

HAYES AND JO ON BALINESE PHONOTACTICS

Jeffrey Heinz



Stony Brook University

LIN 626
Stony Brook University
Apr 30, 2020

Balinese stem phonotactics and the subregularity hypothesis

The argument

- ① Fact1: Copying is not regular.
- ② Fact2: There are phonotactic patterns that are sensitive to copying.
- ③ Conclusion: Some phonotactic patterns are non-regular; the subregularity hypothesis is wrong.

Balinese stem phonotactics and the subregularity hypothesis

Additional points along the way:

- 1 Formal Language Theory/Computational Analysis does not carve phonology at its joints.
- 2 Real questions of phonology lie elsewhere.
- 3 Arguments relying on overgeneration are arguments ‘from silence’; hypotheses placing limits are later overturned because of our small sample size of languages.
- 4 Proper application of FLT ought to employ a principle of the ‘extendable alphabet’.
- 5

Guardian interview with an epidemiologist Carl Bergstrom

(Published 4/28/2020)

G: It requires a lot of discipline to argue really hard for something but also be scrupulously open about all of the weaknesses in your own argument.

CB: But it's more important than ever, right? A really good paper will lay out all the most persuasive evidence it can and then in the conclusion section or the discussion section say, 'OK, here are all the reasons that this could be wrong and here are the weaknesses.'

QUESTIONING THE “FACTS”

- ① ?Fact1: Copying is not regular.
- ② ?Fact2: There are phonotactic patterns that are sensitive to copying.

I will focus on putative Fact 2, but have something to say about Fact 1.

COPYING

- The copy language $\{ww \mid w \in \Sigma^*\}$ is not recognizable with $\text{MSO}(<)$ or a finite-state acceptor (regardless of whether the acceptor is 1way, 2way, deterministic, or non-deterministic).
- The copy function $\forall w \in \Sigma^*, f(w) = ww$, is recognizable with $\text{FO}(\text{succ})$ logical transformation and with a 2way deterministic finite-state transducer.
- In other words, recognizing ww words may be in a sense more complicated than taking w as input and outputting ww .
- Does the logical perspective on the latter arguably make ‘copying’ a primitive of *regular* transformations?

COPYING

- The claim that “Copying is not regular” is based on a view of stringsets, not string transformations.
- A synthesis of copying operations with finite-state/logical/regular transformations is not only possible, it already exists.
- Is there a way to understand Balinese in terms of transformations?

?FACT2: BALINESE PHONOTACTICS ARE SENSITIVE TO COPYING.

The basic argument:

(1)a. “Normal” stems with medial NC

[mp]	<i>dampin</i>	‘side, edge’
[nt]	<i>lontar</i>	‘palm-leaf book’
[ŋg]	<i>pungal</i>	‘to break off’

b. *Pseudoreuplicated stems*

[pd]	<i>dapdap</i>	‘tree species’
[gb]	<i>bugbug</i>	‘pile up’
[ml]	<i>lumlum</i>	‘yellowish-white’

Basically only NT clusters are allowed medially (1a). Other clusters are allowed only in pseudo-reduplicated stems. So evaluating the well-formedness [dapdap] requires checking that the [pd] cluster is in a pseudo-reduplicated word.

NOT A CLEAN DIVISION

(3) Counts for medial clusters in “normal” stems

		Second consonant																	
		p	t	c	k	b	d	j	g	s	h	m	n	ɲ	ŋ	l	r	w	j
First consonant	p															10	1	1	
	t	1			1							1	1				7	3	4
	c															2			
	k		5							5						2	3	1	
	b															2	4		1
	d																3		2
	j																		
	g							1					1			2	5		3
	s	1	16	1	1							1	1				1	3	
	h	1										1							1
	m	56				73										1			1
	n		60		1		84												
	ɲ			17				24											
	ŋ		1		83	1			50	41	1					5		1	
	l					1						1							1
	r		5	2		2	1	1	3	4		4	1					1	
	w																		
j																			

