MALAY VERBAL REDUPLICATION WITH THE *məN-* PREFIX*

Meng Yang & Deborah J.M. Wong
University of California, Los Angeles
mengyang@ucla.edu, deborah.jm.wong@ucla.edu

Abstract
Malay is known for having both full reduplication (e.g. *buŋa* 'flower', *buŋa-buŋa* 'flowers') and nasal substitution when a nasal is followed by a voiceless stop. These two processes interact in verbal reduplication, where the *məN-* prefix attaches either to the first or second component of the reduplicated verb. When *məN-* attaches to the first component, the copy that doesn't bear the prefix surfaces with a nasal homorganic to the underlying voiceless stop (e.g. *mənari+nari* 'dance (continuous)'). When the *məN-* attaches to the second component, the copy that doesn't bear the prefix surfaces faithfully with the underlying voiceless stop (e.g. *tari+mənari* 'dance (reciprocal)'). This pattern is difficult to account for under any derivational analysis. We propose that this pattern of reduplication is best accounted for in Parallel OT. Our OT analysis includes a new markedness constraint $*N_{[word\text{-}initial]}$, which reflects the phonotactics of Austronesian languages of Southeast Asia.

Keywords: Malay, reduplication, *məŋ*-prefix

ISO 639-3 codes: may/msa

1. Introduction

This paper looks at the interaction between two common processes found in Malay: reduplication and nasal substitution. Reduplication is a highly productive process in Malay and enables the formation of new words from root words. According to an early study of reduplication in Malay by Asmah (1975), the process can be classified into six categories based on differences in the morphological changes of the word. The six categories are as follows:

(1) a. Whole-word reduplication
b. Rhyming and chiming reduplication
c. Partial reduplication
d. Reduplication with 'em' infixation
e. Reduplication with root that does not stand alone
f. Reduplication with affixation (prefix, suffix, circumfix) process

The reduplicative pattern we are interested in is (1f), reduplication with affixation, specifically the affixation of *məN-*. In this type of reduplication, the reduplicative process can be seen interacting with nasal substitution at the morpheme boundary between the stem and the prefix when the stem begins with a voiceless stop. When the verb is reduplicated, *məN-* can attach either to the first component of the

* We would like to thank Kie Zuraw and the audiences of the UCLA Phonology Seminar, the Austronesian Formal Linguistics Association (AFLA), and the International Symposium on Malay/Indonesian Linguistics (ISMIL) for their valuable feedback and suggestions. Many thanks to our consultants, Leila Louise Fitton, Sarah Amalina Ab Wahid, and Yasmin Rizal for the data and judgments. All remaining errors are our own.

reduplicated verb, giving it a continuous meaning, or to the second component of the reduplicated verb, giving it a reciprocal meaning. We show that the pattern arising from the interaction between the processes, morphological and phonological, cannot be accounted for under any derivational account. Instead, we propose that this interaction is best accounted for in Parallel Optimality Theory (OT), which takes as its input a morphologically complete form and evaluates the overall phonological well-formedness of candidates over a set of ranked constraints (Prince & Smolensky 1993).

The paper is organized as follows: In Section 2, we present the data and describe the phenomenon. In Section 3, we show the problem that the data poses for derivational accounts, as well as for McCarthy & Prince's (1995) Parallel OT analysis. Drawing from Malay phonotactics, we then propose a revised OT analysis in Section 4. Finally, we provide a conclusion for the paper.

2. The Data

2.1 Full reduplication and məN- prefixation in Malay

In full reduplication, the entire word, whether simplex or complex, undergoes reduplication. In Malay, this process occurs with different classes of words and serves many different functions as shown in the examples below.

(2)

Word	*Gloss*	*Reduplicated word*	*Gloss*
buŋa	'flower'	*buŋa-buŋa*	'flowers'
oraŋ	'person'	*oraŋ-oraŋ*	'scarecrow'
lambat	'slow'	*lambat-lambat*	'slowly'
apa	'what'	*apa-apa*	'anything'

The full reduplication of Malay nouns such as *buŋa* 'flower' usually indicates plurality, though there are exceptions like *oraŋ* 'person', which, when reduplicated, results in a new lexical item *oraŋ-oraŋ* 'scarecrow'. Reduplicating an adjective usually results in a change of category. For example, *lambat* 'slow' reduplicates to become the adverb *lambat-lambat* 'slowly'. Reduplicating a *wh*-word such as *apa* 'what' results in the indefinite *apa-apa* 'anything'.

