Phrasal stress in Mandarin disyllabic phrases: an investigation using focus

Hao Yi

Department of Linguistics, Cornell University
hy433@cornell.edu

Abstract

While primarily a tone language, Mandarin Chinese has been claimed to have phrasal stress which falls on a nonhead constituent: on the modifier in a modifier-noun phrase, and on the object in a verb-object phrase (MODNt and VtOBJ, respectively; the subscript h stands for head, and the stressed constituent is underlined). This Nonhead Stress Rule is motivated by greater information load carried by the nonhead than its syntactic head [1]. Taking Nonhead Stress Rule as a point of departure, the current study investigated Mandarin phrasal stress by using focus as a diagnostic tool. Fifteen pairs of homophonous disyllabic phrases, each consisting of a MODNt phrase and a VtOBJ phrase, were elicited under both BROADFOCUS and NARROWFOCUS. The phonetic correlate of phrasal stress – duration – were measured. The results showed that there was a significant difference in duration between a MODNt and a homophonous VtOBJ, consistent with the interpretation that MODNt exhibits initial stress and VtOBJ exhibits final stress. Moreover, the duration difference was amplified under NARROWFOCUS. In sum, the contrastive stress patterns of MODNt and VtOBJ support the information-motivated Nonhead Stress Rule.

Index Terms: phrasal stress, focus, duration, Mandarin Chinese

1. Introduction

While primarily a tone language, Mandarin Chinese has been claimed to have phrasal stress. The distribution of stress, according to [1], is governed by Nonhead Stress Rule: phrasal stress falls on the nonhead constituent of a phrase, because the nonhead carries more information than its syntactic head. Therefore, stress falls on the object in a verb-object phrase VtOBJ, and on the modifier (adjective or noun) in a modifier-noun phrase MODNt. (The subscript h stands for head; the stressed constituent is underlined.)

This distribution of stress has been addressed by several acoustic studies, all taking [1] as their point of departure. One of the studies based on spoken corpus investigated the rhythmic patterns in Mandarin polysyllabic words [2]. It concluded that there was no difference between disyllabic VtOBJ and MODNt in stress pattern on the basis of acoustic measurements, therefore could not confirm Nonhead Stress Rule.

Another study used identical MODNt and VtOBJ disyllabic pairs in a production experiment [3]. During the elicitation, a disyllabic phrase was overtly preceded its part of speech in the carrier sentence so that a MODNt and an identical VtOBJ can be differentiated. The results showed that while VtOBJ exhibited final stress, MODNt showed no initial stress, which, again, did not confirm Nonhead Stress Rule.

A third study used homophones that differed in terms of syntactic structures, i.e., each pair of homophonous disyllabic phrases consisted of one VtOBJ and one MODNt (which differed in orthography) [4]. Target phrases were elicited in isolation. It was concluded that most of the disyllabic phrases in the study exhibited final stress, and that therefore syntactic structure did not govern stress allocation in Mandarin.

While these studies lent great insight into the distribution of Mandarin phrasal stress, none of them confirmed Nonhead Stress Rule. Moreover, they raise methodological concerns, such as potential complications due to the unfounded reliance on Mandarin speakers’ judgement of parts of speech [3] or due to phrase final lengthening [4].

In the current study, such methodological drawbacks are carefully controlled for. Focus is used as a diagnostic tool to look for prosodic regularities in Mandarin disyllabic phrases. Specifically, this study investigates the phonetic correlates of phrasal stress in Mandarin Chinese, by measuring the duration under both BROADFOCUS and NARROWFOCUS. The effects of focus and syntactic structure on duration are tested in fifteen homophonous pairs of MODNt and VtOBJ. If a pair of homophones (MODNt and VtOBJ) displays contrastive stress patterns, focus-introduced prominence will apply differently: the duration changes induced by NARROWFOCUS for the stressed constituents (the MO of MODNt and the OBJ of VtOBJ) will be of greater magnitude than for their unstressed counterparts (the Nt of MODNt and the Vt of VtOBJ).

2. Methods

2.1. Participants

Two female speakers (F01 and F02) and one male speaker (M01) who are native speakers of Beijing Mandarin participated in this experiment. All three speakers were born and raised in Beijing, and were graduate students at Cornell University at the time of recording. The recording took place in the sound-proof booth in Cornell Phonetics Lab in Department of Linguistics at Cornell University. The participants were naïve to the purpose of the study.

