Approaching templates from below

Eric Raimy
English Language and Linguistics
April 15, 2008
Department of Linguistics, UW-Madison

1.0 Minimalism in phonology

- Chomsky (2007:4)
  "Throughout the modern history of generative grammar, the problem of determining the character of FL (faculty of language) has been approached from the 'top down': How much must be attributed to UG to account for language acquisition? The M(inimalist) P(rogram) seeks to approach the problem from the 'bottom up': How little can be attributed to UG while still accounting for the variety of I-languages attained, relying on third factor principles? The two approaches should, of course, converge, and should interact in the course of pursuing a common goal"

- Third factor principles (Chomsky 2005, 2007)
  –efficient computation
  –structural architecture

- McCarthy (1988:84)
  "Simply put, if the representations are right, then the rules will follow. The entire theory or research program known as nonlinear phonology is based almost entirely on this idea. The second premise is that modular theories are generally more constrained than homogenous ones. For this reason, nonlinear phonological theory is segregated into distinct but interacting subcomponents dealing with stress, syllabification, and segmental phonology."

- Goal of this talk is to derive aspects of reduplication from third factor principles. Specifically:
  –the existence of reduplication
  –the shapes of templates
  –the distribution of patterns

2.0 Prosodic Morphology

- McCarthy and Prince (1986/1996:3)
  "The fundamental goal of a template representation system must be to characterize the shape-invariant that unites the various allomorphs"

- McCarthy and Prince (1993/2001:1)
  "Prosodic Morphology Hypothesis
  Templates are defined in terms of the authentic units of prosody: mora (μ), syllable (σ), foot (Ft), prosodic word (PrWd)."

- Templates are stipulated (even when derived from constraint interaction) as surface targets for 'well-formedness'

3.0 Reduplication in Pangasinan

- Pangasinan is an Austronesian language spoken in the Phillipines
- All data is drawn from Benton (1971)
- Multiple patterns of reduplication with a 'mixed' semantics

(1) Total reduplication-nominal affixation

toó	'man'	tóo-tóo	'figure of a man',
ogáw	'child'	ogáw-ogáw	'figure of a child',
abóng	'house'	abóñg-áboñg	'toy house'
man-bása	'(will) read',	man-bása-bása	'reading anything and everything',
man-pasiár	'(will) go around',	man-pasiár-pasiár	'going around all over the place with no special destination in mind, etc.'

(2) Initial (C)VCV

Nouns Plural

tamuró	'forefinger'	tamu-tamuró
pañgánsi	'ring finger'	pañgá-pañgánsi
lusór	'cup'	lusó-lusór
otót	'mouse, rat'	otó-otót

(3) Initial (C)VC reduplication *Nouns plural* (B 100-101)

báley	'town'	bal-báley
balíta	'news'	bal-balíta
paltóg	'gun'	pal-paltóg
lúpa	'face'	lup-lúpa
áteñg	'parent'	at-áteñg

(4) Infixing(?) CVC reduplication *Numerals of limitation*

sakéy	'one'	sak-sakéy	'one only'
taló	'three'	tal-taló-ra	'three only'
apát	'four'	a-pat-pátíra	'four only'

(5) Infixing(?) CV reduplication *Noun Plurals*

kanáyon	'relative'	ka-kanáyon	'relatives'
dalikán	'clay stove'	da-ralikán	'clay stoves'
amígo	'friend'	a-mi-mígo	'friends'
báso	'glass'	ba-báso	'glasses'

(6) Initial (C)V reduplication *Verbs*

likét	'be happy'	maí-li-likét	'always happy'
ermén	'be sorrowful'	maí-e-ermén	'sentimental'
akís	'cry'	maí-a-akís	'cry babyish'

- Near minimal pairs indicate at least 6 distinct reduplication patterns

(7) Summary of patterns in Pangasinan (still partial)

a.	total	man-pasiár	man-<u>pasiár</u>-pasiár
b.	prefixing foot	otót	<u>otó</u>-otót
c.	prefixing heavy syllable	áteñg	<u>at</u>-áteñg
d.	infixing heavy syllable	apát	a-<u>pat</u>-pát-íra
e.	prefixing light syllable	akís	maí-<u>a</u>-akís
f.	infixing light syllable	amígo	a-<u>mi</u>-mígo