In Malay, the voicing of the stem-initial stop usually determines the phonological outcome of *məN-* prefixation. Given that Malay does not permit NÇ (nasal/voiceless stop) clusters or non-homorganic NC (nasal/voiced stop) clusters, these illicit sequences formed through the prefixation process are repaired by nasal substitution and nasal assimilation respectively (e.g. Hassan 1974, Farid 1980). Nasal substitution refers to the process whereby the stem-initial voiceless stop coalesces with the nasal in the prefix to produce a nasal that has the same place of articulation as the stop (e.g. Pater 2001)[1]. On the other hand, nasal assimilation refers to the process in which the nasal in the prefix assimilates in place to the stem-initial stop but does not replace it. Examples of these two processes are given in (3).

(3) a. Nasal substitution when stem begins with a voiceless stop

/mə**N**+**p**ukul/	mə**m**ukul	'hit'
/mə**N**+**t**ari/	mə**n**ari	'dance'

[1] In this paper, we assume Pater's (2001) analysis of nasal substitution, in which the final nasal of the prefix and the initial voiceless stop of the stem coalesce into (i.e. are substituted by) a single nasal segment that is homorganic to the stop. However, we remain agnostic between this analysis and others, for example those in which the final nasal of the prefix assimilates to the place of the initial stop, then the initial stop deletes (e.g. Onn 1973)

/mə**N**+**k**əd͡ʒar/ mə**ŋ**əd͡ʒar ‘chase’

b. Nasal assimilation when stem begins with a voiced stop

/mə**N**+bunʊh/	me**m**bunʊh	‘kill’
/mə**N**+duga/	mə**n**duga	‘suspect’
/mə**N**+ganti/	mə**ŋ**ganti	‘change’

Examples in (3a) show the phonological output of *məN-* prefixation when the stem begins with a voiceless stop. In these cases, the final nasal of the *məN-* prefix and the initial stop are both replaced by a nasal that is homorganic to the stop. For example, when *məN-* attaches to the stem *pukul*, we get a bilabial nasal, /m/, at the morpheme boundary. In contrast, examples in (3b) show the output of *məN-* prefixation when the stop at the beginning of the stem begins with a voiced stop. For these words, the final nasal in *məN-* takes the place feature of the following stem-initial stop, and the stop is still pronounced. For example, when *məN-* attaches to the stem *bunʊh*, we get the sequence /mb/ at the morpheme boundary.

2.2 Verbal reduplication with məN- prefix

Under verbal reduplication, the *məN-* prefix can attach to either the first or the second component of the reduplicated word. When *məN-* appears on the first component, the reduplicated word carries a ‘continuity’ meaning; when *məN-* appears on the second component, the reduplicated word carries a reciprocal meaning.

When the stem begins with a voiced stop, the copies in the reduplicated form with *məN-* prefixation are identical since the addition of the prefix does not change the form of the stem. Examples are given below in (4).

(4) Verbal reduplication when stem begins with voiced stop

məN+X-X	X-*məN*+X	Stem	Stem gloss
məmbunʊh-bunʊh	*bunʊh-məmbunʊh*	*bunʊh*	‘kill’
mənduga-duga	*duga-mənduga*	*duga*	‘suspect’
məŋganti-ganti	*ganti-meŋganti*	*ganti*	‘change’

From the examples in (4), we can see that regardless of whether *məN-* attaches to the first or second component of the reduplicated verb, both components are pronounced the same. However, when the root word begins with a voiceless stop, the copies in the reduplicated form with *məN-* prefixation are no longer identical, as shown in (5).

