ISCA Archive
http://www.isca-speech.org/archive

4th International Symposium on
Tonal Aspects of Languages (TAL-2014)
Nijmegen, The Netherlands
May 13-16 2014

On the elicitation of focus – Prosodic differences as a function of sentence mode of the context?

Frank Kügler [1], *Susanne Genzel* [1]

[1] Department of Linguistics and Research Centre on Information Structure (SFB 632), Potsdam University, Germany

kuegler@uni-potsdam.de, susonne7@gmail.com

Abstract

This paper investigates whether the sentence mode of a context leads to differences in the prosodic realization of focus. Semantically, a communicative reaction in terms of a correction is expected to differ if the context is an assertion or a polar question. In two production studies using a picture naming task, 12 speakers of Mandarin Chinese were recorded reacting to context questions and context assertions which elicited focus on the subject. Experiment 1 examines the syntactic preference, i.e. the use of clefts, in responses to different contexts eliciting corrective and counter-presuppostional focus. In experiment 2, the same setting was used, but speakers were primed for their answers by test trials in order to receive comparable speech material for prosodic analysis. Results show (i) the well-known prosodic effect of focus, and (ii) no difference of elicitation method. Subject focus was preferably realized in basic SVO word order.

Index Terms: focus, context, elicitation method, Mandarin Chinese

1. Introduction

The central issue addressed in this paper is whether different types of context have an effect on the prosodic realization of focus. Consider the conversations in (1) and (2) [1]. The assertion (1a) can be answered with the confirmation (1b) or the rejection (1c). The polar question (2a) triggers identical conversational moves, cf. (2b) and (2c). The crucial difference is that the rejection (1c) leads to a "conversational crisis" [1], while (2c) does not. This is because the polar question comes with two presuppositions [2], which in (2) include the positive alternative 'Sam is home' and the negative one 'Sam is not'. The according answers (2b/c) represent thus a confirmation of one of the presuppositions. In contrast, the rejection (1c) represents a case where the presupposition of the assertion in (1a) is denied or subject to correction. The general assumption pursued here is that a speaker prosodically marks such a "crisis" employing more emphasis, e.g. [3, 4] to express an unexpected discourse move. Note however that 'more emphasis' does not necessarily mean higher F0 [5]. Based on this assumption, the question we want to investigate is whether the type of context has an effect on the prosodic realization of focus.

(1) a. Sam's home.
b. Yes/Yeah, he's home.
c. No, he isn't home.

(2) a. Is Sam home?
b. Yes/Yeah, he's home.
c. No, he isn't home.

1.1. Focus

Semantically, a focus defines a set of alternatives from which a speaker chooses one element [6]. Pragmatically, focus highlights relevant information within an utterance, e.g. [7]. The focus thus represents the most important information of a sentence, which receives its prominence by means of either syntax, morphology, or prosody, or a combination thereof, and languages vary in the use of these means.

Narrow corrective focus, which is of interest here, requires an antecedent in the previous discourse that the focus of the sentence would correct, hence restricting possible alternatives (3). According to [6, 8] the proposition of a previous context belongs to the set of focus alternatives. The correction arises if the proposition differs from the context proposition, hence excluding the context proposition from the set of possible alternatives.

(3) a. Mary stole the cookie.
b. (No,) [Peter]$_F$ stole the cookie. [6]

Languages differ with respect to the prosodic marking of focus. Tone languages like Mandarin express focus in terms of pitch register changes [9–11]. Other tone languages do not employ prosodic means at all to express focus, e.g. [12, 13].

1.2. Mandarin Chinese

Mandarin Chinese (henceforth MC) is a four tone language of the Sino-Tibetan language family [14], distinguishing between tone 1, a high tone, tone 2, a rising tone, tone 3, a low tone, and tone 4, a falling tone. Tones will be indicated in the transcription after each syllable. The phonetics of tone has been intensively studied, cf. [15] and references therein. Sentence level pragmatic meanings such as focus and topic were found to affect intonation in MC (e.g. [9, 16–19]). In particular, MC expresses focus in terms of pitch register expansion [9, 18]. In addition, duration increases under focus [17]. In a recent dissertation [20] also found a difference in tone scaling between different types of corrective focus. Counter-presuppositional focus on an element, which roughly may correspond to a contextual situation as in (1c), has been shown to be realized with more emphasis, which is instantiated as a raised pitch span and increased segmental duration.

Syntactically, MC is an SVO language [21]. Focus may be realized in situ as the examples in [9] show. However, cleft structures may signal corrective focus [20, 22] and counter-presuppositional focus [20].

