

# Prosodic Organization and Focus Realization in Taiwan Mandarin

**Yu-Yin Hsu**

Hong Kong Polytechnic University  
Hung Hom, Hong Kong  
yyhsu@polyu.edu.hk

**James S. German**

Aix Marseille Université, CNRS, LPL  
Aix-en-Provence, France  
james.german@lpl-aix.fr

## Abstract

Cross-linguistically, the way that focus is marked through prosody can depend on a variety of factors, including local constraints on prosodic organization or the position of a word within the larger focus constituent. Here we report on a production study that explores the possible influence of prosodic organization and position on focus realization in Taiwan Mandarin. The materials consisted of sentences in which the syntactic subject consisted of a monosyllabic numeral, classifier, and noun. The context was manipulated to elicit narrow information focus (i.e., using *wh*-questions) on either the numeral, the noun, or the entire NP. The resulting target syllables were then analyzed in terms of their F0 characteristics, duration, and amplitude. The results revealed clear asymmetries in how the numeral and noun were realized in their corresponding single-word narrow focus condition versus in the NP focus condition, though confiding to the intrinsic tone-patterns (e.g., Tone 1 versus Tone 3 words showed different contours). Moreover, the classifier did not always conform to its expected status when external to the focus constituent. These results thus show that focus effects in Taiwan Mandarin cannot be explained in terms of single parameter acoustic enhancement on individual focused words, and they suggest that focus marking in that variety is conditioned by independent constraints on local prosodic structure.

Chinese varieties has emphasized the relative phonetic prominence of a single word that constitutes either the subject or the object of a sentence (e.g., Jin, 1996; Xu, 1999; Chen & Gussenhoven, 2008; Chen, Wang, & Xu, 2008). It has been reported, for example, that narrow new-information focus is associated with prominence in individual acoustic parameters, including longer duration (Xu, 1999), higher mean F0 (Chen et al., 2008), or a larger F0 range (Xu, 1999; Chen & Gussenhoven, 2008). Conflicting findings have been reported regarding whether focus constituents have a higher mean intensity (Chen et al., 2009 vs. Jin, 1996). Very little is known, however, about how prosodic modification may apply to larger and more complex phrasal domains in varieties of Chinese (cf. Chen, 2010).

While prosodic correlates of focus are well-studied for Beijing Mandarin, different patterns of focus realization have been reported for other varieties of a Chinese (see Xu, Chen, & Wang, 2012). For example, whether post-focal compression of pitch range applies to single simple nouns in a sentence may be different between Beijing Mandarin and Taiwan Mandarin (Chen et al., 2014), closely related Taiwan Min varieties (Chen et al., 2009), and between Melaka and Penang Min varieties (Huang & Hsieh, 2016). Unlike Mandarin and Min varieties, it is reported that Cantonese foci do not show post-focus F0 compression and only show greater duration and intensity (Wu & Xu, 2010).

In this study, we take a cross-representational approach, wherein we not only extend the search space from purely prominence-based marking to other phonological features such as phrase boundaries; we also consider how such marking may be simultaneously conditioned by syntax. To do this, we conceptualize focus in terms of focused constituents, which may consist of one or more lexical words in a more complex syntactically and

## 1 Introduction

Formally, focus concerns the way that a sentence is divided into pieces of information which relate to the discourse context in distinct ways (Krifka, 2007). While focus constituents larger than words are well-attested in many languages, existing research on the prosodic marking of focus in

semantically related morpho-syntactic structure (numeral-classifier-noun) (cf. Hsu, 2018 and Hsu & Xu, 2017). We use this structure to study one variety of Mandarin (Taiwan Mandarin, henceforth TwM).

This approach is in contrast to most existing approaches which consider only whether a single focused lexical word is highlighted in some way relative to its neighbors. Given the relative lack of research on phrasing levels intermediate to the prosodic word on the one hand, and the intonation phrase on the other, this approach seeks to explore how focus marking is situated within the overall organization of the prosodic system in Taiwan Mandarin.

To do this, we conducted a production study involving sentences in which the syntactic subject consisted of a monosyllabic numeral, a classifier, and a noun. The context was manipulated to elicit narrow information focus (i.e., using wh-questions) on either the numeral, the noun, or the entire NP. We then used various acoustic measures, including mean F0, duration, and intensity, to test whether focus marking involved simple highlighting of words within the constituent, or whether (i) there was evidence for systematic marking of the edges of the focus constituent, or (ii) a possible influence of syntactic constituency that would lead to asymmetries in how words were prosodically realized inside versus outside of the focus constituent.

