CONSTRAINING PHONOLOGY COMPUTATIONALLY: EXPERIMENTAL EVIDENCE

NECPHON 2011

Regine Lai

10/15/2011

rlai@udel.edu

About this study

Goal

- To explore the universal restrictions for phonology

How?

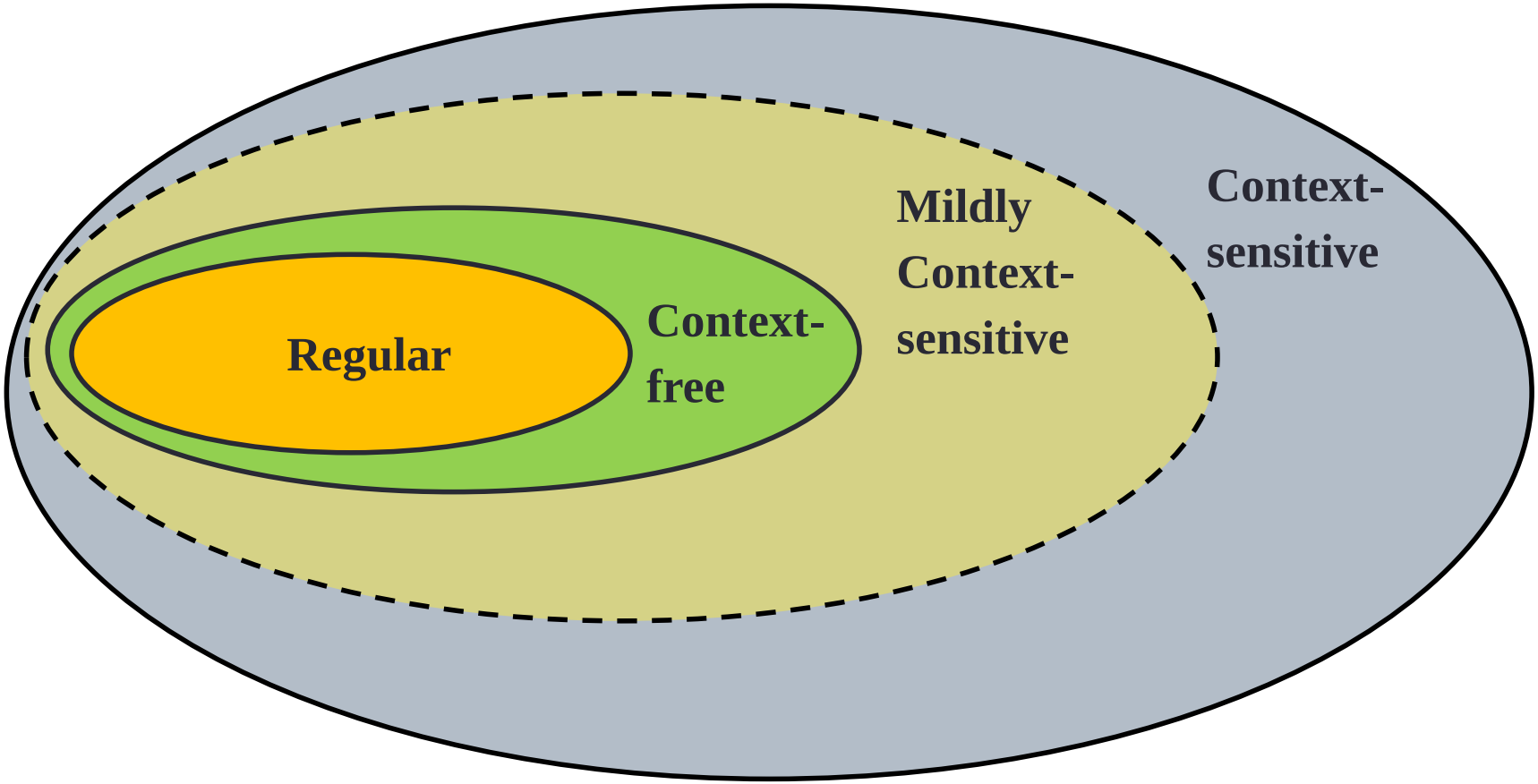
- Test the learnability of a particular phonologically plausible sound pattern which is not found in any natural languages and not within the identified computational regions.
- Artificial language learning experiments.

Implications

- What constitutes a possible phonological pattern
- Provides insights into human's phonological learning mechanisms.

