

# A UNIFIED ANALYSIS OF SOME VIETNAMESE REDUPLICATION FORMS

Sonny X. Vu

*Massachusetts Institute of Technology*

<sonny@sonny.vu>

## 1 Introduction<sup>1</sup>

In Vietnamese, one of the main forms of derivational morphology is reduplication. Although reduplicative forms in Vietnamese have been extensively cataloged (Hoàng Văn Hành<sup>2</sup> (1997)) and at least partially described (Emeneau (1951), Thompson (1965), Ngô Thanh Nhân (1984), Hoàng Văn Hành (1985), Nguyễn Tài Cẩn (1996), Nguyễn Kim Thân (1997), et seq.), they have been subjected to limited analysis (Ngô Thanh Nhân (1984), Agbayani (1997)) and not examined for general principles. A crucial distinction that often has been overlooked in previous studies is that only a few of the patterns identified as reduplications are productive. In this study, I describe, characterize, and present an OT analysis of four major *productive* patterns. I show that rerankings of a core set of well-attested constraints provide a natural account for the data that I present.<sup>3</sup> Most importantly, I show that a number of seemingly disparate properties in these four processes follow from differences in the prosodic structures (Selkirk (1978) et seq.) of the output forms.

The four processes I describe<sup>4</sup> and analyze are: Full Reduplication (FR), reduplication involving Emergence of the Unmarked effects (TETU) (McCarthy and Prince (1994b)), and two processes involving Melodic Overwriting (Alderete, et al. (1997))

---

<sup>1</sup> This paper was supported in part by the Vietnamese Advanced Summer Institute 1997 Program. Special thanks to Nguyễn Thị Minh-Phuong for clarifying judgments on all the data in this paper, Cheryl Zoll for guiding this study, and my classmates Calixto Agüero-Bautista and Philippe Schlenker for extensive discussions. Thanks to the participants in the 1997 Workshop in Phonology and Morphology at MIT, the audiences at the 14<sup>th</sup> Northwest Linguistics Conference and at the 8<sup>th</sup> Southeast Asian Linguistics Conference, Brian Agbayani, Mark Alves, Cao Xuân Hào, ñinh Lê Thu, Ken Hale, Morris Halle, Michael Kenstowicz, Alec Marantz, Ngô Như Bình, Rob Pensalfini, Phạm Hoa, Shinko Tamura, and Joseph Thai for useful comments and discussions. Usual disclaimers apply.

<sup>2</sup> For the sake of clarity, all Vietnamese names mentioned in this paper will be written in the traditional format and orthography: last, middle, first name. All Vietnamese data will also be presented in the traditional Vietnamese orthography for ease of reference.

<sup>3</sup> Hence, Vietnamese presents a case of conflicting rankings of constraints within a language.

<sup>4</sup> In the data presented, subscripts denote tones (see below for numbering scheme) and superscripts denote stress levels on the following syllable. For example:

<sup>0</sup>nhà <sup>1</sup>toán <sup>0</sup>hoc = ‘mathematician’

bears the stress pattern 010. The base (BASE) is underlined when it can be determined. Other abbreviations: emph = emphatic; pej = pejorative; att = attenuative; int = intensive; RED = reduplicant. Some of the data in this paper were taken from a number of sources such as Hoàng Văn Hành (1997)’s *Từ Điển Từ Láy* (a dictionary of reduplicative forms produced by the National Linguistics Institute of Vietnam) or from Ngô Thanh Nhân (1984). All data in this paper are either provided or confirmed by native speaker consultants including, Cao Xuân Hào, Nguyễn Thị Minh-Phuong, Phạm Hoa, and myself.

---

Mark Alves, Paul Sidwell & David Gil, eds. *SEALSVIII: papers from the 8th meeting of the Southeast Asian Linguistics Society* (1998). Canberra, Pacific Linguistics, 2007, pp.165-191.

© Sonny X. Vu

which I will refer to as a-reduplication (a-Red) and iec-reduplication (iec-Red). A summary of some of the differences is provided in Table 1.

