Acoustic and perceptual cues for compound - phrasal contrasts in Vietnamese

Thu Nguyen and John Ingram

Linguistics program,
University of Queensland
thunguyen@uq.edu.au & j.ingram@uq.edu.au

Abstract

This paper reports two experiments that examined the acoustic and perceptual cues that Vietnamese use to distinguish between compounds and noun phrases. Fifteen minimal sets of the two patterns classified into three different word/phrase types (head noun-adjective modifier (*hoa [flower] hồng[pink]: pink flower*), head noun-verb modifier (*bò[ox] cày[plough]: ox ploughing*), and head noun-noun modifier (*bàn[table] giấy[paper]: paper table*) were recorded in two experimental conditions: one with a picture-naming task and one with a minimal pair sentence task by forty five Vietnamese native speakers of three dialects (Hanoi, Hue, and Saigon). In a perception task, the meaning of the patterns is identified in a forced choice test by fifteen listeners. The results showed that while there is evidence that Vietnamese use juncture (pausing) and pre-pausal lengthening to distinguish between compounds and phrases, no significant acoustic and perceptual evidence was found to support a claim for contrastive stress patterns between compounds and noun phrases in Vietnamese.

1. Introduction

Compounding is a highly productive process in Vietnamese and in many languages. Compounding may be characterised as a device for creating words (new lexical items) from phrases (compositional syntactic constructions). Phonologically, this involves a prosodic contrast in which the prosodic characteristics of the phrasal construction are altered to conform to the prosodic template of the word. Typically, in a stress accent language such as English this involves a change in temporal structure, such that the first element of the compound takes on the accentual characteristics of primary word stress and the compound as a whole takes on the rhythmic properties of a lexical word, as a domain for rhythm-induced temporal compensation effects. The phonology of compounding may be seen to form a bridge between the lexical phonology of the word and the post-lexical phonology of the phrase. As a consequence, in English, compounds and phrasal constructions are phonemically contrastive, as is evidenced by minimal pairs such as "blackberry" vs. "black berry". Phonetically, in terms of acoustic cues, compounds are left headed (i.e. pitch and intensity of the first stressed syllables are higher, the first stressed syllables are longer) while broad focus noun phrases are right-headed (i.e. pitch and intensity of the second stressed syllables are higher than those of the compounds) [2, 5]. In terms of temporal structure, compounds are compressed (significantly shorter than their phrasal counterparts) to conform to the temporal template of a word counterpart [1, 5]. In tone languages such as Shanghai Chinese [8], and Jingpho Chinese [4], compounds are prosodically distinguished from phrases on the basic of lexical tone sandhi. In Vietnamese, tone sandhi is only restricted to reduplications [7, 10]; therefore, with the absence of tone sandhi in Vietnamese compounds, it is questionable whether there is an independent prosodic mechanism for distinguishing Vietnamese compounds from their phrasal counterparts or whether listeners are required to rely exclusively on context to disambiguate phrasal from compound constructions.

Vietnamese noun phrases and compounds are contrastive, at least in terms of meanings (compositional vs. non-compositional) and morpho-syntactic structure (phrases vs. words). Phonologically, it is generally claimed that the pattern of prominence in Vietnamese is the reverse of the English pattern; that is *weak-strong* for compounds and *strong-weak* for noun phrases [9, 10]. Nevertheless, whether the contrastive pattern of prominence of Vietnamese compounds and noun phrases is acoustically supported and whether Vietnamese compounds are subject to the rhythm-induced temporal properties of a lexical word, as in English, needs to be verified empirically. Studies on acoustic correlates of stress in tone languages such as Mandarin have shown that duration and surface fundamental frequency (F0) range are more important acoustic cues than intensity, while F0 has no effect on stress perception [8, 12]. In Vietnamese, a tone language, no system of culminative word stress has been found; nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level [9]. With pitch height, pitch direction and voice quality as more important correlates of phonemic tone contrasts than duration and loudness in Vietnamese [6, 11], one is curious to investigate what acoustic cues are used for stress distinctions (if any there be) and whether pitch plays any important role in stress contrasts.