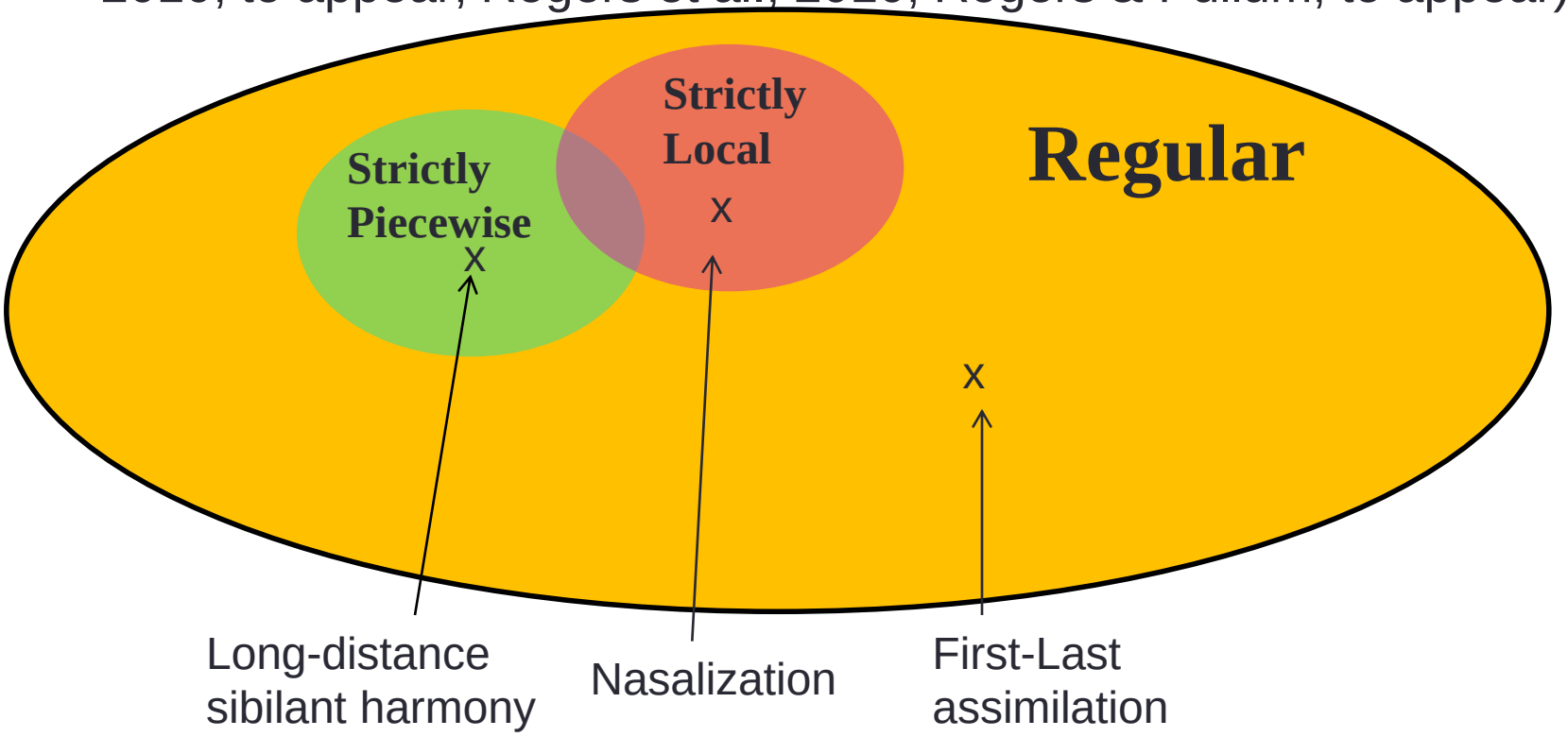
The Chomsky Hierarchy

- Phonology is regular (Kaplan & Kay, 1994)



The Regular Region

- Some identified subregular classes:
  - ***Strictly-Piecewise (SP)*** and ***Strictly-Local (SL)*** (Heinz, 2009, 2010, to appear; Rogers et al., 2010, Rogers & Pullum, to appear)



Sibilant Harmony (SH)

- Attested in Chumash.
- If 2 or more sibilants appear in the same word, they have to be agree in anteriority.

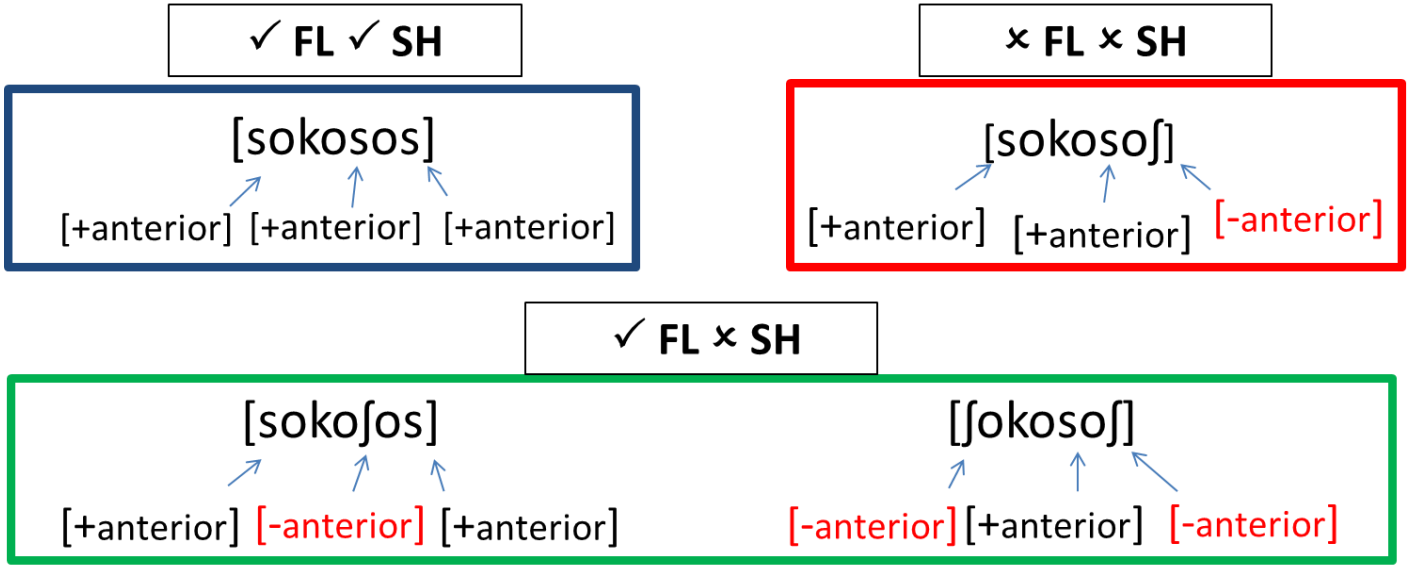
- For example:
- [**s**oko**s**o**s**] ✓
- [**s**oko**ʃ**o**s**] ✗
- [**ʃ**oko**s**o**s**] ✗

