Dual-focus Intonation in Standard Chinese

Bei Wang[1] *Caroline Féry*[2]

bjwangbei@qq.com caroline.fery@gmail.com

1Minzu University of China, Beijing, China 2Goethe-Universität Frankfurt，Frankfurt, Germany

ABSTRACT

This paper examines prosodic encoding of sentences with two foci in Standard Chinese, henceforth *dual focus*. Five speakers read 10 SVO sentences with two lengths (short vs. long) in four focus conditions: initial, final, dual (initial+final) and neutral. Following results were obtained for the dual focus sentences: [1] There was an increase in F0 and word duration in both foci, and this to almost the same degree as their initial and final focus counterparts respectively; [2] The word following the first focus did not differ from its neutral and final counterparts in F0 and duration. This held for both the short and the long sentences. We conclude that no prosodic boundary was inserted after the first focus and that the two foci were realized in one intonational phrase. From a theoretical perspective, culminativity is violable in Chinese and focus assignment and phrasing are largely independent of each other.

***Index terms*:** dual focus, phrasing, intonation

1. INTRODUCTION

Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions[1]. In prosody, focus is usually the most prominent part in a sentence[2, 3]. Both in English and Standard Chinese, focus is realized with on-focus F0 rising, and post-focus F0 compression [2-5]. Longer duration on focused words has also been found cross-linguistically[6-9].

A rarely asked question is how two foci are realized in one sentence. Experimental studies in English showed that the realization of two foci in one sentence is not the same as the one arising from simply concatenating two single foci. Interestingly, post-focus F0 compression is lacking after the first focus in dual-focus sentences[2], and it holds for both short and long sentences [10].

In a study on dual focus in Mandarin Chinese, Kabagema-Bilan, López-Jiménez and Truckenbrodt found that the second focus showed higher F0, longer duration and post-focus F0 lowering, whereas the first focus did not show any phonetic variation[11]. The two foci were in the initial and medial position of a sentence. They argued that the phonetic effects on the second focus are triggered by the fact that this focus carries the sentence accent, and thus the highest prominence in the intonational phrase. The phonetic effects are thus, according to them, due to intonation phrase stress, rather than to the F-feature that marks focus. A bit problematic was the fact that the distance between the two foci only betrayed one syllable (a verb). It could be that the short distance between the two foci is the reason why the first focus was not realized.

Jia, Li and Xiong [12] used a single 9-syllable sentence with only Rising (R) tones, except for the very last syllable (‘le’, a neutral tone). The two foci were located on the initial and final syllables of the sentence. Contrary to Kabagema-Bilan et al., they found higher F0 and longer duration on both foci.. In addition, similar to the English results[2, 10], the first focus did not show post-focus lowering.

A possible explanation for the diverging results obtained by the two studies just reviewed is their different design. We assume that there are two strategies for realizing dual focus. First, if a sentence is short, and the two foci are close to each other, only the second focus is fully realized. Second, when a sentence is long, both foci are fully realized, with higher F0 and longer duration.

In our study, we investigated whether sentence length bears on how dual focus is realized. Length relates to another issue, that is, prosodic structure or phrasing. The previous studies assumed that both foci are realized in one intonation phrase. In mainstream theories on prosodic structure, the formation of prosodic phrases is a necessary step for the assignment of accents: each prosodic constituent has a single head on the prosodic level immediately below[13-16]. In narrow focus cases, the focused word carries nuclear accent and the post-focal words are often deaccented [15-19], which is in agreement with the metrical culminativity principle, i.e. one metrical prominence per prosodic constituent. Some researchers propose that a focus intonation is created by modifying the prosodic structure of the sentence, see for instance [20, 21]. This is the model assumed by Kabagema-Bilan et al.[11] for interpreting their results.

