MANDARIN QUESTION INTONATION PATTERNS IN THE PRODUCTION OF HUNGARIAN LEARNERS OF CHINESE

Kornélia Juhász & Huba Bartos

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

ABSTRACT

In this acoustic analysis we aim to compare Mandarin Chinese (MC) syntactically marked (trisyllabic) and unmarked (disyllabic) interrogative intonation patterns contrasted with their declarative counterparts in the production of two L2 learners' groups with different level of language experience whose L1 is atonal Hungarian. We hypothesize that the synchronization of tones and intonation in MC poses problems for L2 learners and when the analysed L2 tonal patterns sharply differ from the native L1 pattern, L2 learners favour the L1 pattern. We recorded short dialogues of interrogative and declarative sentences, and the extracted f_0 contours were compared by GAMMs. Our results show that L2 tone sequences containing T1 & T2 pose more difficulties for L2 learners than those with T1 & T4. Our hypothesis regarding the presupposed difficulties of effects of L1 transfer has been partially confirmed, although L2 learner groups with different levels of L2 experience differed in producing L2 interrogative patterns.

Keywords: Mandarin Chinese, tonal and atonal languages, L2 intonation production

1. INTRODUCTION

In this acoustic analysis, we aim to explore how Hungarian learners contrast Mandarin Chinese syntactically marked and unmarked yes-no interrogative intonation patterns to declaratives. In tonal languages, such as Mandarin Chinese (MC), f_0 serves for the realization of both lexical tones and intonation [1]. This means that, on one hand, f_0 modulation is locally dependent primarily on tone values or tonal contexts. On the other hand, these local effects also interact with intonation patterns, yielding the actual f_0 contour [2]. Taking a broader view, in distinguishing MC statement and question intonation patterns, local (e.g., terminal rise on the last (tonal) syllable) and global acoustic cues (raised f_0 over the whole utterance) have been identified concerning f_0 register and f_0 range [3]. Regarding f_0 register, according to Shen's MC intonation model [1], statements display a gradually descending pattern, while unmarked yes-no questions feature a significantly higher f_0 throughout the whole utterance (compared to the declarative contour), complemented by a terminal rise. This terminal rise, in the case of syntactically unmarked ('bare') interrogatives, can be attributed to the absence of lexical/syntactic cues (such as the particle *ma*) of interrogative force, thus prosodic cues are used exclusively to express this force [3, 4, 5]. In the case of syntactically marked ('particle-final') interrogation, an additional particle, the particle *ma*, verbalizes interrogative mood. Lacking an inherent tonal value, *ma* is realized tonally as the prolongation of the contour of the preceding tone: following a rising Tone 2 (T2), the toneless particle has a high target, while following a falling Tone 4 (T4) it has a low target [1, 3].

In contrast to MC, Hungarian is a non-tonal language. While in Hungarian the declarative pattern is realized with a descending contour similar to MC, the prosodic structure of character contour in yes/no questions differs in MC and Hungarian: In the latter, the f_0 contour is characterized by a rising structure followed by a fall (L*HL) [6, 7, 8]. However this character contour only appears at the terminal part of the intonation phase, more precisely, it initiates on the last accented syllable of the utterance [6]. Thus we can define three relevant positional alternants for the realization of yes/no questions [8]: (i) if the final stress group contains only one syllable, then truncation of the final falling phase of the character contour occurs, in this manner the interrogation is realized exclusively with a rising phase (L*H); (ii) If the final stress group contains two syllables, then L* appears on the 1st syllable, and the falling HL is displayed by the 2nd syllable; (iii) if the final stress group contains at least 3 syllables then each target of the character contour has its own syllable to convey. It should be further added that prior to the terminal rising-falling contour, a scale and/or preparatory contour appears, most frequently realized as a half-fall, that is, a slightly decreasing pattern [6: 52].