## 2 Method

### 2.1 Materials

The target items consisted of 48 sentences involving three-syllable complex nominals (NP) in subject position containing a monosyllabic numeral, followed by a monosyllabic measure word, and then a monosyllabic noun. For half of the items, all three target syllables were Tone 1 (T1), and for the other half, the syllables were all Tone 3 (T3). These items were directly adopted from part of the experimental items in Hsu (2018). Examples of target NPs are given in Table 1. To avoid possible effects of utterance-initial boundary phenomena, the target NPs were preceded by a three-syllable adverbial phrase. Target NPs were

followed by a two-syllable verb phrase and a sentence final particle. All target sentences therefore consisted of nine syllables.

Table 1: *Examples of target NPs/sentences.*

	Target	Pinyin	Gloss
<b>Tone 1</b>	三枝花	san1 zhi1 hua1	“three flowers”
	e.g. 陽台上三枝花枯萎了。 ‘On the balcony, three flowers withered away.’		
<b>Tone 3</b>	九本譜	jiu3 ben3 pu3	“nine copies of sheet music”
	e.g. 下雨天九本譜打濕了。 ‘On a rainy day, nine copies of sheet music got wet.’		

The span of the focus constituent within the target NP was controlled using short contexts consisting of a wh-question. These contexts were pre-recorded by a native speaker of TwM and presented auditorily. The wh-element of these questions targeted either (i) the entire NP (NP-FOCUS), (ii) the numeral only (NUM-FOCUS), or (iii) the noun only (N-FOCUS). Example glosses are given in Table 2. None of the target sentences included other potentially focus sensitive lexical items (e.g., adverbs such as *no*, *certainly*, *only*). There were therefore 72 target items in total (24 sentences  $\times$  3 focus conditions).

Table 2: *Example contexts and target sentences.*  
*Focus constituent indicated by underlining.*

Focus	Context	Target sentences
NP-FOCUS	A:陽台上什麼枯萎了? ‘On the balcony, <u>what</u> withered away?’	B:陽台上三枝花枯萎了。 ‘On the balcony, <u>three flowers</u> withered away’
NUM-FOCUS	A:陽台上幾枝花枯萎了? ‘On the balcony, <u>how many flowers</u> withered away?’	B:陽台上三枝花枯萎了。 ‘On the balcony, <u>three flowers</u> withered away.’
N-FOCUS	A:陽台上三枝什麼枯萎了? ‘On the balcony, what of three units withered away?’	B:陽台上三枝花枯萎了。 ‘On the balcony, <u>three flowers</u> withered away.’

## 2.2 Participants

30 native speakers of Mandarin from Taiwan, who were university students born and raised in Taiwan and who were visiting Hong Kong for one month (15 male, 15 female, aged between 20 and 26) participated in our study. Based on a 7-point self-report scale, 8 participants reported non-fluency in Taiwanese Min (1-3 points), 15 reported intermediate fluency (4-5 points), and 7 reported high fluency (6-7 points). None reported any history of hearing problems. Ethics approval for data collection and basic demographic information were obtained before the experiment. Each participant was paid HK\$60 compensation after the experiment.

## 2.3 Procedure and Analysis

The experiment was conducted in a sound-attenuated room on the campus of Hong Kong Polytechnic University. Each participant first signed an informed consent form and filled out a language background questionnaire. During the experiment, participants were seated in front of a computer screen and wore headphones. Stimuli were presented one at a time (self-paced) on the screen. The order was pseudo-randomized using three separate lists, such that no target item occurred immediately adjacent to itself. For each item, participants were asked to first listen to the context question, and then read the target sentence aloud as casually and naturally as possible; no instructions were given regarding focus or emphasis. Participants produced each sentence twice; additional repetitions were allowed in cases of mispronunciation or hesitation. Productions were recorded in .wav format at a sampling rate of 44.1 kHz with 16-bit encoding. Three practice trials were completed before the experiment. A 5-minute break was given after each set of 18 trials. The experiment lasted around 30 minutes.

