Using language-wide phonotactics to learn affix-specific phonology

Brian W. Smith
UC Santa Cruz
brwismit@ucsc.edu
Today

• A case of phonologically-conditioned suffixation, in which...

  • learners have very little data

  • the distribution of affixes can be learned through a combination of...
    (1) extending language-wide phonotactics
    (2) learning small affixal differences

• No sublexicons are necessary for this data
Case study

- Two suffixes, each with two allomorphs

- licious  - alicious
  [lɪʃəs]    [əlɪʃəs]

- thon  - athon
  [θən]    [æθən]
Phonological conditioning

- Both suffixes conditioned by phonology
  - Schwaful variant is more likely after stressed syllables, consonants
  - Schwaless variant is more likely after unstressed syllables, vowels
- Shown in the following slides using data from GloWbE (Davies et al. 2013)
  - GloWbe: data from 2012-2013, 60% blogs
Effect of segment: -(a)licious

• -**alicious** / C_
  
  appoliciious   good-a-licious
  craftiliciious bookaliciious

• -**licious** / V_
  
  royliciious   bow-licious
  skaliciious   rawlicious
Effect of stress: -(a)licious

- **-alicious / ɔ_**
  - spookalicious
  - nomalicious
- **-licious / ō_**
  - dietlicious
  - Twilightlicious

swoon-a-licious
meadilicious
summerlicious
Jerseylicious
Rate of schwa in -(a)licious

V vs. C

Proportion schwa

V-final ₀

C-final ₀

C-final ₀

0.90

0.79

0.22

0.10

0.21

0.78

-licious

-abolicious
Rate of schwa in *(a)licious*

\[
\begin{array}{c|c|c}
\text{Proportion schwa} & \ddot{\sigma} vs. \dot{\sigma} \\
\hline
\text{V-final } \ddot{\sigma} & 0.90 & 0.10 \\
\text{C-final } \ddot{\sigma} & 0.79 & 0.21 \\
\text{C-final } \dot{\sigma} & 0.22 & 0.78 \\
\end{array}
\]
Rate of schwa in -(a)thon

V vs. C

Proportion schwa

<table>
<thead>
<tr>
<th></th>
<th>V-final</th>
<th>C-final</th>
<th>C-final</th>
</tr>
</thead>
<tbody>
<tr>
<td>ṡσ</td>
<td>0.79</td>
<td>0.29</td>
<td>1.00</td>
</tr>
<tr>
<td>0.71</td>
<td>0.21</td>
<td>-thon</td>
<td>-athon</td>
</tr>
</tbody>
</table>
Rate of schwa in -(a)thon

\[ \ddot{\sigma} \text{ vs. } \acute{\sigma} \]

<table>
<thead>
<tr>
<th></th>
<th>V-final $\ddot{\sigma}$</th>
<th>C-final $\ddot{\sigma}$</th>
<th>C-final $\acute{\sigma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thon</td>
<td>0.21</td>
<td>0.71</td>
<td>1.00</td>
</tr>
<tr>
<td>Athon</td>
<td>0.79</td>
<td>0.29</td>
<td></td>
</tr>
</tbody>
</table>
Idiosyncratic differences

- Despite the fact that phonological conditioning is similar across the suffixes, the suffixes differ in their overall rate of schwa.
- This difference holds across *all* phonological contexts.
-(a)licious and -(a)thon
Summary

• -(a)licious and -(a)thon are conditioned by phonological context in the same ways

\[
\begin{align*}
\text{-thon} & \quad \{ C \_ \} \\
\text{-licious} & \quad \{ \ddot{o} \_ \}
\end{align*}
\]

• But!
  -athan is used more often than -alicious across all phonological contexts
Whence phonological conditioning?

