Prosodic Realization of Information Structure Categories in Standard Chinese

Yiya Chen & Bettina Braun

Center of Language Studies, Radboud University Nijmegen
Yiya.Chen@let.ru.nl

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Bettina.braun@mpi.nl

Abstract

This paper investigates the prosodic realization of information structure categories in Standard Chinese. A number of proper names with different tonal combinations were elicited as a grammatical subject in five pragmatic contexts. Results show that both duration and F0 range of the tonal realizations were adjusted to signal the information structure categories (i.e., theme vs. rhyme and background vs. focus). Rhemes consistently induced a longer duration and a more expanded F0 range than themes. Focus, compared to background, generally induced lengthening and F0 range expansion (the presence and magnitude of which, however, are dependent on the tonal structure of the proper names). Within the rhyme focus condition, corrective rhyme focus induced more expanded F0 range than normal rhyme focus.

Introduction

There has been a long tradition of research on the proper classification and characterization of information structure divisions (see Kruijff, 2001, for review). Generally speaking, utterances contain elements that advance the discourse and often an element that links it to the discourse. Various terms have been adopted to denote this distinction. Following Firbas (1964) and Steedman (2000), among others, we will label these elements as rhyme and theme, respectively. Recently, some researchers have also argued for two information structure layers (Vallduví & Engdahl, 1996; Steedman, 2000). Steedman, for instance, draws a primary distinction between theme and rhyme and a secondary distinction between focus and background. Focused elements are intonationally marked while backgrounded elements are not. Within such a framework, sentence constituents can therefore be classified into one of the following four categories:

- Theme background
- Theme focus
- Rheme background
- Rheme focus

For clarity, these four categories, together with the contexts that trigger them, are presented in examples (1) – (5). In (1) and (2), the proper name “Mona” is thematic information, once background (1) and once focus (2). In (3) – (5), the proper name is rhematic, either background (3) or focus (4,5), and (4) and (5) differ in the specific type of focus elicited (which we will refer to as normal vs. corrective rhyme focus respectively). Most prosodic research on focus elicits either type. However, Oviatt et. al, 1998, for example, have shown that in intonational languages, corrective focus is marked by longer duration and more expanded F0 range. In a language such as English, the different information categories are said to be realized intonationally, indicated mainly via F0 changes (Steedman, 2000, among others). Assuming that such information structure partitions (as illustrated in 1-5) are universal, it is interesting to study how tone languages, in which F0 changes are employed for lexical contrasts, signal the partition prosodically.

Note that most previous research on the prosodic realization of information structure has concentrated on the division between rhyme focus vs. theme background (e.g., Xu 1999 for Standard Chinese) or theme focus vs. theme background (e.g., Braun, 2005, for German), in the terms adopted here. This study was designed to examine the prosodic realization of all four categories of the information structure in Standard Chinese. Standard Chinese has four lexical tones and a neutral tone (Chao 1968). Previous research on the prosodic realization of information structure in Standard Chinese has shed light mainly on the difference between rhyme focus vs. theme background (e.g., Xu 1999; Chen 2003 for contrastive/corrective focus; cf. Jin 1996). Results show that speakers rely on both lengthening and distinctive tonal implementation with expanded F0 range to differentiate focused rhematic materials from background thematic ones; but it is unclear whether the observed differences are due to the difference between theme vs. rhyme (hereafter Discourse Context difference), background vs. focus (hereafter Contrast difference), or both dimensions interactively.

In this paper, the first question we would like to address is whether Standard Chinese distinguishes the four information structure categories prosodically and if so, how. Our second question is whether different types of focus (normal or corrective), with the same information structure partition, may induce differences in the prosodic realization of the utterance.
This will also make our work more comparable to earlier studies since many phonetic experiments have elicited focus by simply asking for corrections or contrasts.

**Method**

Five different information structures were elicited by means of appropriate context questions. Subjects read answer pairs comparable to those in examples 1 – 5.