Trials were segmented using Praat (Boersma & Weenink, 2018). Syllable boundaries were manually labeled through visual and auditory inspection of the waveform and spectrogram. The vocal pulses were manually checked and corrected in cases of creaky vocalization or pitch halving/doubling. For each syllable, duration,

mean intensity (rms amplitude), and mean F0 were measured automatically in Praat using a custom script developed by the research team. The F0 normalization was realized by dividing each syllable into 10 equal temporal intervals, and F0 was converted from Hz to semitones by relativizing values to 1 Hz via the formula:  $12 \ln(x / 1) / \ln 2$ .

Linear mixed effects models were fit to the data for each target syllable using the lme4 package (Bates et al., 2015) in R (R Core Team, 2014, Version 3.1.0). Fixed factors included only FOCUS condition. Random intercepts included Speaker and Item. The inclusion of random slopes led to non-convergence for all models we tested. Pairwise post-hoc comparisons for the three Focus conditions were conducted using the *glht* function in the *multcomp* package (Hothorn et al., 2017).

## 3 Results

### 3.1 Duration

Boxplots for duration by syllable and FOCUS condition are shown in Figure 1 for T1 and T3. These plots suggest that the duration of a syllable is greatest when it is the sole narrow focus (i.e., NU, numeral; NO, noun), least when it is outside of the focus, and intermediate when the entire NP is focused. Descriptively, this effect appears stronger for the noun than for the numeral. Focus appears to have little effect on the duration of the classifier, when it is after (NU) and before (NO) narrow foci, and even when it is internal to the focus constituent (NP-FOCUS).

Post-hoc comparisons confirm some of these observations, in that for T1, the numeral was significantly longer in NUM-FOCUS than in NOUN-FOCUS ( $p < 0.001$ ). The noun was significantly longer in NOUN-FOCUS than in NP-FOCUS ( $p < 0.01$ ) or NUM-FOCUS ( $p < 0.01$ ), and longer in NP-Focus than in NUM-FOCUS ( $p < 0.01$ ). For T3, the numeral was longer in NUM-FOCUS than in NOUN-FOCUS ( $p < 0.01$ ), and the noun was longer in Noun-Focus than in either NP-Focus ( $p < 0.001$ ) or NUM-FOCUS ( $p < 0.001$ ). Focus condition had no significant effect on classifier duration for either T1 or T3.

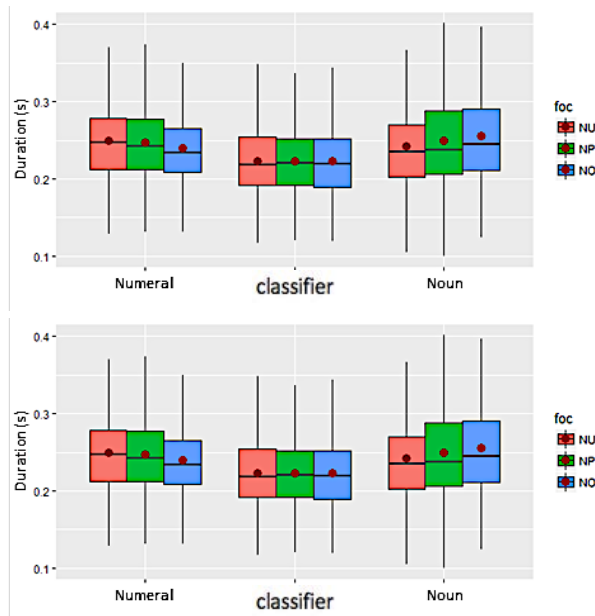


Figure 1: *Boxplots and means (dark red) of syllable duration by focus condition for T1 (top) and T3 (bottom) items.*

### 3.2 Mean F0

Boxplots for mean F0 were not particularly informative for this data set, so time-normalized plots of F0 (semitone) averaged across trials are given by FOCUS conditions in Figure 2 for T1 and T3. These plots suggest that while the basic tonal contours for T1 and T3 were maintained across FOCUS conditions, when the noun or numeral was within the focus constituent, either because it was the sole focus or within the focused NP, it had a higher overall F0. By contrast, F0 on the classifier appears higher for both NP-FOCUS and NUM-FOCUS (NU) than for NOUN-FOCUS (NO). In other words, it appears to pattern with the F0 level of the numeral. The plot for T3 also suggests that third tone sandhi applied between the numeral and classifier, in that there is a dramatic lowering toward the end of noun, but not between the numeral and classifier.

