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]	[<u>**ə**</u>lɪʃəs]
-thon	-athon
[θɑn]	[<u>**ə**</u>θɑn]

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

- *-<u>a</u>licious* / C_

appolicious good-a-licious
craftilicious bookalicious

- *-licious* / V_

roylicious bow-licious
skalicious rawlicious

Effect of stress: -(a)licious

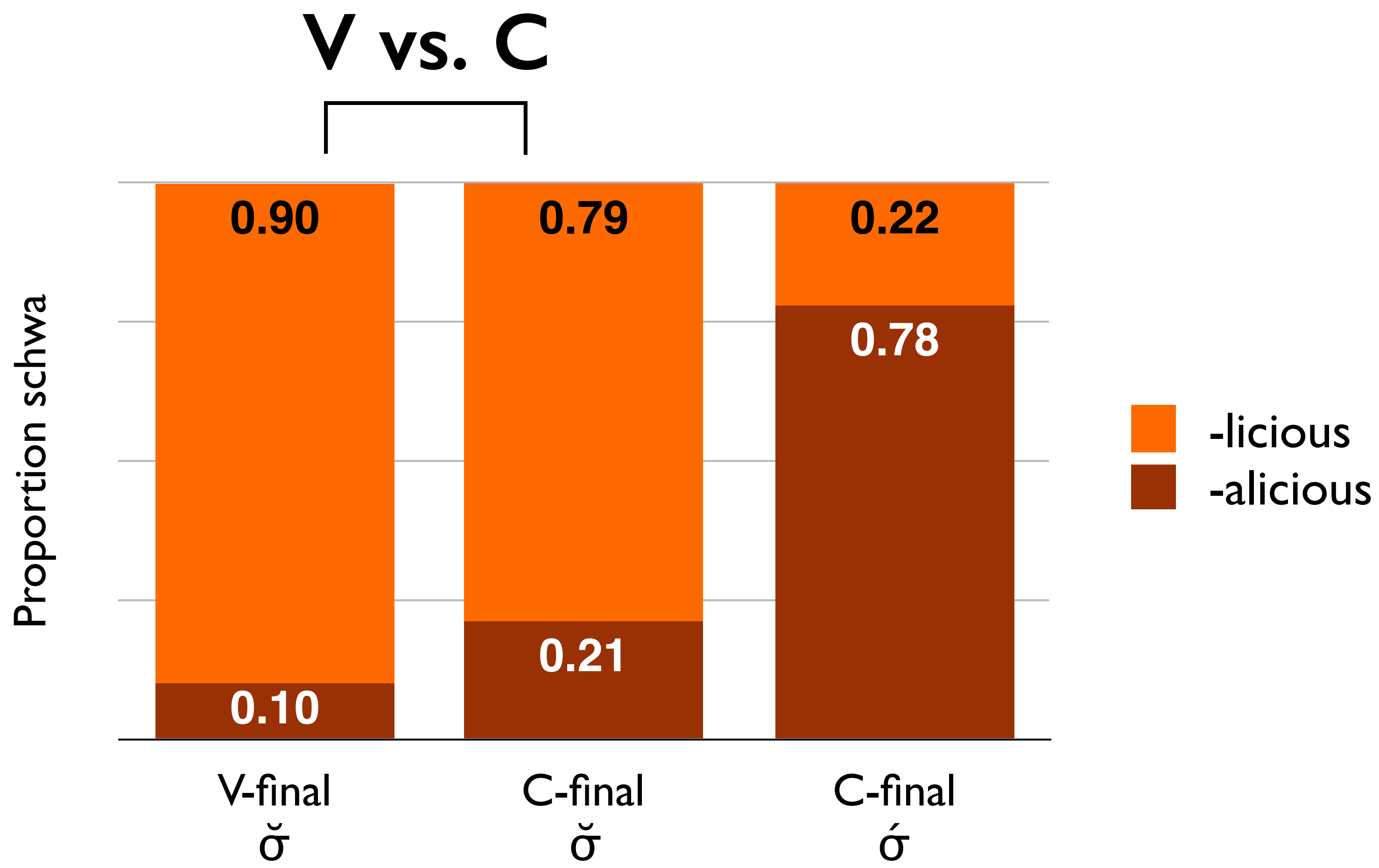
