The phonological grammar is probabilistic: New evidence pitting abstract representation against analogy

Claire Moore-Cantwell

Yale university

October 9, 2015

Introduction

Speakers extend probabilistic trends in their lexicons to new words

- **Example:** Initial stress in English
  - a majority of 2-syllable words have initial stress (about 75%)
  - but stable exceptions are plentiful: *guitár*, *garáge*, *devíce*
  - English speakers prefer initial stress in novel words (Guion et al., 2003)
  - Probabilistically:
    - They sometimes produce finally-stressed nonwords as well
    - The rate of initial stress can be influenced by other factors
      - Part of speech
      - Syllable weight

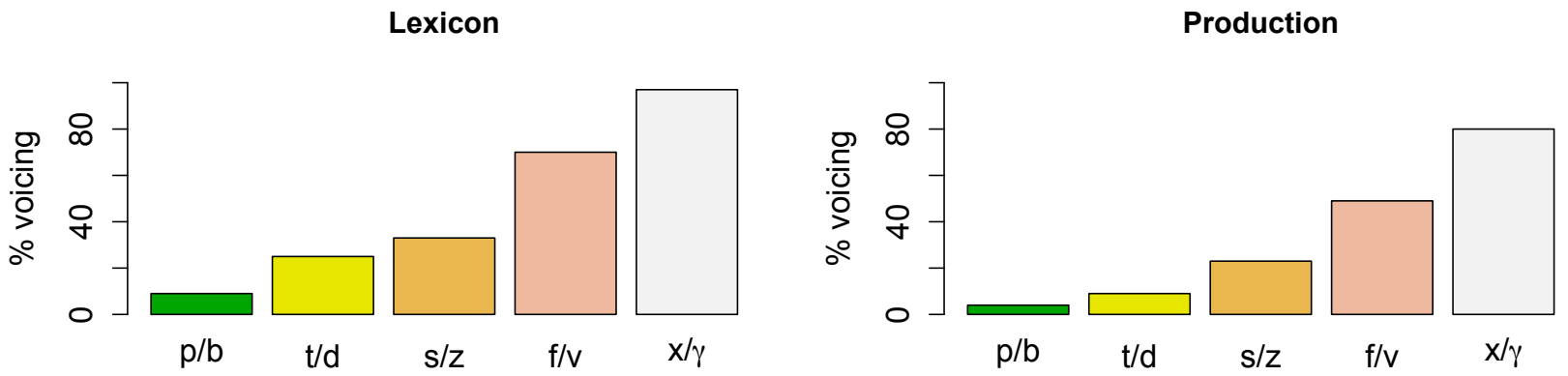
- **What is the cognitive mechanism that underlies this ability?**

Introduction

Speakers extend probabilistic trends in their lexicons to new words

- **Example:** Dutch voicing alternations (Ernestus and Baayen, 2003)

  [vɛrʊɛi**d**ən], [vɛrʊɛi**t**ən] → [vɛrʊɛi**t**]



- Similar results: Hayes et al. (2009); Becker et al. (2011); Zuraw (2000, 2010) and many others

Introduction

Speakers extend probabilistic trends in their lexicons to new words

- They 'probability match'
- Rather than categorically choosing the most common pattern
- ? **Grammar contains probabilistic generalizations?**
  - Represents not just what to do, but also how often to do it
- Or are these trends represented some other way?
  - Analogy to existing items
  - Statistical learning: Cognitively general mechanism

Introduction

Speakers extend probabilistic trends in their lexicons to new words

- They 'probability match'
- Rather than categorically choosing the most common pattern
- ? **Grammar contains probabilistic generalizations?**
  - Represents not just what to do, but also how often to do it
- Or are these trends represented some other way?
  - ? **Analogy to existing items**
  - Statistical learning: Cognitively general mechanism

Introduction

1. Case of probability matching in the English stress system
2. **Is analogy happening?**
   - Use nonwords with no near lexical neighbors
   - Ask participants to provide potential analogical bases
   - Compare: Stress of analogical base to produced stress
   - Guion et al. (2003): Effects of analogical base AND phonological generalizations

Analogy

How do you choose what to analogize to?

- Randomly choose a word
  - No guarantee that your word will have the necessary properties
- Use the entire lexicon
  - Divide the lexicon up into categories; choose the one where all the words match your nonword in some relevant way (Skousen, 1989)
  - Calculate the phonetic similarity between your nonword and each actual word (Nakisa et al., 2001)
- Choose a word based on similarity
  - Lookup words using feature(s) of the nonword
  - Use Lexical access mechanism?

  e.g. TRACE (McClelland and Elman, 1986)

English Stress

Chomsky and Halle (1968); Halle and Vergnaud (1987): 'Latin Stress Rule'

(A) Stress a heavy penultimate syllable (*amálgam*)

- Very few exceptions in the lexicon (*ga<u>l</u>axy*,*char<u>a</u>cter*)
- Obeyed in speakers' productions (Domahs et al., 2014; Olejarczuk, 2014)

(B) else stress antepenult (*cánopy*)

- Exceptions abound (*vanílla, banána, spaghétti, canáry* . . .) (Pater, 1994)
- Not obeyed in speakers' productions (Domahs et al., 2014)

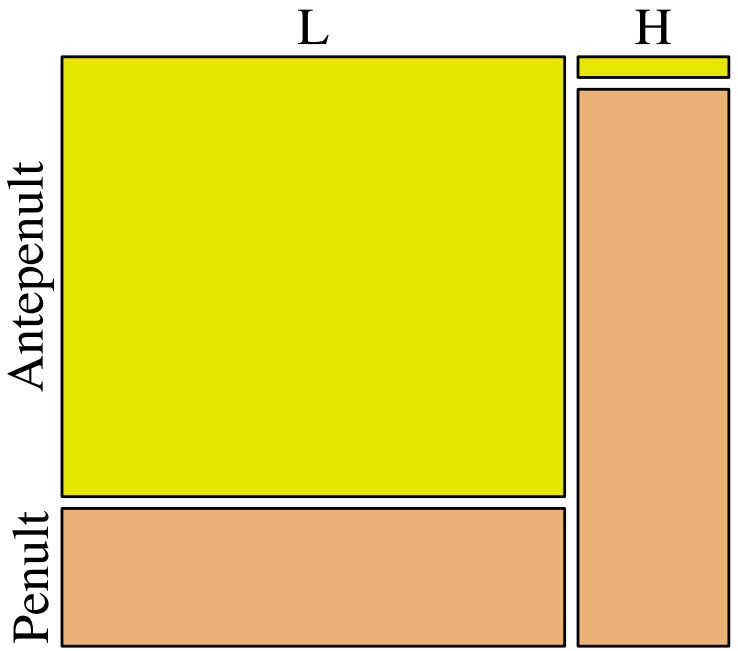
English Stress

Corpus search:

- **Corpus:** CMU pronouncing dictionary (Weide, 1994)
- Frequency threshold: SubtLex (Brysbaert and New, 2009)
- All words 3+ syllables
- Automatic annotation: syllable structure, vowel qualities, stress pattern

English Stress

Chomsky and Halle (1968); Halle and Vergnaud (1987): 'Latin Stress Rule'



H: CVV,CVC*

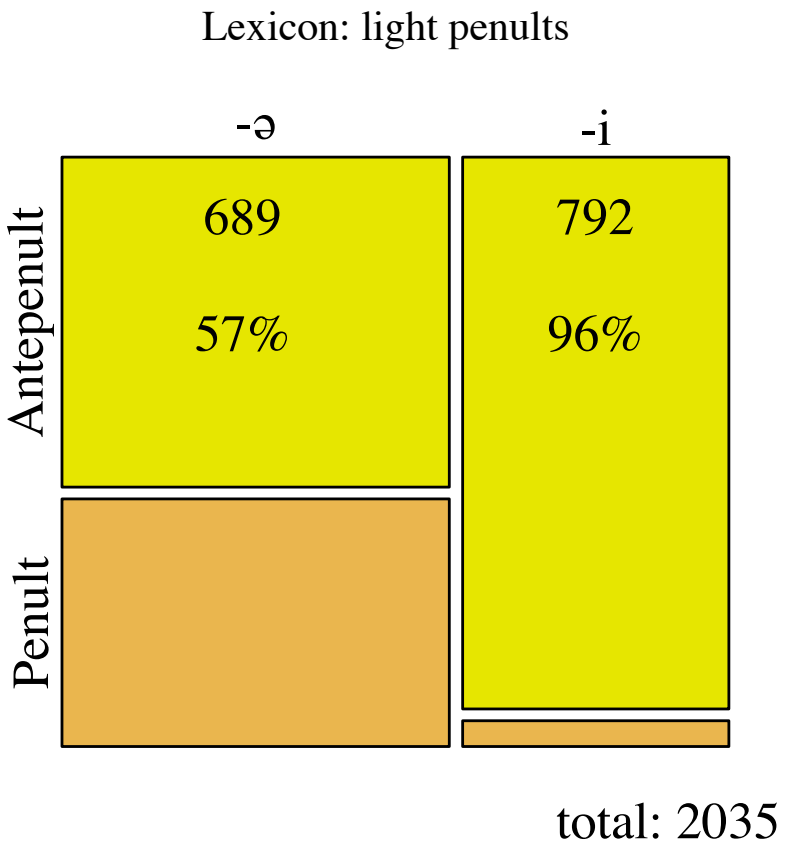
L: CV

Heavy penult: *aróma*, *bonánza*

Light penult: *tobóggan*, *elícit*

English Stress

Stress is partially conditioned by the final vowel



If [ə]-final, no preference

If [i]-final, then Antepenultimate

Final [i] drives stress leftward (Hayes, 1982; Liberman and Prince, 1977)

Strategy

(1) Does the i-final trend generalize to nonwords?

(2) Do speakers use analogy to do so?

- Methods very similar to Guion et al. (2003)
- Part 1: wug test
- Part 2: same nonce words again, this time fill-in-the-blank

  What real word does it remind you of?
- Web-based experiment using Amazon Mechanical Turk