First-Last Assimilation (FL)

- If there are sibilants in both the initial and final position of a word, they have to agree [anterior].

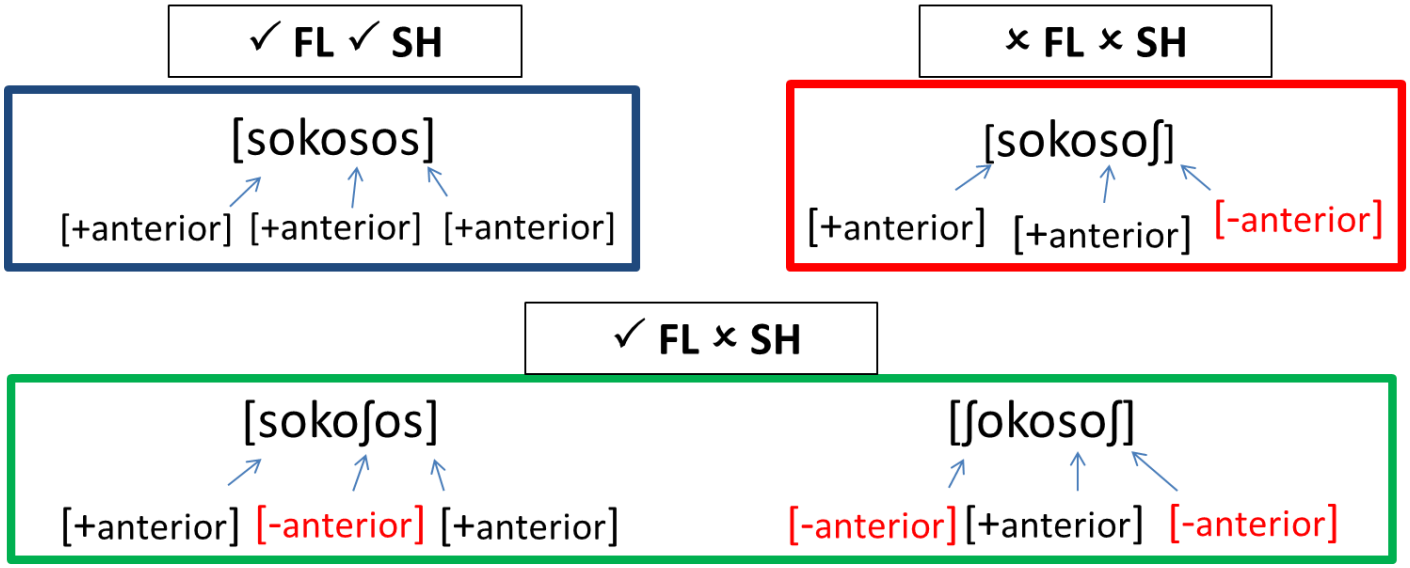
- For example:
- [**s**oko**s**o**s**] ✓
- [**s**oko**ʃ**o**s**] ✓
- [**ʃ**oko**s**o**s**] ✗

Sibilant Harmony vs. First-Last Assimilation



Note: **×FL✓SH** is not present because anything that obeys SH, also obeys FL.

Sibilant Harmony(SH) vs. First-Last Assimilation(FL)



Sibilant Harmony

	[s]	[ʃ]
[s]	✓	×
[ʃ]	×	✓

[_…_]

First-Last Assimilation

	[s]	[ʃ]
[s]	✓	×
[ʃ]	×	✓

[#_…_#]

FL is not a random choice

- 1. LD sibilant harmony is attested;
- 2. word edges are relevant in phonology;
- 3. initial and final positions of a word are salient positions;
- 4. there is an example in natural language that looks very similar to FL: C'Lela

C'Lela (Detteriler, 2000; Pulleyblank, 2002; Archangeli & Pulleyblank, 2007)

- Niger-Congo, ~90,000 speakers.
- Vowel height of suffix agrees with base.
- Direct object 1st person pronoun: **mi/me**

High base	*Nonhigh base*
b**u**zᵊkᵊ **mi** 'chased me'	**ɛ**pkᵊ **me** 'bit me'
s**i**pkᵊ **mi** 'grabbed me'	w**e**g**a**k**a** **me** 'indicated me'
f**u**mtᵊkᵊ **mi** 'pulled me'	b**a**tkᵊ **me** 'released me'

C'Lela

- If more than 1 suffix, only word-final suffix alternates.
- Word –medial suffix is transparent.