This study examined the acoustic and perceptual cues that distinguish between Vietnamese compounds and noun phrases. The aim of the study is two-fold: (1) to verify whether previous claims about Vietnamese compound and noun phrase stress patterns (weak-strong in compounds vs. strong-weak in noun phrases) are supported with acoustical evidence; (2) if there is no acoustical evidence for the contrastive stress patterns, what acoustic and perceptual cues are employed to distinguish between Vietnamese compounds and noun phrases? In order to pursue the aim of the study, two experiments were carried out: a production experiment with two production tasks (a picture naming task representing spontaneous natural speech and a minimal pair sentence one to elicit maximal contrastive patterns) and a perception experiment. Two competing hypotheses are postulated;

Hypothesis 1*:* This hypothesis is to test whether there is acoustic evidence for the contrastive stress patterns (strong-weak in noun phrases vs. weak-strong in compounds). This hypothesis predicts that the vowel of the first element (V1) in a noun phrase will have longer duration, stronger intensity, larger F0 range, possibly higher F0 and fuller vowel quality than the same vowel in a compound and vice versa, the vowel in the second element (V2) of the compound will have longer duration, stronger intensity, larger F0 range, possibly higher F0 and fuller vowel quality than the same vowel of the phrase.

Hypothesis 2: This hypothesis predicts that in case no strong evidence of duration, intensity and pitch as a cue to the contrastive stress patterns is found, a junctural pause between two constituents of the phrase is the acoustical and perceptual cue to the distinction between compounds and noun phrases in Vietnamese. If hypothesis 2 turns out to be correct, it raises the question of whether a compound – phrasal prosodic contrast is part of the phonology of Vietnamese or whether Vietnamese speakers are making use of a prosodic device for syntactic phrase disambiguation (juncture) which is available universally. 'Juncture' as a phonological category is probably a complex phonetic construct, cued by pre-boundary lengthening, pausing and possibly also pitch contour modification (though this may well be much less prominent in a tone language). In this study, the pausing between the two constituent syllables of the test items and the pre-pausal lengthening of the first elements are examined in order to decide whether there is a juncture or not.

2. Experiments

2.1. Linguistic materials

A list of minimal sets of two-syllable compounds and noun phrases was first compiled. This list of minimal sets was classified into three types on the basis of their grammatical structures: noun-adjective (N-A type), noun-verb (N-V type), and noun-noun (N-N type). The items in the N-A type are made up of a head noun and a descriptive adjective (e.g., hoa hồng: flower pink). Those items in the N-V type group are made of a head noun and a verb which has the progressive meaning in the noun phrase (bò cày: ox ploughing). The N-N type consists of items that have two nouns with various meaning relationships in the noun phrase such as gender (bạn trai: boy friend), material (nhà đá: a house made of stone) or possession (chân vịt: a duck's foot). There were in total 15 sets of words/phrases in which five are in each word/phrase type.

2.2. Task types

There are two experimental task types performed by two different groups of speakers, a picture-naming task and a minimal pair sentence task. The aim of the picture-naming task was to elicit spontaneous natural speech produced by speakers without contrastive focus in mind. The aim of the minimal pair sentence task was to prompt speakers to produce contrastive patterns.

Minimal pair sentences: the minimal sets of testing items were embedded in minimal pair 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)

Noun phrase***: Hoa hồng*** thì đẹp (A ***pink flower*** is beautiful)

Picture-naming task: The same suggested sentence frame was used to describe the picture of the object denoting the meaning of the compound/phrase. The aim was to make sure all testing items appeared in non-final position of the same contextual sentence having the same number of words so as to avoid final lengthening, tone coarticulation and speaking rate effects. For examples,

Có(there's)+classifier**+ *compound/phrase***+ở đây (here)

Có một bông ***hoa hồng*** ở đây (there is a rose here)

Có hai con ***cá mập*** ở đây (there are two sharks here)

2.3. Subjects and recording procedures

In the *minimal pair sentence task*, thirty first-year university students in Vietnam (speakers of Hanoi [n=10], Hue [n=10], and Saigon dialect [n=10); half males and half females in each dialect group, age ranged 18-22) were given the list of the minimal pair sentences with distinctive meanings in parentheses and asked to read the pair of sentences in such a natural way that listeners can distinguish the meaning between a compound and a noun phrase.

In the *picture-naming task*, a different group of fifteen international students (dialect: 6 Hanoi, 4 Hue and 5 Saigon; gender: 9 females and 6 males) at the university of Queensland were asked to describe the picture using the given key word (i.e. the test item) and the suggested sentence frame. They were in the age range of 22-45 and had been in Australia from four months to one year. The picture and the key word were presented via a power-point slide show on a desktop computer. The key word was presented visually on the picture slide to make sure the subjects use the target test items. Before the recording, they were given instruction by the experimenter and visually on the computer screen that they had to use the key word in each slide and the same sentence frame for describing the pictures.