(5) Verbal reduplication when stem begins with voiceless stop

məN+X-X	X-*məN*+X	Stem	Stem gloss
məmukul-mukul	*pukul-məmukul*	*pukul*	*‘hit’*
mənari-nari	*tari-mənari*	*tari*	*‘dance’*
məŋəd͡ʒar-ŋəd͡ʒar	*kəd͡ʒar-məŋəd͡ʒar*	*kəd͡ʒar*	*‘chase’*

When *məN-* attaches to the first component, nasal substitution occurs on the first component. Additionally, the initial segment on the second component also surfaces as the homorganic nasal (eg. *mənari-nari* instead of **mənari-tari*), even though the context for nasal substitution is not met, thus constituting a case of overapplication. When *məN-* attaches to the second component, nasal substitution applies normally, only occurring on the second component, while the first component continues to have

as its initial consonant a voiceless stop (eg. *tari-mənari* instead of **nari-mənari*). In the next section, we discuss the difficulties in accounting for this pattern of nasal substitution.

3. Other Accounts of Reduplication with *məN*- Prefixation

3.1 Derivational accounts

In any derivational account of phonology, the underlying representation along with a set of phonological processes contain all the information necessary to generate the appropriate surface form. Our case involves two morphological processes: full reduplication and *məN*- prefixation. The full reduplication of the verb involves copying the phonological material of the root verb in its entirety. Affixing the *məN*- prefix to a stop-initial verb stem creates an illicit sequence at the morpheme boundary, which must be repaired by the phonological process of either nasal assimilation or nasal substitution.

In the case of verbs with roots beginning with a voiced stop, applying the two processes in either order yields the same surface form since the form of the stem does not change. Examples of both orders of process application on the verb *bunʊh* 'kill' are given in (6).

(6) a. Reduplication > *məN*- prefixation

Process	'continuous' *məN*+X-X	reciprocal X-*məN*+X
reduplication	*bunʊh-bunʊh*	*bunʊh-bunʊh*
məN- prefixation	*məmbunʊh-bunʊh*	*bunʊh-məmbunʊh*

b. *məN* prefixation > Reduplication

Process	'continuous' *məN*+X-X	reciprocal X-*məN*+X
məN- prefixation	*məmbunʊh*	*məmbunʊh*
reduplication	*məmbunʊh-bunʊh*	*bunʊh-məmbunʊh*

In (6a), reduplication is shown applying before *məN*- prefixation. The reduplication process produces a form with two copies of the verb. The *məN*- prefix is then free to attach to either copy with no phonological consequence to the stem. In (6b), the order of reduplication and *məN*- prefixation is reversed, with the same outcome. The *məN*- prefix first attaches to the verb stem, and the nasal assimilates to the place of the following stop, but the stop remains unchanged. The phonological material from the unchanged verb stem is then copied in the reduplication process. However, when the root verb begins with a voiceless stop, neither order of the processes will derive the correct result simultaneously for both the continuity meaning, where *məN*- affixes to the first component, and the reciprocal meaning, where *məN*- affixes to the second component.

(7) a. Reduplication > *məN*- prefixation

Process	'continuous' *məN*+X-X	reciprocal X-*məN*+X
reduplication	*tari-tari*	*tari-tari*
məN- prefixation	***mənari-tari***	*tari-mənari*

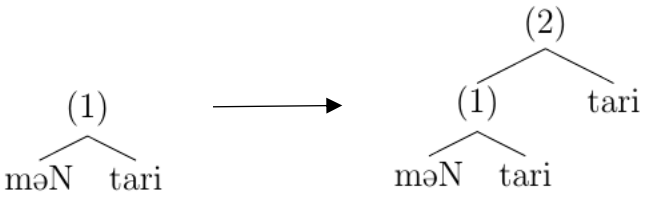
b. *məN*- prefixation > Reduplication

Process	'continuous' *məN*+X-X	reciprocal X-*məN*+X
məN- prefixation	*mənari*	*mənari*
reduplication	*mənari-nari*	***nari-mənari***

If reduplication occurs before *məN-* prefixation, as in (7a), then the voiceless stop would be copied in the reduplicant. Thus when *məN-* is affixed, only the component to which it is attached will undergo nasal substitution while the other non-prefixed component will always surface with the underlying voiceless stop. The result is that the correct form is derived for the word with reciprocal meaning in which *məN-* attaches to the second component, but the incorrect form is derived for the word with the continuous meaning in which *məN-* attaches to the first component. If *məN-* prefixation occurs before reduplication, as in (7b), the initial voiceless stop on the stem would be substituted by a homorganic nasal. When the stem is then reduplicated, both copies will begin with a nasal, thus deriving the correct form for the word in which *məN-* attaches to the first component, but not for the word in which *məN-* attaches to the second component. Clearly, a derivational account like this cannot account for the forms of reduplicated verbs with the continuous and reciprocal meanings.