High base			
i-z**i**s-**i**	'CM-long-CM'	i-z**i**s-i-n**i**	'CM-long-CM-ADJM'
u-p**u**s-**u**	'CM-white-CM'	u-p**u**s-u-n**i**	'CM-white-CM-ADJM'
Nonhigh base			
i-r**e**k-**e**	'CM-small-CM'	i-r**e**k-i-n**e**	'CM-small-CM-ADJM'
u-gʲ**ɔ**z-**o**	'CM-red-CM'	u-gʲ**ɔ**z-u-n**e**	'CM-red'CM-ADJM'

C'Lela

- A very similar to FL assimilation
- But: possibly within-base harmony
- Base-final suffix harmony.

High base	*Nonhigh base*
b**u**zᵊkᵊ **mi** 'chased me'	**ɛ**pkᵊ **me** 'bit me'
s**i**pkᵊ **mi** 'grabbed me'	w**e**g**a**k**a** **me** 'indicated me'
f**u**mtᵊkᵊ **mi** 'pulled me'	b**a**tkᵊ **me** 'released me'

C'Lela

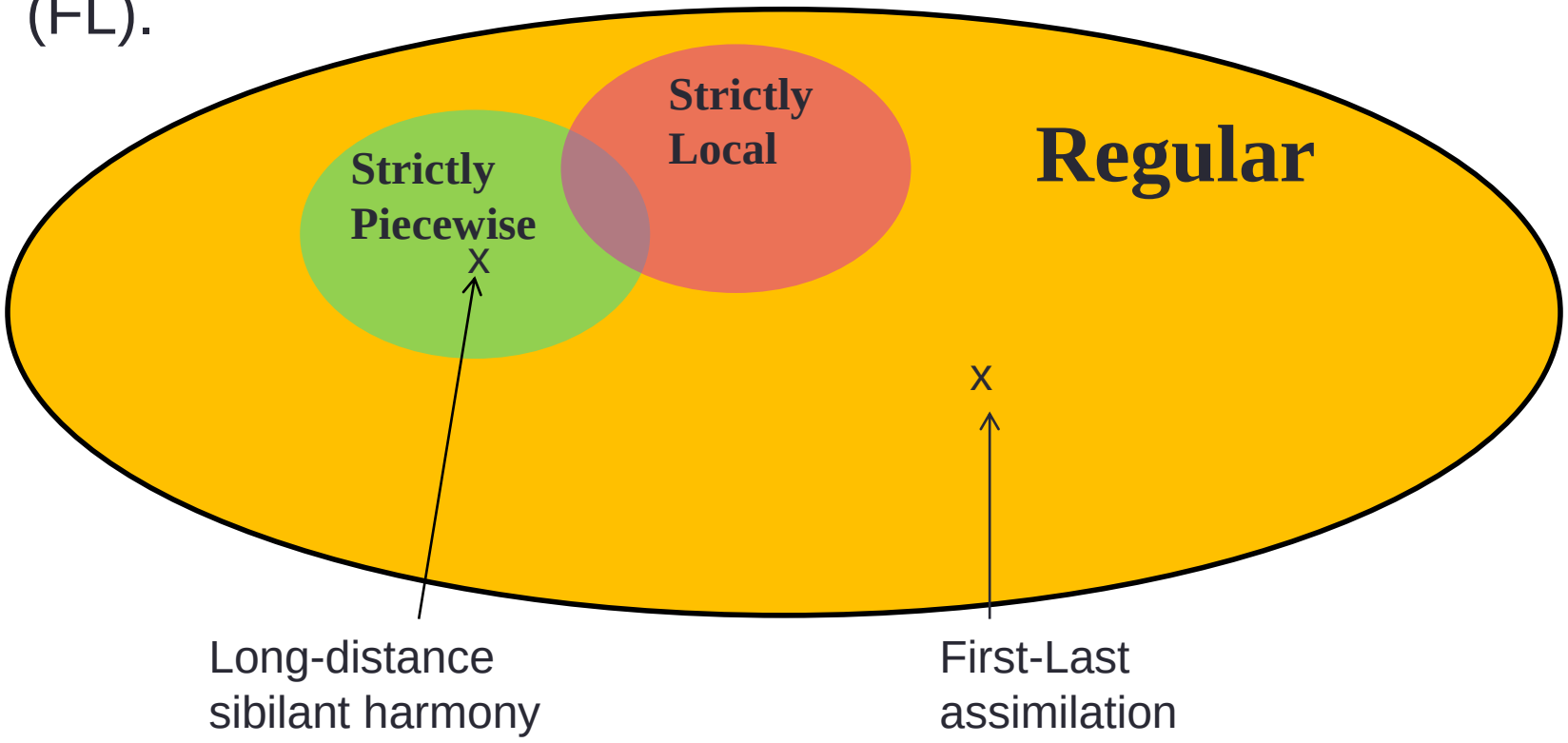
- Prefixes are allowed, but are transparent.
- Target: final position→ not exactly FL