The recording was made in a quiet room using a sound recording and editing computer software (Praat) at 20 kHz sampling rate and 16bit precision

2.4. Measurement

The following acoustic parameters were measured:

- Duration of first and second vowels and syllables (*V1 duration, V2 duration, S1duration, S2 duration*); duration of the pause (if longer than 100ms) between the syllable constituents of the phrases/compounds
- First and second vowel formant at vowel mid point (*F1 and F2*)
- Fundamental frequency (F0) at *10 equidistant points* on the tone contour of each syllable rime; mean F0 value of these 10 points (*F0 mean*); and tone range (*F0 range*)
- Mean of vowel intensity (db) at four equidistant points (*V1 intensity mean and V2 intensity mean*)
- 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).*

2.5. Perception experiment methodology

Stimuli of four Hanoi female speakers from the production experiment (two speakers from the picture-naming task and two from the minimal pair sentence task) were used for the perception study. Fifteen subjects (6 Hanoi, 2 Hue and 7 Southerners; 10 females) with no known auditory deficiencies, participated in the perception experiment. The listening identification test consists of two parts: in part one, listeners listened to the target items presented in isolation; in part two, they heard a contextual sentence containing the target item. Part one was performed before part two to avoid carry over effect of contextual sentence. Two different meanings of the target item were given in the answer sheet: one as a compound and one as a noun phrase. The subjects' task was to choose the meaning which they think was

expressed by the speaker by circling the letter corresponding to their response in the answer sheet.

The experiment was delivered online so that subjects could either perform the test via their personal computer at home or at the researcher's office. There were instructions on the webpage on how to download the Microsoft word answer sheet and how to listen to the sound files which were presented in order of links. They were instructed to perform part one before part two and to listen to the sound files only once. Subjects returned their answer sheets to the researcher via email or post. One problem is how one can be sure the instruction was well followed; however, among the fifteen participants, the results of the four participants performing the test at their own space was not different from the eleven listeners who came to do the test at the researcher's office.

2.6. Analysis

The statistical analysis involves twenty two acoustic parameters (the italicized in section 2.4 above) as the dependent variables and five factors: (1) conditions (compounds [CP] vs. noun phrases [NP] as the main effect of interest), (2) word types (NA, NN, NV), (3) dialects (Hanoi, Hue and Saigon); words (15 different minimal pairs of test items) and speakers (30 speakers in the minimal pair sentence task and 15 speakers in the picture naming task). In order to account for the effect of speakers' differences and intrinsic segmental as well as tonal differences among the 15 test items, a restricted maximum likelihood (REML) applied to mixed model ANOVA methodology was performed on each of the twenty two acoustic parameters. The fixed effects included conditions (CP vs. NP), word types (NA, NN, NV), dialects (Hanoi, Hue and Saigon), and their two- and three-way interactions. The random effects were speakers and words. A Tukey post-hoc test was then conducted to determine the significant differences among levels of the main fixed factors and their interaction effects.

2.7. Statistical results of the production experiment

The mixed model ANOVA results showed no significant effects (neither main nor interaction effects) for any of the twenty two acoustic parameters of the picture-naming task. Significant effects were found for only six acoustic parameters of the minimal pair sentence task, namely intensity mean, F0 mean, F0 range, syllable duration, vowel duration and vowel first formant. Thus only these significant effects are discussed below.

Juncture pauses: The spectrographic analysis showed that there was a juncture pause (longer than 100 ms) between syllable constituents of the noun phrases elicited under the minimal pair sentence task while this juncture was not found in compounds of the minimal pair sentence task and neither in the compounds nor phrases elicited under the picture-naming task. Nevertheless, juncture appeared mainly in two types of noun phrase constructions: noun-adjective and noun-verb, not in noun-noun phrases. Juncture pauses occurred approximately above 80% of the noun-adjective phrase constructions, about 65% of the noun-verb phrase constructions but only 15% of the noun-noun phrase constructions across dialect groups.

Spectral tilt and intensity and vowel formant: No significant effect was found for spectral tilt values. There was a very marginally significant difference (CP vs. NP) in intensity (db) for the first syllable of the noun-verb word/phrase type only ($p<0.05$). Similarly, there was a significant condition effect for first formant of the first vowel ($p=<0.0001$); however, post-hoc multiple comparison showed negligible significant effect for the formant value of the first syllable of the noun-adjective construction only ($p<0.05$).