- *-alicious* / σ́_

spookalicious swoon-a-licious
nomalicious meadilicious

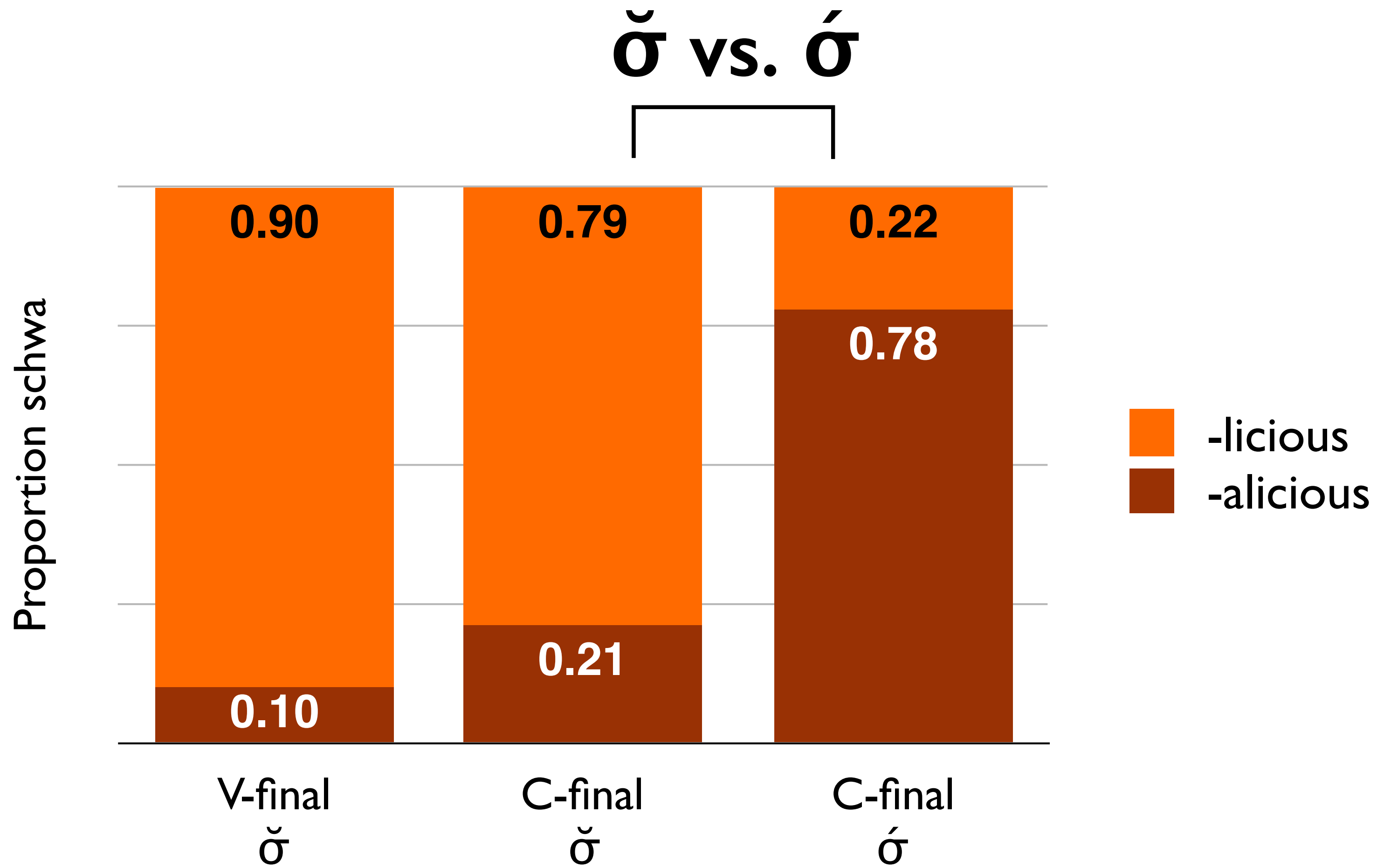
- *-licious* / σ̆_

dietlicious summerlicious
Twilightlicious Jerseylicious

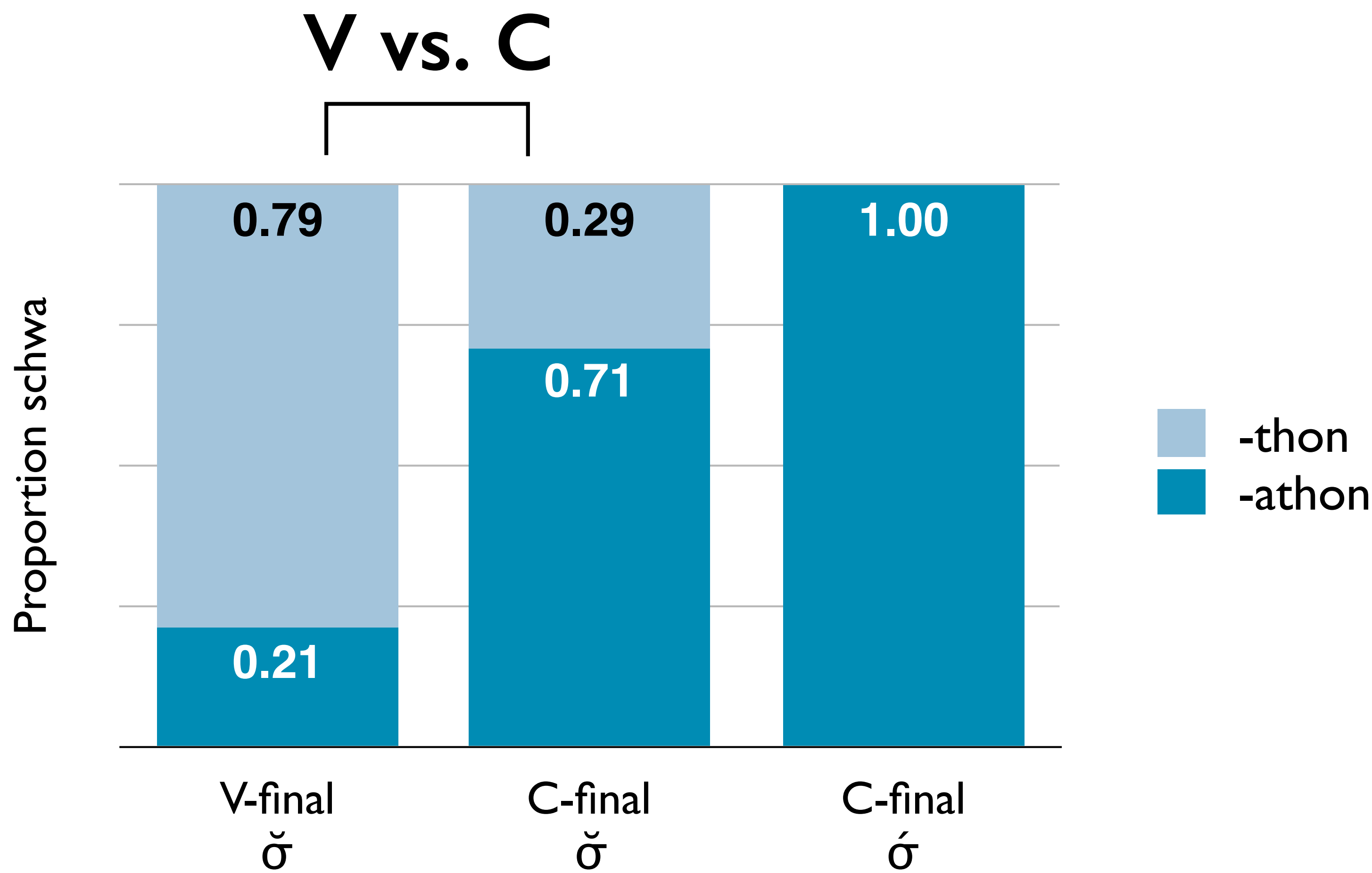
Rate of schwa in *-(a)licious*



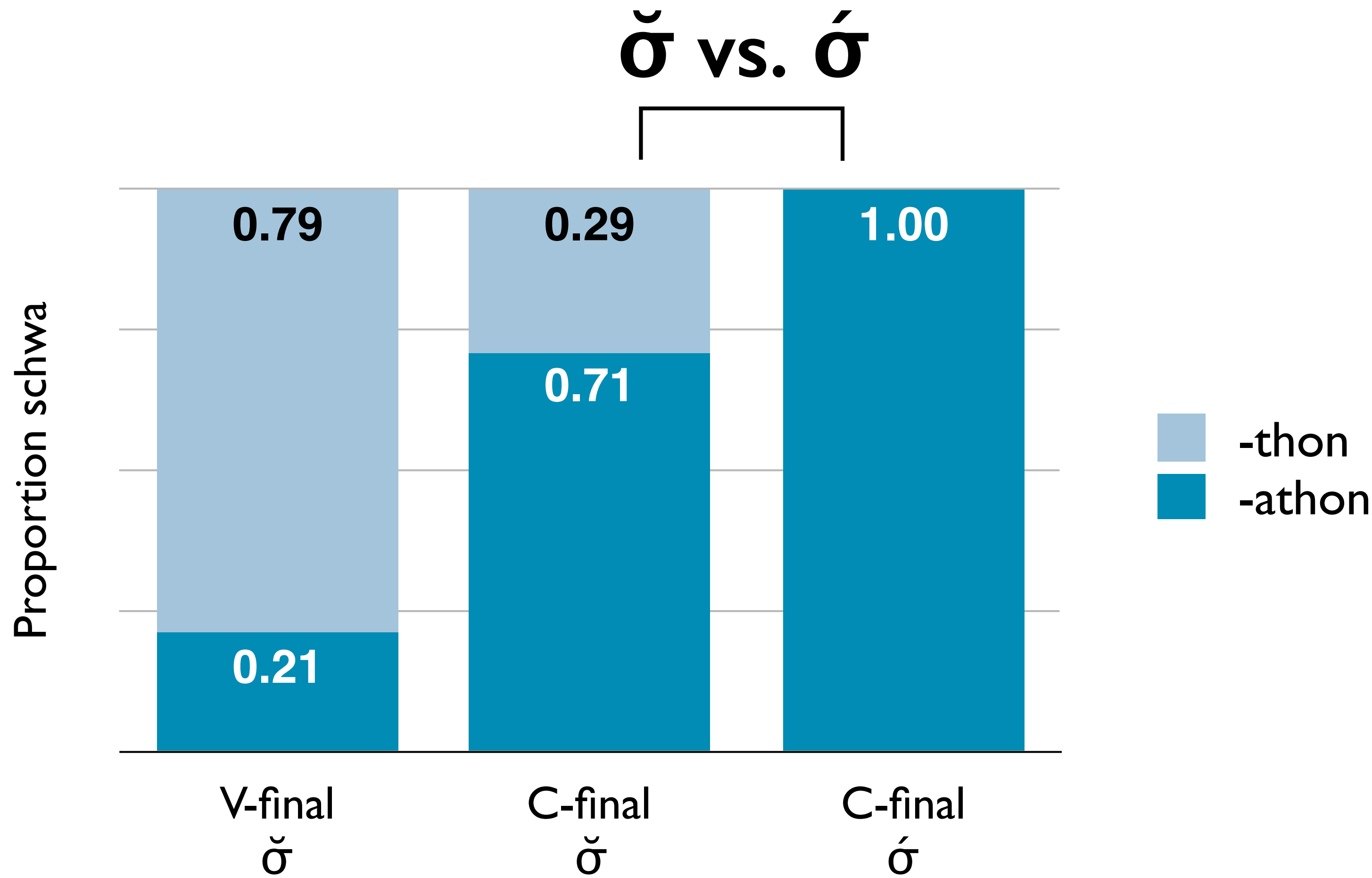
Rate of schwa in *-(a)licious*



