

Stress, tone and word prosody in Vietnamese compounds

John Ingram & Thu Nguyen

Linguistics Program E.M.S.A.H.
University of Queensland,
Australia
j.ingram@uq.edu.au

Abstract

Vietnamese compounds and phrasal constructions are investigated for phonetic correlates of lexical stress. Two series of experiments on the acoustic and perceptual characteristics of Vietnamese compound words and their phrasal counterparts are reported here and in a companion paper (Nguyen & Ingram, this volume), involving paradigmatic (same position within a syntactic frame) and syntagmatic comparisons (between adjacent elements), on five likely acoustic correlates of stress or prominence (f₀ range and contour, duration, intensity and spectral slope, vowel reduction). We conclude that Vietnamese has lexical stress as a phonetic tendency, but not as an active phonological contrast.

1. Introduction

At the level of word prosody, stress and tone are usually regarded as competing prosodic systems, whose phonetic exponents cannot dutifully serve two masters, though mixed systems of lexical stress and tone are certainly attested. Thus, a language which employs voice pitch contours contrastively for tone may be inhibited from using this as a cue for contrasting patterns of word stress. Stress and tone also have competing organizational or phonological properties which suggest that a language has one or the other system. Vietnamese and English may be cited as classical exemplars of polar opposites on a continuum of prosodic types.

Tone	←	Pitch accent	→	Stress accent
Vietnamese		Japanese		English

Yet traditional linguistic descriptions of Vietnamese are divided on the questions of whether word stress and a distinctive level of word prosody exist in Vietnamese. We approach these questions by examining the nature of Vietnamese compound words and their phrasal counterparts, with particular reference to the special case of reduplications.

1.1 The prosody of compounds

All languages possess a class of compound words whose members are (more or less) segmentally homophonous with a corresponding phrasal construction and whose etymology and meaning may or

may not be transparently related to the phrase (compare: ‘hot-dog’ vs. ‘hot dog’ with ‘hot-pants’ vs. ‘hot pants’). Compounding is a word formation device, whereby habitual word collocations acquire new, non-compositional meanings from phrases originally generated by the grammar of the language. In normal usage, compounds may be disambiguated from their corresponding phrases by linguistic context.

Many languages signal the word-like status of compounds by distinguishing them prosodically from phrases. The particular way that a language distinguishes a compound from an otherwise homophonous phrase provides a window on the properties of the prosodic word. It is controversial whether such a prosodic unit can convincingly be demonstrated in Vietnamese. Some argue that Vietnamese, a highly syllabic, contour tone language, falls on one end of a typological continuum of word prosody and lacks a system of word/lexical stress. Others argue that bi-syllabic word forms in Vietnamese are prosodically headed constructions, not equivalent in their prominence relations, that Vietnamese compound and phrasal constructions have a contrasting pattern of word prominence, as does a stress-accent language like English (Thompson, 1965; Tran Huong Mai, 1969; Ngo, 1984).

Many languages also mark the distinction between compound words and their phrasal counterparts prosodically, in ways that are language specific and possibly related to their prosodic type. In English, a Germanic language, where words tend to take initial

stress, this is achieved by assigning left-edge prominence to compounds, in contrast to the right-edged prominence of a phrase. Along with their characteristic left prominence, most English compounds also take on the prosodic characteristics of a word, de-accenting and temporally compressing the second element so as to conform to rhythmic constraints of stress-timing (e.g.: ‘*hót-dốg véndốr*’ vs. ‘*hót dốg véndốr*’; Faure, Hirst, & Chafcouloff, 1980; Farnetani & Cosi, 1988; Nguyen, Ingram & Pensalfini, submitted).

1.2 Vietnamese compounds

Compounding is a highly productive word formation process in Vietnamese. The word order in a noun phrase and a nominal compound is the reverse of that of English: *hoa hồng* (flower pink) vs. *hoa-hồng* (rose). Also, it is claimed by some linguists that the pattern of prominence in Vietnamese compounds, is the reverse of the English pattern; that is *weak-strong* for compounds “with weak stress on their first base/syllable” (Thompson, 1965; Tran Huong Mai, 1969; Ngo, 1984) and *strong-weak* for phrases (Thompson, 1965, p. 121). On the other hand, native speakers of Vietnamese are disinclined to recognize any pronunciation differences between compounds and their corresponding phrases. We sought therefore to test: a) whether there is any prominence related contrastive pattern between compounds and phrases in Vietnamese and b) whether compounds in Vietnamese as disyllabic word forms show any evidence of word-internal prosodic constituency, such as de-accenting one of the syllables of a bisyllabic word.