2.2. Speech materials and data collection

The stimulus set consisted of 15 homophonous pairs of MODNt and VtOBJ. Homophones were chosen because segmental variations within each minimal pair can be controlled. The stimulus set exhausted the possible combinations of four lexical tones (i.e. Tone1, Tone2, Tone3, and Tone4) in Mandarin Chinese to the exclusion of the Tone3+Tone3 combination because of third tone sandhi. The target stimuli were elicited in two discourse contexts: BROADFOCUS and NARROWFOCUS.

(i) In each trial, the speaker was first presented with a sentence in Chinese characters as the background information. The information was presented in black.
(ii) Five seconds later, the speaker was presented with a related question based on the above background information. The question was presented in red.

(iii) The speaker was instructed to answer the prompted question based on the given information.

For a given background sentence (BACKGROUND), there were two types of questions: the BroadFocus question and the NarrowFocus question, which are listed in (BroadFocus) and (NarrowFocus), respectively. The speech materials were exemplified in Pinyin, the official phonetic system for transcribing Mandarin Chinese into Latin alphabet. Tones are omitted. The disyllabic target stimulus is represented as $\sigma_1 \sigma_2$.

Table 1: Example of elicitation under BroadFocus and NarrowFocus.

<table>
<thead>
<tr>
<th>(BACKGROUND)</th>
<th>ta juede shuo $\sigma_1 \sigma_2$ shun henduo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(BroadFocus)</td>
<td>‘He thinks it’s a lot more fluent to say $\sigma_1 \sigma_2$.’</td>
</tr>
<tr>
<td>(NarrowFocus)</td>
<td>ta juede shenme shun henduo?</td>
</tr>
<tr>
<td></td>
<td>‘What does he think is more fluent to say?’</td>
</tr>
</tbody>
</table>

In every block of elicitation, there were 30 (= 15 tone combinations $\times$ 2 syntactic types) BroadFocus trials, 30 NarrowFocus trials. The trials were presented in a random order. The blocks were separated by 5-minute breaks. The experimenter would ask the speakers to repeat the answer if the experimenter failed to perceive the intended focus. In total, 833 trials were collected.

2.3. Data analysis

The start and the end of both syllables of the target stimuli were manually labelled in Praat [5]. Durations at the syllable level were obtained in MATLAB. I took into consideration that durations of the target syllables vary with their syllable structures, therefore deriving the DurationRatio—the ratio between the durations of the first syllable ($\sigma_1$) and the second syllable ($\sigma_2$) within a disyllabic phrase.

The effects of syntactic structure (TYPE: ModN and VObj) and discourse context (DISCOURSE: BroadFocus and NarrowFocus) on DurationRatio were tested using Linear Mixed Models (lme4 [6] in R version 3.2.0). Other variables of fixed effects included tone types of both syllables (Tone1 and Tone2). Stimuli (STIM) and speakers (SPK) were included in the mixed model as variables of random effects.

3. Hypothesis and predictions

Hypothesis i: The nonheads have phrasal stress.

Prediction i: The nonheads will have greater durations under both focus conditions. Therefore, the DurationRatio of ModN is larger than that of VObj. Consequently, ModN and VObj will exhibit different stress patterns.

Hypothesis ii: Under NarrowFocus, focus-introduced prominence applies only to the stressed constituent, leading to stronger production of the nonheads in both ModN and VObj phrases.

Prediction ii: Under NarrowFocus, durational increase of the nonheads will be greater than their syntactic heads. Therefore, the DurationRatio of ModN will increase from BroadFocus to NarrowFocus, whereas the DurationRatio of VObj will decrease from BroadFocus to NarrowFocus.

4. Results

Globally, there was an effect of Type on DurationRatio. The DurationRatio of ModN was significantly larger than that of VObj ($t(822) = 4.3767$, p < 0.00001) (Figure 1).

In particular, under BroadFocus, the DurationRatio of VObj was significantly larger than that of ModN ($t(420) = 2.3043$, p < 0.05); under NarrowFocus, the DurationRatio of VObj was significantly larger than that of ModN ($t(394) = 3.9462$, p < 0.0001) (Figure 2). Moreover, the DurationRatio difference between ModN and VObj was more pronounced under NarrowFocus (0.097) than under BroadFocus (0.057).

