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

The structured lexicon

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)

$\rightarrow$ 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 $\sim$ 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:

$$\emptyset \to d \ / \ [\qquad\qquad \text{ʃ} \qquad \text{ain} \quad \_ \quad]_{[+past]}$$
$$\emptyset \to d \ / \ [\quad \text{kən} \quad \text{s} \qquad \text{ain} \quad \_ \quad]_{[+past]}$$
$$\Rightarrow \ \emptyset \to d \ / \ [\quad \text{X} \quad [\text{vcls}] \quad \text{ain} \quad \_ \quad]_{[+past]}$$
$$\vdots$$
$$\Rightarrow \ \emptyset \to d \ / \ [\qquad\qquad \text{X} \qquad\qquad \_ \quad]_{[+past]}$$

- Islands of Reliability (IOR's)
  - Words of a similar shape all take the same past
  - Both irregulars and regulars (e.g. $\emptyset \to t / [\text{ X f}\_]_{[+past]}$)

More structure in the lexicon?

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

$\rightarrow$ **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:

- +PAST: $i \rightarrow æ$ (e.g. *ring* $\rightarrow$ *rang*)
- +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
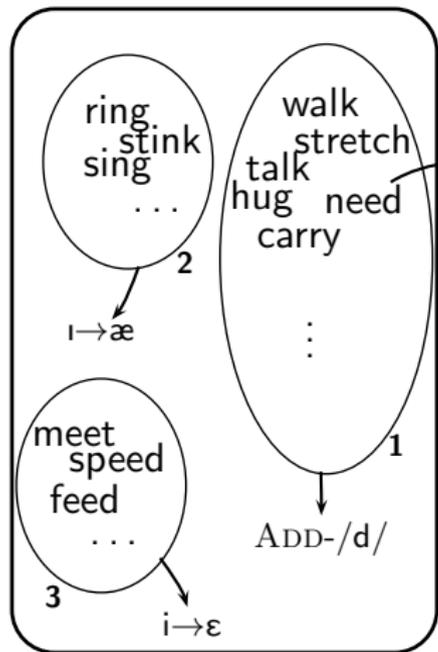
Integrating the lexicon and morphology

- 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)
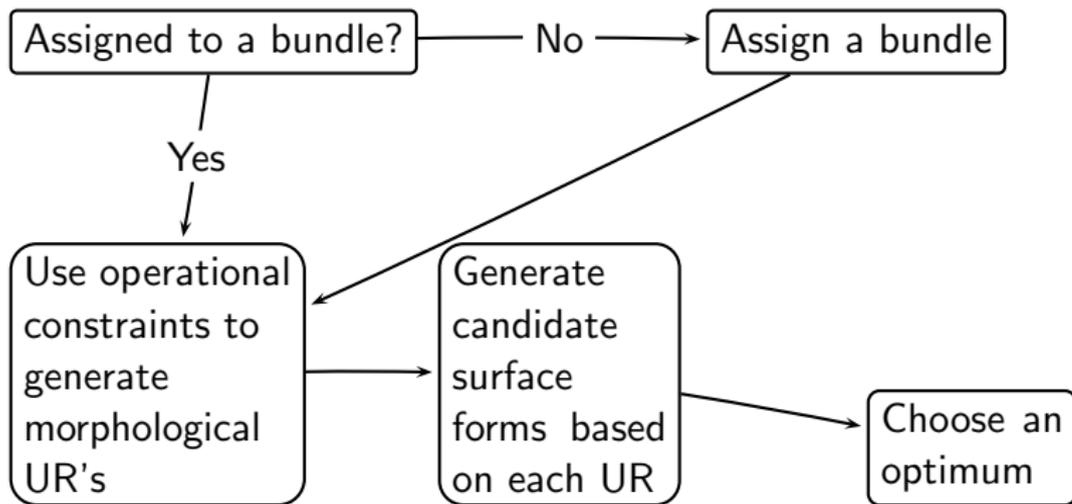
Structure of the model

How the model generates output

Assigned to a bundle? ── No ──▶ Assign a bundle

Yes

Use operational constraints to generate morphological UR's → 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

2. Try to merge that bundle with existing bundles:

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
    $distance = e^{-c \sum_{Con}(v_1 - v_2)^2}$
  - A bundle is chosen based on distance: more similar bundles are more likely to be chosen
    $P = \frac{distance(base, gp)}{\sum_{Bundles}(distance)}$