- Affixes prefer some roots over others
- Two explanations:
  - Language-wide grammar
  - Subcategorization
Language-wide grammar

• Phonological conditioning comes from the phonological grammar (e.g., Mester 1994, Kager 1996, Mascaró 1996)

• In OT, markedness constraints: one set of constraints for allomorphy, alternations, and phonotactics

• E.g., choice of suffix avoids hiatus and stress clash, driven by *HIATUS and *CLASH
Subcategorization

- Lexical listing / subcategorization frames (e.g. Paster 2006, Embick 2010)
  - **-alicious** $\leftrightarrow$ C __
  - **-licious** $\leftrightarrow$ V __

- Sublexicons: every suffix can have its own GateKeeper grammar (Becker, earlier)
Three arguments for language-wide grammar approach

• 1. Cross-suffix similarity
   Many suffixes are subject to the same phonological conditions

• 2. Poverty of the stimulus
   Both suffixes are very rare, with uneven distributions in a corpus

• 3. The same constraints that condition the suffixes also play a role in alternations and the distribution of words in the lexicon
A solution

- Problem: the phonological conditioning of the suffixes persists despite a lack of learning data.
- Using the pre-existing phonotactic grammar to choose between suffixal forms solves this problem — learners don’t need many -(a)licious words data to learn -(a)licious!
Cross-suffix similarity

Experiment:
-(a)licious and -(a)thon
Experiment

• Goals
  • Test for effects of segment and stress beyond words in corpus
  • Estimate probabilities of -(a)licious and -(a)thon across contexts
Experiment

• Web-based forced choice presented through Ibex Farm
Item design

- Four stress and segmental contexts

<table>
<thead>
<tr>
<th>EXAMPLE</th>
<th>FINAL SEGMENT</th>
<th>STRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>cactus</td>
<td>C</td>
<td>10</td>
</tr>
<tr>
<td>police</td>
<td>C</td>
<td>01</td>
</tr>
<tr>
<td>hero</td>
<td>V</td>
<td>10</td>
</tr>
</tbody>
</table>

- 10 roots of each type (plus 40 fillers)
Item design

- Four stress and segmental contexts

<table>
<thead>
<tr>
<th>EXAMPLE</th>
<th>FINAL SEGMENT</th>
<th>STRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>cactus</td>
<td>C</td>
<td>10</td>
</tr>
<tr>
<td>police</td>
<td>C</td>
<td>01</td>
</tr>
<tr>
<td>hero</td>
<td>V</td>
<td>10</td>
</tr>
</tbody>
</table>

- 10 roots of each type (plus 40 fillers)
Item design

- Four stress and segmental contexts

<table>
<thead>
<tr>
<th>EXAMPLE</th>
<th>FINAL SEGMENT</th>
<th>STRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>cactus</td>
<td>C</td>
<td>10</td>
</tr>
<tr>
<td>police</td>
<td>C</td>
<td>01</td>
</tr>
<tr>
<td>hero</td>
<td>V</td>
<td>10</td>
</tr>
</tbody>
</table>

- 10 roots of each type (plus 40 fillers)

Effect of segment
Experiment

- 109 participants after exclusions
  - All self-identified as native speakers of English
  - Only included data for American participants
Summary for -(a)licious

V vs. C  ̃σ vs. ̄σ

<table>
<thead>
<tr>
<th></th>
<th>Proportion schwa</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-final ̃σ</td>
<td>0.93</td>
</tr>
<tr>
<td>C-final ̃σ</td>
<td>0.55</td>
</tr>
<tr>
<td>C-final ̄σ</td>
<td>0.07</td>
</tr>
</tbody>
</table>

-licious
-atical
Summary for -(a)thon

V vs. C  ṣ vs. ́ ṣ

Proportion schwa

V-final Ṣ  

C-final Ṣ  

C-final ṣ  

0.29  0.79  0.71

0.21  0.01  0.01

0.71  0.21  0.99

-thon
-athon
Comparing -(a)licious and -(a)thon we find roughly the same phonological conditioning. We also see that schwa is used more often in -(a)thon.
-(a)licious and -(a)thon
-(a)licious and -(a)thon

- The greater preference for schwa in -(a)thon holds across items and participants
  - Including fillers, which have other stress patterns, e.g. 102, 201, 010
  - True for 78/80 words in the experiment
  - True for 87% of participants (95/109)
By item
Experiment part 2: the Rhythm Rule

In this section, I present the rest of the experimental results, focusing on the interaction between -(a)licious, -(a)thon, and the Rhythm Rule, a phonological alternation that resolves stress clash.

