Psych229: Language Acquisition

Lecture 20
Syntax Learning

Baker (2001): Complex Systems

Similarities & Differences: Parameters
Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages.

Idea: A relatively small number of parameters yields a large number of different languages.

English
French
Japanese
Tagalog
Navajo

Baker (2001): Complex Systems

Similarities & Differences: Parameters
Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the environment determines what values these parameters should have.

Baker (2001): Complex Systems

Similarities: Greenberg’s Generalizations
Word Order Generalizations

Navajo
Japanese

Basic word order: Subject Object Verb
Basic word order: Subject Object Verb

Postpositions: Noun Phrase Postposition
Postpositions: Noun Phrase Postposition

Ashkii átédé bahííbitá
Boy girl saw

Jareth-ga Hoggle-o butta
Jareth hit Hoggle

The boy saw the girl

Jareth went to London with Sarah by car.

Toby-no imooto-ga
Toby’s sister

Chidí bi-jáád
the wheel of a car

Postpositions: Noun Phrase Postposition
Postpositions: Noun Phrase Postposition

Possessor before Possessed Possessor Possession

Chidí imo-leg
Car

London ni ita
London to went

Jareth went to London with Sarah by car.
Baker (2001): Complex Systems

Similarities: Greenberg’s Generalizations
Word Order Generalizations

Navajo
Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition
Possessor before Possessed
Possessor Possession

Japanese
Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition
Possessor before Possessed
Possessor Possession

Navajo
Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition
Possessor before Possessed
Possessor Possession

Japanese
Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition
Possessor before Possessed
Possessor Possession

English
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase
Sarah found Toby.

Edo (Nigeria)
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase
Ozó mién Adésuwá Ozo found Adesuwa.

English
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase
Sarah found Toby.

Edo (Nigeria)
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase
Ozó mién Adésuwá Ozo found Adesuwa.

English
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase
Sarah found Toby.

Edo (Nigeria)
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase
Ozó mién Adésuwá Ozo found Adesuwa.

Jareth gave the crystal to Sarah.

Ozo rhié néné ebé né Adésuwá Ozo gave the book to Adesuwa.

Possessed before Possessor
Possession
Possessor

Ozo Ozó child Ozo

Ozo’s child

Possessed before Possessor
Possession
Possessor

Ozo Ozó child Ozo

Ozo’s child

Point: Forty-five “universals” of languages found - patterns overwhelmingly followed by languages with unshared history (Navajo & Japanese, English & Edo)

Not all combinations are possible - some patterns rarely appear

Ex: Subject Verb Object language (English/Edo-like) + postpositions (Navajo/Japanese-like)
### Baker (2001): Complex Systems

<table>
<thead>
<tr>
<th>More Similarities &amp; Differences</th>
<th>French vs. Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject Verb</strong></td>
<td><strong>Subject Verb</strong></td>
</tr>
<tr>
<td>Jareth arriverà.</td>
<td>Jareth verrà.</td>
</tr>
</tbody>
</table>

**Embedded Subject-question formation (easy to miss)**

*Qui veux-tu que Marie épouse Jean?*
Who want-you that Marie marries Jean?

*Che credi che Jareth verrà?*
Who think-you that Jareth will-come?

**Expletives**

Pluv. It-rains.

Tu veux que Marie épouse Jean. You want that Marie marries Jean.

Che credi che Jareth verrà. You think Jareth will-come.

All these involve the subject in some way - coincidence?

**Idea:** No! Parameter involving the subject.
Yang (2004): Learning Complex Systems

Language is a complex system

Only humans seem able to learn human language.
Something in our biology must allow us to do this. Chomsky: Universal Grammar is innate biases for learning language.

But obviously language is learned, not just prespecified beforehand.
Constrained variation across languages: phonology, lexicon, structure.

The point: need innate biases & probabilistic learning abilities
Need to explicitly integrate them with each other.

Yang (2004): Learning Complex Systems

Statistics for word segmentation (remember Gambell & Yang (2006))

"Modeling shows that the statistical learning (Saffran et al. 1996) does not reliably segment words such as those in child-directed English. Specifically, precision is 41.6%, recall is 23.3%. In other words, about 60% of words postulated by the statistical learner are not English words, and almost 80% of actual English words are not extracted. This is so even under favorable learning conditions."

Unconstrained (simple) statistics: not so good.

If statistical measure is constrained by language-specific knowledge (words have only one main stress), performance increases dramatically: 73.5% precision, 71.2% recall.

Constrained statistics - much better!

Yang (2004): Learning Complex Systems

Combining statistics with Universal Grammar

A big deal:
"Although infants seem to keep track of statistical information, any conclusion drawn from such findings must presuppose that children know what kind of statistical information to keep track of."

Ex: Transitional Probability
...of rhyming syllables?
...of syllables with nasal consonants?
...of syllables of the form CV (ba, ti)?

Universal Grammar: Principles & Parameters

Principles: Apply to all human languages.
Ex: Language has hierarchical structure.
Smaller units are chunked into larger units.