However, other experimental studies showed the independent mechanisms of focus and phrasing. Ishihara showed for Japanese that prosodic boundaries remain unchanged when a focus is added to a sentence[22]. There is also experimental evidence showing that the durational adjustment for boundary and focus are largely independent of each other. For instance, Turk and Shattuck-Hufnage found that pre-

boundary lengthening applies in both pitch-accented and unaccented phrase-final words in English[23]. Horne, Strangert and Heldner found that focus does not affect pauses in Swedish[24]. Chen's study on Chinese analyzed durational variation of onset and rhyme of syllables in the initial and medial positions of an intonational phrase, comparing them with focused and unfocused conditions [25]. She found that the onset is longer when a syllable is initial than medial, whereas focus induces lengthening more consistently in rhyme than in onset in both positions. She suggested that focus does not insert an intonational phrase boundary to the left edge of a focused constituent.

To get a better understanding of dual focus realization in Standard Chinese, we tested sentences with all four tones. Moreover, sentence length was varied to test whether it affects how dual focus is realized and whether a longer sentence triggers a prosodic boundary between the two foci.

2. METHODS

2.1. Materials

The procedure used in this experiment was similar to Eady et al.[2]. The speakers were asked to read aloud sentences in which the location of focus was systematically varied by different preceding questions. Four focus conditions were investigated: initial, final, dual (initial + final) and neutral.

The short (6 syllables) and the long (12 syllables) sentences had a simple SVO word order. The short sentences contained three disyllabic words, and the long sentences were constructed by adding modifiers before the object, leaving the subject and the verb unchanged, see (1a) and (1b), for an example. The two focused words in the dual focus condition are marked with bold letters.

(1a)**Wang.Ying** can.guan **che.jian**.
Wang Ying visit workshop.
'Wang Ying visited the workshop.

(1b) **Wang.Ying** can.guan Shan.Xi Qing.Xiang yi.jie **che.jian**.
Wang.Ying visit Shan.XiQing.Xiang first-street workshop.
'Wang Ying visited the Shan.Xi Qing.Xiang First-street workshop.'

Four focus questions are listed in (2a-2d).

(2a) Initial focus: shui can.guan che.jian?
Who visited the workshop?

(2b) Final focus: Wang Ying can.guan shen.me?
What did Wang Ying visit?

(2c) Dual focus: shui can.guan shen.me?
Who visited what?

(2d) Neutral focus: Shen.men shi?
What happened?

We used five base sentences with different tone combinations. The four lexical tones, High (H), Rising (R), Low (L) and Falling (F), have the pitch contours of high-level, mid-rising, low-dipping and high-falling respectively. We constructed five sentences similar to the sentences used in [4], which contained the disyllabic words with either HH, HL, LH, FR or RF tone combinations.

The total number of utterances for each speaker was 80, that is, 2 (sentence length) × 4 (focus conditions) × 5 (tone combinations) × 2 (repetitions) = 80 (utterances).

2.2. Participants

Three female and two male native speakers of Standard Chinese, aged between 22 to 30, participated in the experiment. All had no self-reported speech and hearing impairments. They were paid for their participation.

2.3. Recording procedure

The participants were recorded individually in a sound-proof booth at the Phonetic Laboratory at the Institute of Linguistics, University of Potsdam. They were asked to read aloud the target sentences after hearing the questions. The signal was digitized at 48 kHz and the sampling format was one channel 16-bit linear in real time. The target sentences were presented on a computer screen using a custom-written Java program[26], which randomized the sentences for each recording.

The background questions were recorded by the author BW, a 32-year-old native Chinese speaker. There were 160 fillers that were used for another linguistic study. The recording lasted about 1.5 hours per speaker.

2.4. Acoustic measurement

The target sentences were extracted and saved as separate wav files. The acoustic analysis procedures were similar to those in [4]. ProsodyPro[27], a Praat script was used to take F0 and duration measurements from the sentences. To extract continuous F0 contours, the vocal cycles were first marked by Praat [28] and then hand checked. Segmentation labels were added to mark the syllable boundaries manually. The vocal periods were converted into F0 values by ProsodyPro. The vocal pulse marking, segment labels, and F0 values for each utterance were saved in text files. ProsodyPro also obtained the highest and lowest F0 and duration of each syllable. The F0 values were converted from Hz to semitone (st) using the formula: $Fst = 12 \times \log2(F0/50)$, in which 50 Hz is the reference F0.