Vietnamese also possesses a productive class of more specialized compounds; reduplicative word forms, themselves made up of several subclasses (e.g.: *đậm đậm*, ‘*quite strong*’ [strong strong]). These reduplicated forms enabled us to make syntagmatic comparisons between the base and its reduplicant in order to test for prominence asymmetries within the disyllabic word. The paradigmatic and syntagmatic comparisons of interest are shown in Table 1 below.

Table 1. Prosodic contrasts

Phrase:	hoa hồng <i>pink flower</i> (flower pink)
	S1 S2
	← PARADIGMATIC C.
	S1 S2
Compound:	hoa hồng <i>rose</i>
Reduplication:	đậm đậm <i>quite strong</i> (strong strong)
	S1 – S2 ← SYNTAGMATIC C.

2. The Experiments

Two series of experiments are reported here. In the first series, acoustic evidence from native speaker productions and perception was sought as to whether there is any prosodically distinctive marking of compound words in Vietnamese (in relation to their corresponding phrases). Native speaker production and perceptual discrimination of compounds and phrases was tested under two speaking conditions: 1) one of ‘maximal contrast’, where subjects were asked to produce minimal prosodic pairs of target stimuli in a manner that would ‘bring out the differences in meaning’ between a compound and its corresponding phrase; and 2) in a ‘picture naming task’, where there was no particular focus upon the possible differentiating phonetic features of the compound - phrasal contrast.

The ‘maximal contrast’ condition was intended to test whether the language provides a way of prosodically marking distinctions in meaning between syntactically generated phrases and their corresponding lexical compounds. The picture naming task asks the same question under conditions that more faithfully reflect normal speech production. It is found that while speakers and listeners can reliably distinguish two of three types of compound from their corresponding phrasal constructions under the maximal contrast condition, they fail to do so under conditions of normal elicitation. We explain these results in terms of a language universal juncture marking strategy to which all speaker/listeners have access and conclude that Vietnamese has no contrastive pattern of compound word stress, and by implication, no prosodic template for disyllabic word forms.

In a follow-up experiment, we examined the acoustic properties of two types of reduplicative compound, which enable a sensitive test of prominence asymmetries between the reduplicant and the base elements of the compound. Here we obtained clear evidence of prominence asymmetry; that reduplicative compounds are phonetically right-headed, and by implication, that Vietnamese shows phonetic evidence of prosodic constituency at the level of the disyllabic word.

2.1. Series 1ⁱ. Compound and phrasal stress

2.1.1 Subjects and test conditions

Two groups of native speakers ($n_1=30$, $n_2=15$), aged 18-22, balanced for selection by gender and by the 3 major dialect regions of Vietnam (northern, central and southern) were separately tested for production of compound and phrasal constructions under two elicitation conditions.

Under the maximal contrast condition, subjects were asked to read minimal sentence-pairs in a natural way, such that listeners could distinguish between the meaning of a compound and its corresponding phrase. The test items were embedded in sentences having the same grammatical structure and word order. All test items occurred in utterance non-final position. Examples of these sentences are:

Compound: **Hoa hồng thì đẹp**

A rose is beautiful

Phrase: **Hoa hồng thì đẹp**

A pink flower is beautiful

For the picture naming task, subjects were asked to describe a picture, using a constant carrier sentence to ensure that the target phrase or compound word appeared in non-final position in an utterance with the same number of words, tone coarticulation and speaking rate effects. Two examples:

There + classifier + CP/PH + here

Có một bông **hoa hồng** ở đây

There is a rose/pink flower here.

Có hai con **cá mập** ở đây

There are two sharks/fat fish here.

Recordings were made in a quiet room using sound recording and editing computer software (Praat), at 20 kHz sampling rate and 16bit precision.

2.1.2 Linguistic materials:

Fifteen pairs of two-syllable compounds and their corresponding phrases were constructed from three types of compounds formed on the basis of their grammatical structures: noun-adjective (N-A type), noun-verb (N-V type), and noun-noun (N-N type), illustrated in Table 2.