1. **Test Materials**

Twelve bi-syllabic proper names were used as targets. Half of the names had a Rising tone followed by a Falling tone and half of them a Falling tone followed by a Rising tone (see Table 1). These test items served as the grammatical subject; they were combined with 12 different verb-phrases (starting with a Falling tone and consisting of six syllables) to avoid boring repetitions, see example (6). The assignment of these assertions to the test items was via a pseudo Latin-square, so that each verb-phrase was associated with five different pragmatic contexts for the same name and with 12 different names for the same pragmatic contexts.

<table>
<thead>
<tr>
<th>1st syllable</th>
<th>2nd syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>/mo4/</td>
<td>/mo2/</td>
</tr>
<tr>
<td>/mo4/</td>
<td>/mo2/</td>
</tr>
<tr>
<td>/mi2/</td>
<td>/mi4/</td>
</tr>
<tr>
<td>/ni2/</td>
<td>/ni4/</td>
</tr>
<tr>
<td>/lo4/</td>
<td>/li2/</td>
</tr>
<tr>
<td>/la4/</td>
<td>/li2/</td>
</tr>
<tr>
<td>/li2/</td>
<td>/na4/</td>
</tr>
<tr>
<td>/li2/</td>
<td>/li4/</td>
</tr>
<tr>
<td>/na4/</td>
<td>/li2/</td>
</tr>
</tbody>
</table>

Table 1: Test Items

(6) zhuàng le yī liàng qìche hit -asp one measure word car 'hit a car'

2. **Subjects and Recording**

Five male and five female subjects were paid to participate in the experiment. All were born and grew up in Beijing. The recordings were done in a sound-treated booth at the Phonetics Lab of the Chinese Academy of Social Science in Beijing. Subjects read question-answer pairs (exemplified in (1-5)) which were presented on a computer screen in Chinese characters. They were explicitly told to produce the answers according to the questions but were naive of the purpose of the study. The test stimuli were randomized into two different orders. Two filler items were added at the beginning and end of the lists respectively. Two repetitions, each with a different randomized order, were recorded, at the sampling rate of 22050 Hz, directly onto a computer. If subjects produced an answer that did not fit the question well, they were asked to repeat this item and to try to make the answer appropriate to the question (this happened in less than 5% of the items).

3. **Acoustic analysis**

Sentences were analyzed both segmentally and prosodically. One speaker was excluded as her F₀-patterns were very different from all other speakers.

For the target words, the onset of the first syllable (C1), the onset of the first vowel (V1), the onset of the second syllable (C2), the onset of the second vowel (V2), and the end of the second syllable (C3) were manually annotated in Praat. As tones are mainly realized over the later half of the syllables, F₀-maxima and minima in the two vowels were labeled (H1, L1, H2, and L2). The following dependent variables were analyzed:

- Duration of the first and second syllable in ms (C2-C1, C3-C2)
- The F₀ range in the two tones in semitones (range1=12*(ln(H1/L1)/ln(2)))

**Results and Discussion**

Univariate analyses of variance were conducted on the duration and F₀ range of the target syllables with Subject as the random factor and two other factors as fixed ones: Pragmatic Context (5 levels which correspond to the five information structures exemplified in (1-5)) and Lexical Tone (2 levels; one with the Rising tone followed by the Falling tone and the other with the Falling tone followed by the Rising tone). Results are summarized in Table II. Context had a significant effect on all dependent variables (F (4, 32) > 20, p < .0001). Tone had a significant effect on all dependent variables except for the F₁ range of the 1st syllable (Duration: 1st syllable [F (1, 8) = 14.17, p < .006]; 2nd syllable [F (1, 8) = 9.26, p < .016]; F₀ range: 1st syllable [F < 2 (n.s.)]; 2nd syllable [F(1, 8) = 44.62, p < .0001]). There was also a significant interaction of Tone and Context on the duration and F₀ range of the 1st syllable (Duration: 1st syllable [F (4, 32) = 3.17, p < .027]; 2nd syllable [F < 3 (n.s.)]; F₀ range: 1st syllable [F (4, 32) = 3.26, p < .024]; 2nd syllable [F < 3 (n.s.)]).