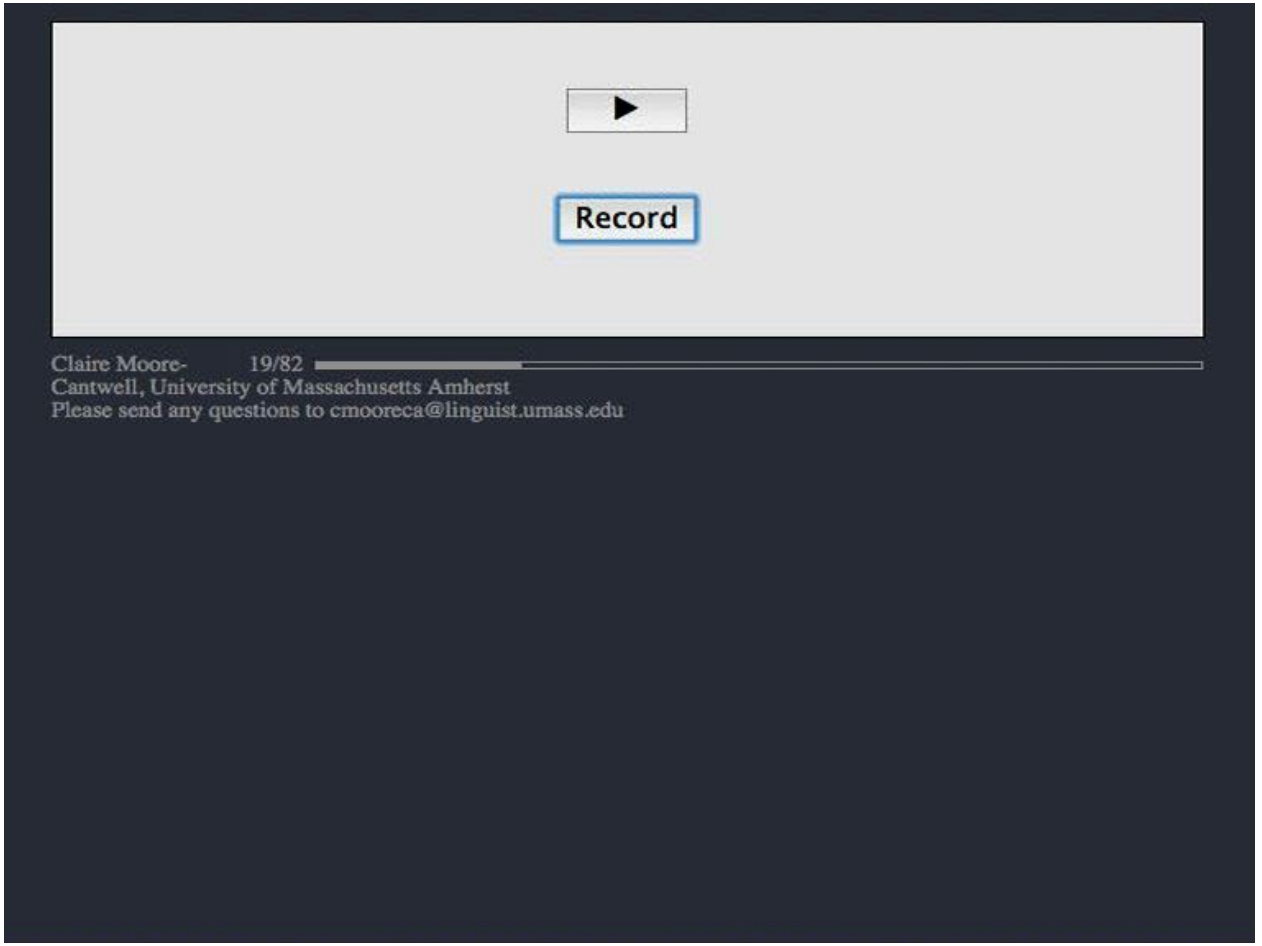
Methods

Wug test

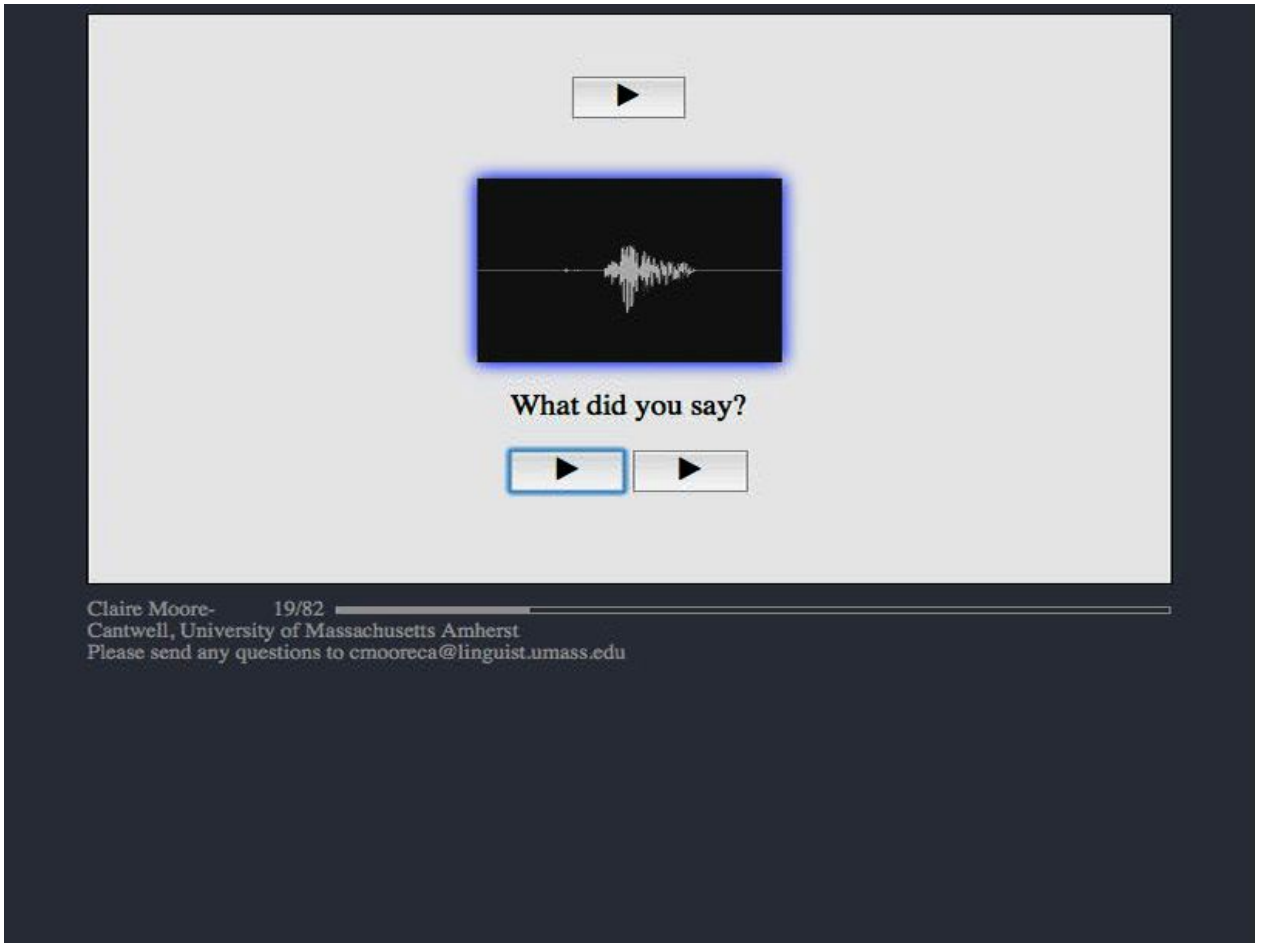
- Isolated syllables presented auditorily: [bǽ] [mǽ] [kǐ]
- Participants speak the word 'fluently'
- Both stress options presented: [bəmǽki], [bǽməki]
- Participants choose one