1.3. Hypothesis

The hypothesis we are pursuing is that both, the effect of focus and the effect of context, are prosodically marked in MC. In particular, we expect that in comparison to a broad focus baseline, narrow corrective focus on an element is prosodically

marked by an expanded pitch register (e.g. [9]), and syntactically marked by the use of more clefts. Further, we expect that a statement-context as (4a) leads to a "conversational crisis" which is marked by more emphasis [20] on the subject in (5d) and the use of more clefts in comparison to a context-question (4b) [20]. In case of (4a), the presupposition 'Mulei has cherries' is denied, and at the same time a corrected presupposition is proposed in (5d).

(4) a. Mulei has cherries.
b. Does Mulei have cherries?

2. Experiments

Two experiments with 12 speakers each tested the effect of sentence mode of the context on the syntactic (experiment 1) and on the prosodic realization of corrective focus in MC (experiment 2); ten females and two males for experiment 1, and nine females and three males for experiment 2. All subjects were students born and raised in Beijing.

In experiment 2, simple SVO structures served as test sentences (5). The first syllable of the subject always carried the falling tone 4. The second syllable was varied from tone 1 to tone 4. The verb was kept constant across the items whereas the object (fruit) varied in tone randomly. The same items were used for experiment 1, see below.

(5) a. Lu4wei1 you3 li4zhi1.
'Luwei has litchis.'
b. Mu4lei2 you3 mang2guo3.
'Mulei has mangos.'
c. Lou4ya3 you3 yang2 mei2.
'Louya has waxberries.'
d. Wei4na4 you3 ying1tao2.
'Weina has cherries.'

Focus was elicited semi-spontaneously on the subject with the help of modified 'focus cards' from the Questionnaire of Information Structure (QUIS) [23], see 2.1. Broad focus was used as a baseline for comparison.

2.1. Material & Procedure

Both experiments were carried out using presentation software. All written instructions, training examples and names of the subjects were presented in Chinese characters. The speakers were digitally recorded with a RØDE NT1-A 1 cardioid condenser microphone and AudioRec [24] at a sampling frequency of 44.1 kHz and 32 bit resolution.

Broad focus was elicited with a card showing a Chinese female holding a fruit in her hand as in Figure 1a, corresponding to the test sentence (5d.). In experiment 2 investigating the prosodic realization, the test trials were preceded by practice trials which primed the participants for a specific structure, e.g. a card showing XiaXia with peaches in her hands was accompanied with a written version of a SVO description of that picture, namely: 'XiaXia has peaches'. The task was to describe the picture as in the practice trial.

Corrective focus was elicited with a card showing three different females holding fruits in their hands as in Figure 1b, corresponding to the test sentence (5d.). The focus card was accompanied by a pre-recorded sound, which was either a statement (4a) or a yes/no-question (4b) (realized as a shi4-cleft by accident). The participant could listen to the sound in a self-paced manner. The task was to respond to the sound according to the pictures. The test trials of the controlled experiment were preceded by practice trials, which primed the participants for the SVO structure.

In experiment 1, investigating the syntactic preferences, the same cards were used. However, the practice trials provided no specific answer. Hence, speakers were free to add e.g. an adverbial or use a cleft construction.

Each picture was repeated 2 times, which resulted in a dataset of 288 sentences for each experiment; 96 broad focus sentences (4 test sentences x 2 repetitions x 12 speakers) plus 192 sentences containing a corrective focus (4 test sentences x 2 repetitions x 2 contexts x 12 speakers).

Figure 1. *a. card used for the elicitation of broad focus; b. card used for the elicitation of corrective focus.*

2.2. Data pre-processing & statistical analysis

Data pre-processing was done in Praat [25]. A native MC speaker labeled sentences at the level of the syllable by using Prosody Pro [26] (experiment 1). The duration of each syllable was extracted in milliseconds. Furthermore, a Pinyin version plus tonal label (1-4) was provided. Based on the tonal label, local F0 minima and maxima, if present, were automatically marked and manually corrected in case of algorithm faults. The F0 analysis was based on a Hanning window of 0.4 seconds length with a default 10 ms analysis frame. The corresponding F0 values were extracted in semi tones (st) using the reference value 100.

Statistical significance relied on linear mixed effect models in R [27] from the lme4 package [28] with speaker, item & repetition as random factors. The maximal model including random slopes for all random factors was reduced in a stepwise fashion [29] with the help of likelihood-ratio tests. A t value above 2 is taken to signal significance.

The syntactic structures of sentences gained from experiment 1 were identified by a native MC speaker. The data contained two error trials, which were discarded from the analysis.

