Mandarin Chinese sentence intonation patterns in the production of Hungarian learners of Chinese

Kornélia Juhász, Huba Bartos

Hungarian Research Centre for Linguistics
Eötvös Loránd University, Hungary
juhasz.kornelia@nytud.hu, bartos@nytud.hu

Abstract

In this acoustic analysis, we aim to examine how Hungarian learners of Mandarin Chinese (MC) contrast and distinguish the intonation patterns of unmarked and ma-particle marked yes-no questions vs. statements. Since Chinese is a tonal language, the issue of intonation is more complicated than in non-tonal languages, such as Hungarian. L1 patterns for question intonation differ from the L2, and we expect learners not to produce MC questions with the elevated characteristic of f0 (compared to statements), due to L1 transfer. Moreover, we expect Hungarians to produce less variability in f0 range than natives. To test these assumptions, first we extracted a grid of f0 realization, i.e., the highest peaks and lowest troughs of each syllable, attaining the spine of the f0 structure. Then we analyzed f0 range, as the difference between maximal and minimal f0. Our results showed that Hungarians produce MC questions without the elevated f0 characteristic of natives’ production, and the expected f0 range differences were partially confirmed.

Index Terms: Mandarin Chinese, intonation, f0 grid, L2 production

1. Introduction

In tonal languages, such as Mandarin Chinese (MC), f0 serves for both the realization of lexical tones and intonation as well [1]. The pattern of each tone can be modulated and realized in different f0 registers, being dependent on sentence modality, and the patterns are further influenced by segmental effects, tonal coarticulation, etc. [1,2,3]. In distinguishing MC statement and question intonation patterns, local (e.g. terminal rise on the last (tonal) syllable) and global acoustic cues (raised f0 over the whole utterance) have been identified concerning f0 register and f0 range [3]. Regarding f0 register, according to Shen’s MC intonation model [2], statements display a gradually descending pattern, while yes-no questions (unmarked or marked) feature a significantly higher f0 throughout the whole utterance (compared to the declarative contour), moreover with a terminal rise.

Intonation patterns, f0, and f0 range variation can also be displayed and framed by two curves indexing highest peaks (top line) and lowest troughs (base line) of the intonational contour [3,4,5]. According to [2], MC yes/no questions do not feature a widened f0 range, because both the top line and the base line are raised in frequency, which means a change in f0 register, rather than in range. However in the case of unmarked questions (UMQ), a terminal rise is localized on the last syllable (e.g. [6,8]), accompanied by significant f0 range expansion there [3,7]. In marked questions (MQ), according to [8], the f0 rise can be observed on the last full-toned syllable. This is because in marked questions, the f0 of the particle ma, lacking lexical tone specification, is defined by the preceding full-toned syllable [6]. In the L2 acquisition process, prosodic patterns, e.g., intonation patterns, are often transferred from L1 [9]. Lexical tone and intonation interaction may pose problems for Hungarians, whose L1 is atonal. Similarly to most languages, in both MC and Hungarian the f0 gradually declines throughout declarative utterances [2,10,11]. However in contrast to MC, Hungarian f0 contours differ in the prosodic structure of yes/no questions: In Hungarian, the initial f0 value is low (compared to declaratives), and the f0 contour features a rising structure peaking on the penultimate syllable (unless fewer than two syllables follow the final phrasal stress), followed by a fall. [10,11]. In sum, in this analysis, we aim to get a scheme of L2 learners’ Q-intonation in MC, more particularly how Hungarian learners of Chinese contrast question and statement intonation patterns. To be specific, by intonation pattern we mean the above-mentioned grid of f0 realization, comprising the highest peaks and lowest troughs of each syllable, thus attaining the “spine” of the f0 structure, without the analysis of independent tone patterns. Regarding yes/no question types in MC, here we analyse the patterns of both UMQs and MQs in comparison with statement intonation patterns.

2. Hypotheses & Method

2.1. Hypotheses

Due to L1 prosodic transfer, MC learners produce both UMQ and MQ MC questions with L1 prosodic patterns, thus without the elevated f0 characteristic featured by natives (compared to the declarative pattern). Thus, we assume that in Hungarians’ production, question and statement intonations are discriminated by a rise in f0 only in the second half of the utterance. Furthermore, Hungarian production places a peak on the penultimate syllable. Regarding f0 range, we hypothesize that Hungarians give priority to the proper production of tonal patterns, thus the given tonal f0 range is expected to be maintained for the whole utterance, and not expected to be widened terminally (as expected in the case of UMQs of natives).