3. RESULTS

3.1. Graphic analysis

Figure 1 shows the mean F0 contours of the short and long HH tone (upper row) and HL tone (lower row) sentences, with the four focus conditions overlaid in one graph. The time normalization procedure of [3, 27] was applied. This procedure remaps clock time onto a sequence of 10 equidistant points within each individual

syllable. For each point, the F0 values were averaged across the five speakers with their two repetitions. The vertical lines indicate the word boundaries.

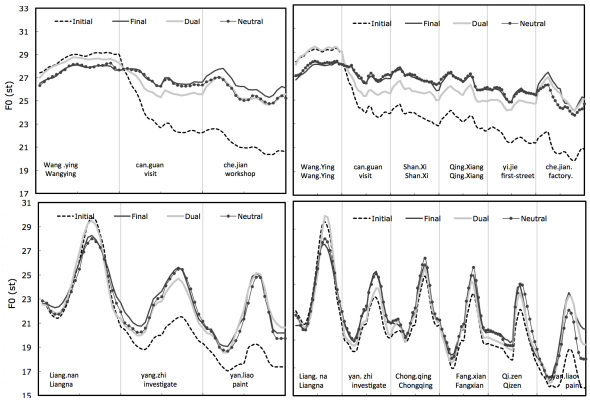

Figure 1. Intonational contours of the HH tone sentence.

The other experimental sentences showed a similar pattern. We do not have space to present them here. For the intonation contours of dual-focus sentences, three visual observations can be made about Figure 1: (1) The first focus has almost the same F0 height in the dual focus condition as in the initial focus, which is higher than in the corresponding neutral and final focus. (2) F0 of the post-focus part after the first focus is not lowered to the same degree as it is in the initial focus condition. (3) The second focus reaches roughly the same height as in the final focus condition. These three observations hold for both the short and the long sentences, and for all tones.

3.2. Analysis of F0

For statistic tests, maximum F0 values of the first, second and last words are graphically displayed in Figure 2 and 3 for the short and long sentences respectively. In the statistic tests, three-way repeated measures ANOVA were carried out for each target word, with focus condition, sentence length, and tone as three independent variables.

We can see in Figure 2 and 3 that the maximum F0 of the target words are the highest when the word is under focus in both single and dual focus conditions. In the second word, which is post-focal, maximum F0 is the lowest in the initial focus condition and the value in the dual focus condition is between initial and neutral focus conditions.

In the *first word*, the 3-factor repeated measures ANOVAs of maximum F0 showed main effects of focus condition ($F(3, 12)=10.85$, $p=0.015$) and tone ($F(4, 16)=7.19$, $p=0.025$) but not sentence length ($F(1, 16)=0.791$, *n.s.*) or any interaction. The post-hoc test (S-N-K) showed that the first word in the initial and dual focus conditions has significantly higher maximum F0 than that in the neutral and final focus conditions, which holds true in both short and long sentences. However, no significant difference between the initial focus and dual focus condition was found. Observation (1) of Figure 1 was confirmed, i.e., the first focus of a dual-focus sentence reaches the same height as an initial focus.

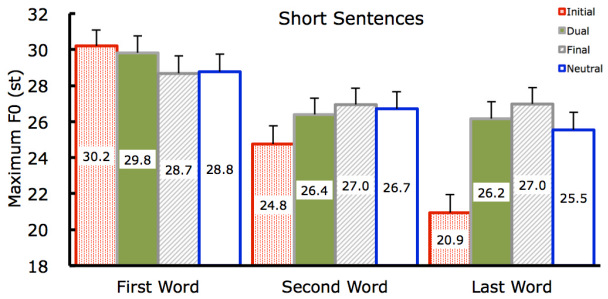

Figure 2. Maximum F0 of the three target words in the short sentences.

