

Contrastive Focus in Mandarin Chinese

Markus Greif¹

¹ SFB 632, Project D2 "Typology of Information Structure", Potsdam University

greif72@yahoo.de

Abstract

The present study investigates the prosodic realization of two types of contrastive (i.e., corrective) foci relative to wh-focus in Mandarin Chinese: the semantic alternative-based and a discourse-pragmatic type of contrastive focus according to which the contrast is primarily yielded by assumptions on speaker-hearer expectations on the Common Ground. It can be shown that the pragmatically oriented corrective focus is marked more distinctively in terms of prosody than is the alternative-based. Based on this empirical evidence, it is suggested to re-evaluate the current definition of contrast which mainly refers to semantic alternatives and to integrate the findings into existing models on the implementation of lexical tones.

Index Terms: corrective focus, wh-focus, lexical tones, speaker-hearer expectations, semantic alternatives

1. Introduction

Mandarin Chinese (MC), also referred to as *Putonghua* ('ordinary speech'), is the largest among the Chinese dialects on mainland China. Despite some differences, it crucially bases on the Beijing dialect [1]. MC is known to be a tone language, i.e., the main acoustic correlate of intonation, F0, is one decisive means for the lexical meanings of words. There are four full tones (T1-T4), and one type of syllable that is generally considered toneless. The lexical tones are sometimes described in different ways in earlier studies, e.g., [2]; in more recent works [3], [4], [5], [6], the tones are mostly classified according to the Autosegmental Metrical notation as developed in [7] and [8] among others: T1: H (high level), T2: LH (rise), T3: L (low "dip"), T4: HL (fall).

Several studies have been carried out in recent years on the interaction of tone and intonation with respect to several aspects, such as overall tonal patterns in interrogative and declarative sentence types as well as downtrend phenomena [9], [10], coarticulatory influences between the respective tones [3], and the information structural concepts topic and focus [11], [4], [5], [6], [12], [13]. Focus is understood, roughly, as that part of an utterance containing the new, non-presupposed information. In non-tone languages like, for instance, German and English, the main prosodic reflexes of focus is F0 in terms of pitch accents among others. In MC, where the F0 is mainly exploited for the proper realization of the lexical tones, Xu has convincingly shown that focus has a considerable effect on the pitch range of the tones as well as on the syllable length of focused lexemes. In specific, the lexical tones of the focused syllables are realized with expanded pitch range by the speakers and increased syllable duration in general. In addition, post-focal syllables are strongly compressed in their F0 such that the lexical tones are pronounced on a much lower pitch level.

Contrastive focus (CF), still being a field of controversies and open questions, is not yet well-studied in MC. Roughly, CF, often equated with corrective focus, is understood as correcting a part of an information in an utterance by substituting it with a possible alternative (for examples see below; for more details on focus and contrast see the surveys in [14] and [15]). A further perspective on CF is discussed in [16]. According to Zimmermann, CF should be defined in terms of the speaker's assumptions about the hearer's expectations on what is part of the Common Ground and what is not. Thus, CF does not primarily refer to more or less salient semantic alternatives but marks a contrast on "the assumed expectation state of the hearer" [ibid.: 8]. Only under this assumption speakers regularly mark contrast. This discourse-pragmatic view will be further pursued in the present study. It will be distinguished between the semantic oriented alternative-based correction and the more pragmatic correction. The latter type of correction is similar to what Gussenhoven calls counterpresupposition focus [17] (for examples, cf. section 2).

Although in [18], [19], [12] corrective focus is part of the studies, nothing is said about the effect of contrastiveness/correction with respect to "normal" wh-focus, or narrow information focus (NF) in MC. The only study directly dealing with a comparison of CF and NF is [20], using the terms 'corrective rheme focus' and 'normal rheme focus', respectively. Chen and Braun found that CF elicited in question-answer pairs like *Q: Martin rented a car? A: (No.) Mona rented a car*, where the underlined *Mona* is in corrective focus, causes a higher pitch range of the lexical tones T2 (LH) and T4 (HL) relative to NF. However, syllable duration does not differ significantly.