Bonferroni Post-Hoc tests separate for the two tonal patterns were further conducted on the effect of these five discourse contexts. The results were quite variable. The only clear pattern that emerged from the complex results was that rhyme focus (i.e. Contexts 4 and 5), compared to theme background/focus (i.e. Contexts 1 and 2), was consistently manifested with a longer duration and a more expanded F₀ range on both syllables of the proper names.
As mentioned earlier, our main interest lies in the relation between the two dimensions of the information structure partitions (i.e. theme vs. rheme and background vs. focus) and the difference between normal rheme focus vs. corrective rheme focus. Thus, we will first compare the effect of these two dimensions and their possible interaction by analyzing data elicited with pragmatic contexts exemplified in 1-4 (§3.1). We then proceed to report on the differences between the two types of rheme focus by comparing utterances elicited in contexts 4 and 5 (§ 3.2). Because tone had either a main effect or a significant interaction with context, separate analyses for the two tonal combinations were conducted.

4. Prosodic realization of the four information structure categories

Univariate analyses of variance were calculated on the duration and F0 range of the target syllables with Subject as the random factor and two other factors as the fixed ones: Discourse Context with two levels (Theme vs. Rheme) and Contrast with two levels (Focus vs. Background). Results are summarized in Table III and illustrated in Fig. 2-5.

For the proper names that start with a Rising tone followed by a Falling tone (Fig. 2-3), both Discourse Context and Contrast had a significant effect on the duration as well as the F0 range of both syllables. (For the duration of the 1st syllable, Discourse Context: [F (1, 8) = 50.02, p < .0001]; Contrast: [F (1, 8) = 17.22, p < .003]. For the duration of the 2nd syllable, Discourse Context: [F (1, 8) = 26.12, p < .001]; Contrast: [F (1, 8) = 42.78, p < .0001]. For the F0 range of the 1st syllable: Discourse Context: [F (1, 8) = 43.38, p < .0001]; Contrast: [F (1, 8) = 20.69, p < .002]. For the F0 range of the 2nd syllable: Discourse Context: [F (1, 8) = 35.93, p < .0001]; Contrast: [F (1, 8) = 17.90, p < .003].) There was, however, no significant interaction between the two factors (though the interaction of Discourse Context and Contrast approaches significance for the F0 range of the 1st Rising tone [F (1, 8) = 4.96, p < .057]), see Fig. 3a.

For the proper names that start with a Falling tone followed by a Rising tone (Fig. 4-5), Discourse Context had a significant effect on the duration as well as the F0 range of both syllables. Contrast had a significant effect on all but the duration of the 1st syllable. (For the duration of the 1st syllable, Discourse Context: [F (1, 8) = 54.50, p < .0001]; Contrast: [F (1, 8) = 3.21, n.s.]. For the duration of the 2nd syllable, Discourse Context: [F (1, 8) = 50.29, p < .0001]; Contrast: [F (1, 8) = 11.62, p < .009].) For the F0 range of the 1st syllable: Discourse Context: [F (1, 8) = 50.45, p < .0001]; Contrast: [F (1, 8) = 16.73, p < .003]. For the F0 range of the 2nd syllable: Discourse Context: [F (1, 8) = 80.89, p < .0001]; Contrast: [F (1, 8) = 52.20, p < .0001].) A significant interaction between Discourse context and Contrast was found in the F0 range of the Rising tone when it is the 2nd syllable of the proper name ([F (1, 8) = 16.01, p < .004]), see Fig. 5b.

Table II. Results of an ANOVA with Subject as random factor; Pragmatic Context and Lexical tone as fixed factors.