Both suffix selection and the Rhythm Rule conspire to avoid stress clash, providing an additional argument for the clash-driven nature of -(a)licious allomorphy.

These results also support a model in which UR selection and the phonological grammar occur in parallel. They provide an example of the chicken-egg effect (McCarthy 2002), a case where two processes must both apply first, creating an ordering paradox.

(33) The chicken-egg effect, a consequence of parallelism (McCarthy 2002)

The application of process A depends on knowing the output of process B, and the application of process B depends on knowing the output of process A.

(34) Chicken-egg effect in -(a)licious and the Rhythm Rule
Conspiracies

- Experiment shows that speakers use roughly the same phonological criteria for both -(a)thon and -(a)licious
- Many suffixes in English seem subject to the same constraints
  - -(a)holic, -(e)teria, -(o)rama, etc.
- And well-established derivational suffixes
  - -(e)ry, -ese, -al, -eer, -ee, -ette, -ize, -ify
    (Raffelsiefen 2005 and earlier work)
Subcat?

- Under subcategorization, similarity across suffixes and alternations is a coincidence

- alicious $\leftrightarrow$ C __
- licious $\leftrightarrow$ V __
- athon $\leftrightarrow$ C __
- thon $\leftrightarrow$ V __
- ery $\leftrightarrow$ C __
- ry $\leftrightarrow$ V _
Poverty of the stimulus
Poverty of the stimulus

- Speakers agree on the phonological conditioning of the suffixes
- But if the corpus data is representative: data is scarce
  - -(a)licious and -(a)thon are not very common to begin with
  - Especially uncommon with roots of certain shapes
Uncommon…

- In GILOWbE, licious- and thon-words are uncommon. Out of 500 million words for American and Canadian English (combined):
  - 933 tokens for -(a)licious
  - 1866 tokens for -(a)thon
- Assume a speaker hears 30,000 words/day…
  - 30 licious/year, 60 thon/year
Variety in root shape...

- For both suffixes, more than half of the types are hapaxes (182/310 for -(a)licious)
- The top 10 most-frequent words account for >50% of the tokens
And most of the frequent \(-(a)thon\) words have schwa
For -(a)licious, the most frequent words don’t have schwa
Poverty of the stimulus

- If a speaker gets 30–60 of these tokens per year and doesn’t get a variety of phonological contexts, learning the “correct” subcategorization frames will be difficult at best
The same constraints are active in suffixation, alternations, phonotactics
MaxEnt model

- A model of -(a)licious and -(a)thon with handpicked constraints
- Fit on experimental probabilities
  - Note: probably not how learners acquire the distribution in the real world
- Using MaxEnt Grammar Tool (Wilson & George 2009)
Markedness constraints

- **CLASH:**
  Assign a violation for every óó sequence

- **LAPSE:**
  Assign a violation for every ōō sequence

- **HIATUS:**
  Assign a violation for every V.V sequence
Constraints to capture preference for schwa

- Analyze schwa alternation as listed allomorphs
- Constraints encode which listed allomorph is default (UR constraints, Pater at al. 2012)

  - UR = /əɪʃəs/  (-alicious)
  - UR = /lɪʃəs/  (-licious)
  - UR = /əθɑn/  (-athon)
  - UR = /θɑn/  (-thon)
## Learned weights