Rate of schwa in *-(a)thon*



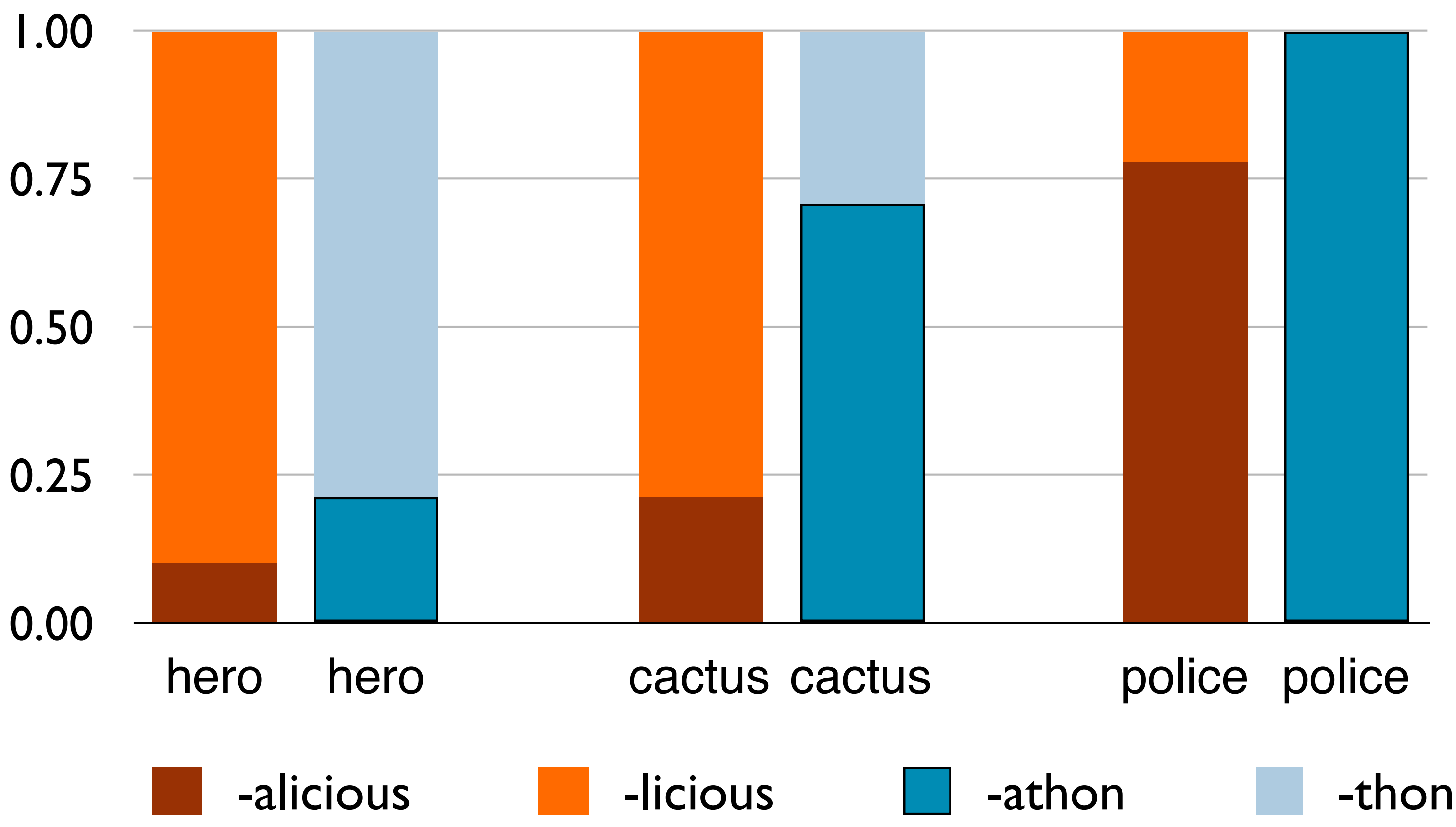
Rate of schwa in *-(a)thon*



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

-*athon* / -*alicious*	} C _ / ́σ _	-*thon* / -*licious*	} V _ / σ̆ _

- But!
  -athon 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* ↔ C __

      -licious ↔ 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:

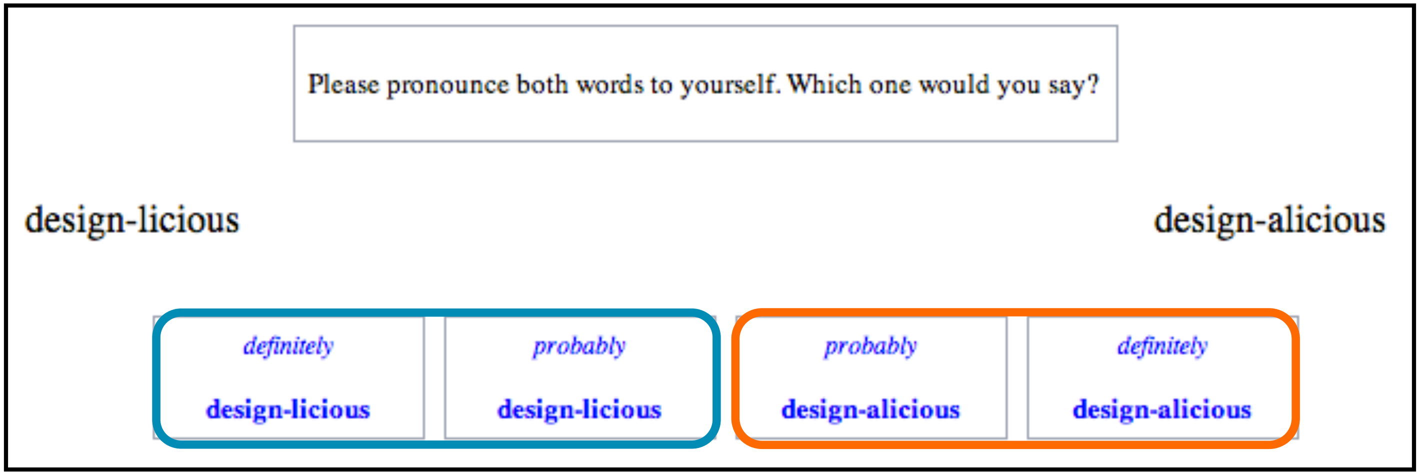
-(a)licious and *-(a)thon*

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

EXAMPLE	FINAL SEGMENT	STRESS
cactus	C	10
police	C	01
hero	V	10

- 10 roots of each type (plus 40 fillers)

Item design

- Four stress and segmental contexts

EXAMPLE	FINAL SEGMENT	STRESS
cactus	C	10
police	C	01
hero	V	10

Effect of final stress

- 10 roots of each type (plus 40 fillers)

Item design

- Four stress and segmental contexts

EXAMPLE	FINAL SEGMENT	STRESS
cactus	C	10
police	C	01
hero	V	10

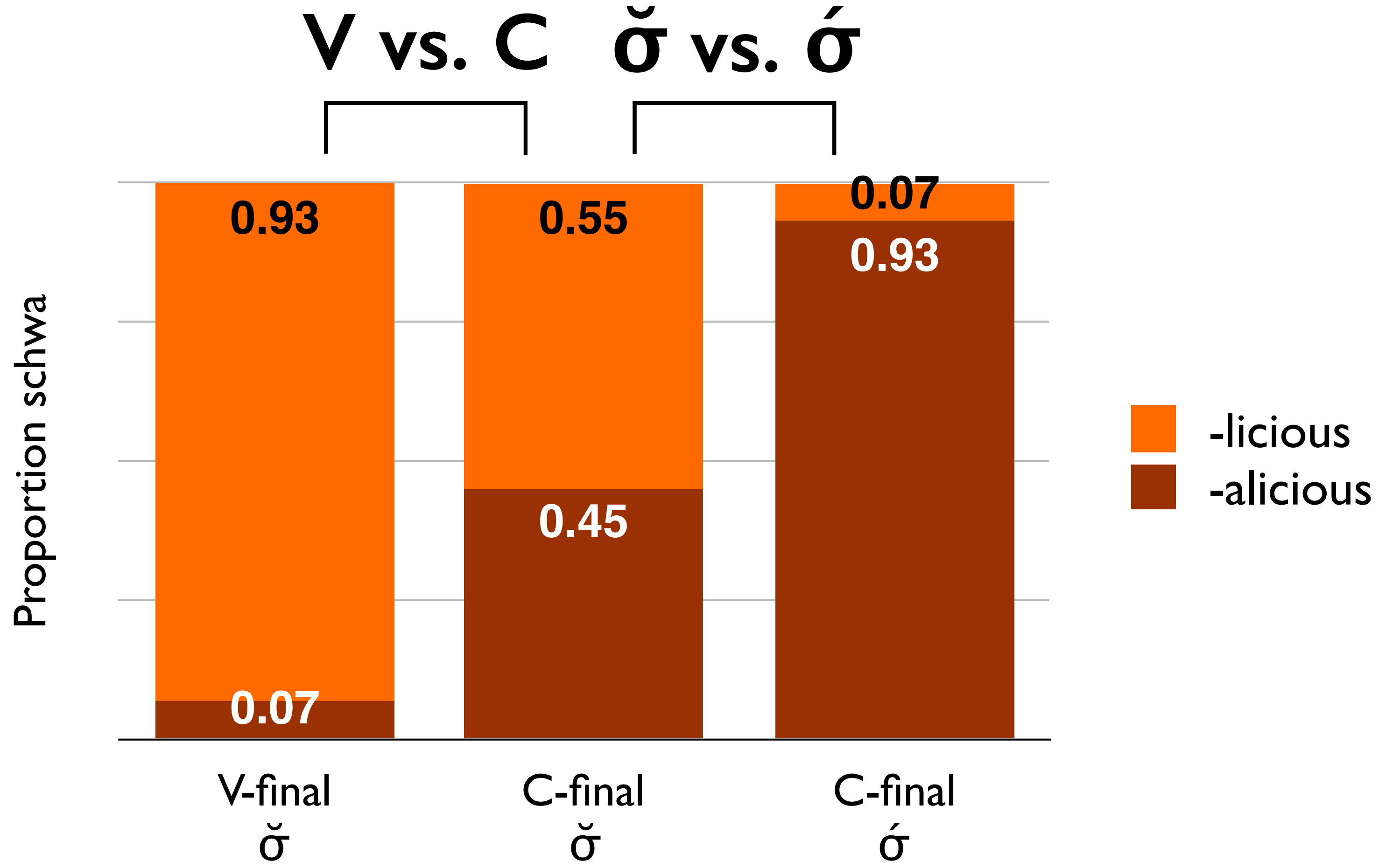
Effect of segment

- 10 roots of each type (plus 40 fillers)