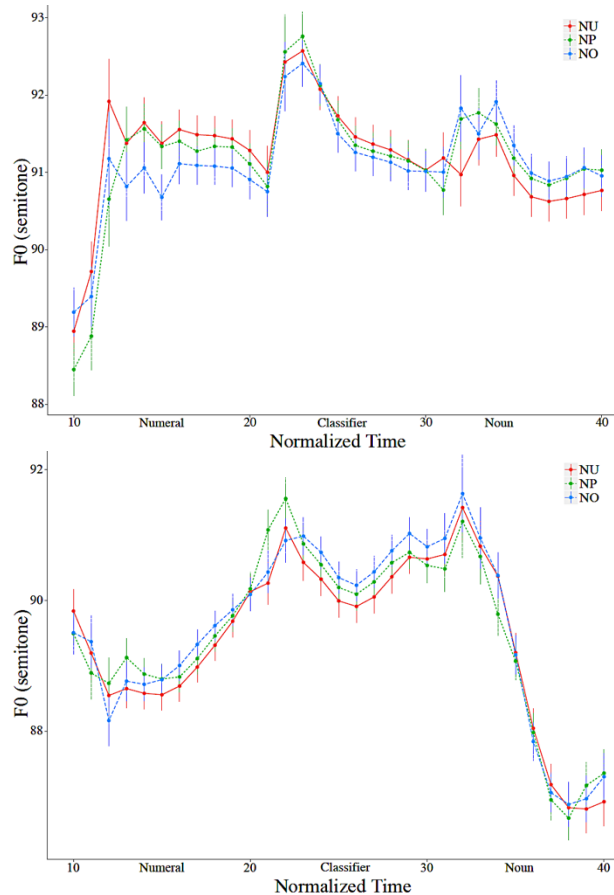


Figure 2: *Time-normalized mean F0 (semitone) with standard error by FOCUS condition for T1 (top) and T3 (bottom).*

Pairwise post-hoc comparisons of mean F0 confirmed some of these observations. For T1, the numeral had a marginally higher mean F0 in NUM-FOCUS than in NOUN-FOCUS ( $p = 0.058$ ), while the noun had a marginally higher mean F0 in NOUN-FOCUS than in NUM-FOCUS ( $p = 0.065$ ). Crucially, the classifier had a significantly higher mean F0 in NUM-FOCUS than in NOUN-FOCUS ( $p < 0.05$ ). No mean F0 comparisons for T3 approached significance.

### 3.3 Intensity

Boxplots for intensity by syllable and FOCUS condition are shown in Figure 3 for T1 and T3. These plots suggest that if intensity is influenced by focus, it is not a particularly strong or systematic effect. The only noteworthy descriptive observation is that for T3, intensity is sharply lower on the noun compared to the two preceding

syllables, while this pattern is absent or at best weak for T1.

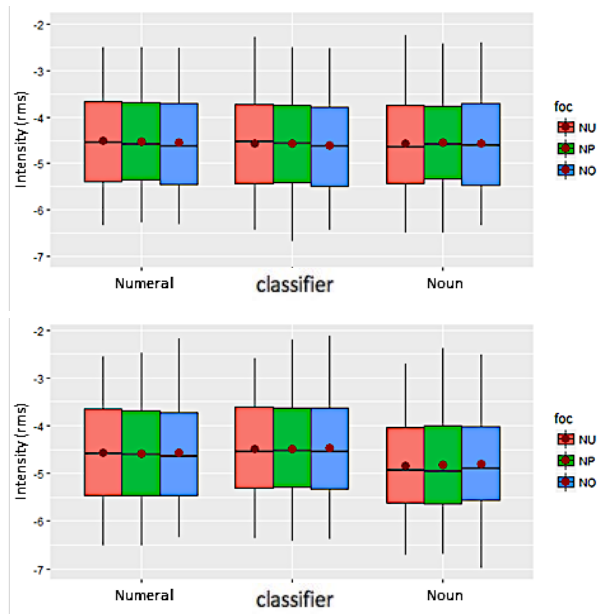


Figure 3: Boxplots and means (dark red) for intensity by syllable and FOCUS condition for T1 (top) and T3 (bottom).