Figure 1: DurationRatio of ModN and VObj. Globally, the DurationRatio of ModN was larger than that of VObj.

In particular, under BroadFocus, the DurationRatio of VObj was significantly larger than that of ModN ($t(420) = 2.3043$, p < 0.05); under NarrowFocus, the DurationRatio of VObj was significantly larger than that of ModN ($t(394) = 3.9462$, p < 0.0001) (Figure 2). Moreover, the DurationRatio difference between ModN and VObj was more pronounced under NarrowFocus (0.097) than under BroadFocus (0.057).

Figure 2: DurationRatio of ModN and VObj, grouped by discourse (BroadFocus and NarrowFocus). The DurationRatio difference between ModN and VObj was more pronounced under NarrowFocus than under BroadFocus.

Figure 3 shows the DurationRatio grouped by Spk. While there were some consistent global patterns indicative of the Type effect, there also existed speaker-specific patterns. Under NarrowFocus, both female speakers (F01 and F02) produced ModN with significantly larger DurationRatio than VObj ($t(167) = 3.6001$, p < 0.001; $t(113) = 2.2586$, p < 0.01). However, the male speaker (M01) did not differentiate between ModN and VObj with DurationRatio under NarrowFocus.
Out of three speakers, only F01 differentiated between ModN$_i$ and V$_{Obj}$ with DURATION$_R$ under BroadFocus ($t$(173) = 3.4725, $p < 0.01$).

Figure 3: DURATION$_R$ of ModN$_i$ and V$_{Obj}$ grouped by S$^3$P$^3$. Global Type effect was observed across speakers; with DURATION$_R$, ModN$_i$ and V$_{Obj}$ were better differentiated under NarrowFocus than under BroadFocus, though there existed cross-speaker variations.

Figure 4 shows the DURATION$_R$ grouped by tone combination ($\text{Tone}_1 + \text{Tone}_2$). Consistent with the previous results, for the majority of the tone combinations, the global patterns were: 1) the DURATION$_R$ of ModN$_i$ was larger than that of V$_{Obj}$; 2) ModN$_i$ and V$_{Obj}$ were better differentiated under NarrowFocus than under BroadFocus. However, there were also anomalies that ModN$_i$ and V$_{Obj}$ were not differentiated under either Discourse condition in terms of DURATION$_R$ (e.g., Tone1+Tone1), or that the DURATION$_R$ difference was larger under BroadFocus than under NarrowFocus (e.g., Tone2+Tone3).

Figure 4: DURATION$_R$ of ModN$_i$ and V$_{Obj}$ grouped by tone combination ($\text{Tone}_1 + \text{Tone}_2$). Global Type effect was observed for the majority of the tone combinations; with DURATION$_R$, ModN$_i$ and V$_{Obj}$ were better differentiated under NarrowFocus than under BroadFocus, though there existed variations across tone combinations.

The above results shown in Figures 1-4 are in line with Prediction i in that the nonheads have greater duration, therefore the DURATION$_R$ of ModN$_i$ is larger than that of V$_{Obj}$ under both BroadFocus and NarrowFocus.

In Figure 5, DURATION$_R$ was grouped by Type to better examine the DURATION$_R$ change from BroadFocus to NarrowFocus. A two-way ANOVA (factors: Type and Discourse) showed that DURATION$_R$ was conditioned by both Type ($F$(1, 829) = 19.232, $p < 0.0001$) and Discourse ($F$(1, 829) = 4.13, $p < 0.05$). Tukey’s HSD post-hoc tests showed that for V$_{Obj}$, the DURATION$_R$ decrease (0.056) from BroadFocus to NarrowFocus was marginally significant ($p < 0.1$), which is consistent with Prediction ii. However, for ModN$_i$, the DURATION$_R$ decrease (0.015) from BroadFocus to NarrowFocus was not only non-significant ($p > 0.1$), but also departs from Prediction ii, which suggests a significant DURATION$_R$ increase. Consequently, as also observed in Figures 2-4, ModN$_i$ and V$_{Obj}$ were better differentiated under NarrowFocus; the DURATION$_R$ difference between ModN$_i$ and V$_{Obj}$ was more pronounced under NarrowFocus.