2.2. Method

We analysed three speaker groups (5 women per group): 1. Hungarian 2nd year undergraduate learners of MC (undergrads); 2. Chinese studies master’s degree students (grads), and 3. Chinese natives. Regarding the material, we compared (i) declarative, (ii) syntactically unmarked and (iii) syntactically marked ma-particle SVO structured sentences (matching in syllable structure and the number of voiced
segments). Each syllable in a sentence has the same tonal value (either T2 (rising) or T4 (falling) tone). Three interrogative and three declarative sentences for each tone were read three times, presented as short question-answer dialogues, projected on a screen, with both Chinese characters and pinyin (Table 1). For UMQ & statement dialogue pairs we recorded altogether 3600 syllables (2 tones × 3 sentences × 5 repetitions × 2 modality × 15 speaker). For MQ & statement pairs, we recorded 2250 syllables in question intonation and 1800 syllables in statement intonation.

Table 1: The UMQs, MQs and Statements that were read in short question-answer dialogues

<table>
<thead>
<tr>
<th>Q/A</th>
<th>Tone 2</th>
<th>Tone 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMQ</td>
<td>Wang Jue mi tx chun?(max)?</td>
<td>Wang Jue mi tx chun?(max)?</td>
</tr>
<tr>
<td>MQ</td>
<td>Doesn’t Wang Jue have a boat?</td>
<td>Does Lang Que sell bracelets?</td>
</tr>
<tr>
<td>Statement</td>
<td>Wang Jue does not have a boat.</td>
<td>Lang Que sells bracelets.</td>
</tr>
<tr>
<td>UMQ</td>
<td>Wang Jue mi tx san?(max)?</td>
<td>Wang Jue mi tx san?(max)?</td>
</tr>
<tr>
<td>MQ</td>
<td>Doesn’t Wang Jue have a land?</td>
<td>Does Lang Que sell electricity?</td>
</tr>
<tr>
<td>Statement</td>
<td>Wang Jue does not have a land.</td>
<td>Lang Que sells electricity.</td>
</tr>
<tr>
<td>UMQ</td>
<td>Wang Jue mi tx tan?(max)?</td>
<td>Wang Jue mi tx tan?(max)?</td>
</tr>
<tr>
<td>MQ</td>
<td>Doesn’t Wang Jue have a spring?</td>
<td>Does Lang Que sell tickets?</td>
</tr>
<tr>
<td>Statement</td>
<td>Wang Jue does not have a spring.</td>
<td>Lang Que sells tickets.</td>
</tr>
</tbody>
</table>

Segmenting and f0 extraction were carried out in Praat [12], the minimum and maximum values of f0 were extracted automatically for each syllable, with the standard settings of To Pitch command. Minimum f0 was set to 100 Hz. Before the statistical analysis, f0 values were converted into semitones with 50 Hz as reference value [13], with Hymisc package in R [14]. In the statistical analysis linear mixed effects models were used [15]. Data was subsetted to analyse different tones and unmarked and marked Q-As separately. In this manner for UMQ-A pairs 3 models were built, where f0min, f0max, f0 range were the independent variables, and syllable number, modality (Q vs. A) and speaker group were the independent variables. The models were further complemented by a random variable, which was the speaker. Speaking of MQs, the 4th syllables of MQ answers were excluded from the analysis due to the high proportion (38%) of irregular voicing. Thus, one model was built for comparing only Q syllables, and another one for comparing the first 3 syllables of Q and A. In models for analysing f0 range, syll. no. and modality were combined to a single variable coding both features. F0max and f0min were only compared within groups, however f0 range was compared across adjacent syllables and across groups as well.

3. Results

In the case of f0max in T2 and T4 unmarked Q&A sequences, the mixed model showed a significant interaction effect between syll. no., modality and group (T2: F(6,1765) = 2.6, p < 0.01; T4: F(6,1784) = 4.4, p < 0.001). Native speakers produced both T2 and T4 Qs with significantly higher f0 compared to As (p < 0.001) and a rising pattern could be observed in this case: the last syllable always had significantly higher f0max relative to the 1<sup>st</sup> syllable of the utterance (Fig. 1. top). Native speakers’ As, in contrast, showed a significant downstep on the last syllable of T2, and a tendency to lower f0max in T4 sequences. Compared to these patterns, neither Hungarian speaker group distinguished between Qs and As in f0max, except for the last syllable, where undergrad produced Qs with significantly higher f0max than As. Grads also produced the same pattern for T2 sequences, however in T4 sequences they did not contrast Qs and As on either syllable. But it should be noted that although we found no statistically confirmed difference, Qs nevertheless had a tendency to have a higher f0max mean than As.