Pairwise post-hoc comparisons nevertheless revealed some noteworthy patterns. For T1, the numeral had a significantly greater intensity for NUM-FOCUS (NU) and NP-FOCUS than for NOUN-FOCUS (NO) (respectively,  $p < 0.001$ ,  $p < 0.05$ ). Parallel to the results for mean F0, the intensity of the classifier was greater for NUM-FOCUS and NP-FOCUS than for NOUN-FOCUS (respectively,  $p < 0.01$ ,  $p < 0.01$ ), suggesting for this acoustic dimension as well, the classifier patterns with the numeral. For T3, the results were less informative; the intensity of the noun was marginally greater in NOUN-FOCUS than in NUM-FOCUS ( $p = 0.067$ ).

## 4 Discussion

Our study sought to explore whether focus realization in Taiwan Mandarin extends beyond general acoustic highlighting of the focus constituent. Our results show support for this idea in several respects. First, we found that both the numeral and noun were longer when inside the focus constituent as compared to outside of it, but the duration of the classifier was not affected by

focus condition. This could be explained if focus-related lengthening concerns primarily the edges of a constituent. Consider that in many languages, edge marking is one of the primary prosodic exponents of focus, and this can occur at either or both the left or right edges (French: Di Cristo, 1998, German & D’Imperio, 2016; Japanese: Pierrehumbert & Beckman, 1988, Gussenhoven, 2004; Basque: Gussenhoven, 2004). If duration is an edge-based means of marking focus in TwM, this would explain the lack of lengthening by the classifier in NP-Focus, since it is further from an edge than the other two elements.

We also found that durational effects of focus were stronger for the noun than for the numeral. Many languages mark the left and right edges of focus constituents differently. In French, for example, the right edge is marked with one category of pitch accent involving primarily rhyme lengthening, while the left edge is marked with a different phonological category associated with onset lengthening (Astesano, 2001). More detailed analysis and further studies are needed to assess whether the different effects of focus on duration for different positions in TwM are due to a left-right asymmetry in general lengthening, or whether it is due to more fine-grained differences in how lengthening applies to different phonological elements.

It is also worth noting that syntactically, the noun is special, since it lies at a strong juncture between the syntactic subject and the VP in our study. Across a wide variety of languages, this location holds a privileged status with respect to prosodic phrasing, in that higher level prosodic boundaries are generally required there than for most other types of syntactic juncture. An alternative explanation could therefore be related to how focus marking interacts with pre-existing prosodic boundaries, or else the tendency for boundaries to be strengthened at specific locations as a result of focus

Finally, we found that the classifier did not pattern with respect to focus in the same way across the three acoustic measures. Specifically, its duration was not affected by focus condition, but it showed a higher mean F0 and greater intensity in the Num-Focus and NP-Focus conditions as compared to Noun-Focus. If this asymmetry across measures involved only NP-Focus, this would

already be surprising, since it would suggest that mean F0 and intensity, but not duration, correspond to a general strategy of prosodic highlighting of the focus constituent. What is most surprising, however, is that the classifier actually patterns with the numeral in this respect, since mean F0 and intensity were similar (and higher) for Num-Focus and NP-Focus. This suggests that a level of prosodic constituency may be influencing how precisely focus marking can target the actual focus constituent. Previous studies have argued for the prosodic word as a phonologically relevant unit in at least some varieties of Mandarin, and that this unit is minimally disyllabic. On the assumption that the numeral and the classifier form a prosodic word independently of focus effects, our results suggest that focus marking targets entire prosodic words, even if this leads to non-optimal enhancement of syllables outside of the focus constituent, or “overfocusing”.

Our current findings are closely in line with those from a similar study on Beijing Mandarin (Hsu, 2018), in showing a similar trend in the prosodic marking of focus. They also suggest that complex internal organization at different structural levels (e.g., prosodic phrasing) interacts with the system of focus prosody. Our results showed, in particular, a discrepancy between prosodic constituents and focus constituents, in that (i) the prosodic realization of the monosyllabic classifier patterned with the focus status of the monosyllabic numeral, and (ii) the application of third tone sandhi between the numeral and classifier complied with prosodic constituency but not focus constituency. In another study on Beijing Mandarin (Hsu & Xu, 2017), the same syntactic structure was used, but unlike Hsu (2018) or the present study, the numeral consisted of two syllables instead of one. The results showed that when the disyllabic numeral itself was the sole narrow focus, it performed as a prosodic constituent independent of its following classifier, and third tone sandhi did not apply between the numeral and classifier. Considering the interaction of prosodic phrasing and focus prosody, we expect future studies to reveal a similar pattern for TwM.