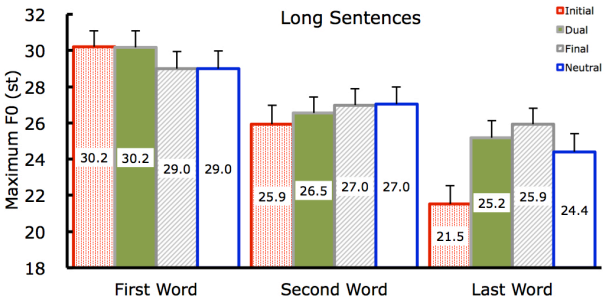

Figure 3. Maximum F0 of the three target words in the long sentences.

In the *second word*, a similar statistic test of maximum F0 showed main effects of focus condition ($F(3, 12)=10.289$, $p=0.021$), tone ($F(4, 16)=33.742$, $p<0.001$) and length ($F(1, 4)=21.2$, $p=0.01$). The post-hoc tests (S-N-K) showed that the maximum F0 of the second word was much lower in the initial focus condition than in the other three conditions, and no difference was found among dual, final and neutral focus conditions. Thus, the post-focus lowering after the first focus in a dual-focus sentence was not as large as that after an initial focus. Two interactions were also significant (Length×Context: $F(3, 12)=16.47$, $p=0.004$; Tone×Context: $F(12, 48)=9.125$, $p=0.005$), which came from the fact that post-focus lowering after an initial focus was larger in a short sentence than in a long one. Observation (2) of Figure 2 is thus confirmed.

In the *last word,* applying a similar statistic analysis of maximum F0, we found that all three factors had main effects (Context: ($F(3, 12)=38.275$, $p<0.001$); Length: ($F(1, 4)=8.428$, $p=0.044$); and Tone: ($F(4, 16)=38.644$, $p<0.001$), but no interaction was found. The post-hoc tests (S-N-K) showed that the last word in the initial focus condition had significantly lower F0 than the other three conditions, and neutral focus had lower F0 than dual and final focus conditions. Moreover, no significant difference between dual focus and final

focus conditions was found. In short words, the second focus in a dual-focus sentence was realized similarly to a final focus. This holds in both short and long sentences. Observation (3) of Figure 1 was also confirmed.

The results of minimum F0 of the three target words in the short and long sentences are presented in Figure 4 and 5. Sentences with H tones are excluded here.

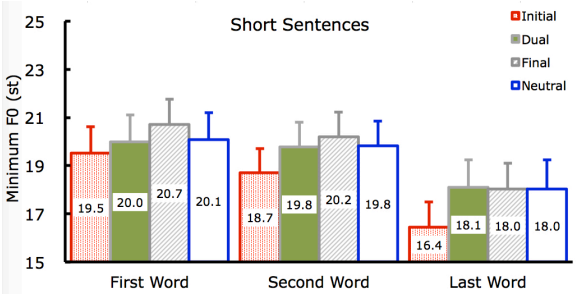

Figure 4. Minimum F0 of the three target words in the short sentences.

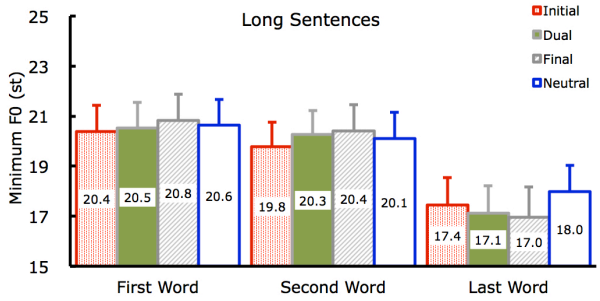

Figure 5. Minimum F0 of the three target words in the long sentences.