In this particular case, we are analysing disyllabic broad focus MC utterances, where the first syllable is the subject (他 *tā 'he'* (with high level Tone 1 (T1))), followed by the verb (来 *lái, 'come'* (rising T2) or 去 *qù 'go'* (falling T4)) and in the case of syntactically marked questions, there is a subsequent syllable, the particle 吗 *ma*. We primarily focus on the verb, because in our examples the verb occurs as the last accented syllable, which is the locus of the terminal rise in bare MC questions, as well as the 'host' defining the pitch value of the particle *ma* in particle-final questions, and in L1 (Hungarian) the last accented syllable is precisely where the character contour is initiated, too.

Additionally, the analysis of the f_0-contour of the 1^{st} syllable is also important, considering the fact that MC elevates the f_0 contour of the whole utterance, while Hungarian does not.

2. HYPOTHESIS

The synchronization of tone and intonation in the production of L2 learners is in the centre of our attention. We chose to use T2 and T4 verbs in our examples because if interrogative and declarative utterances are contrasted, then the interference between the L1 and L2 f_0 curves is expected to be the strongest in these cases. The presupposed interference is induced by the striking disparity of L1 and L2 interrogative and declarative contours with a falling-tone (T4) tonic in questions, and a rising-tone (T2) tonic in statements, and can be attributed to the L1 transfer effect in the L2 acquisition process [9, 10]. In sum we hypothesize that in those cases where the L1 and L2 f_0-patterns differ, L2 learners face difficulties in production, favouring the L1 pattern. The cases, and the presupposed difficulties, are summarized in Table 1. The L1 examples are not analysed in this experiment, just added here as a reference.

	3-syll. syntactically marked interrogation	**2-syll. 'bare' interrogation**	**2-syll. declaration**
L2 T1+T2	Tā lái ma? [tʰa laj ma] H.LH.H *'He comes PRT?'* elevated rising	Tā lái? [tʰa laj] H.LH *'He comes?'* elevated rising	Tā lái. [tʰa laj] H.LH *'He comes.'* rising
L2 T1+T4	Tā qù ma? [tʰa tɕʰy ma] H.HL.L *'He goes PRT?'* elevated falling	Tā qù? [tʰa tɕʰy] H.HL *'He goes?'* elevated falling	Tā qù. [tʰa tɕʰy] H.HL *'He goes.'* falling
L1 pattern	És | jön ma? prep. | L*.HL *'And come-3sg today?'* prep. + rise-fall	És | jön? prep. | L*H *'And come-3sg?'* prep. + rise	És | jön. prep. | HL *'And come-3sg.'* prep.+ fall
Presupposed difficulty	- *ma* particle with H target after *lai,* - elevated character	- producing *qu* with a fall - elevated character	- producing *lai* with a rising pattern

Table 1: The analysed L2 utterances, compared to the exemplary L1 patterns, and the summary of the presupposed difficulties

3. METHOD

We analysed three adult speaker groups (5 female speakers per group): 1. Hungarians with cca. one year language experience of MC: second year Chinese studies undergraduates ('beginners'); 2. Hungarians with 3-4 years of learning MC: Chinese Studies master's programme students ('advanced learners'), and 3. a control group of Chinese natives. We recorded the utterances summarized Table 1., presented as short question–answer dialogues, projected on a screen with both Chinese characters and pinyin transcription. Both 'particle-marked Q + declarative A', and 'declarative + bare echo-Q' utterance-pairs were repeated 5 times. In this manner we recorded a total of 2×150 = 300 utterance pairs, where 'declarative + 2-syll echo-Q' pairs added up to 600 syllables, and pairs with 3-syll 'particle-marked Q + A' added up to 750 syllables (2 sentence types × 2 tones × 5 reps × 15 speakers). Within voiced segments, f_0 was extracted by 5 ms intervals automatically in Praat [11]; since the the onsets of the 2^{nd} syllables of T4-tonic utterances were voiceless segments, they were excluded from the analysis. The extracted f_0 values were converted to semitones (with a reference value of 50 Hz [12] in R [13]) and f_0 curves were analysed by GAMMs [14], using the packages *mgcv* [15] and *itsadug* [16]. We ran 8 models in total: in the case of 'T1+T2' utterances the first two syllables were analysed together since they consisted exclusively of sonorants (after the initial aspirated [tʰ]). Since 'T1+T4' sequences contained a voiceless onset [tɕʰ], we had to analyse the first two syllables separately. The 3^{rd} syllable of interrogatives (the particle *ma*) was analysed in an additional model as well, because it had no counterpart in declaratives. As regards the structure of the models, in each case, f_0 was analysed dependent on the normalized duration of the vocalic section of the syllable, and the models were further complemented by a parametric factor (with contrast treatment), which we defined as an ordered combined variable of speaker group and sentence type. Within the ordered variable, native declaration (Nat D) was chosen as the reference curve, from which difference curves were computed. Additionally, a random smooth function was applied to each f_0-trajectory, respectively. The models were treated for autocorrelation [17].