<table>
<thead>
<tr>
<th>Context</th>
<th>Tone</th>
<th>Context*Tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration-S1</td>
<td>p &lt; .0001</td>
<td>p &lt; .006</td>
</tr>
<tr>
<td>Duration-S2</td>
<td>p &lt; .0001</td>
<td>p &lt; .016</td>
</tr>
<tr>
<td>F0 range-S1</td>
<td>p &lt; .0001</td>
<td>n.s.</td>
</tr>
<tr>
<td>F0 range-S2</td>
<td>p &lt; .0001</td>
<td>p &lt; .006</td>
</tr>
</tbody>
</table>

Results showed clearly that theme and rheme were marked differently in terms of both duration and F0 range. A rhematic name was realized with a longer duration (by 17ms) and a more expanded F0 range (1.1st) than a thematic one. Focused elements were realized with a longer duration (7ms) and a more expanded F0 range (0.8st) than backgrounded elements (except for the duration of the 1st syllable in the names with a Falling tone followed by a Rising tone).
While both duration and $F_0$ were employed by the speakers to indicate the information structure partitions, they were not necessarily concurrent. The durational marking of theme vs. rhyme exhibited a quite consistent pattern (with comparable magnitude) across focus and background conditions (Fig. 2a-b & 4a-b), but the $F_0$ range marking seemed to be sensitive to the property of the lexical tones. The Rising tone exhibited a more pronounced difference between rhyme and theme in the focus condition compared to the background condition (Fig. 3a vs. 5b). Or, to phrase it differently, the focus-background difference was marked more saliently in the rhyme condition than the theme condition. The Falling tone, however, showed no such an effect (Fig. 3b vs. 5a). Note that when following a Rising tone (Fig. 3b), the Falling tone exhibited a much more expanded $F_0$ range than when it preceded a Rising tone (Fig. 5a). This is probably related to the tonal coarticulation effect, in particular the delayed alignment of the $F_0$ peak of a Rising tone relative to the edge of the tone-carrying syllable (Xu 2001). The expanded $F_0$ range of the Falling tone in Fig. 3b could be in part due to the preceding Rising tone.

5. Normal rhyme focus versus corrective rhyme focus

Here, we compare the realizations of normal rhyme focus (4) and corrective rhyme focus (5). Paired t-tests for the 108 tokens were calculated. While there was no difference in duration between the two conditions, the $F_0$ adjustments differ, but show two patterns for the two tonal combinations. For proper names with a Falling tone followed by a Rising tone, the $F_0$ range for corrective rhyme focus was larger in both syllables compared to normal rhyme focus (mean $F_0$ range in the 1st syllable: 4.81st for normal rhyme focus; 5.74st for corrective rhyme focus [$t = 4.9, p < 0.0001$]; mean $F_0$ range for the 2nd syllable: 5.31st for normal rhyme focus; 6.33st for corrective rhyme focus [$t = 5.5, p < 0.0001$]).

For proper names with a Rising tone followed by a Falling tone, there were no significant differences in $F_0$ range in either syllable. However, a significant difference in the $F_0$ maximum of the 2nd syllable was found: it was higher for corrective rhyme focus (248.41Hz) than normal rhyme focus (235.41Hz) [$t = 4.3, p < 0.0001$]. This might also be due to a tonal coarticulation effect. The raised $F_0$ peak of the Falling tone here was in part caused by the delayed $F_0$ peak of the Rising tone. Corrective rhyme focus hence differed from normal rhyme focus, but only in terms of $F_0$ adjustment.

Conclusions

We investigated the prosodic realizations of information structure categories in Standard Chinese. With regard to the difference between normal and corrective rhyme focus, we observed that corrective rhyme focus is marked more prominently than normal rhyme focus, but only in terms of $F_0$. With regard to the difference between the four information structure categories, it was found that both syllable duration and $F_0$ range of the tones were adjusted. More specifically, rhematic names were consistently realized with a longer duration and a more expanded $F_0$ range than thematic names. Focused names in general also showed a longer duration and a more expanded $F_0$ range than backgrounded names, but their effects are smaller than that between theme and rhyme. The data thus suggest that the division between theme and rhyme is marked more distinctively than that between focus and background. This is consistent with the assumption that there are two layers of information structure, and that the theme-rheme distinction is more important than the focus-background distinction.

Acknowledgements

We would like to thank Qiang Fang, Yuan Jia, Li Shifu, and in particular, Aijun Li for their help with the data collection. We are grateful to Toni Rietveld for his advice on the statistics. Usual disclaimers apply here, of course. This work was supported by the VENI grant from the Netherlands Organization for Scientific Research (NWO) to Yiya Chen.

References