The present study aims at, firstly, a general comparison between NF and corrective focus: What impact has corrective focus on the realization of the lexical tones on the focused items? A second type of corrective focus is then considered. It functions as pragmatic or presuppositional correction similar to the ideas of [16] and [17]. This leads us to the second main question, namely, what role does speaker-hearer expectability play in the realization of focus in MC? Both questions concern directly the current debate about the role of semantics and pragmatics in defining contrast with respect to focus. Furthermore, the present paper may contribute further aspects about the encoding of communicative functions in Xu's PENTA model [6].

2. Method

The present work is a semi-controlled production study where three different focus types were elicited by virtue of appropriate questions simulating the corresponding contexts. The focus types are narrow information focus (NF), CF in terms of semantic correction (COR), and CF in terms of pragmatic or presuppositional correction (PCOR). In COR, the speaker corrects the focused part of the question, i.e., roughly,

the central part of the question the hearer (asking person) is uncertain about. In PCOR, the speaker aims at a correction of presupposed background information taken for granted by the hearer (see Example 1 for clarification). Thus, in PCOR, the speaker likely assumes that her correction is highly unexpected by the hearer. The prosodic parameters under investigation are pitch and duration (cf. section 2.2).

2.1. Data and Subjects

The target words are monomorphemic disyllabic person names in sentence-initial position. They are grouped together under three tone combinations as illustrated in Table 1. The period between two syllables, e.g., in *Lai2.ni2*, symbolizes a syllable boundary in the target words and a morpheme boundary in polymorphemic lexemes.

Target Words	Tones
Lai2.ni2 (Leni)	T2.T2 LH.LH
Ma3.long2 (Marlon) Hai3.lun2 (Helen)	T3.T2 L.LH
Mai4.luo4 (Milo) Te4.ya4 (Tea)	T4.T4 HL.HL

Table 1 - Target words and the corresponding tone combinations

The target words were elicited by different types of questions on various randomized visual stimuli. Figure 1 shows one example of a visual stimulus. The name *Tang1.mu3* ('Tom') does not belong to the target items. The three focus types were elicited as outlined in Example (1).



Figure 1 – Example of a visual stimulus

NF: Who has watermelons? (1a)

COR: Is it Milo who has the watermelons? (1b)

PCOR: Has Tom got two or three watermelons? (1c)

(Bu4) [Ma3.long2]_{NF/COR/PCOR} you3 xi1.gua1 (1d)
(NEG) MARLON have watermelon

(No.) [Marlon]_{NF/COR/PCOR} has the watermelons.

While (1a) elicits a wh-focus (NF) on the grammatical Subject *Ma3.long2* ('Marlon'), (1b) and (1c) both require a correction. However, the latter two differ in the following way: (1b) simulates that it is Milo who is in the center of interest because the asking person is uncertain about his role in the discourse. The speaker then updates or corrects the questioner's uncertainty about this information (COR). Whereas (1c) requires the speaker to correct some presupposed background information that is taken for granted by the questioner. In particular, it is not Tom but Marlon who has some watermelons (PCOR).

The elicitation method exemplified above was carried out in a semi-controlled fashion. That is, the subjects were not

asked to read prepared sentences in an imitating way but give spontaneous answers within certain restrictions: They were asked to answer in a standard sentence with canonical word order (SVO), leaving free as to whether or not a short expression of negation is preposed. The stimuli crucially base on QUIS [21]. The pictures were randomly presented and varied with regard to the contents (e.g., melons, bananas, animals, number of persons etc.). Thus, a high degree of natural interaction was guaranteed, since the subjects were always forced to look carefully at the stimuli in order to give the correct answer. The person names were presented in Chinese characters. The context questions were randomized as well, e.g., Agent questions as in (1a), followed by corrective or alternative Patient questions, as well as free questions on every day topics for distraction etc. The semi-controlled elicitation method was preferred over controlled readings of the target sentences, since it is doubtful that the PCOR condition can be achieved in a natural way in prepared reading sessions. Reliable assumptions on speaker-hearer expectations about the Common Ground would hardly be possible.