High base			
i-z**i**s-**i**	'CM-long-CM'	i-z**i**s-i-n**i**	'CM-long-CM-ADJM'
u-p**u**s-**u**	'CM-white-CM'	u-p**u**s-u-n**i**	'CM-white-CM-ADJM'
Nonhigh base			
i-r**e**k-**e**	'CM-small-CM'	i-r**e**k-i-n**e**	'CM-small-CM-ADJM'
u-gʲ**ɔ**z-**o**	'CM-red-CM'	u-gʲ**ɔ**z-u-n**e**	'CM-red'CM-ADJM'

THE PRESENT STUDY

Hypothesis

- Humans can only learn sound patterns that belong to the *Strictly Piecewise* or *Strictly Local* classes (SH). They cannot learn other types of regular sound patterns (FL).



General Experimental Methodology

- Artificial Language Learning Paradigm
- Training Phase
- Testing Phase

Methodology

- All Stimuli (both training and test):
- $C_1V.C_2V.C_3VC_4$ (tryisyllabic)
- Always contain 3 sibilants within a word
- C_1 & C_4: always sibilants
- C_2 & C_3: either sibilant or [k]

	C1	C2	C3	C4
50%	sibilant	sibilant	[k]	sibilant
50%	sibilant	[k]	sibilant	sibilant

- Vowels: [a, i, u, ɛ, ɔ]
- Sibilants: [s, ʃ]
- Stop: [k]

Training

- 40 words x 5 repetitions = 200 words
- Procedure: Listen and repeat each word
- ~ 20 min

3 Training Conditions

- 1. SH: [s…s…s], [ʃ…ʃ…ʃ]
- 2. FL: [s…s…s], [ʃ…ʃ…ʃ], [s…ʃ…s], [ʃ…s…ʃ]
- 3. Control: No training

Testing

- Two alternative forced choice
- Words are presented in pairs (minimally different)
- E.g. [sakisis] vs. [ʃakisis]

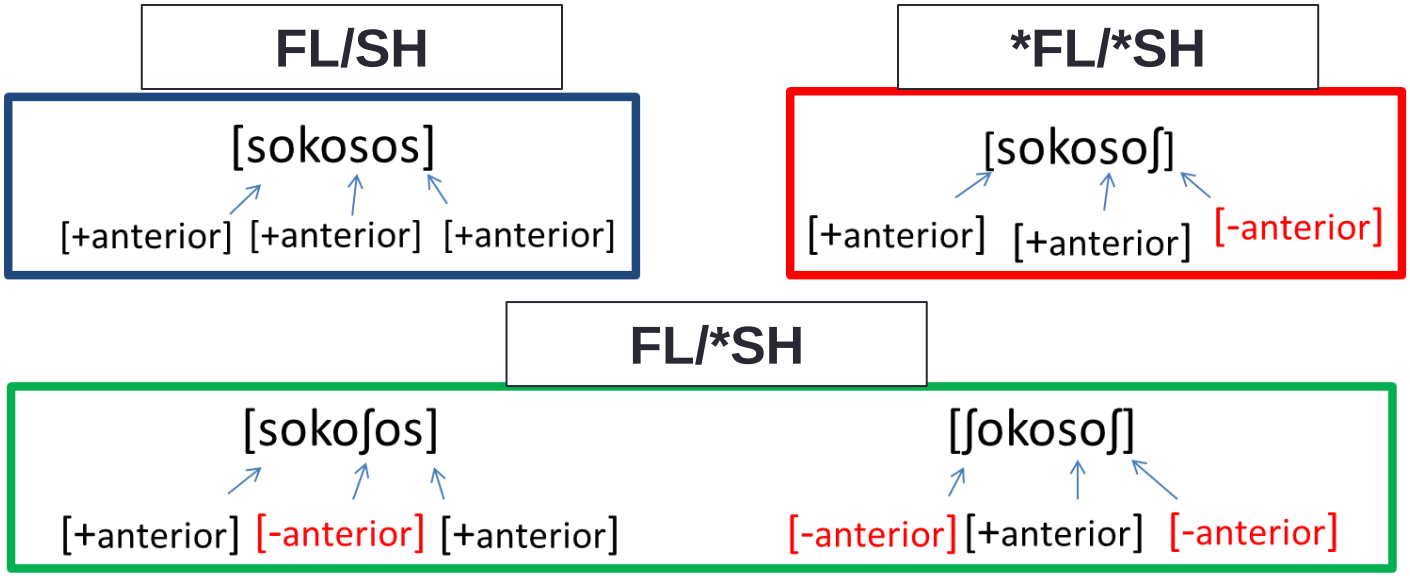
- The different sibilant occurs in either C1, C2, C3 or C4

Testing

- Subjects had to choose a word based on whether they thought the 1st word or the 2nd word within the pair belonged to the language they heard during training.

- 48 pairs in total

Test stimuli



- Note: the logically possible 4th type (*FL/ SH) does not exist because anything that obeys SH also obeys FL.

Test stimuli

- These 3 types of stimuli were pitted against each other and generated 3 types of pairings.