  → Forced choice as proxy for production

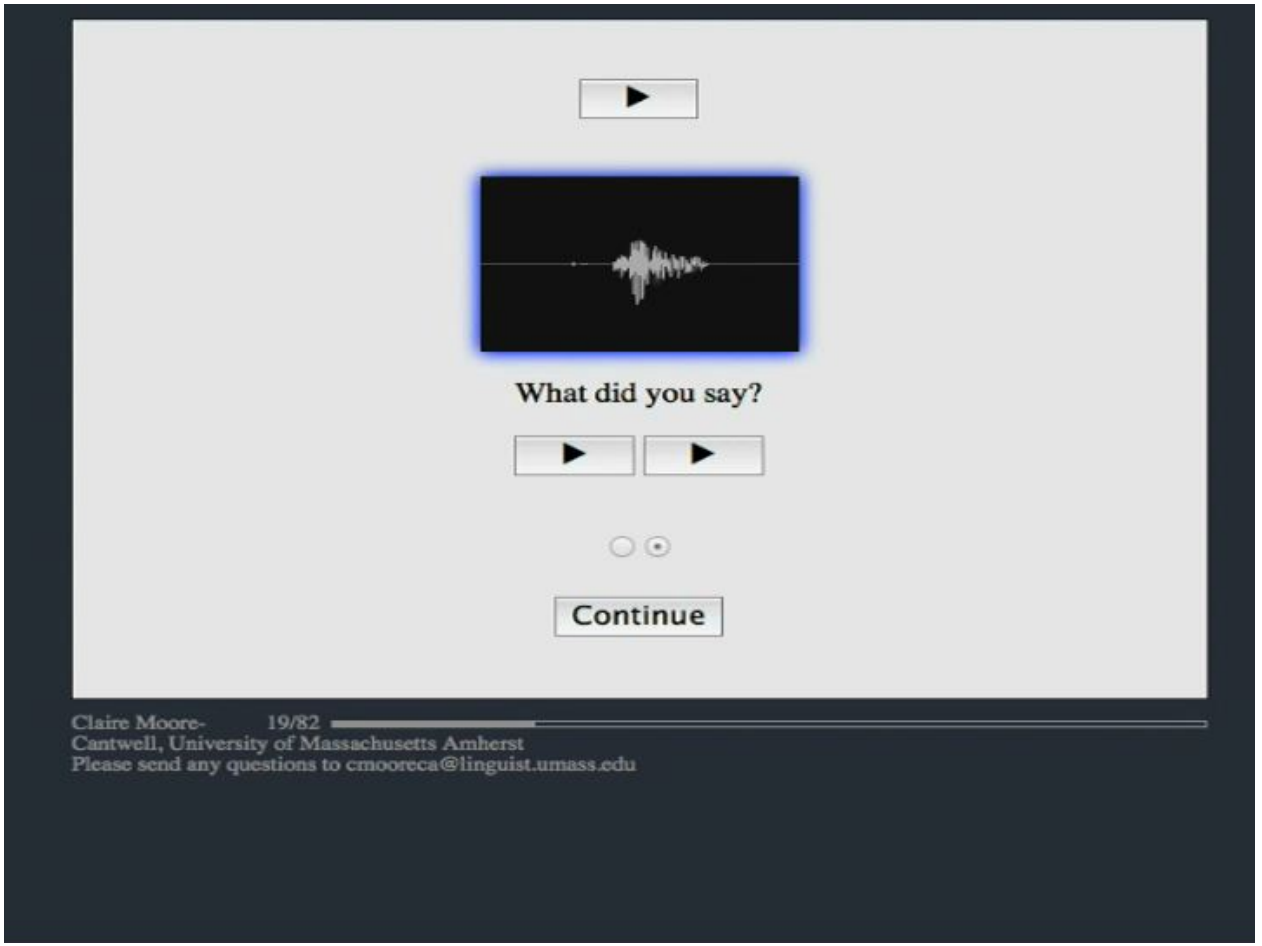
Methods


Record
Claire Moore-
Cantwell, University of Massachusetts Amherst
19/82
Please send any questions to cmooreca@linguist.umass.edu

Methods


What did you say?
Claire Moore-
Cantwell, University of Massachusetts Amherst
19/82
Please send any questions to cmooreca@linguist.umass.edu

Methods


What did you say?
Continue
Claire Moore-
Cantwell, University of Massachusetts Amherst
19/82
Please send any questions to cmooreca@linguist.umass.edu

Methods

Getting potential analogical bases

- Isolated syllables presented again: [bǽ] [mǽ] [kí]
- 'What English word does the sequence of syllables remind you of?'
- Participants filled in a blank

  → Word most likely to serve as analogical base

Methods

Details:

- 48 Participants recruited through Amazon Mechanical Turk
- Presented using Experigen (Becker and Levine) plus a plugin for recording over the web
- 32 nonword items, 8 real word fillers

  Nonwords selected to have very low neighborhood density under (Bailey and Hahn, 2001), GNM value < 0.01
- 20 minutes total

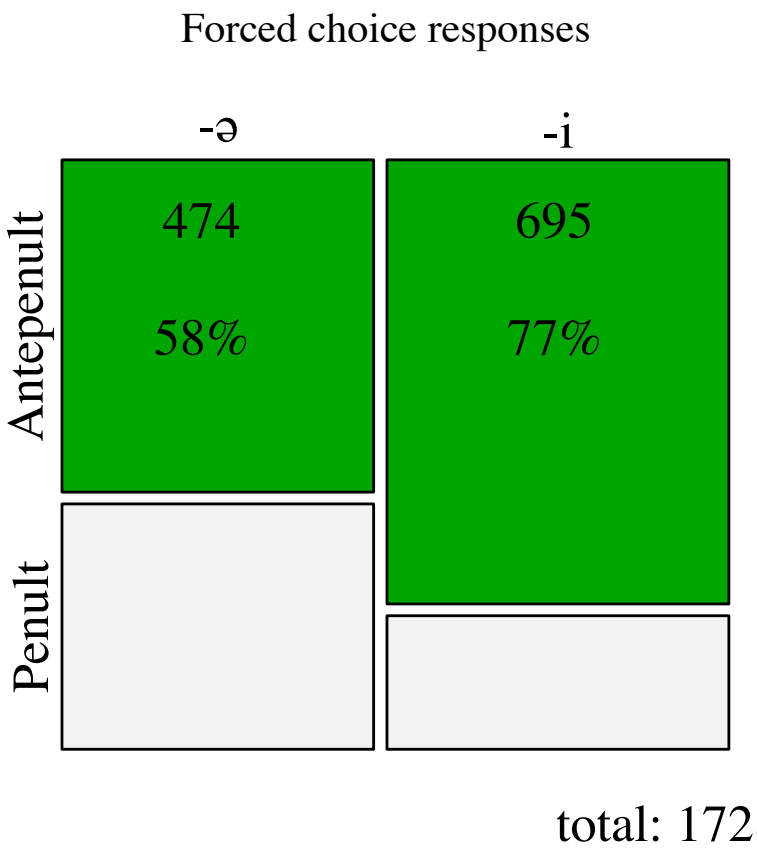
Results

General:

- Most participants succeeded at the production task
  - Produced e.g. [bǽməki] not [bǽmǽkí]
  - Chose the sound file that corresponded to their production
    $\rightarrow$ Can trust forced choice data
- Analogical base task was harder
  - Provided an actual word about 58% of the time
  - Rest of the time: transcribed the nonword
  - or gave no answer

Results

Results of production task

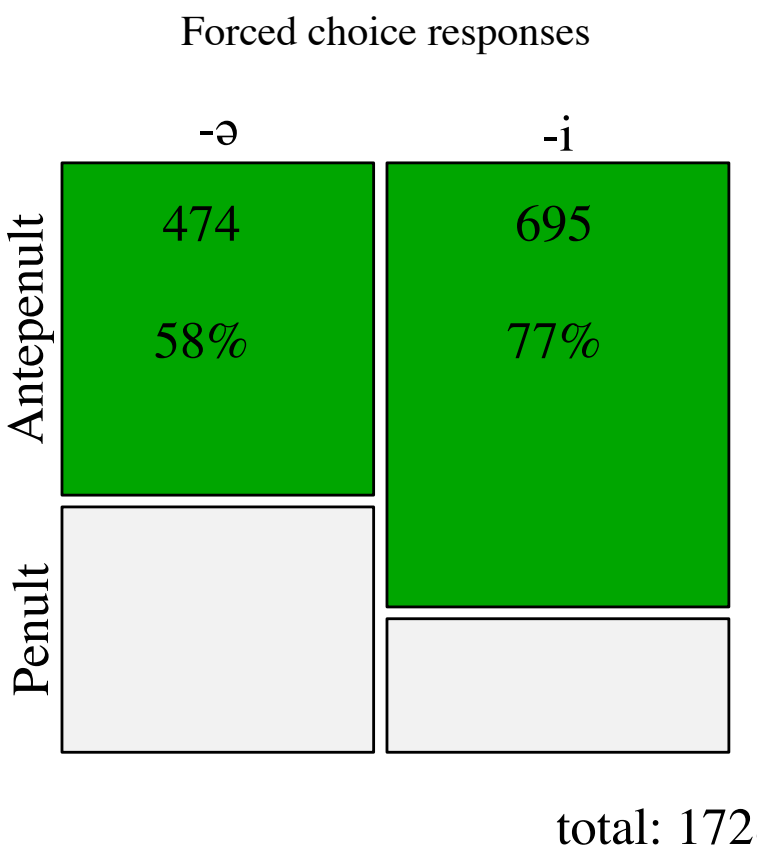


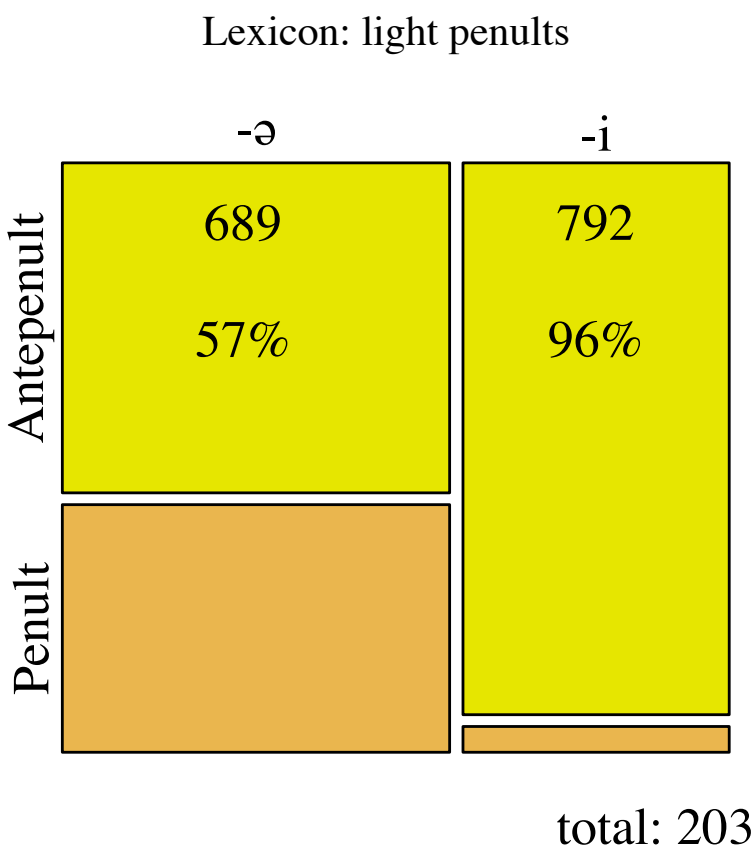
i-final More Antepenult

ə-final Equal

Results

Compare





Results

Properties of analogical bases:

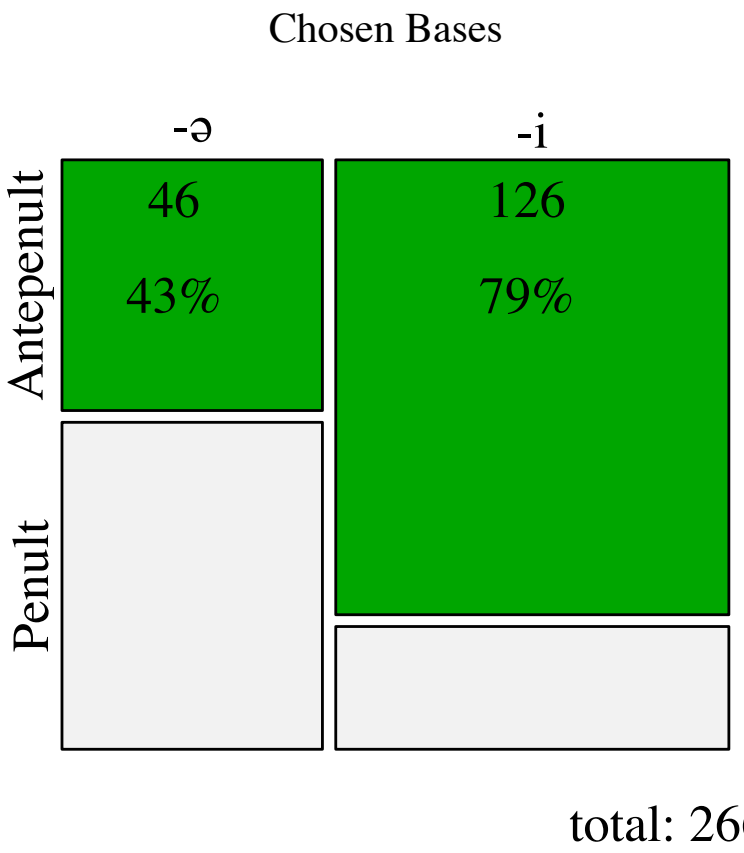
- Favored 3-syllable words

Number of Syllables				
1	2	3	4	5
194	221	411	58	3
22%	25%	53%		

- Matched final vowel 91% of the time

Results

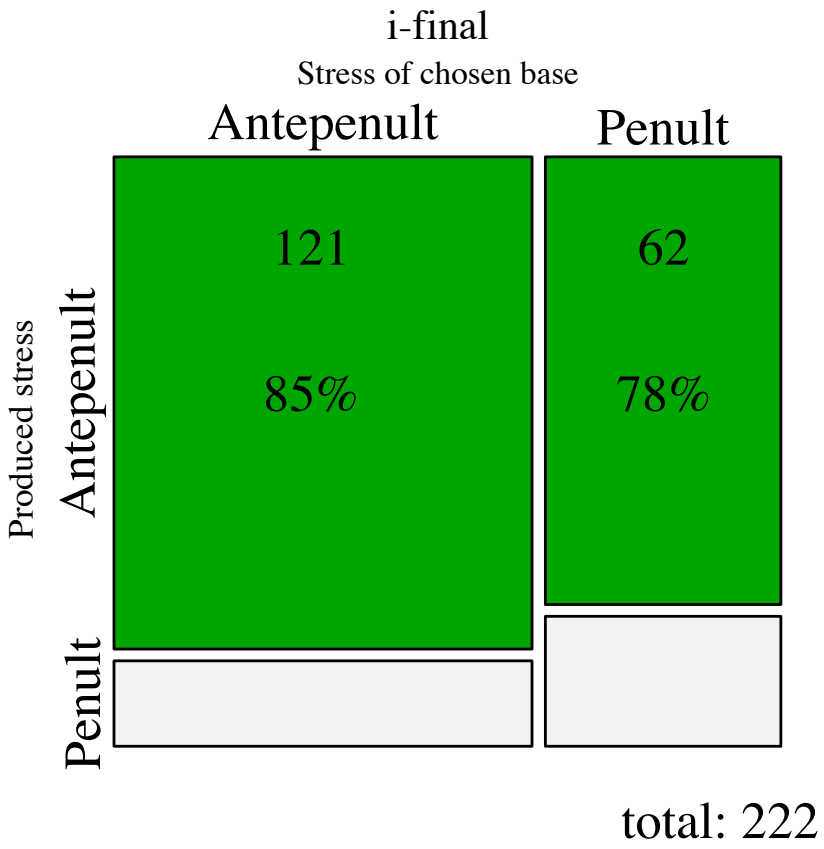
Properties of analogical bases:

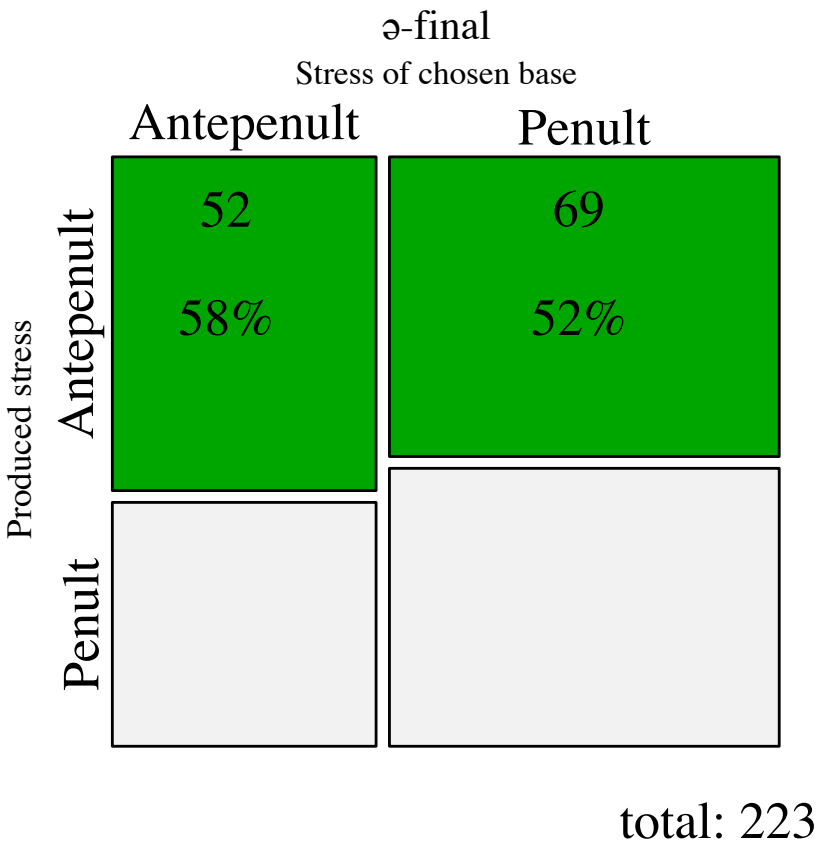


More Antepenult in i-final bases

Results

Does base stress predict produced stress?





Results

Does base stress predict produced stress?

Logistic regression with two factors:

Model: *Produced Stress* $\sim$ *Final Vowel* + *Analogical Base Stress*

	Estimate	p
Intercept	-0.54	0.02
Final Vowel = i	-1.22	0.0001
Analogical Base Stress = Penult.	0.42	0.20
AIC:	290	

remove:	change in AIC	Likelihood ratio	p
Final Vowel	+13	15.66	0.0001
Analogical Base Stress	0	1.7	0.20

Results

- What if participants access a different real word each time they hear the nonword stimulus?
- But they're still using analogy
- ? **What behavior is predicted for each nonword based on the set of nearby real words?**

Stimulus	[rɛ vɛ si]						
Analogical Base	légacy	lívery	prívacy	régistry	rémedy	revéal	receive
no. Responses	1	1	1	1	1	1	1

83% Antepenult, 17% Penult

Stimulus	[sɛ fɛ ni]						
Analogical Base	sýmphony	fámily	sésame	safári	sapphire	say	save
no. Responses	8	1	1	1	1	1	1