There are 3 conditions (NF, COR, PCOR), and 3 tone combinations (the target names, cf. Table 1). The number of tokens per condition and speaker were at least $n=4$, typically varying from 4-7 utterances that are suitable for the analyses. Six subjects were included for the present study. They were all female native Beijingers, undergraduate students and naive as to the goal of the study. In sum, the number of utterances to be analyzed is 6 (speakers) \times 3 (conditions) \times 3 (tone combinations) \times 4 (tokens) ≥ 216 .

The interviews were performed by a native speaker of the same age. The recordings were carried out in a speech lab at the University of Minority Languages in Beijing.

2.2. Analyses

Two main prosodic parameters were analyzed. These were pitch, in particular, F0 curves of the tones as well as pitch span, and duration. To analyse pitch, F0 curves were extracted for time-normalized syllables with the Praat software [22] and an updated script by Xu [4]. The syllables were divided into 20 equally long parts where the F0 was plotted over. Mean values of these F0 curves were calculated in Hertz (Hz) for every single speaker to gain an impression about the realizations of the lexical tones under the three focus types. More specifically, the respective pitch levels and the alignment of the tones within the target words were inspected. The pitch span was determined by extracting the maximal and minimal F0 values of the full words (2 syllables) in order to calculate the differences between them. Furthermore, word duration of the target lexemes was measured. The aim was to find out, whether the two types of corrective foci (COR, PCOR) have an effect on the pitch (span) of the tones and the duration of the words.

When inspecting the F0 values, edge effects caused by the voiceless obstruents in *Te4.ya4* and *Hai3.lun2* (cf. Table 1) were manually eliminated. Thus, no outlying values as a result of such F0 perturbations were included in the statistical analyses. For the duration measures, target words were excluded if they were followed by intonational phrasing. Intonational boundaries are generally known to correlate with an increase in syllable duration in many languages, at least so on the phrase final syllable. Since intonational phrasing may occur for several reasons, the potential final lengthening effect would probably weight the results in an undesirable manner. Prosodic phrasing as a separate parameter was not relevant,

since the speakers only rarely demarcated prosodic phrase boundaries after the target words.

The data were analysed in a repeated measure ANOVA with the two three-level factors Focus (NF as the reference condition, COR and PCOR), and Tone (i.e., the target names subsumed under the tone combinations T2.T2, T3.T2, T4.T4, as shown in Table 1).

3. Results and Discussion

3.1. F0

Focus yielded a significant effect on pitch span. In specific, paired comparisons between NF, COR and PCOR revealed that the pitch span in PCOR was significantly wider than in the reference condition NF, whereas COR did not have an effect (PCOR: $F(1,3) = 15.90, p = .01$; COR: $F(1,3) = 0.006, p > .10$). The mean value in NF is 58.4Hz, while the mean values for PCOR and COR are 68.4Hz and 58.5Hz, respectively. Thus, the pitch span in PCOR exceeds the NF value by 10Hz, whereas COR is roughly of the same size as NF (58.4Hz vs. 58.5Hz). Apart from that, there was no significant effect for Tone, nor was there any interaction with Focus and Tone.

These results show, speakers raise the local pitch span of the target names when they express a correction that they assume to be unexpected for the hearer (i.e., PCOR). Whereas when speaker and hearer share the same background information and the speaker wants to correct the focused part of the question, she obviously sees no need in raising the pitch span. This can be observed for all tonal combinations investigated here. Thus, it is only the pragmatic type of corrections, PCOR, but not the semantic COR that is demarcated in terms of F0 span.