- a) FL/*SH vs. *FL/*SH
- b) FL/SH vs. *FL/*SH
- c) FL/*SH vs. FL/SH

- The order of presentation was counter-balanced across types

Data Analysis

- Dependent variable for each category is different, so they were analyzed separately:

- a) FL/*SH vs. *FL/*SH
  - Rate of choosing FL/*SH

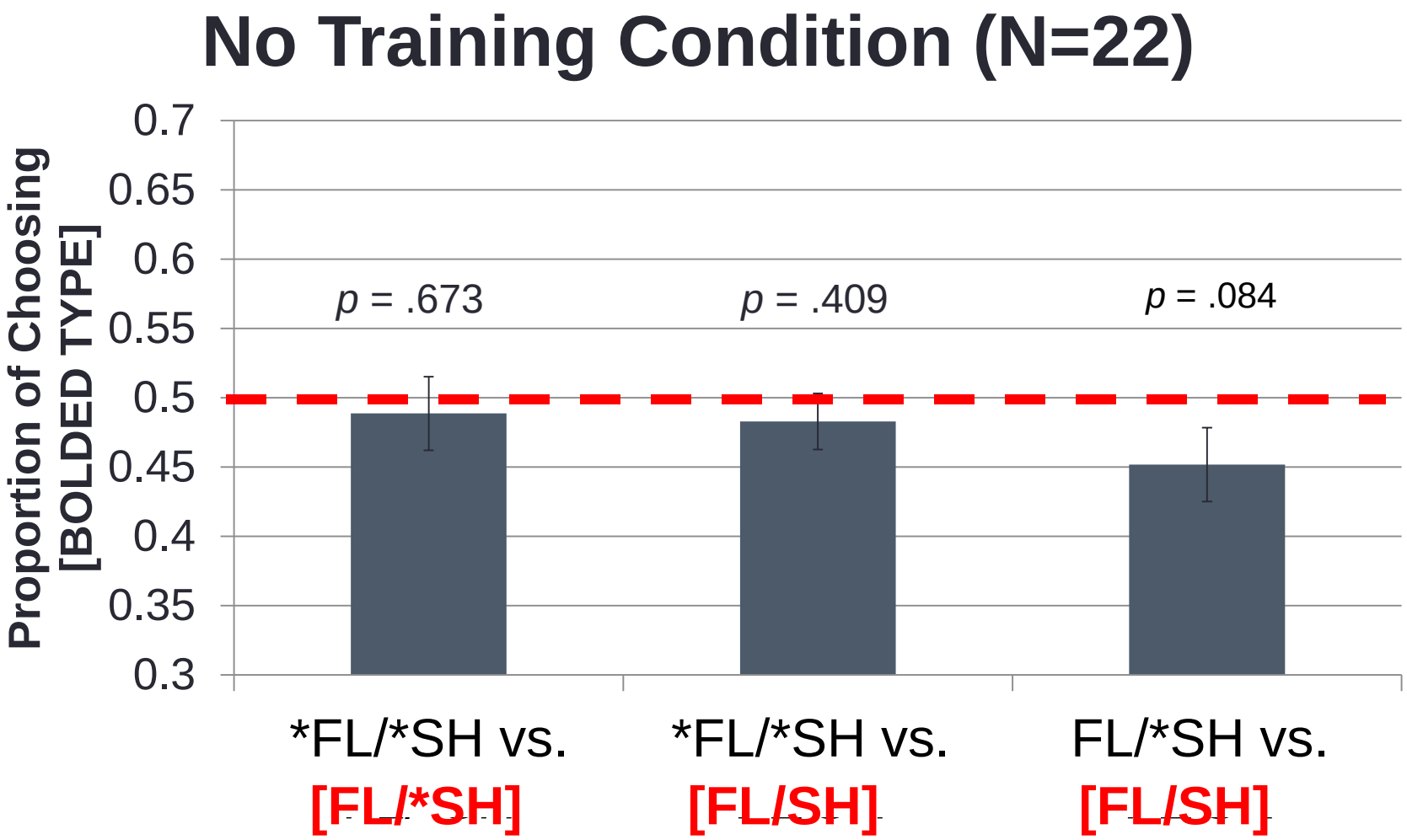
- b) FL/SH vs. *FL/*SH
  - Rate of choosing FL/SH

