Phrasal stress in Mandarin disyllabic phrases: an investigation using focus

Hao Yi
Department of Linguistics, Cornell University

Abstract

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 (ModN$_h$ and V$_h$O$_h$, respectively; the subscript $h$ stands for head, and the stressed constituent is underlined). This NONHEAD STRESS RULE is motivated by the 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. 15 pairs of homophonous disyllabic phrases, each consisting of a ModN$_h$ phrase and a V$_h$O$_h$ phrase, were elicited under both BROAD FOCUS and NARROW FOCUS. The phonetic correlate of phrasal stress—duration—was measured. The hypothesis tested was that the nonheads have phrasal stress. The results showed that at the phrase level, a ModN$_h$ and a homophonous V$_h$O$_h$ differed significantly in duration ratio, consistent with the interpretation that ModN$_h$ exhibits initial stress and V$_h$O$_h$ exhibits final stress. Moreover, the duration ratio difference was amplified under NARROW FOCUS. However, there also existed cross-stimulus variation, which is argued to be idiosyncratic rather than random. In sum, it is concluded that NONHEAD STRESS RULE, despite being a weak universal, is an important component to Mandarin prosody, and underlies the contrastive stress patterns of ModN$_h$ and V$_h$O$_h$. Index Terms: phrasal stress, focus, duration, Mandarin Chinese

1. Introduction

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, V$_h$O$_h$, and on the modifier (adjective or noun) in a modifier-noun phrase, ModN$_h$.

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 a spoken corpus, investigated the rhythmic patterns in Mandarin polysyllabic words [2]. It concluded that there was no difference between disyllabic V$_h$O$_h$ and ModN$_h$ in terms of stress pattern on the basis of acoustic measurements, and therefore could not confirm NONHEAD STRESS RULE.

Another study used identical ModN$_h$ and V$_h$O$_h$ disyllabic pairs in a production experiment [3]. During the elicitation, a disyllabic phrase was overtly preceded by its part of speech in the carrier sentence so that a ModN$_h$ and an identical V$_h$O$_h$ can be differentiated. The results showed that while V$_h$O$_h$ exhibited final stress, ModN$_h$ 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 V$_h$O$_h$ and one ModN$_h$ (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 BROAD FOCUS and NARROW FOCUS. The effects of focus and syntactic structure on duration are tested in 15 homophonous pairs of one ModN$_h$ and one V$_h$O$_h$. If a homophonous pair (ModN$_h$ and V$_h$O$_h$) displays contrastive stress patterns, focus-introduced prominence will apply differently: the duration changes induced by NARROW FOCUS for the stressed constituents (the Mod of ModN$_h$ and the O of V$_h$O$_h$) will be of greater magnitude than for their unstressed counterparts (the N$_h$ of ModN$_h$ and the V$_h$ of V$_h$O$_h$).

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 ModN$_h$ and V$_h$O$_h$. 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 due to third tone sandhi. The target stimuli were elicited in two discourse contexts: BROAD FOCUS and NARROW FOCUS.

(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 transcription Mandarin Chinese in the Latin alphabet. Tones are omitted. The disyllabic target stimulus is represented as $\sigma_1 \sigma_2$.

<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>‘What does he think?’</td>
</tr>
<tr>
<td>(NarrowFocus)</td>
<td>‘What does he think is more fluent to say?’</td>
</tr>
</tbody>
</table>

Table 1: Example of elicitations under BroadFocus and NarrowFocus.

In every block of elicitation, there were 30 ($=15$ tone combinations $\times 2$ syntactic types) BroadFocus trials and 30 NarrowFocus trials. The trials were presented in a random order. The blocks were separated by five-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 DURATION_RATIO—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$_h$ and V$_h$OBJ) and discourse context (DISCOURSE: BroadFocus and NarrowFocus) on DURATION_RATIO 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 (TONE$_1$ and TONE$_2$). Stimuli (STM) 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 DURATION_RATIO of ModN$_h$ is larger than that of V$_h$OBJ. Consequently, ModN$_h$ and V$_h$OBJ 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$_h$ and V$_h$OBJ phrases.