However, the modification of pitch span does not necessarily tell us all about the behaviour concerning the pitch register, i.e., the levels of the tones. Recall, pitch span is defined as the difference between local $F0_{max}$ subtracted by $F0_{min}$, irrespective of the level of the F0. In Figure 2, a representative sample of the target words is presented (due to lack of space, only a part of the F0 extractions are illustrated). We can see the mean F0 of the three tone combinations (T2.T2, T3.T2, T4.T4) from three speakers (LH, Mazb, SQP) in all three conditions on time normalized syllables. The following T3 (*you3* 'have/exist') is not of primary interest. It can be observed that some speakers in fact do not raise the level of the tones on PCOR lexemes (dotted line), although the pitch span is wider relative to NF (black solid line). See, for instance, LH for T2.T2 (left, top): the highest pitch level can be observed for NF, while PCOR and COR (red shaded) are basically on a slightly lower level relative to NF. The T3.T2 (left, mid) words are realized in a similar way, though the PCOR reaches roughly the same level as NF, but starts on a much lower pitch register. That is, the span is higher in both cases although the level is lower to a certain degree. Interestingly, the lexemes in COR are often lowered in their overall pitch level relative to NF with the pitch span staying the same.

Therefore, when we assume that the speakers intend to turn the attention on the corrective item and therefore raise the prosodic prominence of the focused element, it is obviously not a high pitch register per se, but rather the span which is the relevant parameter. In other words, it is not always a high pitch which is the decisive means for rendering a tone more prominent but first and foremost the difference between the

lowest and the highest F0 of the local tone. Hence, the results support the assumption by Xu [6] and Ladd [8] that pitch range is to be divided into the two sub-parameters level and span. Only the latter is the decisive parameter for encoding correction (PCOR) in MC. These findings are partly compatible with what Chen [18] observed for corrective and "more" (i.e., emphatic) focus. She found that it is not primarily a raise in pitch range that marks corrections, but rather the full implementation of the tones with more distinctive rises and falls. A steeper rise or fall as a result of raised prominence, then, may – but needs not – cause a raise in pitch range.

In sum, it can be concluded that speakers consistently exploit F0 to mark contrast when it is connected with the assumption by the speaker that the correction of information is highly unexpected for the hearer. This is the case for PCOR where presupposed background information, which is taken for granted by the hearer, is corrected by the speaker. The sole substitution of an alternative by another, however, does not suffice to mark correction in terms of F0. When speakers mark correction by means of F0, they make use of one single sub-parameter, namely, pitch span.

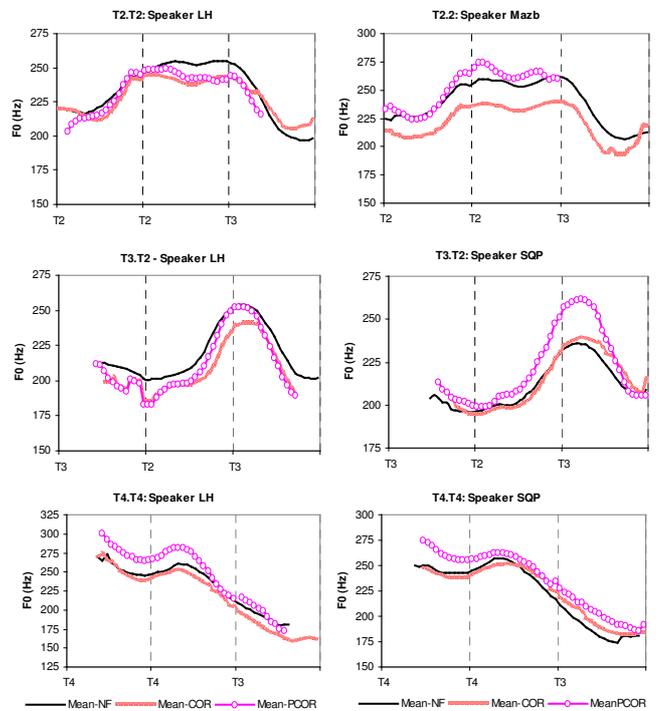


Figure 2 – excerpt of averaged F0 extractions of time normalized T2.T2, T3.T2, T4.T4 for speakers LH, Mazb, SQP; (black solid = NF, red shaded = COR, pink dotted = PCOR)