3. Results

3.1. The effect of focus

Figure 2 illustrates the well-known effect of focus on F0 for sentence (5d). The solid line represents the broad focus baseline and the dashed line the corrective focus (averaged over context; S refers to statement, Q to yes/no-question). The F0 is higher on the subject and lower on the post-focal part (object) under corrective focus compared to the broad focus.

The mean value for the H tone on the first syllable averaged over speakers, repetitions and test sentences amounts to 16.0 st (broad focus) and 16.9 st (corrective focus). A linear mixed effects model was fit for the F0 maximum (H tone) with focus condition as fixed factor. An optimal model fit was achieved by including speakers as random slope. The F0 maximum is significantly higher if the subject is corrected; t = 4.424. A corrective focus leads to an increase of F0 by about 0.4 st ± 0.1 Standard Error (henceforth SE). The mean value for the L tone on the first syllable averaged over speakers, repetitions

and test sentences amounts to 12.3 st (broad focus) and 12.9 st (corrective focus). A linear mixed effects model was fit for F0 minimum (L tone) with focus condition as fixed factor. An optimal model fit was achieved by including speakers and test sentences as random slopes. The F0 minimum is significantly higher if the subject is corrected; $t = 2.478$. A corrective focus leads to an increase of F0 by about 0.3 st ± 0.1 SE.

Table 1 presents the means and standard deviations of the F0 aggregated over speakers and repetitions split by tone of the second syllable of the subject of the different test sentences, see (5). The optimal linear mixed effects models for tones 1, 2, 3 and the L tone of tone 4 did not contain any random slopes. The model for the H tone of tone 4 fitted best with speakers as random slope. The level tones behave as expected. Tone1 (H) is raised under corrective focus, whereas tone 3 (L) is lowered. The L part of tone 2 is lowered slightly, yet not significantly, the H part, however, is raised. Moreover, the H tone part of the contour tone 4 is raised. Again, the L tone does not show any significant effect. Furthermore, duration did not yield any significant results.

Experiment 1 yielded the following results. Under broad focus, all sentences were realized keeping the subject in-situ (n=95/95; 100%), whereas under corrective focus less in-situ variants and more shi4-clefts were produced (n=160/191; 84%). All remaining trials (n=31/191; 16%) exhibited an SVO structure; see 3.2 for further details.

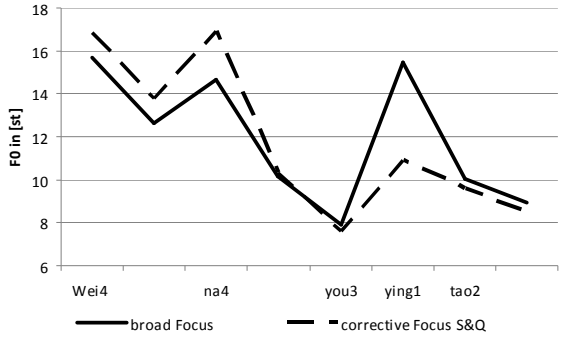

Figure 2. *Interpolated mean F0 maximum and minimum of sentence (5d) 'Weina has cherries', split by focus conditions, n=12.*

Table 1. *Mean F0 (st), standard deviations in parenthesis and t – values of the second syllable of the subject, split by tones and focus conditions, n=12.*

tone	broad focus	corrective focus	*t* – value
1(H)	15.5(5)	17.1(5)	7.59
2(LH)	10.1(4); 13.5(5)	9.9(4); 15.1(5)	0.86.; 6.22
3(L)	9.5(4)	8.8(4)	3.32
4(HL)	14.7(5); 10.1(4)	16.9(5); 10.3(4)	5.76; 0.98

3.2. The effect of context

Figure 3 shows the aggregated F0 maxima and minima of sentence (5d), averaged over speakers and repetitions. The solid line represents the corrective focus elicited with a statement context and the dashed line the corrective focus elicited with a yes/no-question context. Both curves are nearly lying upon each other.

