Input & Universal Grammar

Charles Yang
Department of Linguistics, Computer Science & Psychology
Institute for Research in Cognitive Science
University of Pennsylvania

SLRF Formal and Functional Symposium
Pittsburgh 2012
Input & Universal Grammar

- Input: A statistical look at distributional information
- Output: Quantitative and cross-linguistic patterns of development
- Mechanisms of learning: Why UG can make use of input-driven, probabilistic, and domain general learning processes
- Some speculations on L2 acquisition
Input & Usage Effects?

- Frequency effects, limited extent of diversity ("verb islands"), etc.
- "give me X", a highly frequent expression, is often cited as evidence of the child using formulaic expressions
- From the Harvard children (Adam, Eve, Sarah)
  - give me: 93, give him: 15, give her: 12, or 7.75 : 1.23 : 1
  - me: 2870, him: 466, her: 364, or 7.88 : 1.28 : 1
Input: Very boring

• Zipf’s law: Much of language is repetitions of a few, while most distinct items occur rarely

• Linguistic combinations produce an even large space of possibilities (e.g., bigrams, trigrams, morphology, rules/constructions)
• Rules and their frequencies from the Penn Treebank (log-log scale)
## Verb Islands in adult language (>1Mil)

<table>
<thead>
<tr>
<th>Verb</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
</tr>
</thead>
<tbody>
<tr>
<td>put</td>
<td>401</td>
<td>164</td>
<td>124</td>
<td>15</td>
<td>12</td>
<td>12</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>tell</td>
<td>245</td>
<td>65</td>
<td>49</td>
<td>49</td>
<td>45</td>
<td>36</td>
<td>22</td>
<td>16</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>see</td>
<td>152</td>
<td>100</td>
<td>38</td>
<td>32</td>
<td>28</td>
<td>21</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>want</td>
<td>158</td>
<td>83</td>
<td>36</td>
<td>24</td>
<td>19</td>
<td>15</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>let</td>
<td>238</td>
<td>38</td>
<td>32</td>
<td>23</td>
<td>22</td>
<td>17</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>give</td>
<td>115</td>
<td>92</td>
<td>59</td>
<td>32</td>
<td>31</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>take</td>
<td>130</td>
<td>57</td>
<td>30</td>
<td>21</td>
<td>18</td>
<td>15</td>
<td>14</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>show</td>
<td>100</td>
<td>34</td>
<td>27</td>
<td>21</td>
<td>19</td>
<td>17</td>
<td>12</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>got</td>
<td>58</td>
<td>37</td>
<td>14</td>
<td>12</td>
<td>11</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>ask</td>
<td>45</td>
<td>41</td>
<td>27</td>
<td>24</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>make</td>
<td>67</td>
<td>20</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>eat</td>
<td>67</td>
<td>42</td>
<td>14</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>like</td>
<td>39</td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>bring</td>
<td>43</td>
<td>30</td>
<td>17</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>hear</td>
<td>46</td>
<td>22</td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>total</td>
<td>1904</td>
<td>838</td>
<td>501</td>
<td>301</td>
<td>252</td>
<td>189</td>
<td>137</td>
<td>109</td>
<td>88</td>
<td>75</td>
</tr>
</tbody>
</table>

put: it, your, them, him, my, her, em, you, his, water
PIN Number Analysis

Cumulative 4-digit password usage
Matches and Mismatches

• Roger Brown (1973, *A First Language*): word order errors are “triflingly few”

• Children must be able to learn the basic rules of grammar with 2-3 million sentences

• Yet a great deal of surprises remain, especially if we relate them to the distributions of linguistic patterns in the input
Abundant Input, Late Learning

  - __ want look a man.
- Missing objects as well (Wang et al. 1992)
  - Look at __. __ go a little higher
- Null subject stage last about 3 years but an overwhelming amount of child directed English input do contain the subject, as English is an obligatory subject language (unlike Chinese, Japanese, Spanish, Italian, etc.)
## Lateness is not Universal