3.2. Duration

The results for duration partially head in a similar direction. Paired comparisons revealed that PCOR words are significantly lengthened relative to NF, $F(1,3) = 10.4, p < .05$. COR shows a strong lengthening effect, but this is slightly beyond significance, $F(1,3) = 4.7, p = .08$. (The mean values are 354.9ms for NF, 371.08ms for COR, and 394.87ms for PCOR lexemes.) There was a general effect with Tone, independent of Focus. In specific, T3.T2 was significantly longer than T2.T2, $F(1,3) = 26.7, p < .01$. This effect matches with the findings in [3], according to which T3 is generally the longest tone in duration. No interaction could be found with

Tone and Focus. Thus, there is durational lengthening observable in corrective lexemes in general. However, COR reveals a lengthening tendency that cannot be proved to be significant, whereas PCOR lexemes show robust lengthening being significant relative to NF.

The lengthening effect on COR does not fully correspond to the results observed in [20]. Chen and Braun found that duration does not play a role in marking CF (their corrective rheme focus) as opposed to NF (their normal rheme focus). One might speculate that these differences have to do with the different elicitation methods. In [20], the subjects were asked to read question-answer pairs in a contextually appropriate way. Hence, their results are based on fully controlled data with a low degree of personal interaction between speaker and hearer. Whereas in the present study, an interactive question-answer situation with two speakers was created. It is more likely that speakers vary the speech rate for several reasons when answering "real" questions to another interlocutor. However, the results may look different again in narrative discourse data compared to natural dialogues. Clearly, although the present findings suggest that durational lengthening is of some importance for encoding correction, more studies with near-natural data are necessary to verify this conclusion.

4. Conclusions

The present study shows that, firstly, CF is marked prosodically and is thus distinguished from wh-focus (NF) by the speakers. Secondly, the prosodic marking seems to be more robust when speakers correct presupposed background information. That is, when the speaker expects her correction to be highly unexpected by the hearer, the correction (PCOR) is articulated in a more distinctive way relative to NF with regard to prosody. Under these circumstances, F0 is exploited by means of raising the pitch span, and the corrected words are lengthened considerably. The more semantic type of correction (COR), however, only tends to be lengthened in duration. The effects on F0 are minor in that pitch span is virtually not different from NF. Yet, there appears to be a certain tendency of lowering the overall pitch register in COR. If this can be verified in the future, it is a surprising observation, given the fact that previous studies found a general raise in pitch range in corrective focus.

These findings support the discourse-pragmatic perspective in [16] according to which the assumptions about the interlocutors' expectations on the Common Ground must be taken into consideration when defining CF. Hence, it is not sufficient to define contrast (i.e., correction) only by means of substituting an entity by a semantic alternative. Rather, we are to draw a more precise picture of contrast when we discuss its relation to focus in the future. This, I suggest, will also lead to more uniform results in forthcoming empirical studies.

Furthermore, the effects on F0 have to be considered in a differentiated way as has already been suggested in Xu's [6] PENTA model. The present results may contribute some new facts to this model: The parameter 'pitch span' must be regarded as a primitive that plays a crucial role in implementing the lexical tones under corrective focus. Apart from that, one may follow Chen [18] in considering the parameter 'articulatory strength' as being increased under CF resulting in a more distinctive realization of the tones, e.g., steeper rises and falls. While I am inclined to follow Chen's interpretation, the present study does not directly confirm her conclusions relative to NF. On the one hand, she did not compare narrow wh-focus to corrective focus. On the other

hand, the present study does not systematically inspect velocity or "steepness" of the lexical tones. However, I agree with Chen that it is not sufficient to claim that pitch range is generally raised under PCOR, since it is the local pitch span rather than the register as a whole which must be considered. Furthermore, the PENTA model can be enriched by a communicative function related to contrast defined as PCOR and maybe COR.

