Modeling Morphological Subgeneralizations

Claire Moore-Cantwell    Robert Staubs

December 15, 2013
Overview

1. Overview of our model:
   - Integrated phonology and morphology
   - Probabilistic
   - Explicit representation of subgeneralizations

2. Learning and production in this model

3. Evaluation and comparison to behavioral data
Lexically conditioned morphology

Some morphological patterns are exceptionful and their application is conditioned by the identity of particular lexical items.

- English Past tense:
  - walk → walked
  - sting → stung (∼ swing, string, cling)
  - weep → wept (∼ keep, sleep, sweep)

- This (and many such patterns) cannot be captured as a rule with memorized exceptions.

- The irregular patterns can also be generalized to new forms (Bybee and Moder, 1983; Prasada and Pinker, 1993; Albright and Hayes, 2003).

→ The lexicon and the grammar must interact to determine the output of certain morphological processes.
Processing results motivate models of lexical structure in which similar things are ‘near’ each other

- Semantically related words prime each other: Collins and Loftus (1975)
- Phonologically similar words are competitors in lexical access: McClelland and Elman (1986); Marslen-Wilson (1987)

→ The success of these models in processing has led e.g. Rumelhart and McClelland (1986) to propose a connectionist model of (morpho)-phonological knowledge.
One mechanism or two?

- Rumelhart and McClelland’s model of lexically conditioned morphology has been criticized:
  - On theoretical grounds: (Pinker and Prince, 1988)
  - Failure to capture the generality of the morphology-phonology interaction
  - the t/d/əd ∼ s/z/əz alternation in both plurals, possessives
- ‘Dual-route’ models of lexically conditioned morphology use a connectionist system for irregulars, and a rule for regulars (Pinker and Prince, 1988; Pinker, 1999; Marcus et al., 1995)
  - But Albright and Hayes (2003) argue for a single mechanism:
    - The phonological form of the stem matters for regulars as well as irregulars
One mechanism or two?

- Albright and Hayes (2002, 2003) propose a rules-only account
  - The Minimal Generalization Learner (MGL) uses many rules of varying degrees of generality
  - Ex:
    \[
    \emptyset \rightarrow d / \left[ \begin{array}{c} \int \text{ ain} \_ \end{array} \right][+past] \\
    \emptyset \rightarrow d / \left[ \begin{array}{c} \text{kən} \text{ s ain} \_ \end{array} \right][+past] \\
    \Rightarrow \emptyset \rightarrow d / \left[ \begin{array}{c} \text{X} \text{[vcll s]} \text{ ain} \_ \end{array} \right][+past] \\
    \vdots \\
    \Rightarrow \emptyset \rightarrow d / \left[ \begin{array}{c} \text{X} \_ \end{array} \right][+past]
    \]
- Islands of Reliability (IOR’s)
  - Words of a similar shape all take the same past
  - Both irregulars and regulars (e.g. \( \emptyset \rightarrow t / [\text{X} \text{ f}] [+past] \))
Lexical items can pattern together based on properties that are not directly related to their phonology:

- **Syntactic category, e.g:**
  - Noun vs. verb stress in English (Guion et al., 2003)
  - Word minimality requirements in many languages (Hayes, 1995)

- **Lexical Strata**
  - A cluster of phonological properties causes words to pattern together
  - Ex: Japanese (Moreton and Amano, 1999)
Integrating the lexicon and morphology

We construct a model that integrates the lexicon and morphology:

• Words group together into ‘bundles’
• These ‘bundles’ can be indexed to ‘operational constraints’
  • Similar technology to lexically indexed constraints

→ **Phonology and morphology interact:**
  Operational constraints compete with markedness and faithfulness constraints in Maximum Entropy grammar (Goldwater and Johnson, 2003)
Integrating the lexicon and morphology

Bundles come with ‘operational constraints’ which require that a morpheme be realized via a particular operation

Examples:

- $\text{+Past}: \, i \rightarrow \alpha$ (e.g. ring $\rightarrow$ rang)
- $\text{+Past}: \, \emptyset \rightarrow d$ (e.g. sigh $\rightarrow$ sighed)

These constraints mandate a particular change to a UR ‘prior’ to surface phonology
Integrating the lexicon and morphology

Predecessors include:

- **Anti-faithfulness** (Alderete, 2001)
  - Operational constraints specify a more specific type of “unfaithfulness”

- **Realizational constraints** (Xu and Aronoff, 2011)
  - Operational constraints need not be surface-true
  - Apply to the mapping between input to morphology and its output
Integrating the lexicon and morphology

- Combines ideas from UR constraints (Boersma, 2001), targeted constraints (Wilson, 2013)
  - Also describe properties of UR
  - ...But the mapping between URs, not just the UR itself
- Compare Max-Morph constraints (Wolf, 2008), and their operational version (Staubs 2011)
Integrating the lexicon and morphology

Some departures from the Minimum Generalization Learner:

• Phonotactics of English learned along with its morphology
• The context of a rule is divorced from its application
  • Assignment to a bundle can be based on many factors, not just context (e.g. for lexical strata)
  • Bundle formation can be based on information other than sound (e.g. noun/verb stress in English)
Structure of the model

Lexicon

Grammar

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>/nid+d/</td>
<td>nidd</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>b.</td>
<td>→ /nid+d/</td>
<td>nidΩd</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k.</td>
<td>/nεd/</td>
<td>nεd</td>
<td>-3</td>
<td>-1</td>
</tr>
<tr>
<td>l.</td>
<td>/nεd/</td>
<td>nεdΩd</td>
<td>-5</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
</tr>
</tbody>
</table>
How the model generates output

Assigned to a bundle?