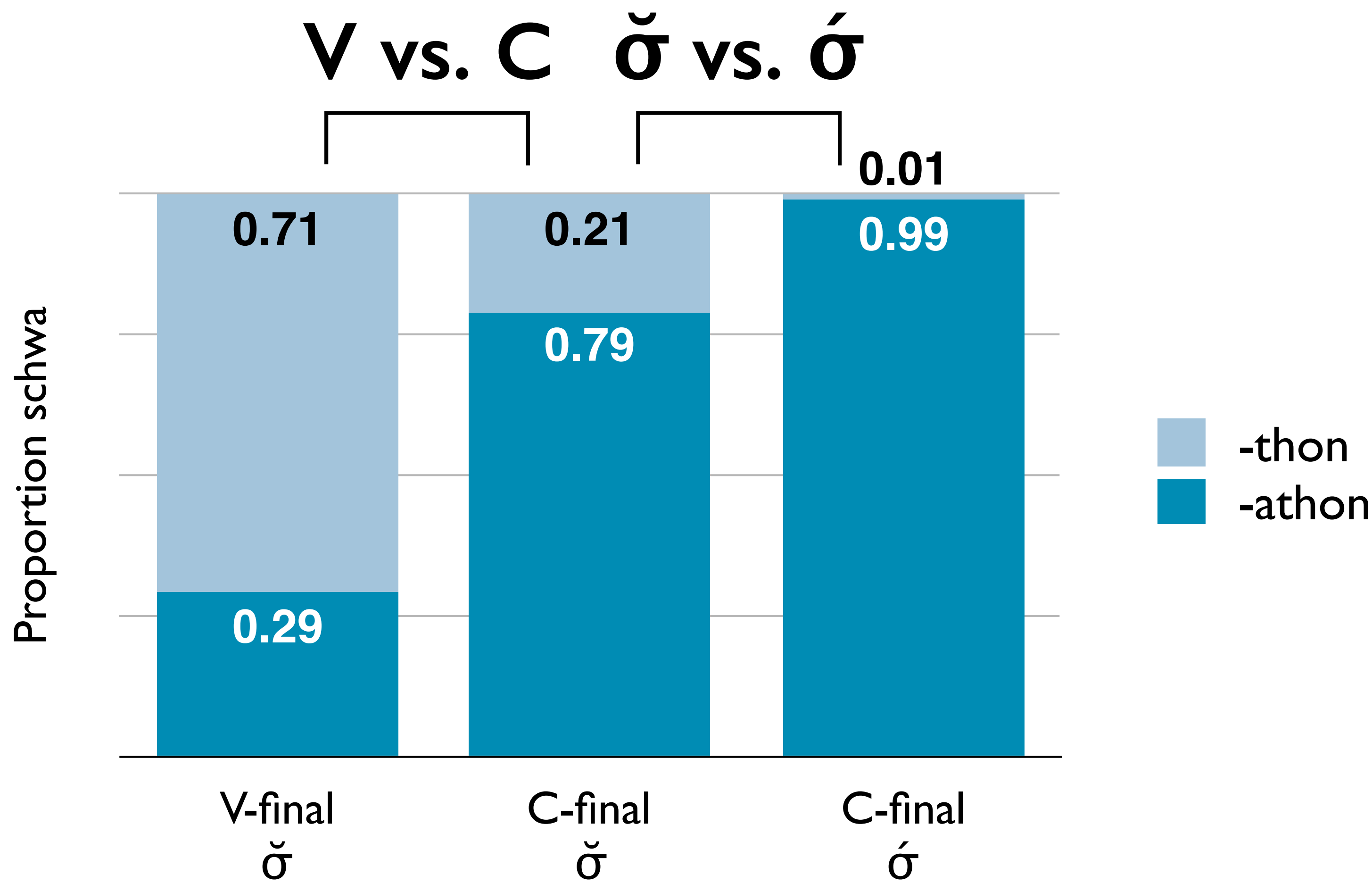
Experiment

- 109 participants after exclusions
  - All self-identified as native speakers of English
  - Only included data for American participants