NOT A CLEAN DIVISION

(7) Counts for medial clusters in pseudoreuplicated stems

Second consonant

First consonant

	p	t	c	k	b	d	j	g	s	h	m	n	ɲ	l	r	w	j
p		1	3			4	1	1	2				1	2	1		
t	3		2	1	5	5		1	4				1	1	1		
c																	
k	7	8	8		7	4	1		5					2			
b		1	3	1		1			2					3			
d	1	2	2	1	1			1	2					2			
j																	
g	2	9	5		3	7	5		7					1	2		
s	7	3	1	1	4	7		4							1		
h	1	1	2	1	2	1	2	1	1				1	1			
m	1	3	1			1		2	3					1			
n		3		2	3	1					1			1		1	
ɲ			1														
l	2	2	5	3	10	3	1	6	9				1	1			
r	3	2		1	1	4	1	2	2		1		1				
w			1	3	1	1			3								
j																	

?FACT2: BALINESE PHONOTACTICS ARE SENSITIVE TO COPYING.

A grammo-statistical analysis is conducted to conclude there are different phonotactics governing normal words and pseudo-reduplicated words.

- Standard constraints are used plus REDUP.
- A MaxEnt grammar is trained which assigns weights.
- The constraints are statistically evaluated to see if they significantly contribute with a leave one-out analysis which uses a likelihood ratio test (appears to be implemented with chi-square).

CONSTRAINT WEIGHTS

(13) *A partial phonotactic grammar for Balinese clusters*

<i>Stratal affiliation</i>	<i>Constraints</i>	<i>Weight in Core</i>	<i>Weight in Redup.</i>
Trans-stratal	*PALATAL IN CODA		∞
	*GLIDE IN CODA		∞
	*GEMINATE		∞
Both	CODA COND	4.4	0.4
	*BRANCHING ONSET	3.6	1.8
	SYLLABLE CONTACT LAW	0.5	2.5
Core only	*CODA VOICED OBSTRUENT	1.7	
	AGREE(voice)	∞	
Reduplicated only	REDUP		∞

RATIO LIKELIHOOD

(12) *Weights and statistical testing of individual stratum-specific constraints*

<i>Constraint</i>	<i>Core Stratum</i>			<i>Reduplicated Stratum</i>		
	<i>Weight</i>	$\Delta(\text{LogLk})$	<i>p</i>	<i>Weight</i>	$\Delta(\text{LogLk})$	<i>p</i>
CODACOND	4.4	970.7	$< 10^{-423}$	0.4	1.7	0.06
*BRANCHING ONSET	3.6	647.8	$< 10^{-283}$	1.8	27.6	$< 10^{-12}$
*CODA VOICED OBSTRUENT	1.7	5.2	0.001	0	—	—
AGREE(voice)	∞	14.0	$< 10^{-6}$	0	—	—
SYLLABLE CONTACT LAW	0.5	1.8	0.06	2.5	49.6	$< 10^{-22}$

QUESTIONS ABOUT THE GRAMMO-STATISTICAL ANALYSIS

- 1 Where do the numbers come from?
 - They estimate the dictionary includes ~ 8 words with normal clusters for every pseudo-reduplicated word with a cluster. But the sample their analysis rests on has ~ 380 psuedo-reduplicated words and ~ 720 normal words, just under a 1:2 ratio. How come?
- 2 Is the chi-square test (or ratio likelihood test) appropriate with very small numbers?
- 3 Is the grammo-statistical analysis meaningful if the frequencies of pseudo-reuplicated words are oversampled?
- 4 What happens if the same leave-one-out statistical test is applied to REDUP?

NEXT ARGUMENT

Suppose the grammo-statistical argument is sound.
Is there another way to understand the facts which does not
jeopardize the subregular hypothesis?

ONE POSSIBILITY

Traditional grammatical descriptions explain the pseudo-reduplicated exceptions to forbidden medial CC clusters as follows.

Consider [dapdap]

- Underlyingly, this is actually /dap²/ which is spelled-out as [dapdap]. The ‘phonotactic’ applies at this underlying level, not the surface level.
- This analysis makes use of an abstract underlying representation which triggers a transformation involving copying.

Is this unreasonable? What is the cost/benefit analysis of this vis a vis the one offered by Hayes and Jo?

ANOTHER POSSIBILITY

This may be too much to expect Hayes and Jo to address, but what about categorical constraints that tolerate so many exceptions?

OTHER THOUGHTS?

1

2

3