4. RESULTS

Our results with the estimated f_0 curves are shown in Figure 1., and the parametric coefficients of the GAMMs are summarized in Table 2. Regarding the pairing of a particle-marked 'T1+T2+*ma*' interrogative (I) and declarative answers (D), native speakers produced these questions with an estimated 3 semitones higher f_0 compared to Ds, and thus they produced a significant difference up until the last 12% of the 2^{nd} syllable, from where the two f_0 curves overlapped. As expected, the rising pattern of the 2^{nd} syllable was extended to the 3^{rd} syllable (the *ma*

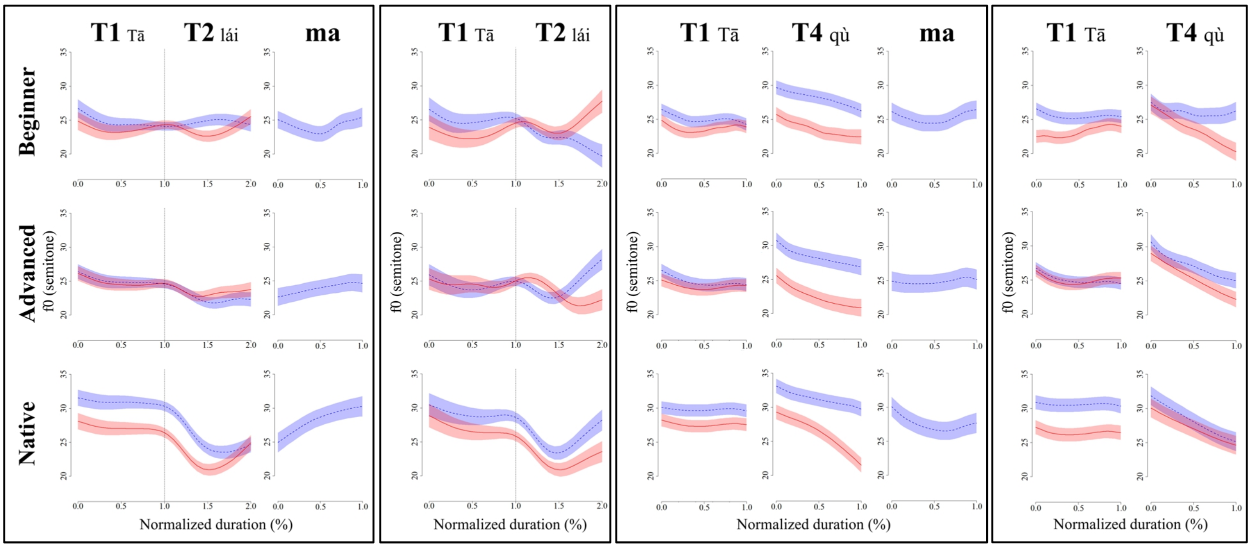

Figure 1: The estimated f_0 curves of the four groups of tri- and disyllabic interrogative (blue dashed curve) + declarative (red solid curve) pairs within the normalized duration of each vocalic section (95% conf. interval)