Table 2. Type of compound, source phrase and syntactic rank

Type	example	Compound	Phrase	Rank
N-A	hoa hồng	<i>rose</i>	<i>flower [is] pink</i>	clause
N-V	bò cày	<i>ox-plough</i>	<i>Ox [is] ploughing</i>	clause
N-N	chân vịt	<i>propellar</i>	<i>duck's foot</i>	NP

There is an important distinction between the phrase-types that correspond to three types of compound used in the experiment. The N-A and N-V compounds derive from syntactic constituents that form a clause and are dominated by an S node in standard syntactic representation. The N-N compounds on the other hand,

derive from noun phrases, dominated by the lower-ranked syntactic constituent NP. It is a well known from prosodic phonology that juncture breaks are much more likely to occur between major clausal constituents than within a phrasal constituent, such as an NP (Selkirk, 1984; Nespor and Vogel, 1986).

2.1.3 Parameter measurements

The test items were segmented and labeled using the Emu Speech Tools, (Cassidy, 1999) in accordance with conventions described elsewhere (Nguyen & Ingram, submitted). Emu-R statistical software was used to extract vowel duration (ms), vowel formant (Hz), intensity (db), spectral tilt and fundamental frequency (Hz). The following acoustic parameters were measured:

- 1) Duration of first and second syllables [S1 S2]); duration of the pause (if longer than 100ms) between S1 and S2.
- 2) First formant and second formant of both first and second syllables (S1F1, S1F2, S2F1 and S2F2). Formant at vowel mid point was taken for monophthongs while formant values at ten equidistant points of the formant trajectory were taken for diphthongs.
- 3) Tone range (F0 range=F0 max-F0min): S1F0range and S2F0range
- 4) Mean of vowel intensity (db) at four equidistant points in S1 and S2
- 5) Vowel spectral tilt (H1-A3): third formant is compared with the first harmonic using Stevens and Hanson's model (1995) (V1 spectral tilt and V2 spectral tilt).

3. Analysis and Results

3.1 Juncture pauses:

An analysis of juncture pauses between syllables S1 and S2 revealed striking differences between the maximal contrast and picture naming elicitation conditions insofar as distinguishing between compounds and phrasal constructions was concerned. Pausing failed completely to distinguish between compounds and phrases under picture naming, but did distinguish between the majority of compound – phrasal pairs for two of the three types under the maximal contrast condition (see Table 3.)

Table 3. Distribution of juncture pauses

Compound Type	Maximal contrast		Picture naming	
	Phrase	Comp.	Phrase	Comp
N-A	80%	0%	0%	0%
N-V	65%	0%	0%	0%
N-N	15%	0%	0%	0%

These results strongly suggest that the speakers were employing a juncture (phrase boundary) marking strategy to differentiate phrases and compounds in the maximal contrast (minimal pair) condition, but were not doing so under the picture naming elicitation condition. We sought to confirm this and to determine whether any *other* mechanism might be operative from a series of paradigmatic comparisons (see Table 1) on the five prosodic parameters described above.

3.2 Statistical analysis of prosodic parameters:

Each of the five prosodic parameters (mentioned above) served as a dependent variable, in a series of analyses of variance conducted on each of the target syllables (S1 and S2) involved in the paradigmatic comparisons of interest for the compound – phrasal prosodic contrast. In addition to the compound – phrasal contrast (the factor of ‘prosodic type’), we were also interested in the factor of ‘compound type’ (N-A, N-V, N-N) and how it interacts with the compound – phrasal contrast under the two production conditions of minimal pair elicitation and picture naming. Each analysis of variance constituted a mixed two-way ANOVA, with prosodic type and compound type as fixed effects and subjects and items as random effects. The restricted maximum likelihood (REML) method was used to estimate variance components.

The pattern of significant main effects and interactions in the ANOVA set (see Appendix) closely paralleled the distribution of juncture pauses (Table 3), providing *prima facie* support for the hypothesis that *all* the prosodic effects which differentiate compounds from phrases are linked to the distribution of juncture pauses (phrase boundary effects). There were no significant effects observed under the picture naming elicitation, whereas robust main effects of prosodic type and interactions between prosodic and compound type were found for the maximal contrast condition for the dependent variables of syllable duration, and F0 range. Intensity yielded a main effect for the phrasal – compound contrast under minimal pair elicitation, but no interaction effect with compound type.

