Learning Models

• All models are error-driven = if the current grammar can parse the sentence, retain it.

• The models differ with respect to what the learner does when its current grammar fails.
  • STLs select a grammar that can parse the current input sentence. (How can a psychologically realistic learner identify such a grammar? See Fodor, 1998b.)
  • The TLA selects a grammar similar to the current grammar, but adopts it only if it can parse the current input.

(a) Structural Triggers Learner: waiting model (Fodor, 1998a)
Parse input sentence with "supergrammar." The supergrammar contains all possible parameter values (in the form of structural treelets; see Fodor, 1998a, for details).
If the parser encounters a choice point, set no parameter values based on the rest of the sentence.

(b) Structural Triggers Learner: guessing models (Fodor, 1998b)
Parse input sentence with "supergrammar." The supergrammar contains all possible parameter values (in the form of structural treelets; see Fodor, 1998a, for details).
If the parser encounters a choice point it selects one grammar:
  (b1) which assigns the input sentence the fewest empty categories.
  (b2) which assigns the input sentence the most compact tree.
  (b3) which differs from the current grammar by the least number of parameters.

(c) Triggering Learning Algorithm (Gibson & Wexler, 1994)
Change any one parameter. Try parsing the sentence. Adopt the grammar if the parse succeeds.

(d) Random choice (baseline)
Guess any grammar.

Simulation procedure

1,000 learning trials (equivalent 1,000 children) per target grammar, per learning algorithm.
For each trial, measure how many sentences consumed by the learning model before attaining the target grammar.
More than 100,000 input sentences consumed = a ‘time-out’ failure
Subset Principle violations were excluded in advance.
1-parameter distant local maxima were excluded. All other local maxima were allowed to time-out.

Learning Efficiency

Data

<table>
<thead>
<tr>
<th>Failure Rate</th>
<th># of sents needed for 99% of learners to attain target</th>
<th>mean # of sents needed for attaining target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural Triggers Learner (STL) : Waiting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waiting STL</td>
<td>75</td>
<td>120</td>
</tr>
<tr>
<td>Minimal Chain</td>
<td>0</td>
<td>1,412</td>
</tr>
<tr>
<td>Local Attachment</td>
<td>0</td>
<td>1,923</td>
</tr>
<tr>
<td>Nearest Grammar</td>
<td>80</td>
<td>180</td>
</tr>
<tr>
<td>Structural Triggers Learner (STL) : Guessing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLA</td>
<td>88</td>
<td>16,990</td>
</tr>
<tr>
<td>Triggering Learning Algorithm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>0</td>
<td>16,663</td>
</tr>
</tbody>
</table>

CUNY CoLAG Language Domain

13-parameters, 3,072 structurally distinct (finite) languages
(For details, see Sakas, ACL 2003)

1. Subject Initial
2. Object Final
3. Complementizer Initial
4. Obligatory Topic
5. Null Subject
6. Null Topic
7. Wh Movement
8. Pied Pipings
9. Topic Marking
10. V to I Movement
11. I to C Movement
12. Affix hopping
13. Question Inversion

Access the languages and grammars online at http://146.95.2.133

Conclusions

1) Waiting for unambiguous triggers is dangerously unreliable. ('The languages it fails on include 'English', 'Japanese', 'German'.)

2) Guessing is better than waiting. Use partly ambiguous triggers!

3) But it is critical which guessing strategy is employed.
   • TLA guessing, which is not input-triggered, is no faster and much less reliable than random guessing. (As anticipated by Berwick and Niyogi, 1996)
   • STL guessing does use partly ambiguous triggers. It is totally reliable, though a little slower, if guided by the parser rather than by domain-search tactics like nearest grammar (= local hill-climbing).