- c) FL/*SH vs. FL/SH
  - Rate of choosing FL/SH

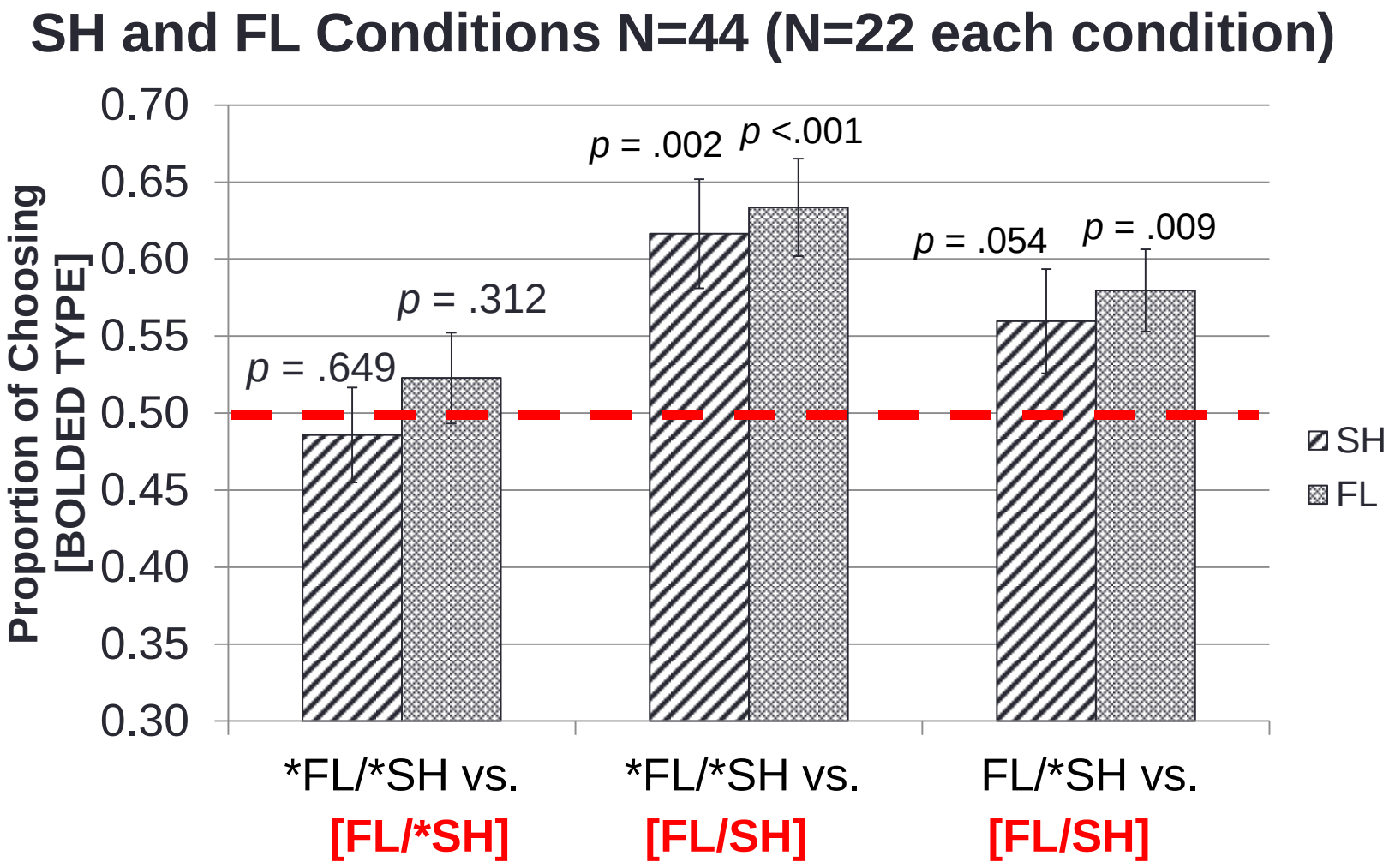
- If subjects learned the grammar that they were exposed to during the training, they should perform as follows:

Training Condition	**FL/*SH vs. *FL/*SH**	**FL/SH vs. *FL/*SH**	**FL/*SH vs. FL/SH**
SH	Chance	Above	Above
FL	Above	Above	Chance
Control	Chance	Chance	Chance

Results


No Training Condition (N=22)
Proportion of Choosing [BOLDED TYPE]
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
p = .673
p = .409
p = .084
*FL/*SH vs. [FL/*SH]
*FL/*SH vs. [FL/SH]
FL/*SH vs. [FL/SH]

Results


SH and FL Conditions N=44 (N=22 each condition)
Proportion of Choosing [BOLDED TYPE]
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
p = .649
p = .312
p = .002
p <.001
p = .054
p = .009
SH
FL
*FL/*SH vs. [FL/*SH]
*FL/*SH vs. [FL/SH]
FL/*SH vs. [FL/SH]

SH results

Types	If SH is learned	Actual SH subjects' performance
a) **FL/*SH** vs. *FL/*SH	Chance	Chance
b) *FL/*SH vs. **FL/SH**	Above	Above
c) FL/*SH vs. **FL/SH**	Above	Above

FL results

Types	If FL is learned	Actual FL subjects' performance
a) **FL/*SH** vs. *FL/*SH	Above	Chance
b) *FL/*SH vs. **FL/SH**	Above	Above
c) FL/*SH vs. **FL/SH**	Chance	Above