The mean value for the H tone on the first syllable averaged over speakers, repetitions and test sentences amounts to 16.7 st (statement context) and 17.1 st (question context). A linear mixed effects model was fit for F0 maximum (H tone) with context as fixed factor. An optimal model fit was achieved by including speakers and repetitions as random slopes. The F0 maximum is not significantly affected by the context; $t = 1.758$. The mean value for L tone on the first syllable averaged over speakers, repetitions and test sentences amounts to 12.8 st (statement context) and 13.0 st (question context). A linear mixed effects model was fit for F0 minimum (L tone) with context as fixed factor. The optimal model did not contain random slopes. The F0 minimum is marginally significantly higher if the correction of the subject is elicited with a question context; $t = 2.008$. A correction elicited by a question leads to an increase of F0 of the L tone by about 0.1 st ± 0.07 SE. Table 2 presents the means and standard deviations of the F0 aggregated over speakers and repetitions split by tones of the second syllable of the subject of the different test sentences. None of the models contain random slopes. The level tones do not show any significant effect of context. The H tone part of tone 2 is slightly significantly raised if the correction is elicited with a question. Moreover, both parts of the contour tone 4 are raised significantly if the correction is elicited with a question. Again, duration did not show any significant results.

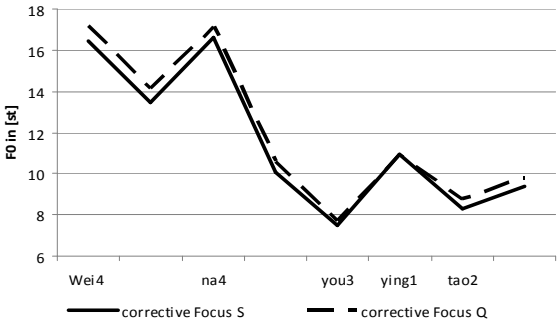

Figure 3: *Interpolated mean F0 maximum and minimum of sentence (5d) 'Weina has cherries', split by context, n=12.*

Table 2. *Mean F0 (st), standard deviations in parenthesis and t – values of the second syllable of the subject, split by tones and contexts (S=statement, Q=yes/no-question), n=12.*

tone	S	Q	*t* – value
1(H)	17.1(6)	17.2(5)	0.14
2(LH)	9.9(4); 14.8(5)	9.9(4); 15.3(5)	0.1; 2.07
3(L)	8.9(4)	8.7(4)	0.83
4(HL)	16.6(5); 10.1(4)	17.1(6); 10.6(4)	2.37; 2.22

Turning to the results of experiment 1, 92% (n=88/96) of the corrections elicited with a statement were produced with an in-situ subject. The remaining 8% (n=8/96) were realized as shi4-clefts. However, the 8 instances were produced only by one speaker. The corrections elicited with a question showed more shi4-clefts (n=23/95; 24%). All other instances were realized in-situ (n=72/95; 76%). Since the question context comprises a shi4-cleft itself, it is likely that the increase in usage of the

cleft construction is due to priming. The 23 instances were produced by 3 speakers including the one that chose the ex-situ variant in the statement context.

4. Discussion & Conclusion

This paper investigates whether the type of context, statement vs. question, affects the prosodic realization of focus. The prediction according to [20] was that a context-statement should result in an answer realized with a higher pitch register compared to a context-question. In other words, context-statements would function as a counter-presuppositional focus. Two experiments were run to elicit syntactic preferences of focus realizations (experiment 1), and to elicit controlled focus structures with identical syntax (experiment 2) for prosodic analysis. In experiment 2, participants were syntactically primed through test trials. The results of this experiment show the well-known focus effect of a raised pitch register [9], cf. Fig. 2. Hence, we may conclude that the experimental setting works to elicit proper focus structures, which enables us to analyze the effect of context on the elicitation of focus.

As for the effect of context, there was no general significant difference between a statement and a question context on the prosodic realization of the focus, cf. Fig 3. The expected higher pitch register in case of context statements because of a prosodic marking of a "conversational crisis" in the answer was not found. Marginal cases of significant differences reported above for tone 4 and the H part of tone 2 are about 0.1 st, which is hardly perceivable. We conclude thus, that the expected effect of context was not borne out. The reported effects along this line in [20] may thus be due to a global effect of pitch register since the test sentences were frequently produced including a negation at the beginning. Also, [20]'s data was not based on a controlled setting as the controlled experiment in this study. Hence, we may conclude that the type of context that elicits focus does not affect the prosodic realization of focus.

Experiment 1 has shown that subject focus, in general, is realized with SVO structures, thus without any particular syntactic strategy as for instance shi4-clefts [22]. This is contrary to [20] who reported a frequent use of clefts. However, one speaker used shi4-clefts in case of focus. Compared to the factor of context, the results showed an increase of the usage of shi4-clefts for context questions. Yet, this increase seems to be due to a syntactic priming effect since the context questions contained shi4-clefts. If increased prominence would yield more clefts, context-statements were expected to induce clefts. Hence, the occurrence of shi4-clefts can be attributed to the syntactic structure of the context rather than to the sentence mood of the context.