Synt. Marked T1+T2+*ma* I + D				
Para-metric	TA LAI		MA	
	Est.	*t*	*Est.*	*t*
D_Nat.	24.4	70.7	-	-
I_Nat.	27.6	20.0	28.2	43.8
D_Adv.	23.9	-1.0	-	-
I_Adv.	23.6	-1.8	23.9	-5.0
D_Beg.	23.7	-1.6	-	-
I_Beg.	24.6	0.4	24.1	-4.8
R^2	96.0%		99.6%	

Bare T1+T2 I + D		
Para-metric	TA LAI	
	Est.	*t*
D_Nat	24.1	68.6
I_Nat	26.8	19.1
D_Adv	23.8	-0.6
I_Adv	24.4	0.7
D_Beg	24.0	-0.2
I_Beg	23.3	-1.5
R^2	94.5%	

Synt. Marked T1+T4+*ma* I + D						
Para-metric	TA		QU		MA	
	Est.	*t*	*Est.*	*t*	*Est.*	*t*
D_Nat	27.6	61.2	26.2	51.4	-	-
I_Nat	29.8	12.8	31.3	40.9	27.5	41.6
D_Adv	24.1	-5.4	22.5	-5.0	-	-
I_Adv	24.8	-4.3	28.2	2.8	24.9	-2.9
D_Beg	23.6	-6.4	23.6	-3.7	-	-
I_Beg	25.0	-4.1	28.1	2.8	25.2	-2.6
R^2	97.8%		98.7%		99.8%	

Bare T1+T4 I + D				
Para-metric	TA		QU	
	Est.	*t*	*Est.*	*t*
D_Nat	26.4	53.4	27.2	41.9
I_Nat	30.6	18.1	28.3	3.5
D_Adv	24.9	-2.4	25.5	-2.0
I_Adv	25.0	-2.1	27.3	0.1
D_Beg	23.1	-4.9	23.7	-4.0
I_Beg	25.5	-1.5	26.0	-1.4
R^2	97.3%		96.2%	

Table 2: The 8 GAM-models' parametric coefficients (estimated average f_0 *(Est.)*, *t*-value) for the two sentence types (interrogative = I, declarative = D) produced by the three speaker groups (Native = Nat, Advanced = Adv, Beginners = Beg), (*t*-value significant if $|t| \geq 1.96$)

particle), which was thus realized in a relatively high f_0 range. In contrast, both L2 learner groups produced more overlapping f_0 curves in this case. Neither group discriminated between the 1st syllable of Is and Ds to a significant extent. Advanced learners produced the 2nd syllable of Ds with a significantly higher f_0 compared to the 2nd syllable of Is (sign. diff: 140–200% of normalized duration). Beginners produced reverse curve patterns for the two moods: domed for Is and concave for Ds (sign. diff: 28–85% of the 2nd syllable). The particle *ma* in the 3rd syllable was produced by L2 learners in a lower f_0 range compared to natives, with a significantly different shape as well.

In unmarked 2-syllable 'T1+T2' I and D pairs, in natives' production the two sentence types were significantly discriminated for the whole normalized duration with an average difference of 2.7 semitones (st), and with Ds positioned in a lower f_0 range than Is. In contrast, both L2 groups produced intersecting I and D curves, and the estimated average difference throughout the two curves was less than 1 st. Advanced learners made no distinction between the two sentence types in the 1st syllable. Additionally, in interrogatives, T2 in the 2nd syllable featured the native-like rising, however Ds showed a strikingly opposite, falling, pattern. In beginners' production, the 1st syllable of Ds was realized in a significantly lower f_0 range compared to Is, however the difference resulted from the gradually rising pattern of declarative T1, meaning that the high level quality of the T1 was missing. In contrast, the T1 in interrogatives was realized with a level pattern, while the rise of T2 in the subsequent syllable was replaced by a falling pattern.