![Figure 1: Maximum f0 (top) and minimum f0 values (bottom) compared for unmarked Qs and As (means&SD)](image-url)

The model for f0min of the T2 Q&A sequences showed a significant interaction effect between the variables in each pairing (modality ~ syll. no.: F(3,1765) = 7, p < 0.001; modality ~ group: F(3,1765) = 23, p < 0.001; syll. no. ~ group: F(3,1765) = 11, p < 0.001). In contrast, in T4 sequences there was a significant interaction effect for all three independent variables (F(6,1784) = 4, p < 0.001). Here, too, native speakers discriminated significantly between Q and A f0mins of each syllable (p < 0.001) (Fig. 1., bottom). T2 Q sequences featured a terminal lowering compared to the 1<sup>st</sup> and 3<sup>rd</sup> syllables, (p < 0.001), while T4 sequences were produced with a terminal rise on the last syllable, again, compared to the 1<sup>st</sup> and 3<sup>rd</sup> syllable. (p < 0.001). In addition to this, As for both T2 and T4 sequences were characterized by a descending pattern. As opposed to native production, neither L2 group distinguished between Qs and As in f0min, except in the last syllable: both groups produced T2 Qs with a terminal rise of f0min on the last syllable. However, in T4 sequences, only undergrads produced significantly higher f0min for Qs than As, and grads tended to apply the same pattern. Both L2
groups tended to produce a descending pattern for both Qs and As in the first three syllables. We should further note that regardless of tone and syllable position, As had a tendency to have lower f0max and f0min means than Qs.

Regarding the f0 range of UMQs, for both T2 and T4 the mixed model showed a significant interaction of group and the combined variable of modality and syll. no. (T2: F(14,1765) = 8, p < 0.001; T4: F(14,1765) = 2, p < 0.05). In natives’ T2 sequences, the last syllable had a significantly wider f0 range than any other syllables of Qs (p < 0.001), and we found almost the same pattern for T2 As, but without any difference between the 3rd and the 4th syllables. Comparing natives’ and L2 learners’ production of the f0 range of the last syllable, L2 learners produced significantly narrower f0 range than natives (p < 0.05). Regarding the f0 range of T2 and T4 in L2 learners’ production, we found differences neither across syllables within groups, nor across groups, with one exception: in T2 Qs of undergrads, the last syllable was produced with a significantly narrower f0 range than the 1st and 2nd syllables (Fig. 2.).

![Figure 2: T2 and T4 sequences’ f0 ranges for each syllable (1-4) of unmarked Qs and As in the production of the three speaker groups](image)

For f0max in T2 and T4 marked Q sequences, the mixed model (which only contained the 5 syllables of Qs) showed a significant interaction effect between syll. no. and group (T2: F(8,1095) = 23, p < 0.001; T4: F(8,1071) = 16, p < 0.001). Natives’ T2 Qs featured a zig-zag pattern with significant differences between adjacent syllables and ended with a terminal rise on the last syllable (Fig. 3., top). In contrast T4 Qs were characterized by a constant f0max on the first 3 syllables, followed by a significantly higher f0max on the penultimate, and a terminal downstep on the last syllable (p < 0.001). The production of both Hungarian groups differed from natives, but a certain pattern was observed: Graduates produced the toneless ma particle with a significantly lower f0max mean than the 1st, or even the penultimate, syllable and undergrads showed the same trends. Finally, both groups of learners produced the penultimate syllable with a higher f0max mean compared to the adjacent values.

For f0min in T2 and T4 marked Q sequences, the mixed model (which only contained the 5 syllables of Qs) showed a significant interaction effect between syll. no. and group (T2: F(8,1095) = 14, p < 0.001; T4: F(8,1071) = 7, p < 0.001). Identically to f0max patterns, natives produced the ultimate syllable depending on tonal context: with higher f0min in T2 sequences, and lower f0min in T4 ones, compared to the penultimate syllable. In L2 learners’ production the same trend was observed as in the case of f0max, that is, the penultimate syllable peaked relatively to the adjacent 3rd and 5th syllables, regardless of tone (Fig. 3.).