## Acknowledgments

This research was made possible through support by the A\*MIDEX project (n° ANR-11-IDEX-0001-02) funded by the Investissements d’Avenir French Government program, managed by the French National Research Agency (ANR), as well as by the general research fund (4-ZZJQ) supported by the Department of Chinese and Bilingual Studies at Hong Kong Polytechnic University.

We would like to thank the three PACLIC-32 anonymous reviewers for their insightful comments. We would also like to thank Anqi Xu, and Sui Li for their technical support. Mistakes remaining are exclusively our own.

## References

- Astesano, C. 2001. *Rythme et accentuation en Français. Invariance et variabilité stylistique*. Paris: Editions L’Harmattan.
- Bates, Douglas, Martin Maechler, Ben Bolker , and Steven Walker. 2015. "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software* 67 (1): 1-48. Accessed 11 10, 2017. <https://cran.r-project.org/web/packages/lme4/lme4.pdf>.
- Boersma, Paul, and David Weenink. n.d. *Praat: doing phonetics by computer*. Accessed 11 10, 2017. <http://www.fon.hum.uva.nl/praat/>.
- Chen, S.-W., Wang, B., & Xu, Y. 2009. "Closely related languages, different ways of realizing focus." *Proceedings of Interspeech*. Brighton, UK.
- Chen, Y, Y Xu, and S Guion-Anderson. 2014. "Prosodic realization of focus in bilingual production of Southern Min and Mandarin." *Phonetica* 71: 249-270.
- Chen, Y. & Gussenhoven, C. 2008. "Emphasis and tonal implementation in Standard Chinese." *Journal of Phonetics* 36 (4): 724-746.
- Chen, Yiya. 2010. "Post-focus suppression: Now you see it, now you don’t." *Journal of Phonetics* 38: 517-525.
- Di Cristo, A. 1998. "Intonation in French." In *Intonation systems: A survey of twenty languages*, 195-218. Cambridge: Cambridge University Press.
- German, J. S. & D’Imperio, M. 2016. The status of the initial rise as a marker of focus in French. *Language & Speech*, 59(2), 165-195.

- Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Hothorn, Torsten, Frank Bretz, Peter Westfall, Richard M. Heiberger, Andre Schuetzenmeister, and Susan Scheibe. n.d. *Simultaneous Inference in General Parametric Models*. Accessed 11 10, 2017. <http://cran.r-project.org/web/packages/multcomp/multcomp.pdf>.
- Hsu, Yu-Yin. 2018. "Prosody and corrective focus within the nominal domain of Mandarin Chinese." *Proceedings of the 43th Annual Meeting of the Berkeley Linguistics Society*. Berkeley.
- Hsu, Yu-yin, and Anqi Xu. 2017. "Focus Acoustics in Mandarin Nominals." *Interspeech*. Stockholm, Sweden. 3231-3235.
- Huang, Ting, and Feng-fan Hsieh. 2016. "Post-focus compression: All or nothing?" *The Journal of the Acoustical Society of America* 140: 3224.
- Jin, S. 1996. *An acoustic study of sentence stress in Mandarin Chinese*. Columbus, OH: Ohio State University.
- Krifka, M. 2007. "Basic notions of information structure." In *Interdisciplinary Studies of Information Structure* 6. Potsdam.
- Pierrehumbert, J., & Beckman, M. 1988. *Japanese tone structure*. Linguistic Inquiry Monographs, (15), 1–282.
- R Core Team. n.d. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria (2014) (Version 3.1.0).
- Wu, W. L., and Y. Xu. 2010. "Prosodic Focus in Hong Kong Cantonese without Post-focus Compression." *Speech Prosody*. Chicago, USA.
- Xu, Y. 1999. "Effects of tone and focus on the formation and alignment of f0 contours." *Journal of Phonetics* 27: 55-105.
- Xu, Y, Szu-Wei Chen, and Bei Wang. 2012. "Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?" *The Linguistic Review* 29 (1): 131-147.