<table>
<thead>
<tr>
<th>Adults</th>
<th>English</th>
<th>Italian</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>~0%</td>
<td>70%</td>
<td>50%</td>
</tr>
<tr>
<td>Object</td>
<td>0%</td>
<td>0%</td>
<td>20%</td>
</tr>
<tr>
<td>Children</td>
<td>Subject</td>
<td>30%</td>
<td>~70%</td>
</tr>
<tr>
<td>Object</td>
<td>8%</td>
<td>0%</td>
<td>~20%</td>
</tr>
</tbody>
</table>

Children age: <3;0  
Data from Wang et al. (1992), Valian (1991), Bates (1978)

“Luke, look at the input ...”
Abundant Input, Late Learning

- Extensive use of Root Infinitives that should be tensed
  - English: Papa have it.
  - Dutch: thee drinken (tea drink-INF)
  - French: Dormir petit bébé (sleep-INF little baby)
  - German: mein Kako hinstellen (my chocolate milk put-INF)
  - Hebrew: Malon lauf (balloon fly-INF)
Optional Infinitives and Null Subjects Together

Large dataset from a Dutch learner (data from Haegeman 1996)
Little Input, Early Learning

• The placement of verbs in French
  • Jean voit souvent/pas Claude. (“John sees often/not Claude”)

• Only 7% of the sentences in child-directed French show this pattern (Yang 2002), yet children learn this property of French by the time of two word combinations (1;8, Pierce 1992)
  • marches pas (“works not”)
  • pas la poupée dormir (“not the doll sleep”)

• Similar findings in similar languages, and languages like English pattern very differently
Same Grammar, Differential Learning

- Germanic languages have Verb Second (V2)
  - Dutch: Dit boek *las* ik gisteren. ("this book read I yesterday")
  - Norwegian: Det *vet* æ ikkje. ("that know I not")
- But Dutch and German children take over 3 years to use V2 reliably as they produce a lot of verb initial utterances (Clahsen 1986, Haegeman 1996), while Norwegian children learn V2 as early as 2;0 (Westergaard 2009)
Central Questions

- Can learning primarily consist of memorization and lexically specific rules?
- What combination of grammar model and learning model will give the best account of child language?
Industrial Lessons

- **Statistical parsing**: Learn from pre-parsed tree structures (e.g., Wall Street Journal, Brown Corpus)

- **Start**: a large set of probabilistic CFG rules
  \[
  S \xrightarrow{p} NP \ VP, \ S \xrightarrow{1-p} NP \ VP
  \]

- **Training**: adjust the probabilities of rule expansions so that they maximize the likelihood of the training data

- **Testing**: run the resulting grammar on new data

- **State of art parsers**: low 90% (impressive but still a long way to go)
Why Google hasn’t solved everything

(a) $VP \rightarrow V \ NP$
(b) $VP \rightarrow V_{\text{drink}} \ NP$
(c) $VP \rightarrow V_{\text{drink}} \ NP_{\text{water}}$

<table>
<thead>
<tr>
<th>Rule Type</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a+b+c$</td>
<td>~89.0%</td>
</tr>
<tr>
<td>$a+b$</td>
<td>~88.4%</td>
</tr>
<tr>
<td>$a$</td>
<td>~84%</td>
</tr>
</tbody>
</table>

- Multiple forms of rules are present in training, ranging from general to lexical
- One can vary certain types to test their effectiveness in generalization (Gildea 2001 *Proc. ACL*, Bikel 2004, *Comp. Ling.*)
- Storing construction or lexically specific rules offers virtually no payoff in data coverage (Yang 2011, *Proc. ACL*)
- The range of grammar (output) is enormous, but the learning data (input) is limited and grows far too slowly
Are the best rules good enough?

• Formal learnability is one thing; the developmental test from child language is ultimately more important

• $S \rightarrow NP\ VP$ will be learned quickly: >95% of the English data

• $VP \rightarrow V_{FIN} \ pas$ will be learned slowly: 7% of the French data

• But French children learn verb placement *early* and English children learn the use of subjects *late*!

• What kind of (grammar, learning) combination would take the input and produce the output like children?
UG + Learning from Input

- Parameters $\approx$ Principal Component Analysis
- “Child competence is identical to adult competence”
- “Parameters are set very early”
- Magic and More Magic ...
- Use parameters
  - a model of language variation and child learning errors
- Do not use Magic
  - use a model of learning that is gradual and takes input into account
From Trigger to Dimmer

Universal Grammar, statistics or both?