- Aspects of the data we want to accout for:
  –existence of reduplication
  –shape of reduplication
  –locatation of reduplication

4.0 Why (not) reduplication? Data structures for phonology

- Considering what we know about phonology, what would be the best data structure to use
- A phonological representation is a 'list' of segments
- The 'list' contains the information of what segments are part of the representation and what the ordering relationships among the segments are
- The ordering relationships are primitives of the 'list' because:
  –there are no 'anagram languages', e.g. /tak/ = /kat/
  –ordering can not be derived between all segments in all languages, e.g. English /æsk/~/æks/~/sæk/~/kæs/~ etc.
- Two common data structure implementations of lists are *arrays* and *linked lists* (Aho, Hopcroft and Ullman 1985)

4.1 Arrays

- An *array* has the following characteristics:
  –elements are stored in contiguous cells
  –insertion and/or deletion of elements at the end of the list is 'easy'
  –insertion and/or deletion of elements to the beginning or middle of list is 'costly'
  –must prespecify possible 'size' of list and always uses this amount of memory regardless of actual content of list
  –identification of elements in the list by position is easy because the positions are indexed via integers (e.g. simply ask what is in position '4')

(8) Array for 'cat'

a. stored representation

segment	k	æ	t		
index	1	2	3	4	5

b. end manipulation

insertion 'suffixation'

k	æ	t		
1	2	3	4	5

insert

k	æ	t	s	
1	2	3	4	5

deletion

k	æ	t		
1	2	3	4	5

delete

k	æ			
1	2	3	4	5

c. beginning manipulation

insertion 'prefixation'

k	æ	t		
1	2	3	4	5

move and insert

k	æ	t		
k	æ		t	
k		æ	t	
ə	k	æ	t	
1	2	3	4	5

deletion

k	æ	t		
1	2	3	4	5

delete and move

k	æ	t		
	æ	t		
æ		t		
æ	t			
1	2	3	4	5

- ‘Cost’ of manipulation is dependent on how many ‘moves’ are required
  –insertion/deletion is ‘cheap’ at the end because only 1 element is manipulated; element is either inserted or deleted
  –insertion/deletion is ‘costly’ at other points because a variable number of elements will need to be manipulated; ‘space’ must be created where the element is to be deleted or inserted and then other elements must be moved in accordance with the modification

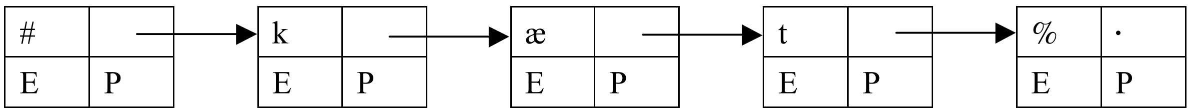
4.2 Linked lists

- A *linked list* has the following characteristics:
  –no need for contiguous storage of elements
  –insertion and deletion of elements is ‘easy’ at all positions
  –‘size’ of list is dependent on elements and links, thus no set size required
  –position is not indexed thus must be calculated
  –data is ‘two part’- element and a pointer to the ‘next’ element

- *Linked list* (for phonology) contain ‘null’ elements (#) and (%) which indicate the beginning and end of the list

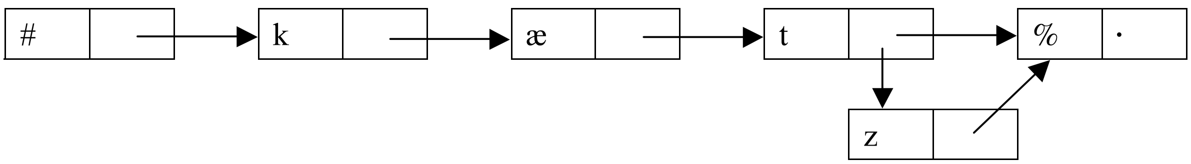
(9) Linked list

a. stored representation

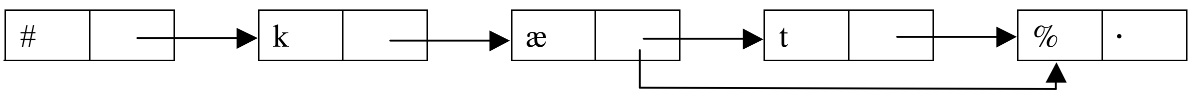


b. end manipulation

‘suffixation’