![Figure 3: Maximum f0 (top) and minimum f0 values (bottom) compared for marked Qs and As (means&SD)](image)
Qs show remarkable similarities to unmarked Qs, that is, natives contrasted Qs and As in f0min in both tone sequences, however neither group of L2 learners did so. In any case, there was a tendency for both graduates and undergraduates to produce As with lower mean than that of Qs at each syllable, however neither group differentiated between them significantly.

Results concerning the f0 range of marked Q’s and A’s, the mixed model showed an interaction of the combined variable modality and syll.no. and group in the case of both tones (T2: F(14,1785) = 15, p < 0.001; T4: F(14,1723) = 6, p < 0.001). Natives produced T2 Qs with significant variability between syllables, and f0 range expanded towards the last syllable (Fig. 4.). Also, the f0 range of natives’ final syllables (the toneless ma particle) was significantly wider, than those produced by both groups of L2 learners. In T4 Qs by natives, the same f0 range was maintained for the whole utterance, in contrast, graduate students produced the last syllable with a narrower f0 range compared to the other syllables.

![Figure 4: T2 and T4 sequences’ f0 ranges for each syllable (1-4) of marked Qs and As in the production of the three speaker groups](image)

4. Discussion & Conclusion

In this acoustic analysis we extracted and compared grids of f0 realization of unmarked and marked MC questions in the speech of two groups of Hungarian learners of MC and one group of native speakers. Hungarians produce MC questions without the elevated f0 characteristic of natives’ production both in UMQs and MQs. In UMQs, Hungarians produced an interrogative f0 pattern identical to statements, but accompanied by a terminal rise. This pattern resembles Hungarian 1- or 2-syllable Qs, where a topic-like part with a slowly descending structure precedes the question intonation tone. In case of one-syllable Hungarian Qs, there is a terminal rising pattern, while two-syllable Q-s are characterized by a rise-fall on the last syllable ([16]: pp. 290–291.). This is one possible explanation for L2 learners’ UMQ pattern in MC, however, further investigation is required to confirm the concrete effect of L1 on production. As regards the f0 range of UMQs, our results correspond to what [1] found: significant f0 range expansion in T2 sequences in question intonation, but the f0 range remained intact in the case of T4. In contrast with our initial hypothesis, undergrads did in fact produce f0 range expansion on the last syllable, but they did not approximate natives’ production, while grats did produce all the syllables with the same f0 range, as we had expected. These results led us to conclude that our hypothesis is partially confirmed in T2 examples, but in contrast to our expectation, surprisingly, it was undergrads who approximated natives’ production better, not grats. A possible reason for this is that tone production becomes more rigid over the years of language learning, and more resistant to the modifying effect of intonation.

Regarding the production of MQs, L2 learners produce Qs and As identically, with a gradually descending pattern (similarly to UMQs), thus differing from natives’ production. Furthermore, learners tend to raise the penultimate syllable regardless of tone, and in this manner they produce patterns distinct from that of natives’ T2 sequences. In native production the f0 value of the toneless particle is tone-dependent, thus T2 requires a relatively high, and T4 a relatively low, f0 for the particle ([17]: pp. 90). Hungarians’ production, with the rise on the penultimate syllable, is overly reminiscent of the L1 question-intonation pattern, however this pattern may also be rooted in the learners’ unawareness of the tone-dependent nature of the realization of the toneless particle, thus L2 learners use the same pattern (that is available for T4) for T2 as well. Concerning f0 ranges of MQs, there is a clear pattern of natives expanding the range towards the end of T2 utterances, thus the particle had an extremely wide-ranged realization. One possible explanation for this is that the f0 of particle is realized as a prolongation of the previous (rising) tone, and due to the sonorant initial in the particle, the whole syllable is analysed, allowing a continuous rise in the contour. In the learners’ production the diametrically opposite pattern was observed: learners produced the particle with a relatively narrow range both for T2 and T4, which may also possibly stem from L2 learners producing the particle with the same pattern regardless of the preceding tone. In sum, our results confirmed the hypothesis that Hungarians encounter difficulties discriminating MC questions from statements, because L2 learners’ production lacks the elevated f0 characteristic of MC questions. The analysis will be extended to f0 curves of syllables in future work. The results of this study contribute to the understanding of the intonation acquisition of a tonal L2 language.

5. Acknowledgements

The authors are grateful to Katalin Mády, Tekla Etelka Gráczi, Andrea Deme and Alexandra Markó for their kind suggestions and support. The research has been supported by ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology, using resources of the Hungarian National Research, Development and Innovation Fund.

6. References