<b>FR:</b>	sạch → <sup>0</sup> sạch <sup>1</sup> sạch = ‘clean’ → ‘clean’(+attenuative, informal) <sup>5</sup>
<b>TETU:</b>	đẹp → <sup>0</sup> đềm <sup>1</sup> đẹp = ‘pretty’ → ‘pretty’ (+attenuative, formal)
<b>a-Red:</b>	<sup>1</sup> lò <sup>1</sup> mò → <sup>0</sup> lò <sup>2</sup> mà <sup>0</sup> lò <sup>1</sup> mò = ‘to grope’ → ‘to grope’(+emphatic)
<b>iec-Red:</b>	bán → <sup>1</sup> bán <sup>1</sup> biệc = ‘to sell’ → ‘to sell’ (+pejorative)

**Table 1:** *Differences between four productive reduplication processes*

Type	Number of syllables in input (BASE)	Stress pattern of output	Syntactic separability of constituents	Avoidance of identical syllables
FR	1	01	no	no
TETU	1	01	no	yes
a-Red	2	0201	no	yes
iec-Red	1 or 2	11, 0101	yes, yes	yes

In Table 1, ‘syntactic separability’ refers to the observation that some of the reduplicative forms in Vietnamese can have syntactic material inserted between BASE and RED if they can be determined. The last column in Table 1 refers to the fact that some processes do not allow the copying of identical material while others do. A shared feature of the four processes is the fact that the number of syllables in the base is equal to the number of syllables in the RED. Thus, for example, the output of a-Red always has four syllables whereas the output for FR is always disyllabic since a-Red always takes a disyllabic base and FR a monosyllabic base. Another shared feature is that all four processes preserve the grammatical category of BASE.

In addition to exploring the differences between these four processes, I also examine the common features shared among them. These include: complete preservation of the base in the input, lack of double reduplication (the application of a reduplicative process on a reduplicated form), and respect of regular tone-coda co-occurrence restrictions that appear elsewhere in the language.

One of the main empirical observations made in this study is the distinction between productive and non-productive reduplication patterns. Although both types obey some of the constraints I posit, only the productive patterns are assumed to be produced by the grammar, whereas the non-productive cases (the majority of reduplicative forms in Vietnamese) are not. Although this distinction has been alluded to in the Vietnamese linguistics literature (e.g., Ngô Thanh Nhân (1984)), many of the descriptions of Vietnamese reduplication (e.g., Nguyễn Tài Cẩn (1996), Nguyễn Kim Thản (1997)) assume that all reduplicative forms are formed in the grammar by various rules; however, a cursory review of the data shows that these rules over-generate the possible forms in the language.

<sup>5</sup> All final obstruents in Vietnamese are unreleased. I do not indicate this in the data for typographical simplicity.

## 2 Vietnamese Phonology and Theoretical Assumptions

### 2.1 Relevant Vietnamese Phonology

**Tones:** In most Northern dialects, the dialects of focus in this study, Vietnamese has six phonologically contrastive tones. Every syllable is associated with one of these tones. Although I will use the same notation as Burton (1992)'s, I depart from his categorization by viewing the tone pairs {1, 4}, {2, 5}, and {3, 6} to be related not by contour shape (since they are clearly not similar in shape, e.g., tone 2 is rising whereas tone 5 is dipping) but rather, in terms of their 'bundles' of phonetic features.<sup>6</sup> See Wannamacher (1997) for a related discussion on voice quality as a potential tonal feature. Tones 1 and 4 are both lax in terms of glottal stricture and medium in terms of length (for more discussion of length as a tonal parameter, see Alves (1997b)). Tones 2 and 5 have tense glottal stricture, are glottalized at the end, and are short. Tones 3 and 6 are also both tense but are long, in fact, longer than all the other tones. Perhaps one way to characterize tonal markedness in Vietnamese is to say that tones that are lax are unmarked and the ones that are not lax are more marked. For a detailed phonetic description of Vietnamese tones, see Han (1969). See Edmondson and Lôi (1997) for a detailed phonetic study of the tones in Northern Vietnamese. The main motivations for the coupling of tones as shown in Table 2 are: their synchronic phonological correlations (e.g., their related distributions in reduplication as described in this paper, see Võ Xuân Hạo (1997) for a detailed discussion) and their tonogenetic history.<sup>7</sup>

**Table 2:** Vietnamese tone registers and contours

		Contour <sup>8</sup>		
		l	h	lh
Register	+High	1	2	3
	-High	4	5	6

There is no phonological tone sandhi except in a few domains of the grammar, e.g., reduplication. This characterization is motivated not only in terms of their phonological 'well-behavedness' but also by a number of tonogenetic accounts such as Haudricourt's hypothesis (Haudricourt 1954) of tonal splitting due to the presence of voiceless obstruent codas lowering the pitch of tones in early Vietnamese.