Fundamental frequency: No significant effect for any of the F0 values on ten equidistant points of the tone contour was found. There was a marginally significant main effect of F0 mean (CP>NP) across three word/phrase types of the first syllable (NA, NN, NV, $p<0.05$) and second syllable of the noun-adjective construction only (NA, $p<0.05$). By contrast, a post-hoc test on F0 range showed a highly significant difference (NP>CP) for first syllables of the noun-adjective and noun-verb constructions ($p<0.001$) but not noun-noun words/phrases

Duration: Significant effects were found for the main factor conditions (CP vs. NP) and the interaction conditions x word-types of the first and second syllable and vowel duration. Effects of conditions: V1: $F(1,850)=902.69$, $p=<.0001$, V2: $F(1,854)=17.95$, $p=<.0001$, S1: $F(1,850)=971.16$, $p=<.0001$, S2: $F(1,854)=8.63$, $p=0.0034$. The interaction of conditions x word-types: V1: $F(2,850)=56.67$, $p=<.0001$, V2: $F(2,854)=19.09$, $p=<.0001$, S1: $F(2,850)=19.21$, $p=<.0001$, S2: $F(2,854)=20.94$, $p=<.0001$. Tukey post-hoc comparison shows highly significant difference (NP>CP) for the first vowels and syllables of all three word/phrase types (NA, NN, NV). Nevertheless, the magnitude of lengthening was much greater for NA and NV than NN constructions. By contrast, significant effect for the second vowel and syllables (CP>NP) was only found for noun-verb construction.

2.8. Discussion of acoustic results

Duration: It is clear that there is a relationship between the statistical results on duration and the juncture pause between the constituents of the noun phrase elicited under minimal pair sentence task: the NN word/phrase type is different from the other two phrase types NA and NV: there is almost no (very few) junctures between two syllables of the N-N type and the lengthening effect on V1 of the N-N type is less than those of the N-A and N-V types. In other words, there is a lengthening effect on the first syllable of the noun phrase which is less in the case of the N-N phrase with the absence of a juncture between its constituent elements. It is argued that there is a ***word-level final lengthening***, that distinguishes the sequence of two independent words in the N-N noun phrases (e.g., nhà đá: house stone, chân vịt: foot duck) from the word-internal morpheme juncture of the compounds (e.g., nhà đá: prison, chân vịt: propeller) and a ***phrase-level junctural effect*** that distinguishes the sequence of a noun (N) and a verb phrase (VP) in the NA and NV noun phrases (e.g., hoa hồng: flower pink, người ở: person living) from the word-internal morpheme juncture of their corresponding compounds (hoa hồng: rose, người ở: servant).

Fundamental frequency: the statistical results showed no significant effect in terms of F0 height across ten points of the tone contour. However, the significant difference in F0 range (NP>CP) on the first syllable of only the NA and NV word/phrase types showed an interrelation with first syllable lengthening effect. In other words, the first syllable of the NA and NV noun phrases was longer and thus allowed its tone to be more fully realized and thus had a larger F0 range than its counterpart in compounds.

2.9. Results of the perception experiment

Generally, the result of a four-way fixed effect ANOVA (2 task types (word/phrase vs. sentence) x 2 conditions (CP vs. NP) x 3 word types (NA, NN, NV) x 4 speakers) conducted on the number of correct responses (the number of listeners over 15 listeners who chose the correct response for a testing item) showed no significant effect between word/phrase and sentence stimuli across speakers except for only one speaker. This indicates that contextual sentence did not have a strong effect on listeners' improvement of discrimination between the compound and noun phrase patterns across natural (picture-naming) and contrastive (minimal pair) speech stimuli.

There are three other important findings from the perception experiment. (1) Target items with no juncture between constituents of the noun phrase (e.g., NN constructions elicited under minimal pair sentence task represented by speakers 1&2 and all test items elicited in the picture-naming task represented by speakers 3&4 tend to trigger more compound responses. This can be due to the carry-over effect of junctural pause in many noun phrase items; however it also indicates that listeners expected a juncture pause between components of phrases as a distinction from compounds. (2) Performance for stimuli with an absence of a juncture between constituents of the noun phrase was at chance level (clustering around 50%), indicating listeners' confusion between the compounds and phrases. (3) In patterns that have a juncture between syllables of noun phrases (NA and NV under the maximal contrast condition represented by speakers 1&2), identification was better though not optimal (approximately 70%) and has an equal rate of success between the two patterns.