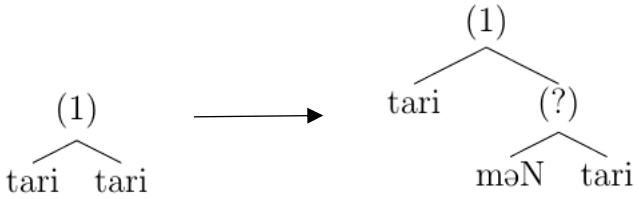
Given that *məN-* prefixation on the first or second component of a reduplicated verb results in semantic differences, we might also entertain the possibility that the order of the two morphological processes could be different for the continuous meaning and the reciprocal meanings. For the continuous meaning, prefixation could precede reduplication, yielding the correct surface form, e.g. *mənari-nari*. From a morphosyntactic perspective, this order is unproblematic if we assume a left-branching structure. The stem and prefix merge first, producing *mənari*, then the root verb is copied on the right, as in (8).

(8) Morphosyntactic structure for *mənari-nari* 'dance (continuous)'



For the reciprocal meaning, reduplication must precede prefixation to yield the correct surface form, e.g. *tari-mənari*. However, since in this case *məN-* affixes onto the second component of the reduplicated verb, the later morphological process of prefixation would require us to merge the prefix between two structures that have been built, as in (9).

(9) Morphosyntactic structure for *tari-mənari* 'dance (reciprocal)'



To avoid the insertion of new structure between the two copies of the reduplicated verb, we would have to stipulate that *məN-* merges to the left of *tari-tari*, then the second moves to the right of *mənari* to derive the correct surface order. Movement in syntax is typically motivated by the need to check certain features. In this case, it is difficult to argue that the movement of the reduplicant is motivated by anything other than the need to derive the correct surface order.

In this section we examined two different derivational analyses. In the first analysis, we assumed that the order of application for the two morphological processes, reduplication and *məN-* prefixation, was the same for whether the prefix attaches to the first or the second component of the reduplicated verb. This type of analysis was able to derive the correct surface form for one of the two types of verbs but not both. In the second analysis, we allowed for each type of verb to have a different order of application for the two processes. However, we would either have to violate basic syntactic assumptions, or stipulate unnecessary movement. Thus, neither derivational accounts adequately capture the phenomenon described. Therefore, following McCarthy and Prince (1995), we adopt a parallel Optimal Theoretic approach in which all the necessary morphology is completed first, and sound patterns are dealt with later in the phonology. We first give their analysis of the nasal substitution patterns for reduplicated verbs beginning with voiceless stops, its successes, and potential issues. Then, in Section 4, we present our revised OT analysis.

3.2 Parallel OT

McCarthy and Prince (1995) uses a simple parallel analysis that account for these over/underapplication problems in reduplication. In Optimality Theory (Prince & Smolensky 1993), the phonological output is determined by the interactions between faithfulness constraints comparing the input and output strings and markedness constraints on surface configurations. Since reduplication involves copying phonological material from the base to the reduplicant, McCarthy and Prince's analysis also employs base-reduplicant faithfulness constraints that militate against differences between the two copies of reduplication. The three types of cover constraints used in their analysis are defined in (10).

(10) Constraints from McCarthy and Prince's (1995) analysis

a. **NASALSUB:** A cover constraint for finer constraints that result in the substitution of the final nasal in *məN-* and the initial voiceless stop in the stem by a nasal homorganic to the stop.[2] One violation is incurred for every NÇ sequence in a candidate.

b. **FAITH-IO:** The cover constraint for all input-output faithfulness constraints. One violation is incurred for every mismatch between a segment in the input and a segment in the base of the output.

c. **FAITH-BR:** The cover constraint for all base-reduplicant faithfulness constraints. One violation is incurred for every mismatch between a segment in the base of a candidate and a reduplicant of the same candidate.