• The Variational model (Yang 2002 Oxford UP)
• Parameter values are associated with probabilities (p: VO, 1-p: OV)
  • try out a value, reward/punish) based on success/failure
  • learning rate: magnitude of change, subject to individual variation
• More tokens of parameter signatures, faster learning
Signatures & Learning

• Verb raising in French: 7% input, very early acquisition

• Learning the use of subject in English
  • Hearing “I eat pizza” doesn’t no good because it does not disambiguate the types of grammars the learner considers

• Expletive subject sentences
  • “There is a cookie on the floor” (1%)

• Signature for Chinese-type topic drop: null objects (12%)

• The most comprehensive study of a realistic parameter domain (Fodor & Sakas 2012 Language Acquisition) shows that most if not all parameters have signatures, which make learning feasible
From Input to Output

**TABLE 1**  Statistical Correlates of Parameters in the Input and Output of Language Acquisition

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Target</th>
<th>Signature</th>
<th>Input Frequency (%)</th>
<th>Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wh fronting</td>
<td>English</td>
<td>Wh questions</td>
<td>25</td>
<td>Very early</td>
</tr>
<tr>
<td>Topic-drop</td>
<td>Chinese</td>
<td>Null objects</td>
<td>12</td>
<td>Very early</td>
</tr>
<tr>
<td>Prodrop</td>
<td>Italian</td>
<td>Null subjects in questions</td>
<td>10</td>
<td>Very early</td>
</tr>
<tr>
<td>Verb raising</td>
<td>French</td>
<td>Verb adverb/pas</td>
<td>7</td>
<td>1.8</td>
</tr>
<tr>
<td>Obligatory subject</td>
<td>English</td>
<td>Expletive subjects</td>
<td>1.2</td>
<td>3.0</td>
</tr>
<tr>
<td>Verb second</td>
<td>German/Dutch</td>
<td>OVS sentences</td>
<td>1.2</td>
<td>3.0–3.2</td>
</tr>
<tr>
<td>Scope marking</td>
<td>English</td>
<td>Long-distance questions</td>
<td>0.2</td>
<td>&gt;4.0</td>
</tr>
</tbody>
</table>

- **Parameters have developmental correlates** (Yang 2012, *WIREs Cognitive Science*)

- **Same parameter, different languages:**
  - **V2 in Norwegian:** 10% of OVS in input→**early**
  - **V2 in Dutch:** 1.2% of OVS input→**late**
Input & Individual Variation


• No need to appeal to unmotivated and unnecessary theoretical machinery to account for the gaps between children and adults

• Optional Infinitives: verbal morphology that mark tense-you are not learning Chinese (Legate & Yang 2007 Lg. Acq.)

• Individual level correlation between length of OI stage and the amount of tensed morphology in CDS (Hadley et al. 2011 JSLHR)

• Suggests that the source of delay in SLI children may be due to (more general) learning: poor morphological learner (Leonard et al 1992, Rice et al. 2000) make less effective use of the tense information to unlearn the RI usage
Learning & Learning Language

• Variational Model uses Reinforcement Learning (Bush & Mosteller 1951), a very general learning mechanism with broad behavioral and neural support

• Strongly demonstrated in human subjects (children and adults)

• See especially the work in the acquisition of sociolinguistic variables (Labov and co.)
L2 Acquisition: Re-turning the dimmer?

- The combination of grammar model and learning model
- The Variational Model provides a precise and testable hypothesis for L2 acquisition research
  - Even if the grammar model is not parameter based
- If initial state is L1-independent:
  - L2 learners mirror the time course trajectories of L1 learners
- If initial state is L1:
  - L2 learners will eschew the time course of L1 learners
- It’s not sufficient to study a single parameter: cross-parameter comparison is necessary
Conclusion

• Input is rich and interesting, but it alone won’t do the job

  • Also need to be cautious about drawing conclusions from input effects

• Input and Universal Grammar are perfectly consistent

  • Previous conception of the learning mechanism needs to be reconsidered

• Input effects in L2 likewise may be assessed accurately by making the grammar-learning interaction very explicit