- sounds: g a b l i n
- syllables: g a b l i n
- words: goblin
- phrases: Noun Phrase (NP) - Verb Phrase (VP)
- sentences: The sneaky goblin stole the baby
Yang (2004): Learning Complex Systems

Universal Grammar: Principles & Parameters
Parameters: Constrained variation across languages. Child must learn which option native language uses.

Japanese/Navajo
Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition
Possessor before Possessed Possessor Possession

Edo/English
Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase Possessed before Possessor Possessor Possession

Learning Parametric Systems: Triggering Grammar = combination of parameter values

Trigger Learning:
At any given time, the child has in mind a single grammar.
If this current grammar can successfully analyze the current data, it stays. Otherwise, the child will shift to a completely new grammar by altering one or more parameter values. This new grammar will (hopefully) be able to analyze the current data.

Learning trajectory expectation: Sudden shifts in performance, not gradual. This is problematic.

The Learning Algorithm
For each data point d encountered in the input
Choose a grammar probabilistically from available grammars
If this grammar can analyze the data point, increase its probability slightly (reward)
Else decrease its probability slightly (punish)

The Basic Idea
If there is a single target grammar (the usual case), the non-target grammars will be chosen to analyze data at some point and be unsuccessful. Each time this happens, they will lose some probability.
Grammars compete against each other to see which can best analyze the available data.

The Basic Idea
If there is a single target grammar (the usual case), the non-target grammars will be chosen to analyze data at some point and be unsuccessful. Each time this happens, they will lose some probability.

The target grammar, in contrast, is always able to analyze the data and so will always increase in probability. It will eventually win out over the non-target grammars. (Probability $\approx 1.0$)

The Main Force
The crucial data is that which is unambiguous for the target grammar: this data is incompatible with non-target grammars. The more unambiguous data there is, the faster the target grammar will win.

Added perk: Learning is then gradual (probabilistic).

Problem: Does unambiguous data exist for entire grammars? This requires data that is incompatible with every other possible parameter of every other possible grammar.

The Learning Algorithm
For each data point encountered in the input
Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward)
The Learning Algorithm
For each data point encountered in the input
Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.
If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward)
Else decrease the probability of all participating parameters values slightly (punish)

Yang (2004): Learning Complex Systems
Learning Parametric Systems: Variational Learning
Grammars compete against each other to see which can best analyze the available data.

The Main Force
The crucial data is that which is unambiguous for the target parameter values: this data is incompatible with non-target parameter values.
The more unambiguous data there is, the faster the target grammar will win.

Added perk remains: Learning is still gradual (probabilistic).
Problem ameliorated: unambiguous data much more likely to exist for individual parameter values instead of entire grammars.

What happens for an English-learning child?
Pro-drop languages depend on rich subject-verb agreement morphology. English doesn’t have that, which is something a child will easily notice. Knock out +pro-drop grammars.
Yang (2004): Learning Complex Systems

Variational Learning: Sample Case
Null subjects: 2 binary parameters, 4 grammars

What happens for an English-learning child?
But this still leaves the +topic-drop option. What data will rule that out?
Answer: Expletive subjects. (Can’t topic-drop them.)

Chinese

What happens for an English-learning child?
But this still leaves the +topic-drop option. What data will rule that out?
Answer: Expletive subjects. (Can’t topic-drop them.)

Chinese

Yang (2004): Learning Complex Systems

Variational Learning: General Predictions
The time course of when a parameter is set depends on how frequent the necessary evidence is in child-directed speech.
Parameters set early: more unambiguous data
Parameters set late: less unambiguous data
Parameters set at the same time: equal quantity of unambiguous data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Early set</th>
<th>Late set</th>
<th>Both at same time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expletive</td>
<td>High</td>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>Noun</td>
<td>High</td>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>Verb</td>
<td>High</td>
<td>Low</td>
<td>Middle</td>
</tr>
</tbody>
</table>

A look at real language properties
Possible categories in a language:
Determiners (“the”, “a”), Nouns (goblin, child), Adverbs (easily), Verbs (steals)

The goblin easily steals the child.

If the child only ever sees this, no way to know how the phrases break up.

A look at real language properties in action with transitional probabilities
Example: Optional phrases
Possible categories in a language:
Determiners (“the”, “a”), Nouns (goblin, child), Adverbs (easily), Verbs (steals)

The goblin easily steals the child.

If the child only ever sees this, no way to know how the phrases break up.

With the optional phrase left out, the probability of (B C D E F) is less than 1. Post a phrase boundary there. AB is a unit, DEF is a unit.

The goblin steals the child.