Summary for *-(a)licious*



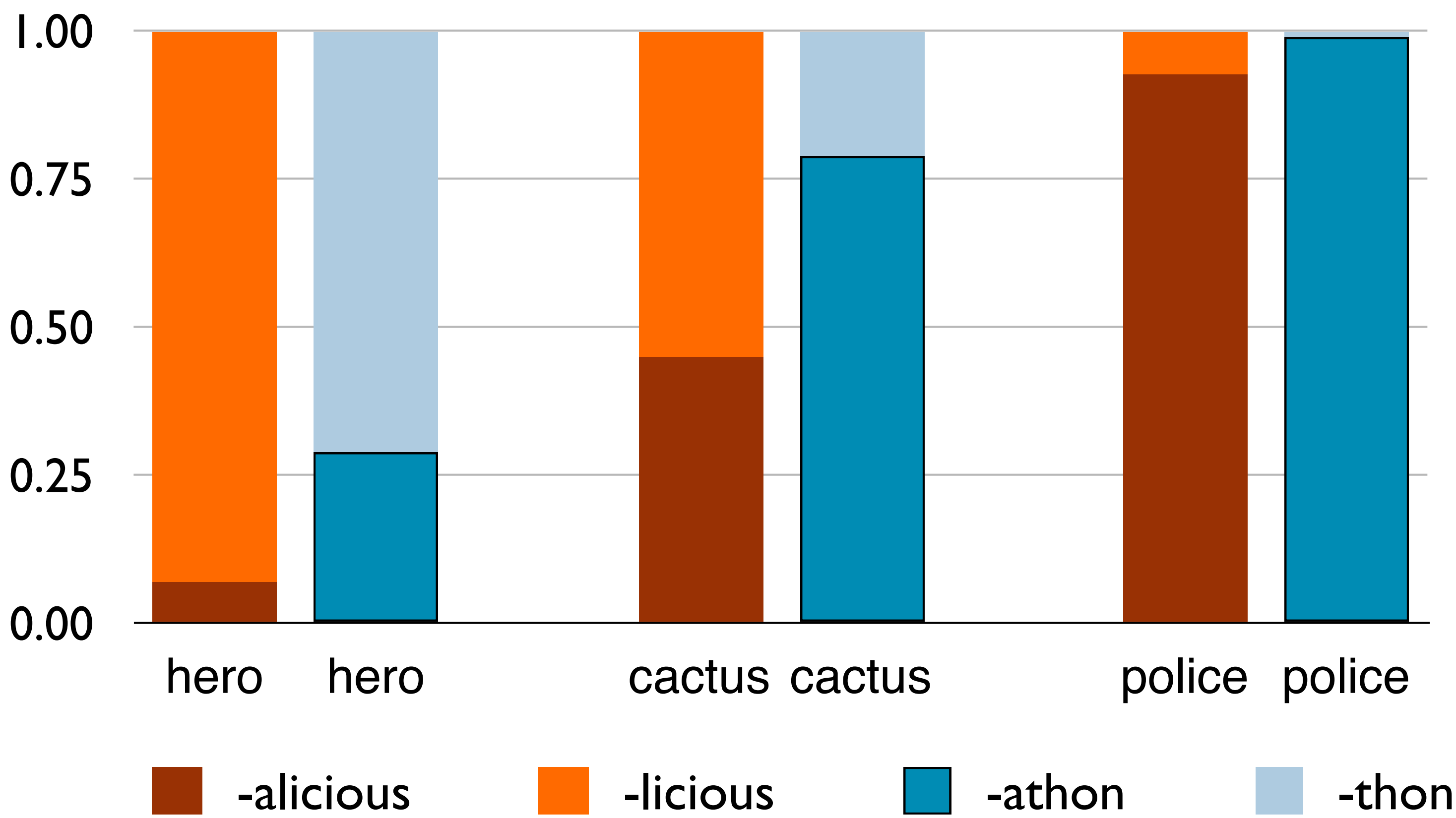
Summary for *-(a)thon*



-(a)licious and *-(a)thon*

- 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*

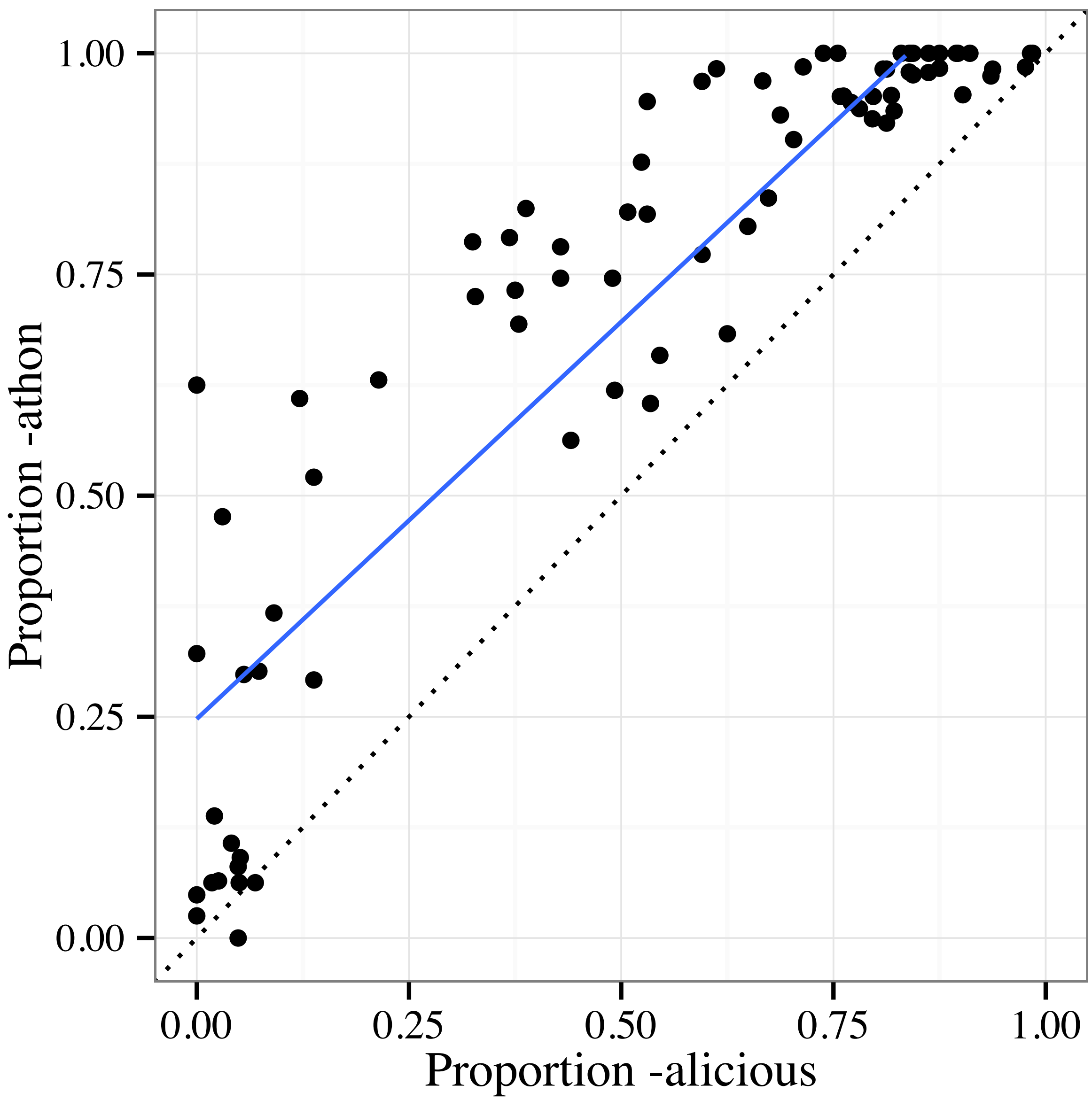
-(a)licious and *-(a)thon*



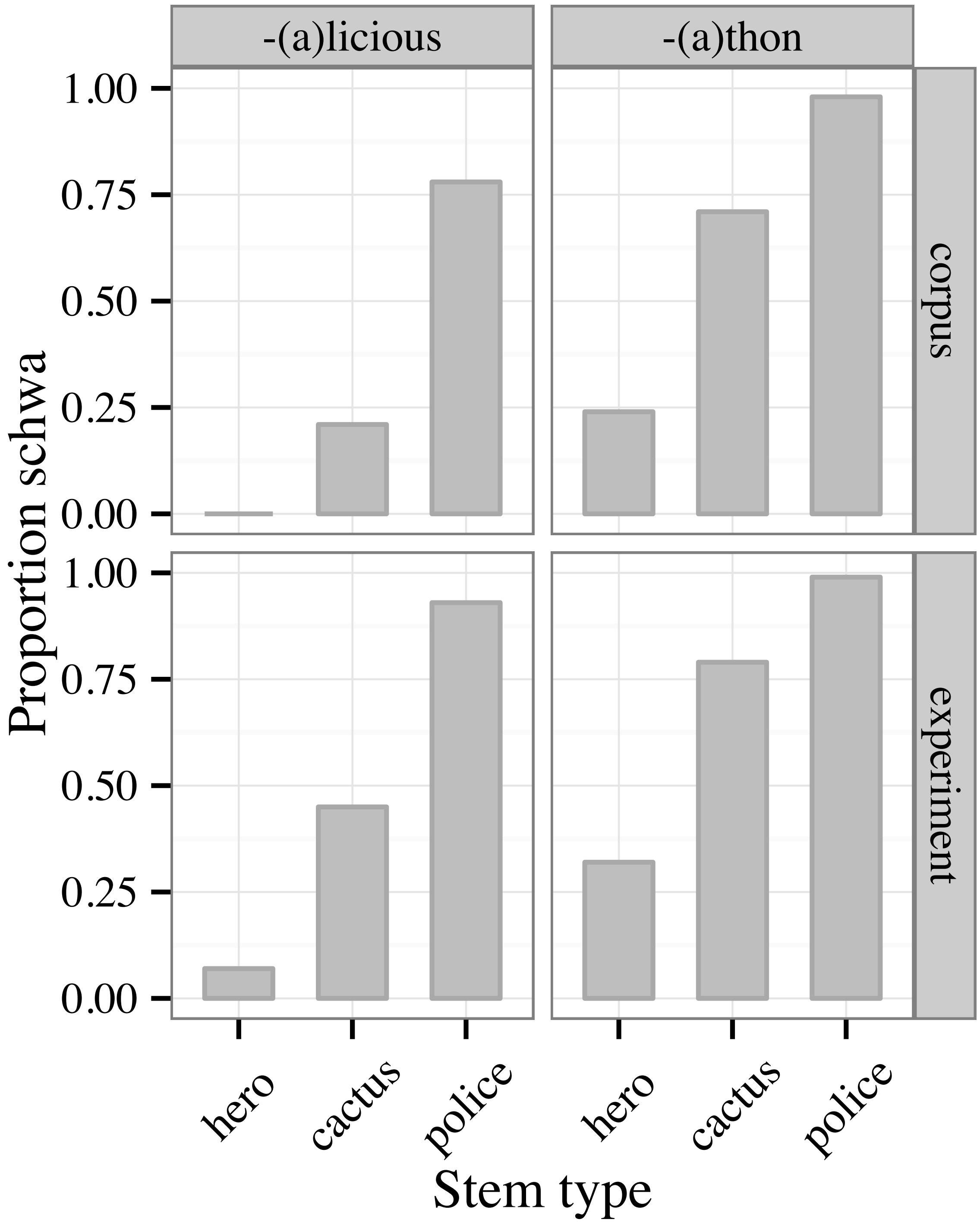
-(a)licious and *-(a)thon*

- 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



1.00
0.75
0.50
0.25
0.00
Proportion -athon
0.00
0.25
0.50
0.75
1.00
Proportion -alicious



-(a)licious
-(a)thon
corpus
experiment
Proportion schwa
1.00
0.75
0.50
0.25
0.00
hero
cactus
police
Stem type

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	↔	C __
-licious	↔	V __
-athon	↔	C __
-thon	↔	V __
-ery	↔	C __
-ry	↔	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…