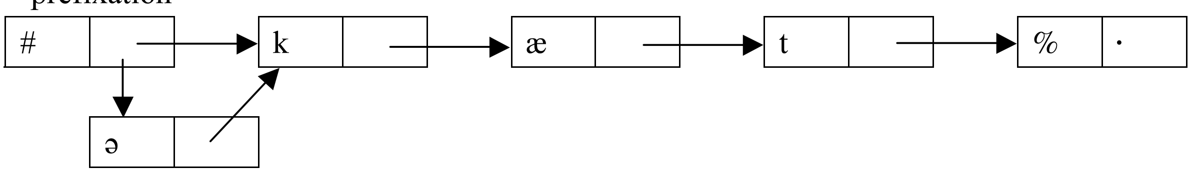


‘deletion’

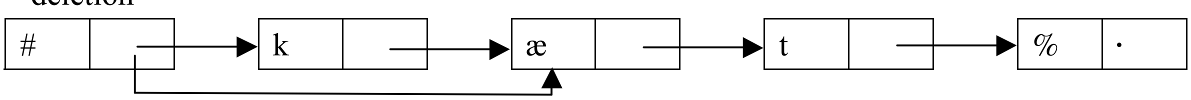


c. beginning manipulation

‘prefixation’



‘deletion’



- Cost of insertion and deletion in a linked list is the same regardless of position –new element and links to and from the element (for insertion) or just a link (for deletion)

4.3 Arrays vs. linked lists

- If phonological representations are 'optimal' for FL, which data structure meets this goal better?

(10) Comparison of array vs. linked list

Phenomenon	Language	Array	Linked List
end manipulation	suffixation	+	+
beginning manipulation	prefixation	-	+
middle manipulation	infixation	-	+
indexing	4th element?	-	+
generativity	longest word?	?	+

- General affixation processes (and phonological epenthesis and deletion) favor *linked list*
- Indexing favors *linked list* because there are no absolute phonological processes and phonology must SEARCH (Mailhot and Reiss 2007)
- Generativity partially favors *linked list* because arrays may or may not be of fixed size depending on computer language
- Conclusion is that computational efficiency favors a *linked list* type of representation

4.4 Linked lists to directed graphs to reduplication

- *Linked list* is technically not able to support the representations proposed in Raimy (1999, 2000, et seq., henceforth Precedence Based Phonology, PBP)
- *Linked list* have a pre-determined number of possible pointers for each element
- PBP requires arbitrary number of possible pointers and the appropriate *abstract data type* for this is a *directed graph*
- A *directed graph* consists of a list of vertices/segments and arcs/precedence relations and this is sufficient to encode our representations

(11) Directed graph

a. graph $\# \rightarrow k \rightarrow æ \rightarrow t \rightarrow \%$

b. list vertices/segments {æ, t, k}

arc/precedence {#k, kæ, æt, t%}

- What types of representations can we encode?

(12) Types of representations

a.	# → X → %	root	no Anchor Points
b.	# → X → {AP}	prefix	1 Anchor Point = {#_}
c.	{AP} → X → %	suffix	1 Anchor Point = {_%}
d.	{AP1} → X → {AP2}	affix	2 Anchor Points
	i.	infix	if AP1 immediately precedes AP2
	ii.	overwrite	if AP1 precedes AP2
	iii.	redup	if AP2 precedes AP1

- How can we prevent representations in (12d)? Stipulation… only allow one Anchor Point per representation
- Wait! We don't want to prevent (12d) because these structures are common in the world's languages
- We have derived reduplication from the computational necessity of dealing with deletion and epenthesis in word initial and medial position
- PBP predicts reduplication, infixation and melodic overwriting

5.0 What to reduplicate?

- Computational efficiency derives the presence of reduplication, can it derive aspects of what patterns of reduplication we find?
- Mailhot and Reiss (2007:37 fn. 9) in their development of a model for long distance dependencies state that:
  "A linear scanning procedure is at least necessary for identifying potential environments of application for phonological rules, or constraint violation locations in OT"
- Fits with phonological representations being *directed graphs* in that environments must be calculated (and are not indexed)

5.1 Finite state aspects of phonological scanning

- Model phonological processes in finite state automata (Idsardi in press) with restrictions on the topology of FSAs