3.2.1 Syllable duration

The main effect of duration under the maximal contrast condition was found to occur on S1 (Figure 1). Lengthening of S1 in the phrasal constructions (compared with the compounds) in the manner indicated is entirely compatible with pre-pausal lengthening and the juncture hypothesis.

The interaction of compound type with the compound – phrasal contrast was reflected in the fact that the

duration differences, though significant for N-N was not as strongly marked as for the other two compound types.

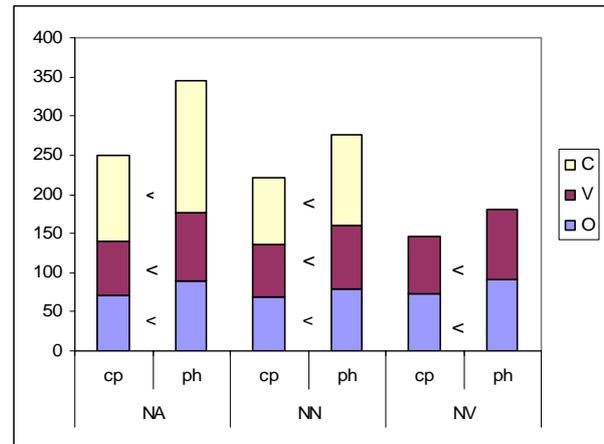


Figure 1. Segment lengthening of PH in S1 under maximal contrast condition. O=onset, C=coda.

3.2.2 F0 range

Evidence of pitch range expansion as a possible marker of compound phrasal contrast was only found under the maximal contrast condition. There was a significantly wider F0 range on S1 (confined to the dynamic ‘falling’ and ‘rising’ tones). This was more apparent on the N-A and N-V compound types than the NN type (as evident in the interaction of prosodic type and compound type). An expanded pitch range on dynamic tones on the right edge of a phrase boundary is consistent with the juncture interpretation of the temporal measurements.

3.2.3 Intensity

Phrasal constructions at both syllables (S1 and S2) had a lower intensity than their corresponding compounds. While a lower intensity at S1 might be expected on grounds of a phrase boundary between S1 and S2, the lower intensity on S2 is hard to account for. *Spectral slope* measurements showed no significant effects under maximal contrast or picture naming elicitation.

3.2.4 Formant frequencies

The distribution of vowels in F1 – F2 space for the compound – phrasal contrast was examined for evidence of vowel reduction, at S1 and S2 target positions. But no significant effects were obtained, under either the picture naming or the maximal contrast conditions. However, there was a significant raising of the first formant of the vowel of noun phrases in comparison to their corresponding compounds under the maximal contrast condition. No such effect was found

under picture naming elicitation. We conjecture that the raised F1 value could be due to greater jaw lowering associated with postural motoric adjustments at a phrase boundary.

4. Conclusions

In summary, all prosodic indicators of a compound – phrasal contrast were only observed under the maximal contrast elicitation condition. Furthermore, (almost) all of the various timing, pitch, and intensity cues were consistent with well-established prosodic phrase boundary effects. There was no evidence of word-level de-accenting, timing adjustment, or syllable reduction, associated with the Vietnamese compound – phrasal contrasts examined in this study. We conclude that Vietnamese has no phonological mechanism for differentiating compound words from their corresponding phrases, by making the former conform to the prosodic template of a word, as in languages, like English, with lexical stress. When required to do so by task demands, Vietnamese speakers have access to a language-universal juncture marking strategy to disambiguate the lexical meaning of a compound from its corresponding phrase – providing the phrase meets syntactic requirements for phrase boundary insertion (met in the case of N-A and N-V, but not N-N compound types).