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Weight</th>
<th>Constraint</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>*CLASH</td>
<td>2.61</td>
<td>-athon</td>
<td>1.97</td>
</tr>
<tr>
<td>*HIATUS</td>
<td>2.31</td>
<td>-alicious</td>
<td>0.37</td>
</tr>
<tr>
<td>*LAPSE</td>
<td>0.43</td>
<td>-licious</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-thon</td>
<td>0.13</td>
</tr>
</tbody>
</table>
*Clash > *Hiatus > *Lapse

- Order of *CLASH, *HIATUS, and *LAPSE is mirrored in English lexicon
- On the next slide: counts from CMU dictionary, number of 3+ syllable words that violate the constraint
- Important: constraint weights were determined using only the experimental probabilities for -(a)licious and -(a)thon
## CMU violators

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Weight</th>
<th>Number of violators</th>
</tr>
</thead>
<tbody>
<tr>
<td>*CLASH</td>
<td>2.61</td>
<td>1,597</td>
</tr>
<tr>
<td>*HIATUS</td>
<td>2.31</td>
<td>2,792</td>
</tr>
<tr>
<td>*LAPSE</td>
<td>0.43</td>
<td>8,702</td>
</tr>
</tbody>
</table>
### Defaults

- For both `-athon` and `-ailicious`, the default is the schwaful form.
- The preference for `-athon` over `-thon` is greater than the preference for `-ailicious` over `-licious`.

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>-athon</td>
<td>1.97</td>
</tr>
<tr>
<td>-ailicious</td>
<td>0.37</td>
</tr>
<tr>
<td>-licious</td>
<td>0.18</td>
</tr>
<tr>
<td>-thon</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Defaults

• Phonotactics alone can’t explain the distribution of -(a)licious or -(a)thon

• No phonotactic motivation for baseline difference in schwa rates

• Speakers then must learn from observation
  • -athon is more common than -thon
  • -licious is more common than -alicious
*CLASH in English

- A small sample
  - Rhythm Rule (Liberman and Prince 1977)
    \[
    \text{thirteen} \rightarrow \text{thirteen mén}
    \]
  - optional *that* (Lee and Gibbons 2007)
    \[
    I \text{ know (that) Lucy went vs. I know (that) Louise went}
    \]
  - genitive alternation (Shih et al. 2015)
    \[
    \text{the car’s wheel vs. wheel of the car}
    \]
  - historical change (Schlüter 2005)
**HIATUS in English**

- English avoids hiatus, especially when the left vowel is lax

- Observable in phonotactics, repairs, and allomorphy
*HIATUS in English

• Radio Rule: no hiatus where $V_1$ is lax
  \[\sqrt{\text{radio}},\ \text{boa}\ *\text{ɹɛdɪ.o, bɔ.ɛ} \ (\text{Chomsky & Halle 1968})\]

• Glottal stop epenthesis
  \[\text{mora-[?]ist (Plag 1999), sea [?] otter (Davidson & Erker 2014)}\]

• Intrusive R and intrusive L
  \[\text{draw[r]ing [dɹɔʃin] (McCarthy 1993); draw[l]ing [dɹɔlɪŋ] (Gick 2002)}\]

• *a/an* allomorphy and function word reduction
  \[\text{an apple, a pear; [ði] apple, [ðə] pear (overview in Smith 2015)}\]
As in the rest of English, hiatus is especially bad when the left vowel is lax.
*HIATUS in -(a)thon

- As in the rest of English, hiatus is especially bad when the left vowel is lax.
Learning with sparse data
How to learn -(a)licious and -(a)thon

• The proposal:

  1. Take the pre-existing phonotactic grammar

  2. For each suffix, learn the rate of allomorphs from available data
BLICK grammar

• Weights of phonological constraints come from BLICK (Hayes 2012)

  • a MaxEnt phonotactic grammar of English based on CMU pronouncing dictionary (Weide 1994)

  • constraints are a mix of hand-picked and machine-generated
Learning the rate of schwa