Yes

Use operational constraints to generate morphological UR’s

No

Assign a bundle

Generate candidate surface forms based on each UR

Choose an optimum
Candidate Generation and Optimization

For a given input:

1. Generate possible URs from morphology based on known operational constraints
2. Assign operational constraint violations to candidates not matching the input’s bundle(s)
3. Apply phonological operations to create surface forms
   - Feature changing
   - Epenthesis
4. Assign faithfulness based on (phonological) operations used
5. Assign markedness based on surface forms
Inducing Operational Constraints

During learning, create a bundle for a new item:

1. Induce an operational constraint by surface string comparison

   **BASE:**
   
   \[
   \begin{array}{c}
   d \ u \ (i) \ \eta \ k \\
   k \ i \ p
   \end{array}
   \]

   **PAST:**
   
   \[
   \begin{array}{c}
   d \ u \ (\ae) \ \eta \ k \\
   k \ \epsilon \ p \ t
   \end{array}
   \]

   \[
   \begin{array}{c}
   i \rightarrow \ae \\
   i \rightarrow \epsilon \\
   \emptyset \rightarrow t
   \end{array}
   \]

2. Try to merge that bundle with existing bundles:

   \[
   \begin{array}{c}
   \text{ring, stink, sing} \\
   \text{drink} \\
   \text{drink, ring, stink, sing}
   \end{array}
   \]

   \[
   i \rightarrow \ae
   \]

   \[
   i \rightarrow \ae
   \]

   \[
   i \rightarrow \ae
   \]
Bundle Assignment

- Sample from bundles based on SIMILARITY
  - We use markedness constraints to assess phonological similarity (a la Golston, 1996)
  - Bundles have a ‘collective’ (average) violation vector
  - Which is compared to the violation vector of the input form
    \[ \text{distance} = e^{-c \sum_{\text{Con}} (v_1 - v_2)^2} \]
  - A bundle is chosen based on distance: more similar bundles are more likely to be chosen
    \[ P = \frac{\text{distance}(\text{base}, \text{gp})}{\sum_{\text{Bundles}}(\text{distance})} \]
Randomly sample a present-past pair:

- Generate an optimum
- Does it match the correct output?
- If not, use delta rule to update constraint weights and:
  - 0.01 induce a new (n-gram) markedness constraint
  - 0.50 Adjust the item’s bundle by MERGER
Bundle MERGER

- Choose a bundle to merge with based on SIMILARITY
- All bundle members are now members of the new bundle
- Update markedness violation vectors accordingly
- Keep the operational constraint of the larger bundle
Testing the model’s performance

**Strategy:** Train on English, test on English and wug-words

- **Training:**
  - data: 4280 present-past pairs from CELEX, lemma freq. $> 10$
  - 10 runs: learning rate of 1, 30 epochs, 1000 test trials per wug
    - $\rightarrow$ 93%-99% accuracy on regulars
    - $\rightarrow$ 69%-99% accuracy on irregulars

- ‘Wug test’:
  - Use Albright and Hayes’ wug-words
  - Does our model behave similarly to experimental participants?
    - ✓ Regulars produced more often than irregulars
    - ✓ More irregulars in irregular IOR’s
    - ✓ More regulars in regular IOR’s
Testing the model’s performance

- Irregular bundles (all runs):
  - **Faithful**: (hurt,split,shed,bet,trust...)
  - $i \rightarrow \emptyset$: (swim,shrink,stink,drink...)
  - $i \rightarrow \Lambda$: (sting,stick,cling,swing...)
  - $i \rightarrow \varepsilon$: (lead,feed,read,meet...)
  - $i \rightarrow \varepsilon$, **Add** -/t/: (deal,mean,keep,sleep...)
  - etc.

- One regular bundle (8/10 runs):
  - 6 runs: **Add** -/əd/: (earn,predict,whisk...)
  - 1 run: **Add** -/d/
  - 1 run: **Add** -/t/

- Multiple regular bundles (2 runs):
  - **Add** -/d/: (earn,prize,smell...)
  - **Add** -/əd/: (predict,cheat,wed...)
  - **Add** -/t/: (whisk,invoke,rip...)

Summary of productions by Island of Reliability

Proportion Forms produced

- IOR
- Non-IOR

Irregular: 22 / 25
Mismatches to the Albright and Hayes data

• When multiple regulars are learned, the phonological alternation is not:
  • [baiz]∼[baizt]
  • [drais]∼[draisd]

• The model’s performance on particular wug items varies a lot
  • It produces the same irregular as subjects sometimes:
    - flip ∼ flɛpt
    - glɪt ∼ glɪt, glæt
    - splɪŋ ∼ splæŋ
    - nold ∼ nɛld
  • But also some weird ones:
    - fro ∼ frɛ (hold ∼ held)
    - nold ∼ nuld (blow ∼ blew)
Summary of results

- Most of the time, the model successfully learns a single regular rule
  - and markedness constraints enforce the t/d/əd alternation
- For wug-words, in concord with Albright and Hayes’ results:
  - Regulars are much more likely than irregulars
  - Irregulars are more likely for word in irregular IOR’s than else
  - Regulars are also more likely in regular IOR’s than else
Future Directions

- German plural
  - Default is not most common
- Arabic broken plural
  - Operations can be non-local
- English noun vs verb stress
  - Indexation to non-phonological properties
- Japanese lexical strata
  - Indexation based on an array of phonological properties


Allan M. Collins and Elizabeth F. Loftus. A spreading-activation


Gary F. Marcus, Ursula Brinkmann, Harald Clahsen, Richard