Learning

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:
    - *.01* induce a new (*n*-gram) markedness constraint
    - *.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
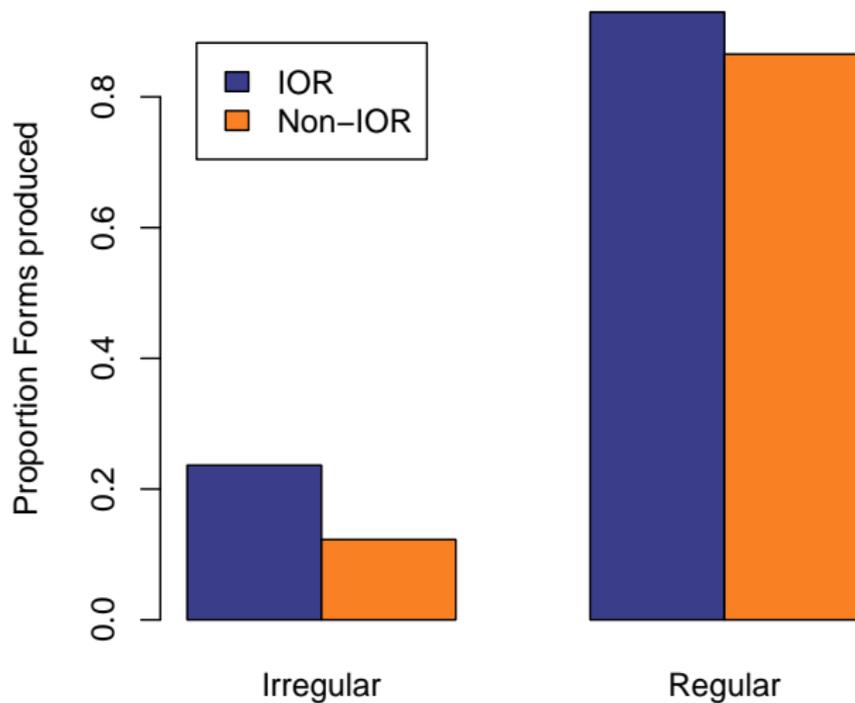
Testing the model's performance

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...)
  - ɪ→ æ: (swim,shrink,stink,drink...)
  - ɪ→ ʌ: (sting,stick,cling,swing...)
  - i → ɛ: (lead,feed,read,meet...)
  - i→ɛ, 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

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

Adam Albright and Bruce Hayes. Modeling english past tense intuitions with minimal generalization. In Mike Maxwell, editor, *Proceedings of the 2002 Workshop on Morphological Learning*, Philadelphia, 2002. Association of computational Linguistics, Association for computational linguistics.

Adam Albright and Bruce Hayes. Rules vs. analogy in english past tenses: a computational/experimental study. *Cognition*, 90: 119–161, 2003.

John Alderete. *Morphologically governed accent in Optimality Theory*. Psychology Press, 2001.

Paul Boersma. Phonology-semantics interaction in OT, and its acquisition. In Kirchner et al, editor, *Papers in Experimental and Theoretical linguistics*, volume 6. University of Alberta, Edmonton, 2001.

Joan Bybee and Carol Lynn Moder. Morphological classes as natural categories. *Language*, 59:251–270, 1983.

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

theory of semantic processing. *Psychological Review*, 82(6): 407–428, 1975.

Sharon Goldwater and Mark Johnson. Learning ot constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson, and Osten Dahl, editors, *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, pages 111–120, 2003.

Chris Golston. Direct optimality theory: Representation as pure markedness. *Language*, 72(4):713–748, December 1996.

Susan G. Guion, J.J. Clark, Tetsuo Harada, and Ratree P. Wayland. Factors affecting stress placement for english nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, 46(4):403–427, 2003.

Bruce Hayes. *Metrical stress theory: principles and case studies*. University of Chicago press, 1995.

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

Wiese, and Steven Pinker. German inflection: the exception that proves the rule. *Cognitive Psychology*, 29:189–256, 1995.

William D Marslen-Wilson. Functional parallelism in spoken word-recognition. *Cognition*, 25(1):71–102, 1987.

James L. McClelland and Jeffrey L. Elman. The trace model of speech perception. *Cognitive Psychology*, 18:1–86, 1986.

Elliott Moreton and Shigeaki Amano. Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. *Proceedings of the 6th European Conference on Speech Communication and Technology*, 1999.

Steven Pinker. *Words and Rules: the ingredients of language*. William Morrow, New York, 1999.

Steven Pinker and Alan Prince. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1):73–193, 1988.

Sandeep Prasada and Steven Pinker. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8:1–56, 1993.

D. E. Rumelhart and J. L. McClelland. *On Learning the past tenses of English verbs*, chapter 18, pages 216–271. MIT press, 1986.

Colin Wilson. A targeted spreading imperative for nasal place assimilation. In *Proceedings of the 41st meeting of the North East Linguistics Society*, 2013.

Matthew Wolf. *Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model*. PhD thesis, University of Massachusetts, Amherst, 2008.

Zheng Xu and Mark Aronoff. A realization optimality theory approach to blocking and extended morphological exponence. *Journal of Linguistics*, 47(03):673–707, 2011.