: Generating feature data

Let $f_i$ be the feature value at node $i$

\[
p(f) \propto \exp \left( -\frac{1}{4} \sum_{i,j} \frac{(f_i - f_j)^2}{d_{ij}} - \frac{1}{2\sigma} f^T f \right)
\]

(Zhu, Lafferty & Ghahramani)
A hierarchical Bayesian model

\[ P(S, F | D, n) \propto P(D | S) P(S | F, n) P(F) \]
Feature data: results

animals

features

judges
cases
Developmental shifts

5 features

20 features

110 features
Similarity data: results
Relational data: results

Primates

“x dominates y”

Bush cabinet

“x tells y”

Prisoners

“x is friends with y”

Diagram:

- Primates graph
- Bush cabinet network
- Prisoners diagram
Universal Structure grammar

Form

Structure

Data

mouse
squirrel
chimp
gorilla

whiskers
hands
tail
Node-replacement graph grammars

Production (Chain)

Derivation
A hypothesis space of forms

Product of two chains

Product of a chain and a circle
The complete space of grammars

1

\[
\begin{array}{c}
1 \\
\vdots \\
4096
\end{array}
\]

\[
\begin{array}{c}
\Rightarrow \\
\Rightarrow
\end{array}
\]
Universal Structure grammar $U$

Form

Structure

Data

<table>
<thead>
<tr>
<th></th>
<th>whiskers</th>
<th>hands</th>
<th>tail</th>
<th>feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>mouse</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>squirrel</td>
<td>●</td>
<td>○</td>
<td>●</td>
<td>?</td>
</tr>
<tr>
<td>chimp</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>?</td>
</tr>
<tr>
<td>gorilla</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>?</td>
</tr>
</tbody>
</table>
Conclusions: Part 2

• Hierarchical Bayesian models provide a unified framework which helps to explain:
  – How abstract knowledge is acquired
  – How abstract knowledge is used for induction
Outline

• Learning structured representations
  – grammars
  – logical theories

• Learning at multiple levels of abstraction
9. STOCHASTIC LEARNING THEORY 1
   by Saul Sternberg, University of Pennsylvania

10. STIMULUS SAMPLING THEORY 121
    by Richard C. Atkinson, Stanford University
    and William K. Estes, Stanford University

11. INTRODUCTION TO THE FORMAL ANALYSIS OF NATURAL LANGUAGES 269
    by Noam Chomsky, Massachusetts Institute of Technology
    and George A. Miller, Harvard University

12. FORMAL PROPERTIES OF GRAMMARS 323
    by Noam Chomsky, Massachusetts Institute of Technology