In the tableaux below, we present an adapted version of McCarthy & Prince's OT analysis of the verbal reduplication phenomenon with these constraints. The base and reduplicant in the candidate set are marked by subscript B and R respectively.

(11) a.

/məN+tari-RED/	NASALSUB	FAITH-IO	FAITH-BR
a. məntari$_B$-tari$_R$	*!		
☞ b. mənari$_B$-nari$_R$		*	
c. mənari$_B$-tari$_R$		*	*!

[2] We will not discuss in detail the constraints involved in nasal substitution since there are multiple possible analyses, and the specifics are not crucial to the current proposal. For one possible analysis of nasal substitution in *məN*-prefixation, we refer you to Pater (2001).

b.

/tari-məN+RED/	NASALSUB	FAITH-IO	FAITH-BR
a. $tari_B$-$məntari_R$	*!		
b. $nari_B$-$mənari_R$		*!	
☞ c. $tari_B$-$mənari_R$			*

In both tableaux in (11), we can see that **NASALSUB** has to outrank **FAITH-IO** to ensure that the candidate in which nasal substitution occurs wins over the faithful candidate. In (11a) when the prefix attaches to the first component, i.e. the base, candidates (b) and (c) incur the same number of violations of **NASALSUB** and **FAITH-IO**, but candidate (c) also incurs a violation of **FAITH-BR**, and is therefore harmonically bound by candidate (b). In (11b) when the prefix attaches to the second component, i.e. the reduplicant, candidate (c) no longer violates **FAITH-IO** since the base is faithful to the input. Thus **FAITH-IO** must outrank **FAITH-BR**, allowing candidate (c) to win over candidate (b).

McCarthy and Prince's analysis is successful in capturing both patterns of the verbal reduplication with a single ranking of constraints. However, their analysis crucially relies on specifying which component is the base and which is the reduplicant. In this case of reduplication, the first component must be the base and the second component the reduplicant. Flipping the base and reduplicant in the case of the reciprocal meaning results in the wrong output being chosen, as shown in the tableau in (12).

(12)

/RED-məN+tari/	NASALSUB	FAITH-IO	FAITH-BR
a. $tari_R$-$məntari_B$	*!		
💣 b. $nari_R$-$mənari_B$		*	
☹ c. $tari_R$-$mənari_B$		*	*!

We argue that this assumption about base and reduplicant is arbitrary and does not reflect speakers' knowledge about base and reduplicant identity. We asked three native speakers of Malay for their judgments on base/reduplicant identity in various types of reduplication. In some cases, these native speakers were able to determine which part of the reduplicated word is the base and which is the reduplicant, while in other cases they were not. For example, in the case of rhythmic reduplication, e.g. *kwɛ-**mwɛ*** 'pastries', our speakers identified the second component, *mwɛ*, as the reduplicant as it does not stand as a word on its own. Similarly, in the case of ***bolak**-balık* 'back and forth', *bolak* is identified as the reduplicant because it does not exist as a stand-alone lexical item. Note here that order does not determine base- or reduplicanthood since there are words in which the reduplicant is the first component and words in which the reduplicant is the second component. Contrastively, in simple full reduplication, e.g. *buŋa-buŋa* 'flowers', speakers were unable to identify one copy as the base and the other as the reduplicant since the two copies are identical.

In the case of the verbal reduplication with *məN-* prefixation, speakers are able to distinguish the base and reduplicant when *məN-* is attached to the first component, but not when it is attached to the second component. For *mənari-**nari*** 'dance (cont.)', *nari* is not a valid word, hence is the reduplicant. For *tari-mənari* 'dance (recip.)', both *tari* and *mənari* are valid words, speakers were hesitant to identify either component as the reduplicant. The uncertainty speakers had in determining the base vs. reduplicant in the latter type of verbal reduplication indicates Malay speakers do not necessarily need to know which copy

is the base and which is the reduplicant in order to produce the correct phonological output. Yet it is precisely in this case, where *məN-* attaches to the second component, that the identity of the base is crucial for McCarthy and Prince's analysis presented above. Thus, in the next section, we present an alternative OT analysis that does not require the potentially stipulative designation of the base and reduplicant, thereby more accurately reflecting speakers' intuition about these reduplicated forms.