<sup>6</sup> Rolf Noyer has mentioned in personal communication that perhaps the characterization based on the contour and register parameters can be maintained. The actual contour shape of tone 5 might then just be a phonetic reflex of the tone being in the lower register and having the glottalization at the end.

<sup>7</sup> Historically, Vietnamese developed phonetically conditioned allophonic features in different syllable types, A (open syllables), B (syllables with final fricatives), and C (syllables with final stops p/t/c/k/?). As the coda consonants in Vietnamese changed (fricatives and glottal stops, which are well attested in other Mon-Khmer languages, were lost in Vietnamese), those phonetic categories became phonemic (i.e., unpredictable based on environment). The next step was the conditioning of height based on the voicing of the initial. Tonal categories A, B, and C each developed two allophones, with voiceless onsets resulting in higher tones and voiced, lower tones. Eventually, Vietnamese initials changed, thereby masking the original voicing, but leaving the tonal height, which became phonologically contrastive. Thus, 3 tonal phonemes became 6. (Haudricourt (1954), Alves (1997b))

<sup>8</sup> Or better put, grouping with respect to voice quality.

In terms of tonal features, I adopt Agbayani's correlation of [ $\alpha$  vocal fold] features with tone contours on vowels and voice distinctions on consonants:

[stiff]  $\Rightarrow$  C[-voice] or V[h] and [slack]  $\Rightarrow$  C[+voice] or V[l]

**Vowels:** Vietnamese has nine vowels. Although vowels obligatorily lengthen in open syllables, two of the vowels /ɤ/ and /a/ are contrastive in terms of length in other environments as well. The place features of the vowels are the most relevant to the present discussion. For a detailed phonetic study of Vietnamese vowels, see Trần (1967) and Han (1966).

Vowel:	[i]	[e]	[ɛ]	[u]	[o]	[ɔ]	[ʉ]	[ɤ]	[a]
Place:	+cor	+cor	+cor	+lab +dor	+lab +dor	+lab +dor	+dor	+dor	+rad

**Tone-coda co-occurrence restriction:** Syllables with final obstruents, i.e., [p, t, k], can only be assigned tones with contour value h, i.e., tones 2 and 5. Otherwise, any of the six tones can be assigned to any other syllable type. For typographic reasons, diphthongs and triphthongs are represented here the two and three vowels, respectively.

**Other:** Vietnamese lacks consonant clusters and has only a limited number of possible codas. The possible syllable types are: C(G)V(:)C and C(G)V: where G is a glide. Note that C(G)V is not allowed, i.e., syllables with only an onset (without or without a glide) and a short vowel. For typographic reasons, diphthongs and triphthongs are represented here the two and three vowels, respectively.

Although any consonant from the phonetic stock can be an onset, the only possible codas are the voiceless unreleased obstruents [p, t, k], nasals [n, m, ŋ] and glides [w, j]. There are also no co-occurrence restrictions on vowels (or vowel clusters) and onsets. The most noticeable phonotactic effects are the co-occurrence restrictions of certain vowels and coda consonants and the co-occurrence restriction on certain tones and codas. The matching of alveolar and palatal place features between main vowels and coda consonants is the most salient pattern of the former: the vowels [i.e., ɛ] cannot occur with the velar codas [k, ŋ].<sup>9</sup> For a detailed discussion of the general nature of the Vietnamese phonological system, see Thompson (1965) and Đoàn Thiện Thuật (1977).

## 2.2 Theoretical Assumptions

**Optimality Theory:** My analysis is presented in the Optimality Theory (OT) framework in which universal but violable constraints are posited. Within this theory, languages differ according to the way these constraints are ranked with respect to one another. The main mechanisms in OT are:

*Gen:* function which generates all the possible candidates for a given input

*Eval:* function which evaluates the optimality of each of the candidates with respect to a language-specific constraint ranking for the given input

<sup>9</sup> Pham (1997) gives an account of the distribution of final consonants in 4 dialects (Hanoi, Saigon, Hue, Quangnam) and for the co-occurrence restriction of velars after front vowels.