- Heinz (2007) argues that restrictions on possible FSAs are derived from the concept of 'neighborhood distinctness' = main upshot is the observation that 'phonology does not count beyond 2' is derived!
- Anchor Point Theory (Raimy 2005, in press) proposes a parameterized model of possible variables used in affixation (both concatenative and non-concatenative, Raimy 2007)

(13) Parameters for an anchor point

Placement:	{at/before/after}
Edge:	{first/last}
Plane:	{x-tier/metrical/syllable/consonantal}
Target:	{plain/stressed/vowel/consonant}

- Parameters for Anchor Point Theory are derived from independent aspects of autosegmental phonology (McCarthy 1988) and the 'count limitation' derived by Heinz (2007)

(14) Anchor Points for Pangasinan

		AP1	AP2
a.	total	last segment	first segment
b.	prefixing foot	2nd vowel	first segment
c.	prefixing heavy syllable	after 1st vowel	first segment
d.	infixing heavy syllable	2nd consonant	first consonant
e.	prefixing light syllable	first vowel	first segment
f.	infixing light syllable	after 1st consonant	first consonant

(15) Parameter settings

a.	last segment	{at, last, x-tier, plain}
b.	first segment	{at, first, x-tier, plain}
c.	first vowel	{at, first, metrical, plain}
d.	first consonant	{at, first, consonantal, plain}
e.	after first vowel	{after, first, metrical, consonant}
f.	after first consonant	{at, first, consonantal, vowel}
g.	2nd vowel	{after, first, metrical, plain}
h.	2nd consonant	{after, first, consonantal, plain}

- All searches are ‘local’ in that an edge tropic element is found on a particular tier and then there is only a local decision
- ‘Long distance’ searches are actually ‘local’ too in that they are just parameterized to a particular plain which will ‘ignore’ classes of segment

(16) Representations for reduplication in Pangasinan

a. total reduplcation man-pasiár >> man-pasiár-pasiár

root *phonological*

→ p → a → s → i → a → r → % # → p → a → s → i → a → r → %

b. prefixing foot otót >> otó-otót

→ o → t → o → t → % # → o → t → o → t → %

c. prefixing heavy syllable áteñg at-áteñg

→ a → t → e → ŋ → % # → a → t → e → ŋ → %

d. infixing heavy syllable apát a-pat-pát-íra

→ a → p → a → t → % # → a → p → a → t → %

e. prefixing light syllable akís maí-a-akís

→ a → k → i → s → % # → a → k → i → s → %

f. infixing light syllable amígo a-mi-mígo

→ a → m → i → g → o → % # → a → m → i → g → o → %

- See Idsardi and Raimy (2008) for discussion of linearization, for our purposes here, the segments ‘within the loop’ will be repeated once

6.0 What not to reduplicate and why

- Common question/misunderstanding: the reduplication patterns in PBP are not prosody based therefore it predicts that non-prosodic reduplication can occur
- There are two parts to this objection:
  –aberrant reduplication patterns
  –prosodic basis of reduplication

6.1 Organic restrictions on reduplication patterns

- McCarthy and Prince (1986/1996:5) state that
  "The fact that the templates are bounded by a language's prosody follows from their being literally built from that prosody"
- Import of this is that reduplicative templates will always conform to the general prosody of a particular language
- Prosodic Morphology/OT stipulates this observation- must restrict possible reranking of constraints via fiat
- PBP derives this observation from the general grammatical architecture!
- Chomsky and Halle (1968) argues that morphology derivationally precedes phonology thus syllabification (prosody) will always follow reduplication (Carrier 1979)

(17) Grammatical architecture of PF (Idsardi and Raimy 2008)

Module		***Characteristics***
Narrow syntax		hierarchy, no linear order, no phonological content
LINEARIZATION =	**Immobilization**	
Morphosyntax		hierarchy, adjacency, no phonological content
LINEARIZATION =	**Spell-out**	
Morphophonology		no hierarchy, directed graph, phonological content
LINEARIZATION =	**Serialization**	
Phonology		no hierarchy, linear order, phonological string

- McCarthy and Prince's original observation is epiphenomenal and thus should be derived not stipulated