**Artificial language**

<table>
<thead>
<tr>
<th>Baseline pattern: ABCDEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsense Words Assigned to Each Form Class</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>jox</td>
</tr>
</tbody>
</table>

**Optional language**

| Phrases to be extracted: A B C D E F |

**Grammatical strings:**

A B C D E F, C D E F A, A B E F, A B C F D

**Example strings heard:**

kof, hox, jes, sot, fal, ker, rel, zor, taf, nav, mer, neb, rud, sib, daz, lev, tid, lum

**Baseline pattern:**

| A | B | C | D | E | F |
|--------------------------|

**Optional control language**

**Control strings:**

A B C D E F, A B C D E, B C D E F, A B E F, A B C F D, A B C D F

**Stimuli:** 68 of possible 94

**Half canonical:** A B C D E F

**Half distributed among other patterns**

**Assessment of learning**

**Sentence Learning**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 0.05</td>
<td>p = 0.028</td>
</tr>
</tbody>
</table>

**Phrase Learning**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 0.004</td>
<td>p = 0.0018</td>
</tr>
</tbody>
</table>


**Artificial language**

<table>
<thead>
<tr>
<th>Baseline pattern: ABCDEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optional language (remove one phrase at a time)</td>
</tr>
<tr>
<td>Phrases to be extracted: A B C D E F</td>
</tr>
</tbody>
</table>

**Grammatical strings:**

A B C D E F, C D E F A, A B E F, A B C F D

**Example strings heard:**

kof, hox, jes, sot, fal, ker, rel, zor, taf, nav, mer, neb, rud, sib, daz, lev, tid, lum

**Baseline pattern:**

| A | B | C | D | E | F |
|--------------------------|

**Optional control language**

**Control strings:**

A B C D E F, A B C D E, B C D E F, A B E F, A B C F D, A B C D F

**Stimuli:** 68 of possible 94

**Half canonical:** A B C D E F

**Half distributed among other patterns**

**Assessment of learning**

**Sentence Learning**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 0.05</td>
<td>p = 0.028</td>
</tr>
</tbody>
</table>

**Phrase Learning**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 0.004</td>
<td>p = 0.0018</td>
</tr>
</tbody>
</table>


**Artificial language**

<table>
<thead>
<tr>
<th>Baseline pattern: ABCDEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optional language (remove one adjacent pair at a time)</td>
</tr>
<tr>
<td>Control strings: A B C D E F, A B C D E, B C D E F, A B E F, A B C F D, A B C D F</td>
</tr>
</tbody>
</table>

**Stimuli:** 68 of possible 94

**Half canonical:** A B C D E F

**Half distributed among other patterns**

**Assessment of learning**

**Sentence Learning**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 0.05</td>
<td>p = 0.028</td>
</tr>
</tbody>
</table>

**Phrase Learning**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p = 0.004</td>
<td>p = 0.0018</td>
</tr>
</tbody>
</table>

Artificial language
Baseline pattern: ABCDEF

Moved phrases language (1 of 6 legal permutations)
Phrases to be extracted: AB, CD, EF
Grammatical strings: ABCDEF, ABDECF, CDABEF, CDEFAE, DEABCF, EFCDAB

Example strings heard:
kofjoxeszor ker daz neb rel zor nav
Stimuli: 8D
Half canonical: ABCDEF
Half distributed among other patterns

Moved control language (move one adjacent pair at a time)
Control strings:
ABCDEF, ABDECF, CDABEF, CDEFAE, DEABCF, EFCDAB

Baselines pattern:
A B C D E F


Artificial language
Baseline pattern: ABCDEF

Class size variation language (2 or 4 words per class) Phrases to be extracted: AB, CD, EF
Grammatical strings: ABCDEF, ABEFCD, ABEDCF, AEFBCD, EABCFD, EFFCAD

Example strings heard:
kofjoxeszor ker daz neb rel zor nav
Stimuli: 8D
Half canonical: ABCDEF
Half distributed among other patterns

Moved control language (move one adjacent pair at a time)
Control strings:
ABCDEF, ABDECF, CDABEF, CDEFAE, DEABCF, EFCDAB

Example strings heard:
kofjoxeszor ker daz neb rel zor nav
Stimuli: 8D
Half canonical: ABCDEF
Half distributed among other patterns

Moved control language (move one adjacent pair at a time)
Control strings:
ABCDEF, ABDECF, CDABEF, CDEFAE, DEABCF, EFCDAB

Baseline pattern:
A B C D E F


Assessment of learning
Sentence, Day 1
T&N say: Experimental groups better than control for basic word order


Assessment of learning
Phrase, Day 1

Assessment of learning

Day 5

All-Comb Control

Mean Percent Correct

Phrases

TAN say: Experimental groups better than control for basic word order, except for class size

Human tendency towards binary groupings

Thompson & Newport (2007):

Artificial language

All-combined language (optional, repeated, moved, class size variation)

Phrases to be extracted: AB, CD, EF

Grammatical strings:

A-B C-D E-F

All-combined control

p = 0.000

Day 1

Day 5

Mean Percent Correct

Thompson & Newport (2007):

Idea for control subjects' sentence performance

Day 5

Day 5

Sentences: What if only 5% of the data are of the canonical form? No memorization possible, but the transitional probability peaks and valleys are still constant, so experimental condition subjects should still do well.

Phrases

p = 0.000

Discussion: Do we believe that this is strong evidence for the discovery of grammatical structure (and rules) via transitional probability?