**Prediction ii:** Under NarrowFocus, durational increase of the nonheads will be greater than their syntactic heads.

Therefore, the DURATION_RATIO of ModN$_h$ will increase from BroadFocus to NarrowFocus, whereas the DURATION_RATIO of V$_h$OBJ will decrease from BroadFocus to NarrowFocus.

4. Results

Globally, there was an effect of TYPE on DURATION_RATIO. The DURATION_RATIO of ModN$_h$ was significantly larger than that of V$_h$OBJ ($t(822) = 4.3767, p < 0.00001$) (Figure 1).

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

Figure 1: DURATION_RATIO of ModN$_h$ and V$_h$OBJ. Globally, the DURATION_RATIO of ModN$_h$ was larger than that of V$_h$OBJ.

Figure 2: DURATION_RATIO of ModN$_h$ and V$_h$OBJ, grouped by DISCOURSE (BroadFocus and NarrowFocus). The DURATION_RATIO difference between ModN$_h$ and V$_h$OBJ was more pronounced under NarrowFocus than under BroadFocus.

Figure 3 shows the DURATION_RATIO 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$_h$ with significantly larger DURATION_RATIO than V$_h$OBJ ($t(167) = 3.6001, p < 0.001$; $t(113) = 2.2586, p < 0.01$). However, the male speaker (M01) did not differentiate between ModN$_h$ and V$_h$OBJ with DURATION_RATIO under NarrowFocus ($t(106) = 0.4983, p > 0.05$). Out of three speakers, only F01 differentiated between ModN$_h$ and V$_h$OBJ with DURATION_RATIO...
under BroadFocus (t(173) = 3.4725, p < 0.01).

Figure 3: DURATION_Ratio of ModN, and V Obj grouped by Speech. Global Type effect was observed across speakers; with DURATION_RATIO, ModN, and V Obj were better differentiated under NarrowFocus than under BroadFocus, though there existed cross-speaker variations.

Figure 4 shows the DURATION_RATIO grouped by tone combination (TONE1 + TONE2). Consistent with the previous results, for the majority of the tone combinations, the global patterns were: 1) the DURATION_RATIO of ModN was larger than that of V Obj; 2) ModN and V Obj were better differentiated under NarrowFocus than under BroadFocus. However, there were also anomalies: ModN, and V Obj were not differentiated under either DISCOURSE condition in terms of DURATION_RATIO (e.g., Tone1+Tone1), and the DURATION_RATIO difference was larger under BroadFocus than under NarrowFocus (e.g., Tone2+Tone3).

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

In Figure 5, DURATION_Ratio was grouped by Type to better examine the DURATION_RATIO change from BroadFocus to NarrowFocus. A two-way ANOVA (factors: Type and DISCOURSE) showed that DURATION_RATIO 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_RATIO decrease (0.056) from BroadFocus to NarrowFocus was marginally significant (p < 0.1), which is consistent with Prediction ii. However, for ModN, the DURATION_RATIO 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_RATIO increase. Consequently, as also observed in Figures 2-4, ModN and V Obj were better differentiated under NarrowFocus: the DURATION_RATIO difference between ModN 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_RATIO of V Obj was 0.06 smaller than that of ModN(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) = -0.024, p = 0.115). Given that ModN 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-effects model, such an interaction effect suggested that the DURATION_RATIO decrease from BroadFocus to NarrowFocus for V Obj was (marginally) significant, whereas for ModN the DURATION_RATIO change was non-significant. This is consistent with Tukey’s HSD post-hoc tests. Also note that when the second syllable (σ2) bore Tone3, the DURATION_RATIO significantly increased by 0.36 (t(13) = 5.237, 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 BROAD:FOCUS to NARROW:FOCUS for MOD:OBJ 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 (σ₁) is already a strong position, NARROW:FOCUS 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 MOD:OBJ, 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 MOD:OBJ and the V₁ of V:OBJ, and between the N₂ of MOD:OBJ 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 MOD:OBJ. It is possible that the majority of MOD:OBJ stimuli in the current study did not exhibit initial stress, therefore disguising the DISCOURSE 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 MOD:OBJ 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 variation 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 MOD:OBJ was larger than that of V:OBJ for M01, the DURATION:RATIO of MOD:OBJ and V:OBJ were not differentiable under either BROAD:FOCUS or NARROW:FOCUS. For some homophoneous pairs, the DURATION:RATIO of MOD:OBJ was larger than DURATION:RATIO of V:OBJ; for others, the DURATION:RATIO of MOD:OBJ was no 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 homophoneous pair of MOD:OBJ and V:OBJ depends heavily on the idiosyncrasies of particular lexical items or individual speakers. The study strongly suggests that the tendency of contrast MOD:OBJ 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. Percept studies are further needed in order to show whether such knowledge of contrast does exist for those homophoneous pairs that do not exhibit overt contrastive phrasal stress patterns in acoustics.