- For each suffix, set the weight of morpheme-specific constraints to match the overall probability of schwa
- Here, fitted on token frequency not type
- Token frequency provides significantly better fit than type frequency for -(a)licious (difference is largely due to booty-licious)
Model’s performance on -(a)thon

- Model captures the relative likelihood of schwa across contexts
- Overpredicts schwa in CACTUS and HERO roots

<table>
<thead>
<tr>
<th>% schwa</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLICE-thon</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>CACTUS-thon</td>
<td>79</td>
<td>94</td>
</tr>
<tr>
<td>HERO-thon</td>
<td>48</td>
<td>87</td>
</tr>
<tr>
<td>SODA-thon</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
Model’s performance on -(a)thon

Grammar contains a constraint against obstruent-θ sequence

<table>
<thead>
<tr>
<th>%schwa</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLICE-thon</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>CACTUS-thon</td>
<td>79</td>
<td>94</td>
</tr>
<tr>
<td>HERO-thon</td>
<td>48</td>
<td>87</td>
</tr>
<tr>
<td>SODA-thon</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
Model’s performance on -(a)thon

Grammar doesn’t contain a general constraint against hiatus

<table>
<thead>
<tr>
<th>%schwa</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLICE-thon</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>CACTUS-thon</td>
<td>79</td>
<td>94</td>
</tr>
<tr>
<td>HERO-thon</td>
<td>48</td>
<td>87</td>
</tr>
<tr>
<td>SODA-thon</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
Model’s performance on -(a)licious

<table>
<thead>
<tr>
<th>%schwa</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLICE-licious</td>
<td>93</td>
<td>94</td>
</tr>
<tr>
<td>CACTUS-licious</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>HERO-licious</td>
<td>9</td>
<td>46</td>
</tr>
<tr>
<td>SODA-licious</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

- Model doesn’t capture a difference between HERO and CACTUS
Model’s performance on -(a)licious

<table>
<thead>
<tr>
<th>%schwa</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLICE-licious</td>
<td>93</td>
<td>94</td>
</tr>
<tr>
<td>CACTUS-licious</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>HERO-licious</td>
<td>9</td>
<td>46</td>
</tr>
<tr>
<td>SODA-licious</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

Grammar doesn’t contain a general constraint against hiatus
Improving the BLICK grammar

• Assume $HERO+\varepsilon$ violates the stress-sensitive constraint against hiatus in the grammar (*\(\acute{V}V\))

• Assume constraint against obstruent followed by $\theta$ doesn’t operate across morpheme boundaries
With improved BLICK grammar

<table>
<thead>
<tr>
<th>Target</th>
<th>Model</th>
<th>Target</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>POLICE-thon</td>
<td>99</td>
<td>99</td>
<td>POLICE-licious</td>
</tr>
<tr>
<td>CACTUS-thon</td>
<td>79</td>
<td>87</td>
<td>CACTUS-licious</td>
</tr>
<tr>
<td>HERO-thon</td>
<td>48</td>
<td>51</td>
<td>HERO-licious</td>
</tr>
<tr>
<td>SODA-thon</td>
<td>10</td>
<td>1</td>
<td>SODA-licious</td>
</tr>
</tbody>
</table>
Conclusion
Take-away

• Sometimes, affix-specific phonology doesn’t require learning much affix-specific information

• In the account here, the only affix-specific information is the rate of schwa

• Using the phonotactic grammar solves sparse data problems, and also accounts for similarities between suffixes and phonotactics
Predictions

• Under the strongest form of the language-wide grammar hypothesis, we should find a single grammar for all licious-like suffixes
  • Same constraints for every one
  • Same relationship between constraints
  • Only differences are in default forms
Predictions

• Any constraint that’s active in English should have an effect on \(-\text{(a)licious}\)
  
  • Liquid OCP (which has effects in derivational morphology)
  
  • Syllable contact
A problem

- Phonotactics aren’t going to work for every case of affixation, e.g. English comparative *-er*
- How does the learner decide which approach to employ?
Thank you