4. Our Analysis

We argue that the verbal reduplication pattern seen here is best accounted for using a revised parallel OT analysis. This analysis does not require us to assume that the first component in the reduplicated verb is the base. Instead, we introduce a new markedness constraint that draws on a phonotactic trait of Malay which is also evident in other Austronesian languages.

4.1 Drawing upon Malay Phonotactics

Surveying the Austronesian languages of Southeast Asia, Blust (2013) finds that initial nasals are rare. For example, root-initial nasals are disfavored in Tagalog (Zuraw 2010). Malay displays a similar dispreference for word initial nasals. In Table 1, we present Malay nasal counts by place of articulation (bilabial, alveolar, and velar) and percentages within each place type based on word-type frequencies[3] using data from the An Crúbadán Malay Corpus (Scannell 2007). Red cells indicate overrepresentation and blue cells indicate underrepresentation, χ^2 (4, N=48759) = 12687.074, $p < 0.00001$. For full results of the chi-square analysis, refer to the Appendix.

Table 1: Counts of nasals by place of articulation and percentages by position in word. Data from the An Crúbadán Malay Corpus

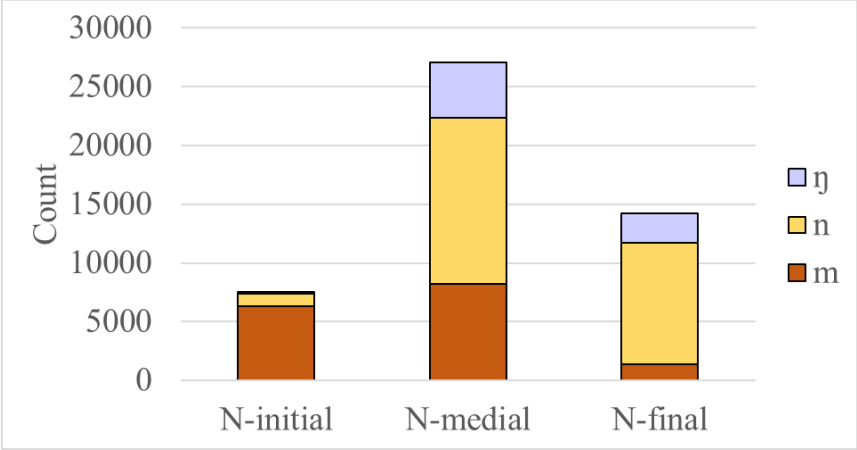
	Initial (%)	**Medial (%)**	**Final (%)**	**Total Count**
m	40%	52%	9%	15,890
n	4%	55%	40%	25,564
ŋ	2%	64%	34%	7,305

As shown in Table 1, word-initial nasals are largely dispreferred with the exception of the bilabial /m/[4]. This dispreference for word-initial nasals is further highlighted in the graph in (13), which displays nasal counts by position.

(13) Nasal counts by position in word.

[3] We chose to use word type frequency rather than token frequency because previous research (e.g. Hayes & Wilson 2008) finds that generalizations drawn from type frequencies better maps onto the phonotactics of a language.

[4] Several instances of 'morpheme-initial m-' in Austronesian languages are prefixes that have fossilized (Blust 2013).



Clearly, the number word-initial nasals is far fewer than nasals in other positions. We draw on this phonotactic tendency in our revised analysis of verbal reduplication with *məN-* prefixation.

Recall that the problematic case for McCarthy and Prince's OT analysis was *tari-mənari* in which *məN-* attaches to the second component of the reduplicated verb. For this type of verbal reduplication, it was necessary to assume that the first component is the base and the second component is the reduplicant. In our revised analysis, we get rid of this arbitrary assumption. Instead, we propose that the first component surfaces with an initial stop because word-initial nasals as phonotactically dispreferred, and this dispreference is stronger than the preference for the base and reduplicant to be identical. However, this phonotactic constraint should not affect words in which the initial segment is a nasal in the input (e.g. *məN-*initial words), only those in which the word-initial nasal is derived. To capture this generalization, we keep the **NASALSUB** cover constraint, which drives nasal substitution in the context of a nasal followed by a voiceless stop. In addition, we introduce a markedness constraint that militates against word-initial nasals. Finally, we replaced the cover input-output and base-reduplicant faithfulness constraints with more specific constraints that militate against changes to nasals. These constraints are defined in (14).