5. References

- [1] Chen, P., *Modern Chinese – History and Sociolinguistics*. Cambridge University Press, 2004 (1999).
- [2] Chao, Y.-R., *A grammar of spoken Chinese*, Berkely, University of California Press, 1968.
- [3] Xu, Y., "Contextual tonal variations in Mandarin Chinese", *Journal of Phonetics* 25 (61-83), 1997.
- [4] Xu, Y., "Effects of tone and focus on the formation and alignment of f0 contours", *Journal of Phonetics* 27 (55-105), 1999
- [5] Xu, Y., and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33 (319-337), 2001.
- [6] Xu, Y. "Speech Melody as articulatorily implemented communicative functions". *Speech Communication* 46 (220-251), 2005
- [7] Pierrehumbert, J. J. B., *The Phonology and Phonetics of English Intonation*. Indiana (Typoskript ¹ 1980, Cambridge, Mass., MIT).1980 (¹ 1980/² 1987)
- [8] Ladd, D.-R., *Intonational Phonology*. Cambridge University Press (1996)
- [9] Shih, C. "A Declination Model of Mandarin Chinese. Botinis, A. (ed.) *Intonation: Analysis, Modelling and Technology* (243-268). Kluwer Academic Publishers.", 2000.
- [10] Shen, X.-N., *The Prosody of Mandarin Chinese*. University of California Press, 1990.
- [11] Jin, S.-D., 1996, *An Acoustic Study of Sentence Stress in Mandarin Chinese*, PhD dissertation, The Ohio State University
- [12] Shih, C., "Tonal Effects on Intonation". *Proceedings of the International Symposium on Tonal Aspects of Languages*. Beijing, China, 2004
- [13] Wang, B., Xu, Y. Prosodic Encoding of Topic and Focus in Mandarin. In *Proceedings of Speech Prosody 2006*, Dresden, Germany, 2006
- [14] Krifka, M., Basic Notions of Information Structure. In Féry, C., Fanselow, G, Krifka, M. (Eds.), *ISIS* (13-56), Vol.6, 2006.
- [15] Molnár, V., Winkler, S., Exploring the architecture of focus in grammar. In Molnár, V., Winkler, S. (Eds.), *The Architecture of Focus*. *Studies in Generative Grammar* 82, Mouton de Gruyter, Berlin, 2006.
- [16] Zimmermann, M., Contrastive Focus. In Féry, C., Fanselow, G, Krifka, M. (Eds.), *ISIS* (147-160), Vol.6, 2006
- [17] Gussenhoven, C., "Types of Focus in English". In Lee, C./ Gordon, M./ Büring, D. (Eds), *Topic and Focus - Crosslinguistic Perspectives on Meaning and Intonation*, 2007.
- [18] Chen, Y., *The phonetics and phonology of corrective focus in Standard Chinese*, PhD dissertation, State University of New York, Stony Brook, 2003.
- [19] Chen, Y., "Durational adjustments under corrective focus in Standard Chinese. *Journal of Phonetics* 34 (176-201), 2006.
- [20] Chen, Y., Braun, B., "Prosodic Realization of Information, Structure Categories in Standard Chinese", In R. Hoffmann, H. Mixdorff (Eds.), *Speech Prosody 2006*. Dresden, Germany: TUD Press 2006.
- [21] Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., Krifka, M., "Questionnaire on Information Structure". *Interdisciplinary Studies on Information Structure (ISIS)*, Vol. 4. Potsdam University, 2006.
- [22] Boersma, P. & Weenink, D. 2009 – *Praat: Doing phonetics by computer (Version 5.1.07)*. Retrieved May 12, 2009, from <http://www.praat.org/>