Three-way repeated measures ANOVAs of minimum F0, with context, length and tone as three independent variables, did not show any effect of focus condition in the *first* (F(3,12)=3.826, *n.s.*) and *last* words (F(3, 12)=2.404, *n.s.*). In the *second* word, however, there were main effects of context (F(3, 12)=4.506, p=0.015), length (F(1, 4)=11.9, p=0.026) and tone (F(3, 12)=4.43, p=0.027). Three interactions were also found in the second word (Length×Condition: F(3, 12)=7.489, p=0.019, Length×Tone: F(3, 12)=5.043, p=0.055, and Length×Condition×Tone: F(9, 36)=3.546, p=0.053). As can be seen in Figure 4 and 5, the minimum F0 of the second word is lower in the initial focus condition than that in the other three focus conditions. Moreover, post-focus F0 lowering after the initial focus was larger in the short sentences than in the long sentences, which was also found in [4].

Altogether, the quantitative analysis of F0 shows that a dual focus sentence raises both foci to relatively the same height as their single focus counterpart, although no post-focus lowering after the first focus is applied. Such a pattern holds for both the short and the long sentences.

3.3. Analysis of word duration

Figures 6 and 7 present word duration of the three target words in the short and long sentences respectively, with the four focus conditions overlaid in one figure.

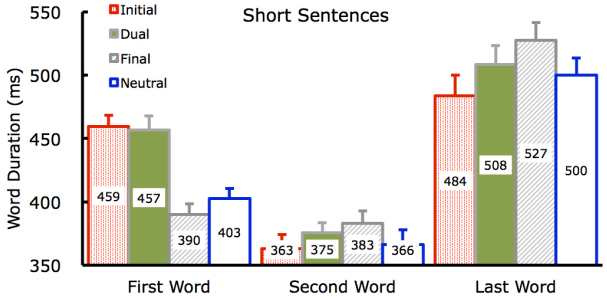

Figure 6. Duration of the three target words in the short sentences.

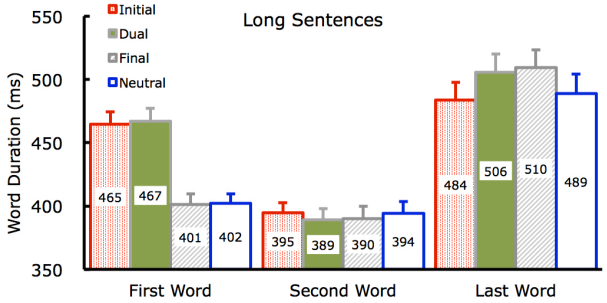

Figure 7. Duration of the three target words in the long sentences.

Similar three-way repeated measures ANOVAs were carried out for each word. As can be seen in Figures 6 and 7, the duration of the first word is longer in the initial and dual focus conditions than in the final and neutral focus conditions (F(3, 12)=0.565, p=0.01). Word duration of the second word is comparable across the four focus conditions (F(3, 12)=0.611, *n.s.*). In the last word, word duration is longer in the dual and the final focus conditions than that in the initial and the neutral focus conditions (F(3, 12)=5.554, p=0.041).

In short words, both foci in a dual focus sentence are lengthened to relatively the same degree as their single focus counterpart. Both pre-focus and post-focus words do not show systematic duration variation.

4. GENERAL DUSCUSSION AND CONCLUSIONS

This paper studied the prosodic encoding of dual focus sentences in Mandarin Chinese. The results showed that on-focus F0 rising and duration lengthening of both foci reached relatively the same value as their single focus counterpart. It differed from the initial focus condition in that the first focus of a dual focus sentence did not trigger post-focus F0 lowering. These results are consistent with Jia et al. [12]. In addition, the prosodic realization of dual focus is similar in short and long sentences. That is to say, the lack of post-focus F0

lowering is not due to time constraint between the two foci. In sum, the prosodic encoding of dual focus sentences in Standard Chinese is very similar to that in English [2, 10].