In pairs of particle-marked 'T1+T4+*ma*' I and D, natives produced Is with significantly higher f_0 compared to Ds for the whole duration of the 1st syllable, with an average difference of 2.2 st. In contrast, L2 groups produced less difference, on average, between the two sentence types in the 1st syllable (beginners: ~1.4 st, advanced learners: < 1 st). Discriminating the falling T4 contours of the two types in the 2nd syllable did not pose difficulties for neither L2 learner groups: similarly to natives (who produced an average of 5.1 st difference), advanced learners featured a difference of 5.7 st, and beginners a difference of 4.5 st on average, between the two sentence types. As for the shape of tonic T4 in declaratives, natives produced a slightly domed falling pattern, but both L2 learner groups showed a concave descending pattern (significantly differing from the native shape), positioned to a lower f_0 range compared to those of natives. The contour of the particle *ma* – here, too – was realized as a prolongation of the preceding falling pattern, however in this case L2 learners positioned the f_0 curve in a lower f_0 range compared to natives.

In disyllabic 'T1+T4' utterances, native speakers significantly discriminated (bare) Is and Ds for the whole duration of the 1st syllable, producing Is with an average 4.2 st higher than Ds. In a similar manner, beginners also distinguished the curves of the two sentence types on the 1st syllable, producing Is an estimated 2 st higher compared to Ds. In the advanced learners' production, the curves for two sentence types intersect in this phase, as Ds feature a rising, and Is feature a quasi-level, pattern. Moving on to the 2nd syllable, the natives' T4 featured a falling structure with the two curves converging up 72% of the curve. However, both L2 groups produced diverging curves here, overlapping in the first ~20%. Advanced learners produced native-like falling T4 patterns, but beginners featured the interrogative T4 with a level structure.

5. DISCUSSION

In this acoustic analysis we compared how L2 learners contrast 'bare' and particle-marked interrogatives with declaratives in MC, using very short example utterances. Native production featured a clear distinction, with question curves positioned to a higher f_0 compared to declaratives and in the case of particle-marked questions, the f_0-level of the particle *ma* appeared as a prolongation of the preceding tonal curve pattern. Concerning L2 learners' production, we hypothesized that due to the interference of L1 intonation and L2 tonal patterns, having to contrast Is with Ds poses problems for the learners – in particular, we suspected L1 transfer in those cases where the L1 and L2 f_0 patterns were inversely related.

Beginners' interrogative contours showed signs of L1 transfer, but not in the expected manner. We hypothesized that the interrogative character contour would begin on the verb in the 2nd syllable, but the interrogative contour apparently initiated on the 1st syllable in beginners' production of particle-marked 'T1+T2+*ma*' and bare 'T1+T2' questions. Thus in the case of particle-marked questions, the final stress group contains 3 syllables, and each target in the L*HL character contour transferred from L1 has its own syllable to convey. In 'bare' questions, on the other hand, the final stress group contains only 2 syllables, so transferred L* appears on the 1st syllable, and the falling HL is displayed by the 2nd syllable. Beginners' 'T1+T4+*ma*' question patterns also confirm this explanation: Since both L1 and L2 question contours require a H target in the 2nd syllable, the production does not pose any problem and approximates the native pattern. However, 'T1+T4' 'bare' questions partially contradict this hypothesis, since the falling T4 is realized with a level contour by the learners, in defiance of the hypothesized falling patterns concurring in L1 and L2. The level-contour realization of T4 might be attributed to some kind of dissimilation applied in order to avoid the complete overlap of contours of the two sentence types.