91% Antepenult, 9% Penult

Results

- What if participants access a different real word each time they hear the nonword stimulus?
- But they're still using analogy
- ? **What behavior is predicted for each nonword based on the set of nearby real words?**

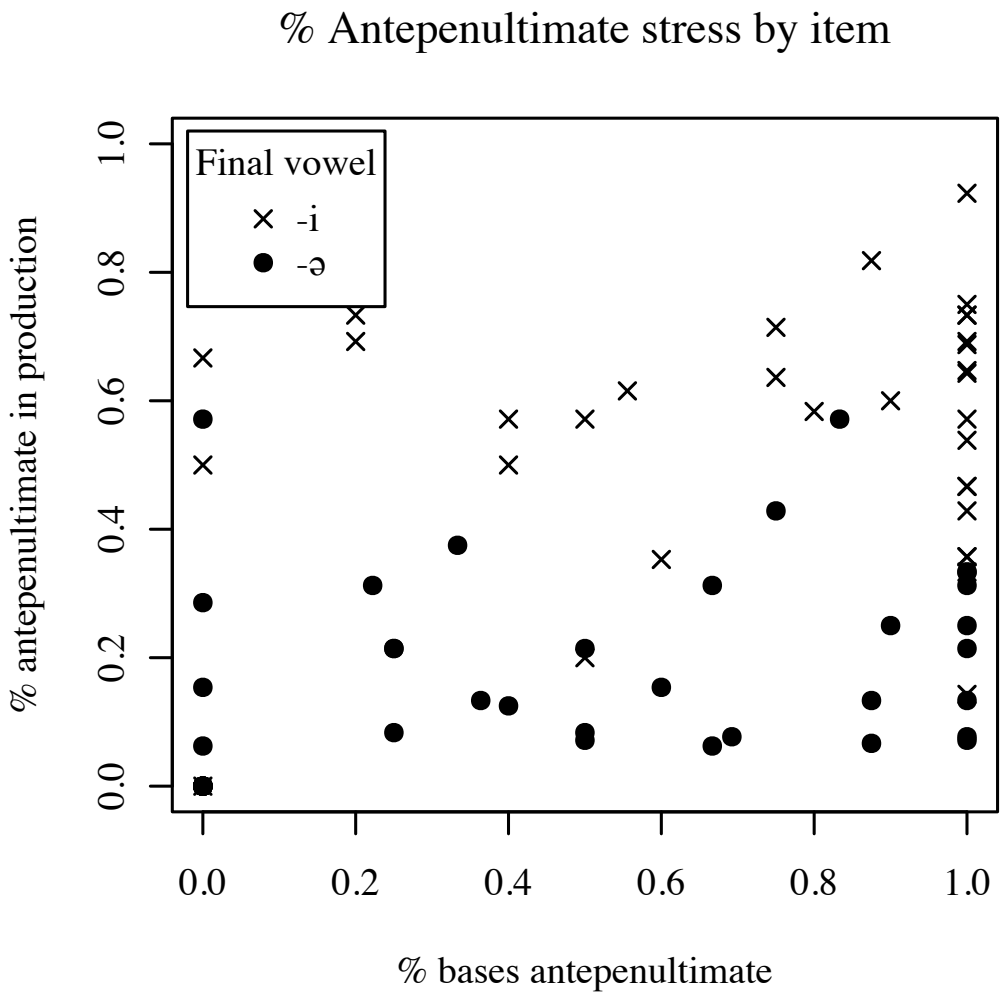
Stimulus	[rɛ vɛ sə]					
Analogical Base	revísion	revérsal	revise	rabbit	vista	vivid
no. Responses	5	1	2	1	1	1

0% Antepenult, 100% Penult

Stimulus	[sɛ fɛ nə]					
Analogical Base	sýmphony	savánna	secondary	seven	saffron	safe
no. Responses	2	2	1	1	1	1

50% Antepenult, 50% Penult

Results


% Antepenultimate stress by item
Final vowel
-i
-ə
% antepenultimate in production
0.0
0.2
0.4
0.6
0.8
1.0
% bases antepenultimate

Results

- Participants 'probability matched' antepenultimate stress on i-final words
- They also observe this trend in their choice of analogical bases

→ But the stress of the base does not predict stress in production

Participants' probability matching seems not to be the result of analogy to exisiting items

Conclusions

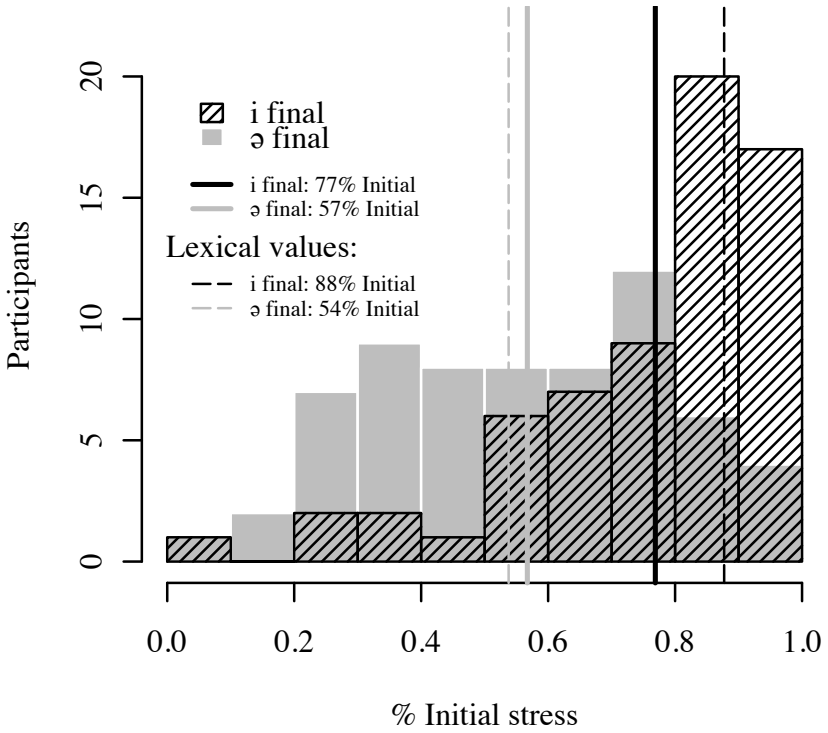
- Analogy is not responsible for the productivity of the i-final trend
- Previous studies (Guion et al., 2003; Baker and Smith, 1976) showed effects of BOTH analogy and abstract generalization