6.2 The (non)prosodic basis of reduplication

- Most common reduplication pattern is total reduplication which is not prosodically defined (no specific number of syllables or feet, etc.)
- Most reduplication patterns do not follow prosody of the base but instead follow surface prosody of the language
- No shortage of reduplication patterns that are not defined in ‘authentic units of prosody’

(18) Non-prosodic reduplication patterns

Language	Reduplicated Forms	Surface Pattern	Source
Temiar- Continuative	kɔ̄w > **kw**-kɔ̄w slɔg > s-**g**-lɔg	CC, C	Benjamin (1976)
Nxa’amxcin- Out of Control	k’ip’ > k’í-**p’**-əp’ ptix̣^{w}-mix > p-**t**-tíx̣^{w}əx^{w}	C	Czaykowska-Higgins and Willet (1997)
Sawai- Reciprocal	ŋamu > fa-**m**-ŋamu duk > fa-**k**-duk	C	Whisler (1992)
Temiar- Simulfactive	kɔ̄w > **k**-a-kɔ̄w slɔg > s-a-lɔg	C, Ø	Benjamin (1976)
Arrernte- Frequentive	unt-em > unt-ep-**unt**-em an’entel̪il-em > an’entel̪il-ep-**il**-em	VCC, VC	Breen and Pensalfini (1999)
Mangarrayi- Plural	jimgan > j-img-imgan yirag > y-ir-irag	VCC, VC	Merlan (1982)
Kannada- Echo Reduplication	nannu [baagil-annu much-id-e] giigilannu muchide anta heeLa-beeDa I-NOM door-ACC close-PST-1S RED that say-PROH ‘Don’t say that I closed the door or did related activities’ verb + object + tense + person/number Lidz (2004)		

- PBP predicts the presence of these non-prosodic patterns while Prosodic Morphology must add additional explanations to account for them

6.3 What was that pattern anyway?

- PBP offers the possibility of multiple surface analyses of reduplication patterns
- Fitzpatrick and Nevins (2004) argue for the utility of this approach in dialectical variation in Tigrinya (Rose 2003)
- Yanti and Raimy (2007) argue for ambiguity of analysis in Tanjung Raden Malay to explain origin of innovative reduplication patterns

(19) Ambiguity of reduplication in Pangasinan

a. Total reduplication or foot reduplication-nominal affixation

toó	'man'	tóo-tóo	'figure of a man',
man-bása	'(will) read',	man-bása-bása	'reading anything and everything',
	{last segment} →	{first segment}	
	{second vowel} →	{first segment}	

b. Initial (C)VC reduplication or Infixing reduplication

balíta	'news'	bal-balíta	
lúpa	'face'	lup-lúpa	
	{after first vowel} →	{first segment}	
	{second consonant} →	{first consonant}	

c. Initial (C)V reduplication or Infixing CV reduplication

kanáyon	'relative'	ka-kanáyon	'relatives'
báso	'glass'	ba-báso	'glasses'
	{first vowel} →	{first segment}	
	{after first consonant} →	{first consonant}	

- Ambiguity of analysis when combined with a selectionist type learner (Yang 2002) produces a model which predicts likely avenues of diachronic change
- Reduplication SYSTEMS can also be studied as an entity because the patterns will all be interrelated (or not)

7.0 Conclusion and future investigations

- Following the minimalist program in phonology can be very fruitful
- Third factor principles of efficient computation and structural architecture derive most if not all aspects of Prosodic Morphology wrt reduplication and infixation

- PBP provides an opportunity to pursue Marr's 3 levels of explanation
  "… one of the keystones of this book is the realization that we have had to be more careful about what constitutes an explanation than has been necessary in other recent scientific developments, like those in molecular biology. For the subject of vision, there *is* no single equation or view that explains everything. Each problem has to be addressed from several points of view–as a problem in representing information, as a computation capable of deriving that representation, and as a problem in the architecture of a computer capable of carrying out both things quickly and reliably." Marr (1982:5)

(20) Marr's three levels of understanding (Marr 1982:25)

Computational Theory	**Representation and Algorithm**	**Hardware Implementation**
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and the output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?
Precedence Based Phonology	*Acquisition modeling* Iba and Nevins (2004), Chinn and Raimy (2006) *Linearization modeling* Iba and Nevins (2004), McClory and Raimy (2007), Idsardi and Shorey (2007)	*Cognitive neuroscience… not even close…*

- Expand this program to general phonology and sociophonetics via the 'hypermodularity' proposal (Purnell, Raimy and Salmons in prep)
- Pursue modular and structural analyses of phenomena to identify efficient computation of the problem at hand

References

Aho, A., Hopcroft, J. and Ullman, J. 1985. *Data structures and algorithms*. Addison-Wesley Publishing Company.