(14) New and revised constraints

a. ***N$_{[word\text{-}initial]}$:** No nasals word-initially. Assign one violation for every word-initial nasal segment.

b. **PRESERVE(NASAL)-IO**: Nasals in the input must be nasals in the output. Assign one violation for every nasal segment in the input that does not correspond to a nasal segment in the output.

c. **IDENT(NASAL)-BR**: Corresponding segments in the base and reduplicant must have the same [nasal] value. Assign one violation for every segment in the base that does not have the same [nasal] value in the corresponding segment in the reduplicant.

The markedness constraint in (14a) captures the phonotactic dispreference for nasals to occur in word-initial position. The modified input-output faithfulness constraint, **PRESERVE(NASAL)-IO**, protects nasals that were in the input from being denasalized in the output. The final base-reduplicant faithfulness constraint militates against differences in the nasality of corresponding segments between the two copies of the reduplicated verb. Note that though we continue to use the term 'base-reduplicant', the analysis we present here does not require us to index identify the reduplicated components as the base or reduplicant. The ranking of these constraints are as follows:

(15) **NASALSUB, PRESERVE(NASAL)-IO >> *N$_{[word\text{-}initial]}$ >> IDENT(NASAL)-BR**

In the following tableaux, we show that these constraints with this ranking are able to account for the pattern of verbal reduplication whether the base is the first component or the second component. The example in (16) shows the analysis for cases in which *məN-* attaches to the first component.

(16) *məN-* attaches to the first component

a. Base as first component, reduplicant as second component

/məN+tari-RED/	NASALSUB	PRES(n)-IO	*N[word-initial]	ID(n)-BR
a. məntari-tari	*!		*	
☞ b. mənari-nari			*	
c. mənari-tari			*	*!
d. bənari-nari		*!		

b. Reduplicant as the first component and the base as the second component

/məN+RED-tari/	NASALSUB	PRES(n)-IO	*N[word-initial]	ID(n)-BR
a. məntari-tari	*!		*	
☞ b. mənari-nari			*	
c. mənari-tari			*	*!
d. bənari-nari		*!		

NASALSUB must dominate **IDENT(NASAL)-BR** to allow for nasal substitution where the context is met, thus ruling out candidate (a). **PRESERVE(NASAL)-IO** must out-rank ***N[word-initial]** so that input nasals are protected, thus ruling out candidate (d). Comparing candidates (b) and (c), both incur a violation of ***N[word-initial]** because of the word-initial [m], however candidate (c) incurs an extra violation of because of the mismatch in nasality between the initial segment in the first and second components of the reduplicated verb. That is, the first component begins with a nasal while the second component begins with a stop.

When *məN-* attaches to the second component of the reduplicant, the same grammar produces the correct winning candidate, as seen in (17).

(17) *məN-* attaches to the second component

/tari-məN+tari/	NASALSUB	PRES(n)-IO	*N[word-initial]	ID(n)-BR
a. tari-məntari	*!			
b. nari-mənari			*!	
☞ c. tari-mənari				*

Here we see the crucial ranking between ***N[word-initial]** and **IDENT(NASAL)-BR**. Candidate (b) is ruled out by the higher ranked ***N[word-initial]** since it has a word-initial nasal, leaving candidate (c) as the optimal

candidate even though there is a mismatch between the base and reduplicant incurring a violation of **IDENT(NASAL)-BR**.

If these patterns are indeed driven by the markedness constraint ***N**$_{[\text{word-initial}]}$, then Malay verbal reduplication with *məN-* prefixation presents a case of the Emergence of the Unmarked (TETU). That is, initial nasals are typically allowed to surface (i.e. are unmarked) in Malay, such as when the word begins with a nasal underlyingly. However, nasals derived from other processes are not allowed when occurring word-initially. TETU effects are commonly found in reduplicative morphology (e.g. McCarthy & Prince, 1994), loanword phonology (e.g. Gouskova 2001), and second language phonology (e.g. Broselow et al. 1998).

