

## Investigating Intonation in Chinese Yes-No Questions: A Corpus Study

Jiarui Zhang, University of Oxford

**Background:** Standard Chinese marks yes-no questions using either sentence-final particles (SFPs), such as *ma5* or *me5* (toneless, T5), or through intonation. Previous research (e.g., Shen, 1990) has shown that the mean F0 of syntactically marked yes-no questions is higher than that of statements but lower than intonational yes-no questions. This study aims to investigate yes-no questions with SFPs using a large-scale speech corpus.

**Methods:** We utilised the MAGICDATA Mandarin Chinese Read Speech Corpus (Magic Data Technology Co., Ltd., 2019), which contains 755 hours of speech recordings from 1,080 native Mandarin Chinese speakers. The recordings, made using smartphones, were transcribed with an annotation accuracy exceeding 98%. Question sentences ending with the SFPs *ma5* and *me5* were identified based on transcriptions, excluding any sentences containing embedded wh-questions. To simplify the analysis, we restricted utterance lengths to 5 to 10 syllables, resulting in 14,347 sentences from 1,008 speakers, including 13,638 sentences ending with *ma5* and 709 ending with *me5*. Furthermore, 12,988 statement sentences were randomly selected, matched by syllable length. The boundaries of both question and statement utterances, as well as the final two syllables (including the SFPs) of the questions, were then aligned and annotated using the Montreal Forced Aligner 3.0 (McAuliffe et al., 2017). These annotations were then manually inspected by an experienced annotator. For the analysis, GAMMs were applied to investigate F0 trajectories across entire utterances and the final two syllables of the questions. In addition, duration, mean F0, maximum and minimum F0, timing of maximum and minimum F0 locations, F0 range, mean intensity, and speech rate (syllables per millisecond) were calculated for the whole utterance, the individual utterance length level, and the SFPs. These measures were analysed using linear mixed-effects models, with intonation as a fixed effect and speaker, gender, and utterance length as random effects. Utterance length was also included as an interaction term in the fixed effects for the whole-utterance analysis.

**Results:** The results showed that, at the whole-utterance level, questions were shorter in duration, had a faster speech rate, a narrower F0 range, lower average and maximum F0, but a higher minimum F0 compared to statements. Furthermore, questions exhibited higher mean intensity and earlier timing of F0 peaks and valleys ( $ps < .05$ ). These patterns were largely consistent at the individual utterance level, where similar trends were observed. Questions remained shorter and faster, although these differences were smaller as utterances became longer. While questions had a lower mean F0 in shorter utterances, this significant difference disappeared in utterances with nine or ten syllables ( $ps < .05$ ). As for the SFPs, no significant differences between *ma5* and *me5* were found except that *me5* had a significantly longer duration and earlier valleys than *ma5* ( $ps < .05$ ). The F0 trajectories across different utterance lengths (Figure 1) showed a general declination tendency in both intonations. Although the F0 of questions was slightly higher than that of statements, this difference was minimal. As utterance length increased, the F0 trajectories of both intonations seemed to converge. Furthermore, the F0 trajectories of SFPs showed a significant influence from the preceding tones (Figure 2). When preceded by T1 (55), the SFPs was a high (H) tone. T2 (35) and T3 (214) caused the SFPs to have rising F0 contours. T4 (51) influenced the SFPs to adopt a low (L) tone, while T5, with its inherently low pitch, resulted in low-level SFP contours.

**Discussion and Conclusion:** We used a large corpus to investigate syntactically marked yes-no questions in Chinese. Our findings suggest that F0 does not necessarily show a significant rise (either globally or locally) or higher frequencies for this type of question. This indicates that intonation may serve a supplementary rather than essential role in marking questions, with syntactic cues functioning as the primary mechanism. However, speech rate appears to be an important feature for signalling questions, along with intensity. But the effect of speech rate weakens as utterances become longer.