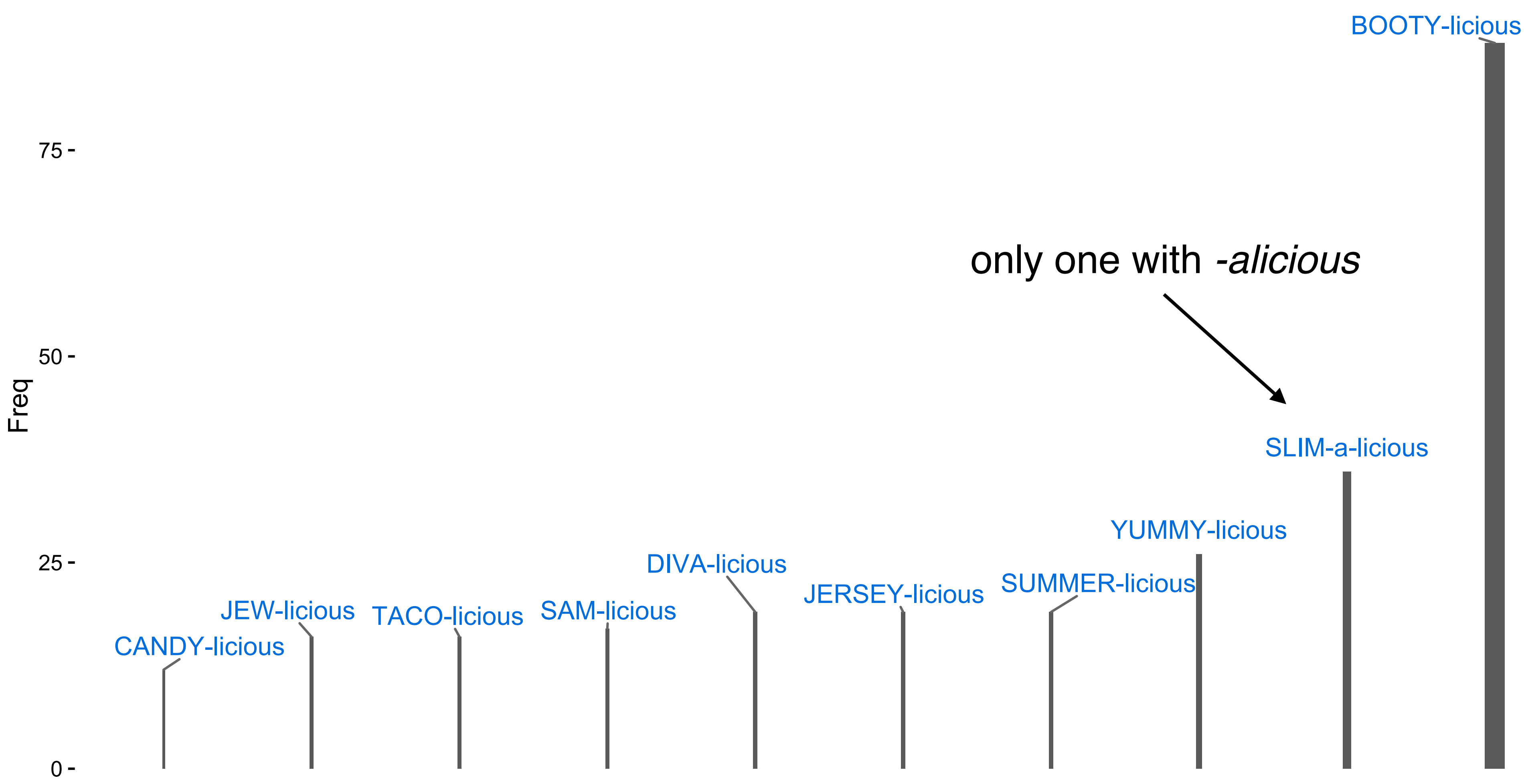
- In GlOWbE, licious- and thon-words are uncommon. Out of 500 million words for American and Canadian English (combined):
  - 933 tokens for *-(a)licious*
  - 1866 tokens for *-(a)thon*
- Assume a speaker hears 30,000 words/day…
  - 30 licious/year, 60 thon/year

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 σ́σ́ sequence
- *LAPSE:
  Assign a violation for every σ̆σ̆ 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 = /əlɪʃəs/ *(-alicious)*
    - UR = /lɪʃəs/ *(-licious)*
    - UR = /əθɑn/ *(-athon)*
    - UR = /θɑn/ *(-thon)*

Learned weights

Constraint	Weight	Constraint	Weight
*CLASH	2.61	-athon	1.97
*HIATUS	2.31	-alicious	0.37
*LAPSE	0.43	-licious	0.18
		-thon	0.13