We believe that this analysis is better suited to the patterns of reduplication presented here than McCarthy and Prince's (1995) original analysis as it i) does away with ungrounded assumptions about base-reduplicant identity, and ii) is grounded in the phonotactics of the language.

A caveat to all phonological analyses of this phenomenon is, of course, whether these patterns are productive or lexicalized, since the latter would obviate the need for any phonological machinery at all. A score against complete lexicalization would be evidence from a text corpus of Malay that there is variation in the production of existing forms. We could also test for productivity by wug-testing Malay speakers to see if and how they generalize these patterns to novel verbs. Since such corpus and experimental studies are not within the scope of this paper, we will not discuss them further here.

5. Conclusion

To summarize, Malay verbal reduplication with *məN-* prefixation presents a unique case of overapplication as well as normal application. This pattern is shown to be difficult to account in any derivational analysis. A previous parallel OT analysis relies on the assumption that the base is the first component. However, this view is not supported by speaker intuitions about these forms. We build upon the previous OT analysis by introducing a new markedness constraint, ***N**$_{[\text{word-initial}]}$, which draws from phonotactic tendencies in Malay and related languages. This new grammar does not require learners to know which is the base and which is the reduplicant, therefore allowing for a richer base. Our analysis calls for the re-examination of other reduplication processes in which the relative position of the base and reduplicant is assumed.

References

Asmah Hj. O. (1975). Reduplication. In *Essay on Malaysian Linguistics* (pp. 185 – 223). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Blust, R. (2013) *The Austronesian languages*. Asia-Pacific Linguistics.

Broselow, E., Chen, S. I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in second language acquisition*, *20*(2), 261-280.

Farid, O. (1980). *Aspect of Malay Phonology and Morphology: A Generative Approach*. Bangi: National University of Malaysia.

Gouskova, M. (2001). Falling sonority onsets, loanwords, and Syllable Contact. *CLS*, *37*(1), 175-185.

Hassan, A. (1974). *The Morphology of Malay*. Kuala Lumpur: Institute of Language and Literature.

Hayes, B. & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39, 379–440.

McCarthy, J.J. & Prince, A. S. (1995). Faithfulness and Reduplicative Identity. In J. Beckman, L. Walsh-Dickey & S. Urbanczyk (eds.). *University of Massachusetts Occasional Papers in Linguistics: Papers in Optimality Theory*, 249–384. Amherst: GLSA.

McCarthy, J. J., & Prince, A. S. (1994). *The emergence of the unmarked: Optimality in prosodic morphology*.

Onn, F. M. (1980). *Aspects of Malay Phonology and Morphology*. Bangi: Universiti Kebangsaan Malaysia.

Pater, J. (2001). Austronesian nasal substitution revisited. In Linda Lombardi (ed.) *Segmental phonology in Optimality Theory: Constraints and Representations*. Cambridge University Press. 159-182.

Prince, A. & Smolensky, P. (1993) *Optimality Theory*. (ms.) Rutgers University and the University of Colorado at Boulder.

Scannell, K. P. (2007). The Crúbadán Project: Corpus building for under-resourced languages. In *Building and Exploring Web Corpora: Proceedings of the 3rd Web as Corpus Workshop* (Vol. 4, pp. 5-15).

Stangroom J. (2018) Chi-Square Calculator. Available from: http://www.socscistatistics.com/Default.aspx

Zuraw, K (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2): 417-472.

Appendix: Chi-square test for nasal counts by position in word and place of articulation.

Notation in each cell: *count* (*expected total*) [*chi-square statistic*]

Results						
	Initial	Medial	Final			***Row Totals***
Bilabial	6310 (2446.45) [6101.53]	8200 (8814.96) [42.90]	1380 (4628.60) [2280.04]			15890
Alveolar	1077 (3935.87) [2076.57]	14146 (14181.60) [0.09]	10341 (7446.53) [1125.08]			25564
Velar	120 (1124.69) [897.49]	4703 (4052.44) [104.44]	2482 (2127.87) [58.94]			7305
Column Totals	7507	27049	14203			**48759 (Grand Total)**

χ^2 (4, N=48759) = 12687.074, $p < 0.00001$

Calculated using the Chi-Square Test Calculator (Stangroom 2018).