Discussion

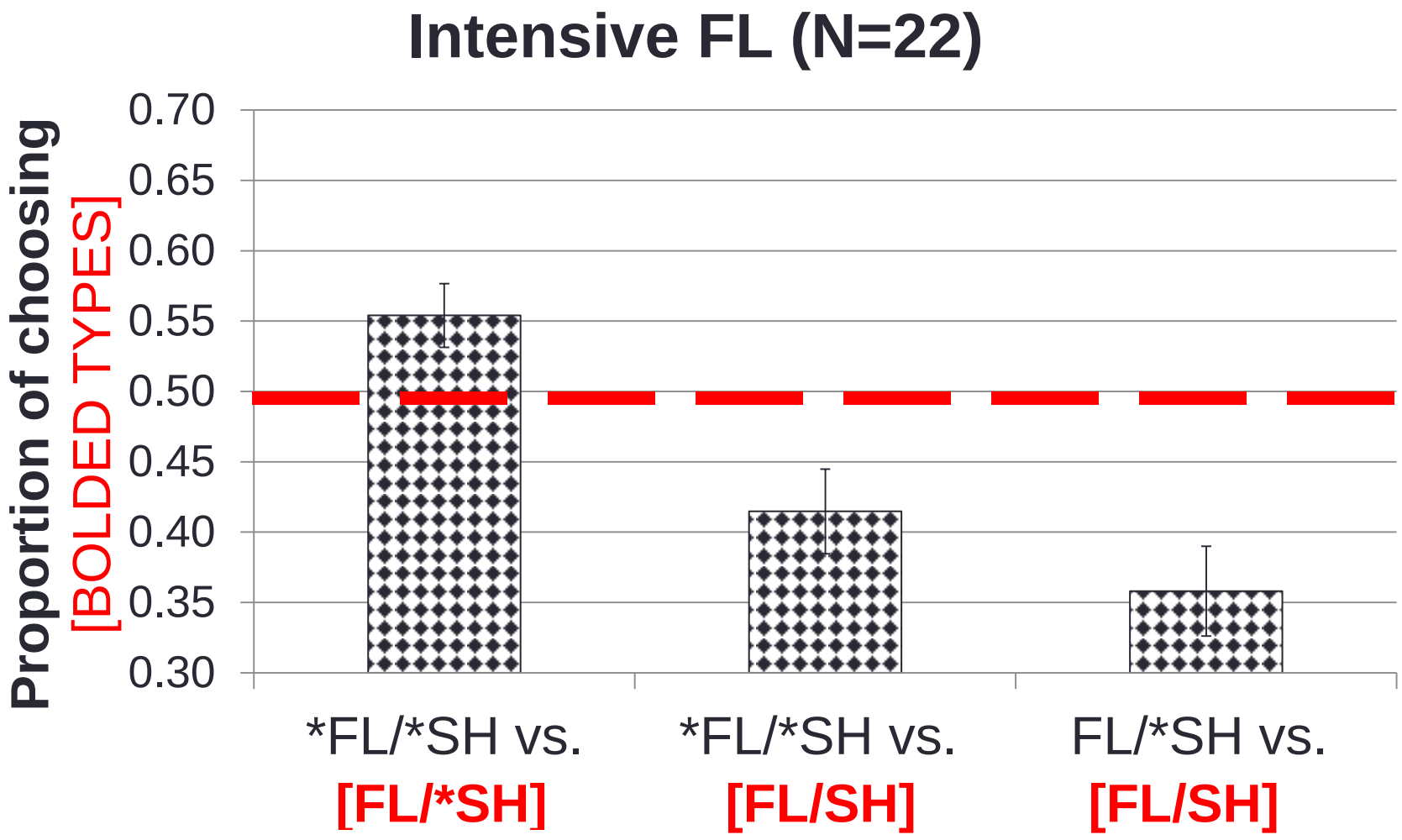
- SH subjects were able to internalize the SH grammar.
- FL subjects were NOT able to internalize FL grammar.
- SH and FL subjects' performance were not significantly different.
- It's puzzling why FL subjects performed so similarly to SH subjects even when they were exposed to stimuli (during training) that did not obey SH (e.g. [s…ʃ…s])

Follow-up condition

- Intensive FL training
- Replaced training stimuli which are consistent with both FL and SH (FL/SH) with ones which are only consistent with FL (FL/*SH).

- The results from Intensive FL were significantly different from FL.

Follow-up results



Discussion

Types	If FL is learned	Actual FL subjects' performance	Actual Intensive FL subjects' performance
a) **FL/*SH** vs. *FL/*SH	Above	Chance	Above
b) *FL/*SH vs. **FL/SH**	Above	Above	Below
c) FL/*SH vs. **FL/SH**	Chance	Above	Below

- Based on these results, we cannot conclude FL is learned in either FL or Intensive FL conditions.

Summary

- The experiments are designed to test the learnability of a regular but not SL or SP pattern (FL).

- If FL is learnable, then it implies the subregular boundaries are not psychologically real.

Summary

- Results indicate that FL cannot be learned in experimental setting with our design.
- Subjects trained with FL performed like SH subjects.
- Subjects were biased towards internalizing SH than FL grammar, even when they were exposed to stimuli that were inconsistent with SH.

Conclusions

- A pattern that belongs to SP group (SH) is learnable in experimental setting, while FL, which is a regular pattern that does not belong to either SP or SL is not learnable.
- The absence of FL pattern in natural phonologies could be due to its unlearnability.
- The current psychological experiment results align with the predictions made by computational theory.
- Support the claim that possible phonological patterns are restricted by certain computational boundaries.

Thank you!

Acknowledgement:
Members of P-lab at UD
Bill Idsardi and Bridget Samuels (UMD)
Sara Finley (U of Rochester)
This project is funded by NSF DDRIG #1123610