Advanced learners often did not differentiate the initial T1 syllable of the two sentence types, and when they did, the distinction did not approximate the native contrast in either disyllabic or trisyllabic utterances. Particle-marked 'T1+T2+*ma*' questions shared similarities with the native pattern, and were exempt from obvious traits of L1 transfer, except for the overlapping sections of the two sentence types. Additionally, in the advanced learners' case, the f_0 range was more compressed than in natives' production. Similarly, the particle *ma* was not produced as high as by natives, just as expected, though it showed a rising pattern. Advanced learners' 'T1+T2' utterances were in proof of our hypothesis: in declaratives the rising curve of T2 is overwritten by the falling pattern of L1 declaratives, while the rise remains intact in interrogatives. It should be noted that the 'bare' questions rarely appear in MC L2 teaching, so in this case the lack of experience on the side of the learners could give rise to direct L1 transfer. Although advanced learners' 'T1+T4+*ma*' and 'T1+T4' questions differed from those by natives both in terms of f_0 level and contour shape, the differentiation of the two sentence types approximated the native contrast. We should further note that the curve of tonic T4 in declaratives differs in its shape from the native pattern, which might again be attributed to L1 transfer [17].

In sum, our results show that Hungarian learners of MC cope better with sequences featuring T4 in the tonic syllable, compared to those with a T2 tonic. Moreover, the two learner groups followed different patterns in establishing a contrast between declaratives and interrogatives. Further testing is desirable with recording spontaneous dialogues and comparing the intonation patterns therein.

6. ACKNOWLEDGEMENT

The research has been supported by ÚNKP-22-3 New National Excellence Program of the Ministry for Innovation and Technology and the National Research, Development and Innovation Fund.

7. REFERENCES

[1] Shen, X. 1990. *The Prosody of Mandarin Chinese*. University of California Press.

[2] Shih, C. L. 2000. A Declination Model of Mandarin Chinese, *Intonation: Analysis, Modelling and Technology* 1, 1–23.

[3] Lee, O. J. 2005. *The prosody of questions in Beijing Mandarin.* Doctoral Dissertation, Ohio State University.

[4] Shen, X. 1989. Interplay of the four citation tones and intonation in Mandarin Chinese, *Journal of Chinese Linguistics* 17, 61–74,

[5] Li, A., Huang, G., Li, Z. 2019. Prosodic Cues in the Interpretation of Echo Questions in Chinese Spoken Dialogues. *Proceedings of APSIPA Annual Summit and Conference 2019* 1056–1061.

[6] Varga, L. 2016. The Intonation of Topic and Comment in the Hungarian Declarative Sentence. *Finno-Ugric Languages and Linguistics* 5/2, 46–77.

[7] Olaszy, G. 2002. The Most Important Prosodic Patterns of Hungarian, *Acta Linguistica Hungarica* 49, 277–306.

[8] Varga L. 2002. The intonation of monosyllabic Hungarian yes-no questions. *Acta Linguistica Hungarica* 49/3–4, 307–320.

[9] Jun, S. A., Oh, M. 2000. Acquisition of Second Language Intonation, *Acoustical Society of America Journal* 107, 73–76.

[10] Ramírez, D., Romero, J. 2005. The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers, *Intercultural Pragmatics* 2/2, 151–168.

[11] Boersma, P., Weenink, D. 2019. *Praat: doing phonetics by computer*. v. 6.1.05, https://www.praat.org/.

[12] Nolan, F. 2003. Intonational equivalence: an experimental evaluation of pitch scales. *Proc. of 15th ICPhS 2003,* 771–774.

[13] R Core Team 2019. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. v. 3.6.1. https://www.R-project.org/.

[14] Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R*. Chapman and Hall.

[15] Wood. S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society* 73/1, 3–36.

[16] van Rij, J., Wieling M., Baayen, R. H., van Rijn, H. 2022. *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.* R package version 2.4.1.

[17] Juhász K.. in press. Atonális és tonális dallammenetek összehasonlítása. [A comparison of atonal and tonal f_0 patterns] A XVI. Alkalmazott Nyelvészeti Doktoranduszkonferencia Tanulmány-kötete. Nyelvtudományi Kutatóközpont.