  $\rightarrow$ Used words with richer neighborhoods, in some cases near neighbors (cinempa)
- Here: no effect of analogy at all
  - Nonwords were very far from any actual word
  - Speakers can extend the i-final trend to nonwords even when analogy is difficult

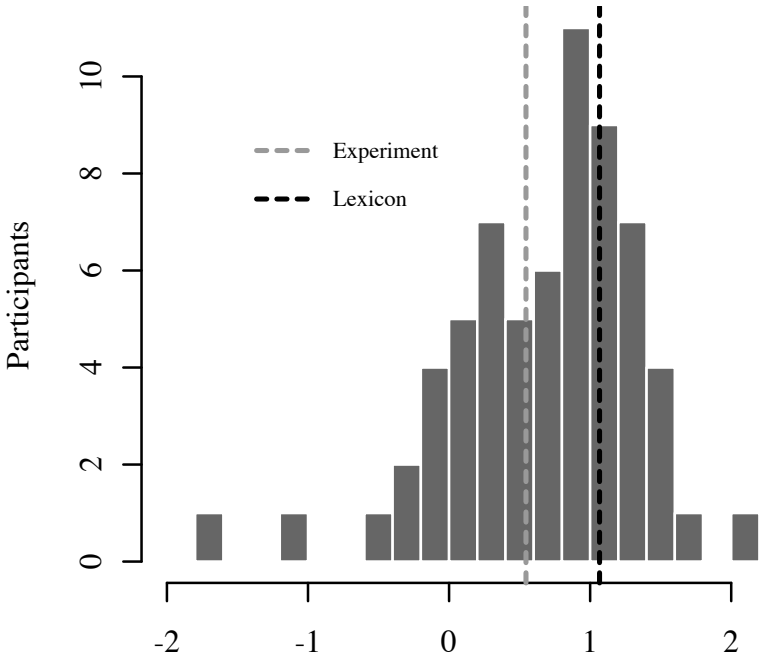
  $\rightarrow$ **Abstract representation of the i-final trend**

Thank You

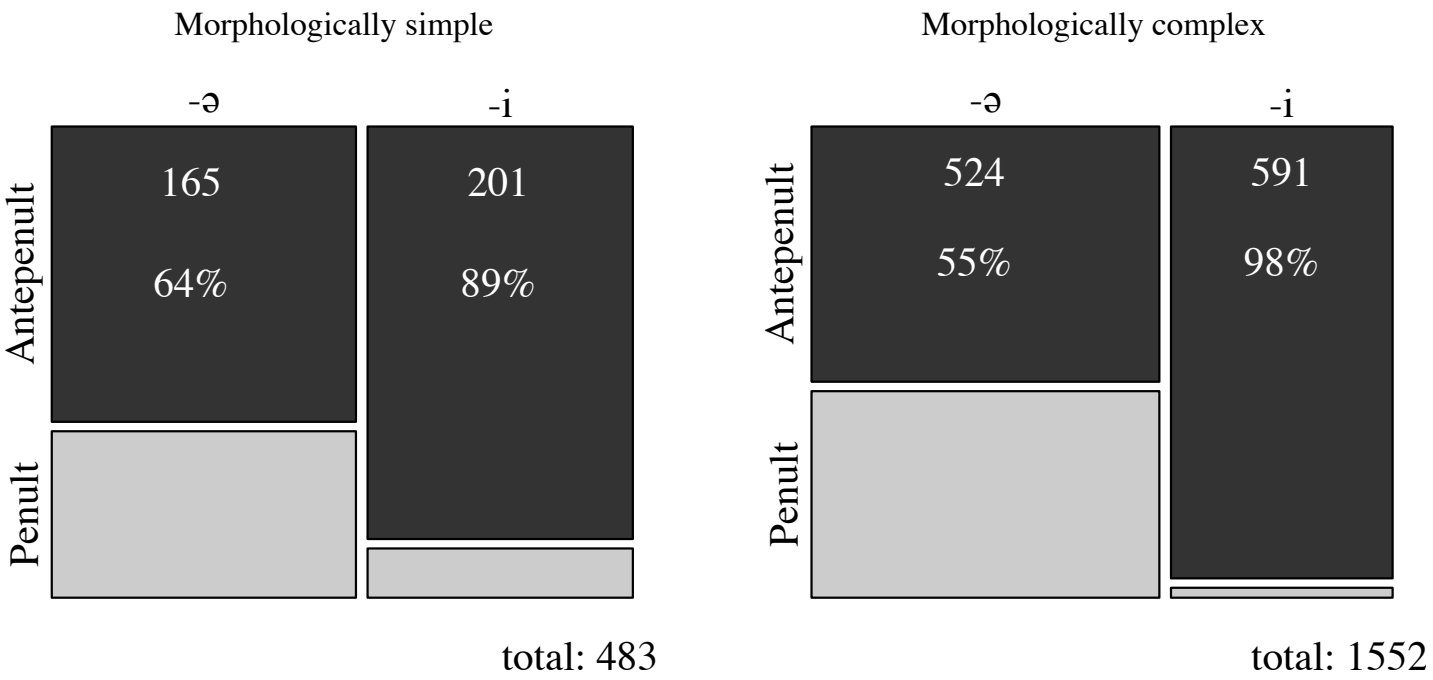
Individual Subjects



d' =Z(% Initial, i-final) - Z(%Initial, ə-final)



Morphology?


Morphologically simple
-ə
-i
Antepenult
Penult
165
64%
201
89%
total: 483
Morphologically complex
-ə
-i
Antepenult
Penult
524
55%
591
98%
total: 1552

Introduction

Categorical phonology: Grammar

- Inexorably applies to new words
  - Regardless of similarity to actual words (Prasada and Pinker, 1993)
  - Speakers cannot viridically perceive violations: [dla] $\rightarrow$ 'gla' (Moreton, 2002; Breen et al., 2013)
- Hard to un-learn
  - Learning the sound pattern of a second language is not simply a matter of learning the words
  - Experimental cases: (Finn and Kam, 2008; Whalen and Dell, 2006)
- Limited range of possible patterns
  - Some categorical patterns are common: Antepenultimate stress
  - Others surprisingly rare: Post-peninitial stress