In [11], the dual focus sentences (initial+medial) did not show clear F0 raising on both foci. One could argue that this is because the sentences were short (6-8 syllables) and there was just one weak syllable between the two foci. However, the short sentences in the current study are comparable to those in [11], as our sentences had 6 syllables and a disyllabic verb between the two foci. We did find on-focus F0 raising in both foci. Because of the numerous differences between the two studies in terms of experimental control, it is hard to judge the sources of the conflicting results. We argue that when two words are under focus, they are both realized with special phonetic cues due to their focused status.

An interesting finding about dual focus intonation is the lack of post-focus compression after the first focus. Eady et al. [2] suggested that this is due to the influence of the second focus on the first one. According to them, the minimized F0 drop after the first focus in a dual-focus sentence facilitates the increase in F0 at the end of the dual-focus sentences. This claim makes some sense when sentences are short or when the distance between the two foci is short. However, the lack of post-focus compression is also found in long sentences in both our study for Chinese and in Liu[10] for English.

Here, we shortly discuss two other possible explanations. First, the lack of post-focus compression in dual focus sentences indicates the upcoming of another important unit of information. In the initial focus condition, on the contrary, post-focus compression implies the absence of other important information in the rest of the sentence. Such a thought is in line with the popular argument that post-focal given information is mostly deaccented[29, 30]. We argue that post-focus F0 compression and on-focus F0 raising are two separate components of focus. The on-focus F0 raising is a correlate of focus proper, whereas the post-focus compression is a correlate of the givenness of the following information.

Second, as suggested by Yi Xu (p.c.), the post-focus part after the first focus in dual focus sentences might be better analyzed as pre-focal material of the second focus. In this analysis, then, it is not subject to post-focal compression. This assumption suggests that the domain of the first focus only lies in the focused word, whereas the rest of the sentence is prosodically grouped together with the second focus, which in turn is realized as a final focus. In [3], it is shown that final focus does not cause any F0 variation in the previous component, which is also the case in the current study. This suggestion is in line with an analysis of the data in the spirit of prosodic phrasing. The initial focus is a subject and forms its separate prosodic phrase. The remainder of the sentence is a verbal phrase, and forms a second prosodic phrase of roughly the same kind. As a result, the material separating the two foci belongs syntactically and prosodically to the second focus, and for this reason, its behavior is pre-focal rather than post-focal.

As last, we noticed that the second word of a dual focus is not different from the neutral and final focus conditions in terms of both F0 and duration. This implies that the phrasing strategy is not changed among these three conditions. Phrasing is determined by syntax and focus does not alter prosodic boundaries, as it is found in previous studies of Chinese[25], Swedish [24]and Japanese[22].

Following conclusions can be drawn for the prosodic correlates of dual focus in Mandarin Chinese:

[1] Both foci increased F0 and word duration to almost the same degree as its initial and final focus counterpart.

[2] The word after the first focus was not different from its neutral and final counterpart in either F0 or duration.

These conclusions hold for both the short and the long sentences.

5. ACKNOWLEDGEMENT

We are thankful to Prof. Caroline Féry and Prof. Yi Xu for their constructive suggestions. The research was supported by the National Natural Science Foundation of China (Grant No. 60905062) and the New Century Excellent Talents in University (NCET-12-0584).

6. REFERENCES

[1] M. Krifka, "Basic Notions of Information Structure," *Acta Linguistica Hungarica,* vol. 55, pp. 243-276, 2008.

[2] S. J. Eady, W. E. Cooper, G. V. Klouda, P. R. Mueller, and D. W. Lotts, "Acoustical characterization of sentential focus: narrow vs. broad and single vs dual focus environments," *Language and Speech,* pp. 233-250, 1986.

[3] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," *Journal of Phonetics,* vol. 27, pp. 55-105, 1999.

[4] B. Wang and Y. Xu, "Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese," *Journal of Phonetics,* vol. 39, pp. 595-611, 2011.

[5] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics,* vol. 33, pp. 159-197, 2005.

[6] Y. Xu, S.-w. Chen, and B. Wang, "Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?," *The Linguistic Review,* vol. 29, pp. 131-147, 2012.