Benton, R. 1971. *Pangasinan reference grammar*. University of Hawaii Press.

Carrier, J. 1979. The interaction of morphological and phonological rules in Tagalog. PhD. Dissertation. MIT.

Chinn, K. and Raimy, E. 2006. A revised and constrained selectionist learner for reduplication. Poster presented at the 80th Annual Meeting of the Linguistics Society of America, Albuquerque, NM, January 5th.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36:1-22.

Chomsky, N. 2007. Approaching UG from below. In *Interfaces + recursion = language?* U. Saurlan and H. M. Gartner (eds), pp. 1-29. Mouton de Gruyter.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Fitzpatrick, J. and Nevins, A. 2004. Linearizing nested and overlapping precedence in multiple reduplication. In *The Proceedings of the 27th Penn Linguistics Colloquium*, S. Arunachalam and T. Scheffler (eds), pp. 75-88.

Heinz, J. 2007. On the role of locality in learning stress patterns. Ms. University of Delaware, Newark, DE.

Iba, A. and Nevins, A. 2004. A selectionist learner for parametrically ambiguous reduplicants. Presented at the 78th Annual Meeting of the Linguistics Society of America.

Idsardi, W. in press. Calculating metrical structure. In *Contemporary views on representations and architecture in phonological theory*, E. Raimy and C. Cairns (eds). Cambridge, MA: MIT Press.

Idsardi, W. and Raimy, E. 2008. Three types of linearization and the temporal aspects of speech. GLOW 31, Newcastle University.

Idsardi, W. and Shorey, R. 2007. Unwinding morphology. Talk at CUNY Phonology Symposium on Precedence Relations http://www.cunyphonologyforum.net/index.php

Mailhot, F. and Reiss, C. 2007. Computing long-distance dependencies in vowel harmony. *Biolinguistics* 1:28-48.

Marantz, A. 1982. Re reduplication. *Linguistic Inquiry* 13:435-482.

Marr, D. 1982. *Vision*. W. H. Freeman and Company.

McCarthy, J. 1988. Feature geometry and dependency: a review. *Phonetica* 45:84-108.

McCarthy, J. and A. Prince. 1986/1996. *Prosodic Morphology 1986*. Ms.

McCarthy, J. and A. Prince. 1993/2001. *Prosodic morphology: Constraint interaction and satisfaction*. Ms. ROA 482.

McClory, D and Raimy, E. 2007. Enhanced edges: Morphological influence on linearization. Poster presented at the 81st Annual Meeting of the Linguistic Society of America. Anaheim, CA.

Purnell, T., Raimy, E. and Salmons, J. in prep. Hypermodularity. Ms. Unversity of Wisconsin.

Raimy, E. 1999. Representing Reduplication. PhD Dissertation, University of Delaware.

Raimy, E. 2000. *The Phonology and Morphology of Reduplication*. New York: Mouton de Gruyter.

Raimy, E. 2003. Asymmetry and linearization in phonology. In *Proceedings from the Asymmetry Conference, vol. 2*, A. M. Di Sciullo (ed), 129-146. John Benjamins.

Raimy, E. 2005. Prosodic Residue in an A-templatic World. Paper presented at Linguistics Colloquium Series, University of Delaware.

Raimy, E. 2007. Precedence theory, root and template morphology, priming effects and the structure of the lexicon. Talk at CUNY Phonology Symposium on Precedence Relations http://www.cunyphonologyforum.net/index.php

Raimy, E. in press. Deriving reduplicative templates. In *Contemporary views on representations and architecture in phonological theory*, E. Raimy and C. Cairns (eds). Cambridge, MA: MIT Press.

Rose, S. 2003. Triple take: Tigre and the case of internal reduplication. *San Diego Linguistic Papers* 1:109-128.

Yang, C. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yanti and Raimy, E. 2007. Reduplication in Tanjung Raden Malay. Ms.