| Fixed effects: | Estimate | df | Pr(>|t|) |
|----------------|----------|----|---------|
| (Intercept)    | 1.17     | 14.5 | 0.0000 *** |
| TYPE:OBJ       | -0.06    | 22.7 | 0.0180 ** |
| DISCOURSE:OBJ  | -0.02    | 798  | 0.3062  |
| TYPE:OBJ       | -0.04    | 798  | 0.1115  |
| DISCOURSE:OBJ  | -0.02    | 798  | 0.3062  |

<table>
<thead>
<tr>
<th>Random effects:</th>
<th>Variance</th>
<th>St.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups</td>
<td>0.0014</td>
<td>0.0378</td>
</tr>
<tr>
<td>SPK</td>
<td>0.0024</td>
<td>0.0495</td>
</tr>
<tr>
<td>Residual</td>
<td>0.0275</td>
<td>0.1660</td>
</tr>
</tbody>
</table>

Number of observations: 833, groups: 30; SPK, 3

Table 2: Results of the mixed model analysis on DURATION:RATIO. Significant factors are shown in bold. Interaction between tones were not shown due to space limit.

The main findings can be summarized as follows: (1) the DURATION:RATIO of MOD:OBJ was larger than that of V:OBJ; (2) the DURATION:RATIO change (decrease) from BROAD:FOCUS to NARROW:FOCUS was significant for V:OBJ, whereas such change was not significant for MOD:OBJ; (3) the DURATION:RATIO difference between MOD:OBJ and V:OBJ was more pronounced under NARROW:FOCUS than under BROAD:FOCUS; (4) there existed cross-speaker and cross-stimulus variations.

5. Discussion & conclusion

The DURATION:RATIO difference between MOD:OBJ and V:OBJ suggested there was a global TYPE effect. Such a difference may arise from one of the following three scenarios: (A) MOD:OBJ stresses the MOD and V:OBJ stresses the OBJ; (B) MOD:OBJ stresses the MOD and V:OBJ has equal stress for both MOD and OBJ; (C) MOD:OBJ has equal stress for both MOD and OBJ and V:OBJ stresses the OBJ.

Focus comes in as a handy diagnostic tool. Finding 2 suggested Scenario (C) was the likely answer. That is, the different behaviors of DURATION:RATIO change in MOD:OBJ and V:OBJ should be mainly attributed to the final stress of V:OBJ. This agrees with the observations in [3] that V:OBJ exhibited final stress whereas MOD:OBJ exhibited no initial stress. Furthermore, such a claim would essentially undermine the validity of NONHEAD:STRESS RULE.

However, rejecting NONHEAD:STRESS RULE as a whole in turn weakens the argument that V:OBJ has final stress, leaving it with no concrete theoretical foundation. Moreover, recall that NONHEAD:STRESS RULE is motivated by the assumption that the information load a constituent carries determines its stress status. This assumption is in line with Finding 3, because the latter shows that under NARROW:FOCUS, the communicative efficiency is facilitated by means of loading more information into the stressed form, i.e., the OBJ of V:OBJ. Therefore, the discrepancy between Scenario (C) (that V:OBJ has final stress and MOD:OBJ has no initial stress) and the information-motivated assumption of NONHEAD:STRESS RULE must be reconciled.
6. References