[7] B. Wang, L. Wang, and T. Qadir, "Prosodic encoding of focus in six languages/dialects in China," presented at the 17th ICPhS, Hong Kong, 2011.
[8] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics,* vol. 36, pp. 680–703, 2008.
[9] S. Zerbian, "Investigating prosodic focus marking in Northern Sotho.," in *Focus strategies: evidence from African languages*, K. Hartmann and M. Zimmermann, Eds., ed Berlin: Mouton de Gruyter, 2007, pp. 55-79.
[10] F. Liu, "Single vs. double focus in English statements and yes/no questions," presented at the Speech Prosody 2010, Chicago (USA), 2010.
[11] E. Kabagema-Bilan, B. López-Jiménez, and H. Truckenbrodt, "Multiple focus in Mandarin Chinese," *Lingua,* vol. 121, pp. 1890-1913, 2011.
[12] Y. Jia, A. J. Li, and Z. Y. Xiong, "A phonetic and phonological analysis of dual and multiple focuses," presented at the Speech Prosody 2010, 2010.
[13] N. Chomsky and M. Halle, *The Sound Pattern of English*. Harper and Row, New York., 1968.
[14] M. Nespor and I. Vogel, *Prosodic Phonology* Foris, Dordrecht, Holland, 1986.
[15] Selkirk and Elisabeth, "Phonology and Syntax. The Relation between Sound and Structure," *Cambridge, MA: MIT Press,* 1984.
[16] E. Selkirk, "Sentence prosody: Intonation, stress, and phrasing," in *Handbook of Phonological Theory*, J. Goldsmith, Ed., ed Oxford: Blackwell, 1995, pp. 550-569.
[17] Gussenhoven and Carlos, "Focus, Mode, and the Nucleus," *Journal of Linguistics,* vol. 19, pp. 377-417, 1983.
[18] Gussenhoven and Carlos, "On the Grammar and Semantics of Sentence Accents," *Dordrecht: Foris,* 1984.
[19] T. Höhle, "Explikation für “normale Betonung” und “normale Wortstellung”," in *Satzglieder im Deutschen: Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung* W. Abraham, Ed., ed Tübingen, 1982, pp. 75-153.
[20] J. B. Pierrehumbert and M. E. Beckman, *Japanese Tone Structure*. Cambridge, MA: MIT Press, 1988.
[21] H. Nagahara, "Phonological phrasing in Japanese," University of California Los Angeles (UCLA), Ph. D. Thesis 1994.
[22] S. Ishihara, "Japanese focus prosody revisited: Freeing focus from prosodic phrasing," *Lingua,* vol. 121, pp. 1870-1889, 2011.
[23] A. E. Turk and S. Shattuck-Hufnagel, "Multiple targets of phrase-final lengthening in American English words," *Journal of Phonetics,* vol. 35, pp. 445-472, 2007.
[24] M. Horne, E. Strangert, and M. Heldner, "Prosodic boundary strength in Swedish: Final lengthening and silent interval duration," in *ICPhS*, 1995, pp. 170-173.
[25] Y. Y. Chen, "Focus and intonational phrase boundary in standard Chinese," presented at the IEEE International Symposium on Chinese Spoken Language Processing (ISCSLP 2004), 2004.
[26] S. White. (2007). http://qaparadigm.sourceforge.net/ *2007*.
[27] Y. Xu, "ProsodyPro. available from," http://www.phon.ucl.ac.uk/home/yi/tools.html, 2005-2012.
[28] P. Boersma, & Weenink, D., "http://www.fon.hum.uva.nl/praat/," ed, 2005.
[29] W. L. Chafe, "Givenness, contrastiveness, definiteness, subject, topics, and point of view " in *Subject and topic*, C. N. Li, Ed., ed: New York: Academic Press, 1976, pp. 25-53.
[30] E. F. Prince, "Toward a taxonomy of given-new information," in *Radical Pragmatics*, P. Cole, Ed., ed: New York: Academic Press, 1981, pp. 223-255.