Linear mixed model analysis (Table 2) confirmed that there was a global effect of Type that on average the DURATION$_R$ of V$_{Obj}$ was 0.06 smaller than that of ModN$_i$ ($t$(22.7) = -2.55, $p < 0.05$). No significant effect of Discourse was found. However, the interaction effect between Type and Discourse bordered on the level of marginal significance ($t$(799) = -1.024, $p = 0.1115$). Given that ModN$_i$ and BroadFocus were assigned the value of 0, i.e., they were the dummy variables, and that V$_{Obj}$ and NarrowFocus were assigned the value of 1 in the mixed-effect model, such an interaction effect suggested that the DURATION$_R$ decrease from BroadFocus to NarrowFocus for V$_{Obj}$ was (marginally) significant, whereas for ModN$_i$ the DURATION$_R$ change was non-significant. This is consistent with Tukey’s HSD post-hoc tests. Also note that when the second syllable ($\sigma_2$) bore Tone3, the DURATION$_R$ significantly increased by 0.36 ($t$(13) = 5.257, $p < 0.0001$). This can be accounted for by the idiosyncrasy induced by Tone3-bearing syllables in that they have shorter durations.
Specifically, the duration ratio change from BroadFocus to NarrowFocus for ModN_h needs to be accounted for. One possible reason is that Mandarin disyllabic phrases have trochaic foot structures in that they show a strong–weak alternating pattern [1]. Because the first syllable (σ_1) is already a strong position, NarrowFocus does not induce any pronounced change in duration ratio (ceiling effect). In this case, the focus-introduced metrical prominence is still associated with the Mod of ModN_h, but is disguised by the underlying strong–weak pattern. Note that the underlying trochaic foot structures do not refute Nonhead Stress Rule. It can be understood as that the underlying strong-weak pattern sets the baseline for all disyllabic phrases, and that the real comparison should be made between the syllables occupying the same positions, i.e., between the Mod of ModN_h and the V_h of V OBJ, and between the N_h of ModN_h and the OBJ of V OBJ. A second possible interpretation is that there might exist stimulus-dependent stress patterns that contribute to the overall non-significant duration ratio change for ModN_h. It is possible that the majority of ModN_h stimuli in the current study did not exhibit initial stress, therefore disguising the discursive effect. Lastly, another possible reason might lie in the choice of acoustic metric in the current analysis. It was found in [2] that, F0, rather than duration, was the phonetic correlate that better reflected the alteration of prosodic strength in both disyllabic and polysyllabic words. More analyses with F0 measurements are under way to look into the distribution of Mandarin phrasal stress.

For these reasons, I will tentatively argue that in line with Scenario (A) (as well as Nonhead Stress Rule), the duration ratio differences between ModN_h and V OBJ reflect the difference between initial stress and final stress, which is further indicative of two different syntactic structures.

Last but not least, the cross-speaker and cross-stimulus variations needs to be accounted for. While there existed variations, no speakers or stimuli showed patterns that went in the opposite direction of Findings 1 – 3. For F01 and F02, the duration ratio of ModN_h was larger than that of V OBJ for MOI, the duration ratio of ModN_h and V OBJ were not differentiable under either BroadFocus or NarrowFocus.

For some homophonous pairs, the duration ratio of ModN_h was larger than duration ratio of V OBJ; for others, the duration ratio of ModN_h was different from the duration ratio of V OBJ. Therefore, I suggest that such variations are more of idiosyncrasies than randomness, and that the information-motivated Nonhead Stress Rule is an important component to the prosodic process in Mandarin as it facilitates communicative efficiency by loading more stress into forms with more information. However, it is also acknowledged that Nonhead Stress Rule is a weak universal in that whether the phrasal stress patterns will surface to differentiate between a homophonous pair of ModN_h and V OBJ depends heavily on the idiosyncrasies of particular lexical items or individual speakers.

The study strongly suggests that the tendency of contrasting ModN_h and V OBJ results from Nonhead Stress Rule. It is argued that Nonhead Stress Rule, despite being weak, exists in Mandarin Chinese, because it helps to facilitate communicative efficiency when needed. Future studies should look into other acoustic correlates such as F0 measurements. Perception studies are further needed in order to show whether such knowledge of contrast does exist for those homophonous pairs that do not exhibit overt contrastive phrasal stress patterns in acoustics.
6. References