*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

Constraint	Weight	Number of violators	
*CLASH	2.61	1,597	8%
*HIATUS	2.31	2, 792	13%
*LAPSE	0.43	8, 702	41%

Defaults

- For both *-athon* and *-alicious,* the default is the schwaful form
- The preference for *-athon* over *-thon* is greater than the preference for *-alicious* over *-licious*

Constraint	Weight
-athon	1.97
-alicious	0.37
-licious	0.18
-thon	0.13

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)

    thìrtéen → *thìrteen mén*
  - optional *that* (Lee and Gibbons 2007)

    I know (that) Lucy went vs. *I know (that) Louise went*
  - genitive alternation (Shih et al. 2015)

    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

  √radio, boa   *ɹedɪ.o, bɔ.ə (Chomsky & Halle 1968)

- Glottal stop epenthesis

  mora-[ʔ]ist (Plag 1999), sea [ʔ] otter (Davidson & Erker 2014)

- Intrusive R and intrusive L

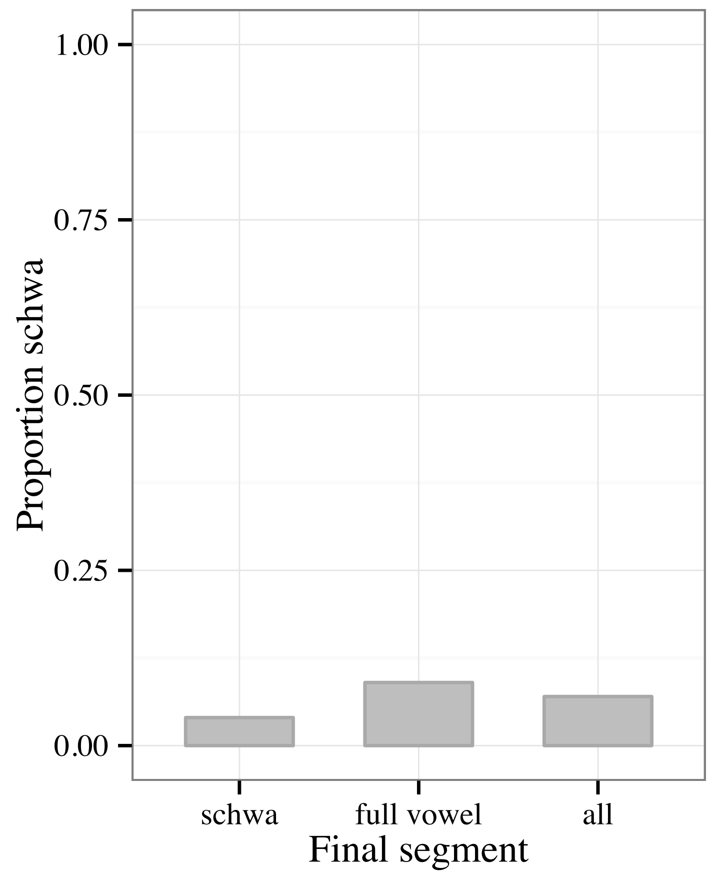
  draw[r]ing [dɹɑɹɨŋ] (McCarthy 1993); draw[l]ing [dɹɔlɨŋ] (Gick 2002)

- *a/an* allomorphy and function word reduction

  an apple, a pear; [ði] apple, [ðə] pear (overview in Smith 2015)

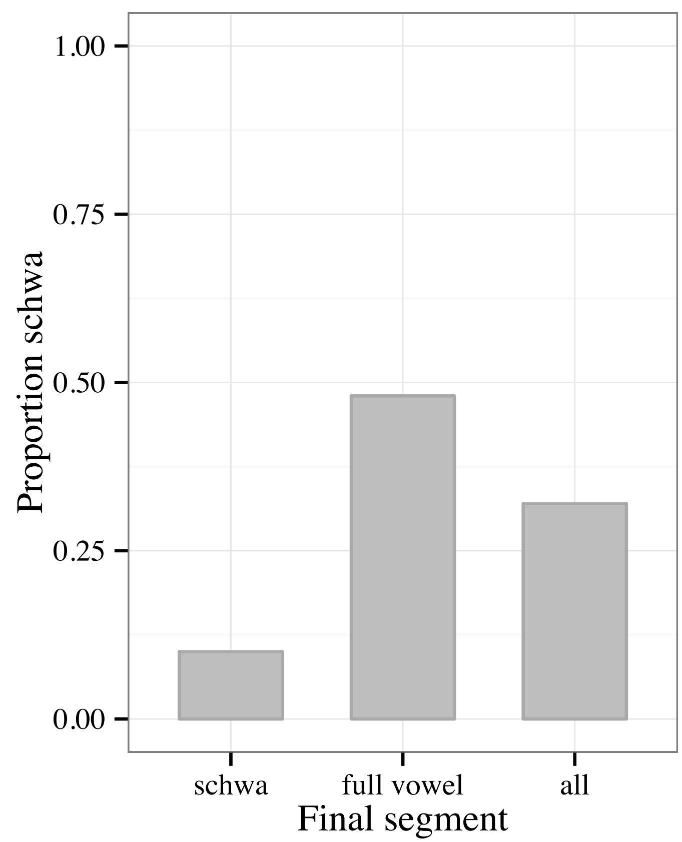
*HIATUS in *-(a)licious*

- 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

%schwa	Target	Model
POLICE-thon	99	99
CACTUS-thon	79	94
HERO-thon	48	87
SODA-thon	10	1

Model's performance on *-(a)thon*

Grammar contains a constraint against obstruent-θ sequence

%schwa	Target	Model
POLICE-thon	99	99
CACTUS-thon	79	94
HERO-thon	48	87
SODA-thon	10	1

Model's performance on *-(a)thon*

Grammar doesn't contain a general constraint against hiatus

%schwa	Target	Model
POLICE-thon	99	99
CACTUS-thon	79	94
HERO-thon	48	87
SODA-thon	10	1

Model's performance on *-(a)licious*

- Model doesn't capture a difference between HERO and CACTUS

%schwa	Target	Model
POLICE-licious	93	94
CACTUS-licious	45	46
HERO-licious	9	46
SODA-licious	4	0

Model's performance on *-(a)licious*

Grammar doesn't contain a general constraint against hiatus

%schwa	Target	Model
POLICE-licious	93	94
CACTUS-licious	45	46
HERO-licious	9	46
SODA-licious	4	0

Improving the BLICK grammar

- Assume *HERO+ə* violates the stress-sensitive constraint against hiatus in the grammar (*V́V)
- Assume constraint against obstruent followed by θ doesn't operate across morpheme boundaries

With improved BLICK grammar

	Target	Model
POLICE-thon	99	99
CACTUS-thon	79	87
HERO-thon	48	51
SODA-thon	10	1

	Target	Model
POLICE-licious	93	94
CACTUS-licious	45	46
HERO-licious	9	12
SODA-licious	4	0

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 *-(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