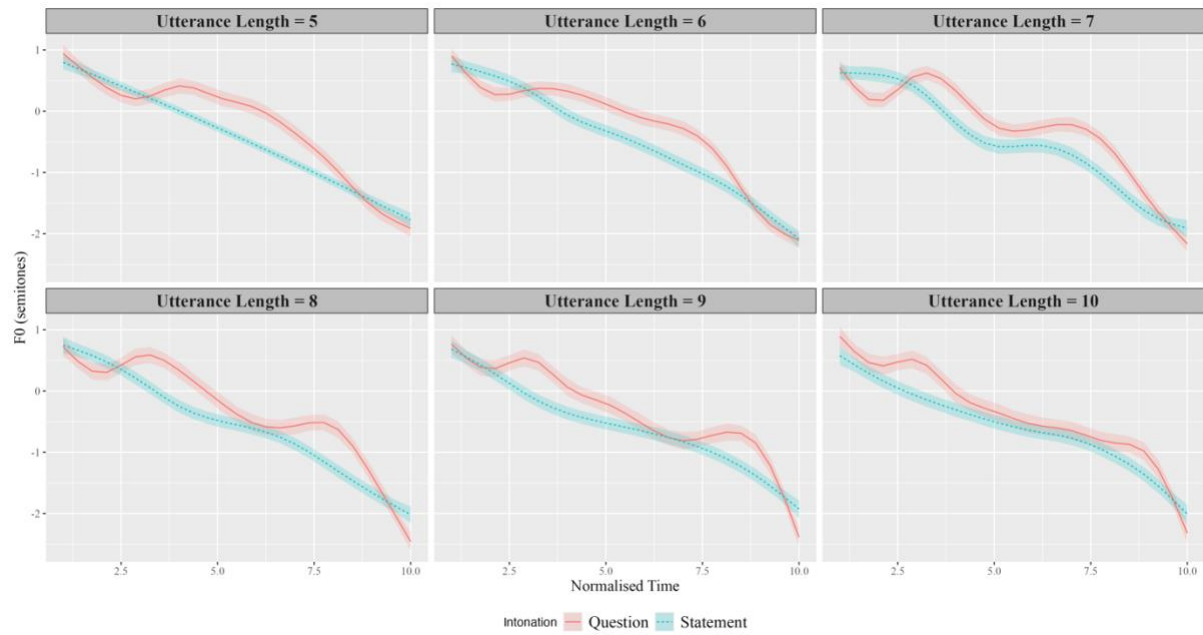


Figure 1 F0 curves of statements and questions by utterance lengths

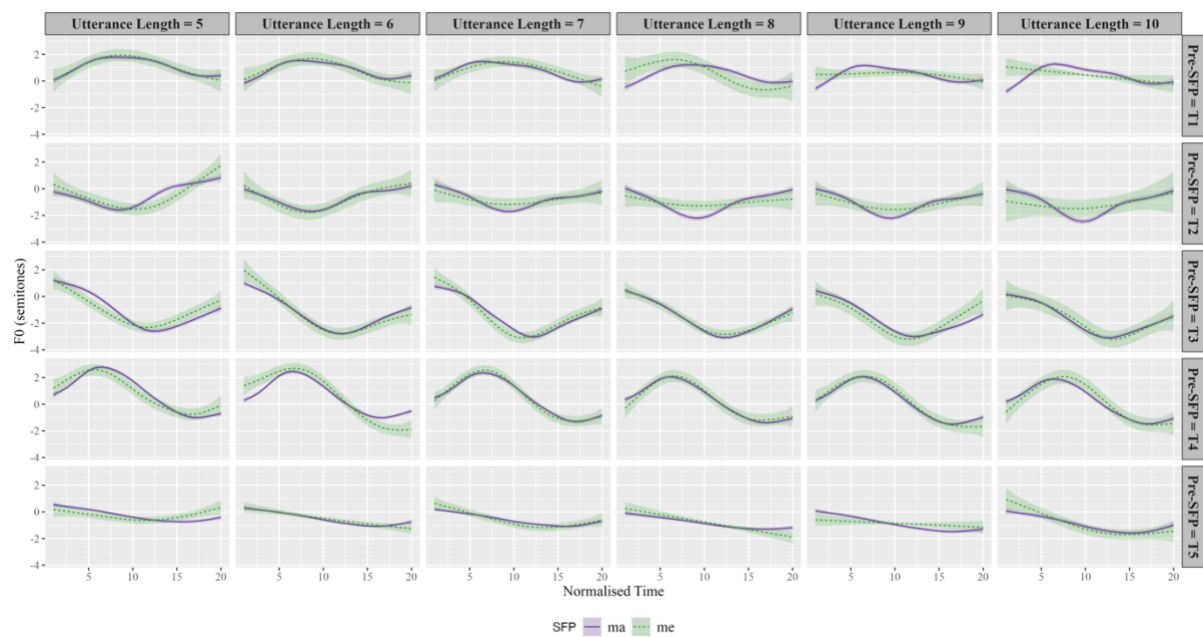


Figure 2 F0 curves of SFPs by utterance lengths and pre-SFP tones

## References:

- Magic Data Technology Co., Ltd. (2019). *MAGICDATA Mandarin Chinese Read Speech Corpus*. Retrieved from [http://www.imagicdatatech.com/index.php/home/dataopensource/data\\_info/id/101](http://www.imagicdatatech.com/index.php/home/dataopensource/data_info/id/101)
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association (INTERSPEECH)*, Stockholm, Sweden.
- Shen, X. N. S. (1990). *The prosody of Mandarin Chinese* (Vol. 118). University of California Press.