This conclusion is reinforced by evidence from a perceptual study (Nguyen & Ingram, submitted), which space restrictions prevent us from describing in detail here. But briefly, it was found that native listeners could reliably distinguish only compound – phrasal tokens culled from the items of this production study, which were elicited under the maximal contrast condition, and only then for the N-A and N-V pairs, not those of type N-N. Compound – phrasal pairs which were elicited via the picture naming task were perceptually discriminated at no better than chance level by native Japanese listeners. Analysis of the perceptual responses revealed that only those pairs in which a phrasal construction was distinguished from its corresponding compound by a juncture break were reliably discriminated by Vietnamese listeners. These results are consistent with native speaker intuitions that phrases and their corresponding compounds are usually pronounced identically in Vietnamese, even in careful speech and that no distinctive compound stress pattern derived from a need to conform to the template of a prosodic word, as in English.

Thus, the acoustic and perceptual evidence appears to converge. Compound and phrasal constructions are normally prosodically homophonous in Vietnamese. Does this mean that Vietnamese lacks a word-level prosodic template, something like that which is

responsible for the distinctive stress and timing characteristics of compound words in English?

It must be conceded that the paradigmatic comparisons upon which the acoustic evidence of the present study relies is less than ideal. Prosodic features, particularly lexical stress patterns, probably rest heavily on syntagmatic comparisons between syllable sized units in the speech stream. Such comparisons were not possible given the phonetic structure of the compound phrasal constructions examined in this study. But Vietnamese provides us with a productive subclass of compound expressions – full and partial reduplications, which provide segmental feature control that enables a sensitive evaluation of asymmetries of prosodic prominence between adjacent syllables of disyllabic words.

This is what motivated the study of Vietnamese reduplicant compounds reported in the companion paper.

5. References

- Cassidy, S. (1999). Compiling Multi-Tiered Speech Databases into the Relational Model: Experiments with the Emu System. *Proceedings of Eurospeech '99, Budapest*, September 1999 2239-2242.
- Farnetani, E. & Cosi, P. (1988). English compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and speech*, 31, 157-180.
- Faure, G., Hirst, D. J., & Chafcouloff, M. (1980). Rhythm in English: Isochronism, pitch, and perceived stress. In Waugh L. R. & van Schooneveld, C.H. (eds.), *The Melody of Language* (pp. 71-79). Baltimore: University of Park Press.
- Nespor, M. & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
- Ngo, Thanh Nhan (1984). *The syllabeme and pattern of word formation in Vietnamese*. Ph.D dissertation. New York University.
- Nguyen, T. and Ingram, J. C. (2006). Acoustic and perceptual cues for compound - phrasal contrasts in Vietnamese, Submitted to: *Journal of the Acoustical Society of America*.
- Nguyen, T.A. T., Ingram, J., & Pensalfini, R. Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. Submitted to *Journal of Phonetics*
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook* 3, 371-405.
- Stevens, K.N. & Hanson H. M. (1995). Classification of glottal vibration from acoustic measurements. In Fujimura O., Hirano H. (Eds.), *Vocal fold physiology: voice quality control*. (pp. 147-170) San Diego: Singular Publishing Group.

Thompson, L. (1965). *A Vietnamese reference grammar*. Honolulu: University of Hawaii Press.

Tran, Huong Mai (1969). *Stress, tone and intonation in South Vietnamese*. Unpublished Ph.D thesis. Australian National University.

APPENDIX: ANOVA RESULTS

Parameter	Factors	Minimal pair		Picture naming	
		S1	S2	S1	S2
S-duration	CP-PH	****	*	ns	ns
	CP-type	ns	ns	ns	ns
	CP-PHxCP-type	****	****	ns	ns
F0-range	CP-PH	****	****	ns	ns
	CP-type	ns	ns	ns	ns
	CP-PHxCP-type	****	ns	ns	ns
V-intensity	CP-PH	****	****	ns	ns
	CP-type	ns	ns	ns	ns
	CP-PHxCP-type	ns	ns	ns	ns
H1-A3	CP-PH	ns	ns	ns	ns
	CP-type	ns	ns	ns	ns
	CP-PHxCP-type	ns	ns	ns	ns
F1	CP-PH	****	ns	ns	ns
	CP-type	ns	ns	ns	ns
	CP-PHxCP-type	ns	ns	ns	ns
F2	CP-PH	ns	ns	ns	ns
	CP-type	ns	ns	ns	ns
	CP-PHxCP-type	ns	ns	ns	ns

F values: **** = $p < .00001$, * = $p < .01$, ns = not sig.

ⁱ See Nguyen and Ingram, this volume for second series of experiments.