The results of this perception experiment indicate that the juncture between components of the noun phrase is essential for listeners' identification between the two patterns in both isolation and sentence context. This also suggests that under normal communication context where the distinction or disambiguation between the compound and phrasal readings is not required, Vietnamese speakers do not usually make a junctural pause between the two components of a noun phrase; however, to listeners, the presence of a junctural break is more important than contextual sentence in assisting the discrimination between meanings.

3. Conclusion

The results of the production experiment shows that hypothesis 1 is not supported: there was no significant effect on any of the acoustic parameters between compounds and noun phrases elicited under the picture naming task which represents spontaneous natural speech. Even under conditions of maximal contrast, there is no significant acoustic evidence to support the claim about the contrastive stress patterns between compounds and noun phrases in Vietnamese. Alternatively, there was a juncture between the two constituents of noun phrases while no juncture was present between components of compounds, which suggests a word-level (for the N-N types) and a phrase-level (for the N-A and N-V types) juncture effect that distinguishes the sequence of two independent words in the noun phrases from the word-internal morpheme juncture of the compounds. The first components of the phrasal constructions were lengthened in comparison with their corresponding compounds as a result of word-final lengthening and was accompanied by an expanded tone range. No conclusive evidence of fundamental frequency, duration or intensity prominence was found on the second components of compounds. Compound words as a whole were not temporally compressed in comparison to their phrasal counterparts as in English, a stressed language.

The results of the perception experiments indicate that listeners relied only on the juncture between the two components of noun phrases as a cue to distinguish between noun phrases and compounds. Listeners' failure to distinguish noun phrases from compounds in the N-N types and in stimuli elicited under picture-naming task where there is no juncture between the two constituents strengthens the evidence that there is neither significant acoustic nor perceptual evidence of contrastive stress patterns (strong-weak vs. weak-strong) between noun phrases and compounds as previously claimed in the literature [9, 10].

It can be concluded that there is evidence that Vietnamese use juncture (pausing) to distinguish between compounds and phrases, consistent with finding in Ingram and Nguyen [3] that juncture and pre-pausal lengthening was used to signal contrasting head/dependency relations in compounds and phrases. By contrast, there is no significant acoustic and perceptual evidence to support the claim about the contrastive stress patterns between compounds and noun phrases in Vietnamese even under conditions of maximal contrasts. In other words, there is no evidence that a compound – phrasal prosodic contrast is part of the phonology of Vietnamese but native speakers can distinguish between phrases and compounds on the basis of meaning or context and that they can mark that distinction in speech by the use of juncture a prosodic device for syntactic phrase disambiguation (juncture) which is available universally.

4. References

[1] Farnetani, E. and Cosi, P. (1988). Emglich compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and speech, 31*, 157-180.

[2] Hardcastle, W.J. (1968). Stress in Australian English. Unpublished M.A. thesis. University of Queensland.

[3] Ingram J.C. and Nguyen T. A. T. (2003) Prosodic typology and compound – phrasal stress contrasts. Proceedings of the 15th International Congress of the Phonetic Sciences, Barcelona, August 3-9, 2003, 479 –482.

[4] Mortensen, D. (2002). Semper Infidelis: Theoretical dimension of tone sandhi chains in Jingpho and A-Hmao. Unpublished, UC Berkerly.

[5] Nguyen, T. A. T. and Ingram J. (2002). Native and Vietnamese Production of Compound and Phrasal Stress Patterns. Proceedings of the Seventh International Conference on Spoken Language Processing. Denver, Colorado, USA.

[6] Pham, Andrea Hoa. (2003). *Vietnamese tone- a new analysis*. New York: Routledge

[7] Seitz, P. (1986). Relationships between tones and segments in Vietnamese. Unpublished Ph.D., University of Pensylvania (UMI).

[8] Shen, X. S. (1993). On the temporal dimension of Mandarin. *Acta-Llinguistica-Hafniensia, 26*, 143-159

[9] Thompson, L. (1987). *A Vietnamese reference grammar*. Honolulu: University of Hawaii Press

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

[11] Vu, Thanh Phuong. (1981). The acoustic and perceptual nature of tone in Vietnamese. Unpublished Ph.D. thesis, (Australian National University, Canberra

[12] Xu, Yi (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics 27*, 55-105.