To conclude, this study has shown that the type of context does not influence the prosodic realization of focus in Mandarin Chinese. An expected higher emphasis due to an unexpected discourse move in case of a statement correction (cf. (1) and (2)) [3, 4] does neither result in a different prosodic realization of focus nor in a different syntactic structure compared to the preferred SVO word order.

5. Acknowledgements

Thanks to the valuable assistance of Bei Wang and Wu Qian. This work was funded by the German Research Association [DFG grant number SFB 632 "Information structure", project D5], http://www.sfb632.uni-potsdam.de/en/cprojects/d5.html.

6. References

[1] D. F. Farkas and K. B. Bruce, "On Reacting to Assertions and Polar Questions", *Journal of Semantics*, 27:81–118, 2010.
[2] C. L. Hamblin, "Questions in Montague English", *Foundations of language*, 10:41–53, 1973.
[3] K. Hartmann, "Focus and Emphasis in Tone and Intonation Languages", in A. Steube [Ed] The discourse potential of underspecified structures, 389–411, Berlin: de Gruyter, 2008.
[4] M. Zimmermann, "Contrastive focus and emphasis", Acta Linguistica Hungarica, 55(3):347–360, 2008.
[5] F. Kügler and S. Genzel, "On the Prosodic Expression of Pragmatic Prominence: The Case of Pitch Register Lowering in Akan", Language and Speech, 55(3):331–359, 2012.
[6] M. Krifka, "Basic notions of information structure", Acta Linguistica Hungarica, 55(3):243–276, 2008.
[7] W. L. Chafe, "Language and consciousness", Language, 50(1):111–133, 1974.
[8] C. Gussenhoven, "Notions and subnotions in information structure," Acta Linguistica Hungarica, 55(3):381–395, 2008.
[9] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours", J Phonetics, 27(1):55–105, 1999.
[10] F. Liu and Y. Xu, "Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation", Phonetica, 62:70–87, 2005.
[11] S.-w. Chen, B. Wang, and Y. Xu, "Closely related languages, different ways of realizing focus," in Proc Interspeech, 2009.
[12] F. Kügler and S. Skopeteas, "On the universality of prosodic reflexes of contrast: The case of Yucatec Maya", in Proc 16th ICPhS, Saarbrücken, Germany, 2007, 1025–1028.
[13] C. Gussenhoven, R. Teeuw, "A moraic and a syllabic H-tone in Yucatec Maya", in E. Herrera et al. [Eds], Fonología instrumental, 49–71, Mexico City: El Colegio de Mexico, 2008.
[14] S. Duanmu, The phonology of Standard Chinese. OUP, 2007.
[15] Y. Xu, "Contextual Tonal Variations in Mandarin", J Phonetics, 25(1):61–83, 1997.
[16] B. Wang and Y. Xu, "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese", J Phonetics, 37:502–520, 2011.
[17] Y. Chen, "Durational adjustment under corrective focus in Standard Chinese", J Phonetics, 34(2):176–201, 2006.
[18] Y. Chen and C. Gussenhoven, "Emphasis and tonal implement-tation in Standard Chinese", J Phonetics, 36:724–746, 2008.
[19] I. Ouyang and E. Kaiser, "Focus-marking in a tone language: Prosodic cues in Mandarin Chinese", LSA Annual Meeting Extended Abstracts, 2012.
[20] M. Greif, Corrective focus in Mandarin Chinese: A question of belief? Muenchen: Lincom Europa, 2012.
[21] C.-T. J. Huang, "Logical relations in Chinese and the theory of grammar", PhD dissertation, MIT, 1982.
[22] K. Li, "Contrastive focus structure in Mandarin Chinese", Proc 20th North American Conference on Chinese Linguistics, 2008.
[23] S. Skopeteas, I. Fiedler, S. Hellmuth, A. Schwarz, R. Stoel, G. Fanselow, C. Féry, and M. Krifka, "Questionnaire on Information Structure (QUIS)", Interdisciplinary Studies on Information Structure ISIS 4, Potsdam: Universitätsverlag Potsdam, 2006.
[24] T. Wang, Audionrec, 2011.
[25] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program], 2013.
[26] Xu, Y., "ProsodyPro. A Tool for Large-scale Systematic Prosody Analysis", In Proc Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, 7-10, 2013.
[27] R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria, 2013.
[28] D. Bates, M. Maechler, B. Bolker, and S. Walker, lme4: Linear mixed-effects models using Eigen and S4, 2013.
[29] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items" J Memory and Language, 59(4):390–412, 2008.