

Prosodic encoding of topic and focus and its influence on speech perception in split sentences of Chinese Putonghua *

WANG Bei¹ Caroline Féry²

(1 *College of Chinese Minority Language and Literature Minzu University of China Beijing 100081*)

(2 *Johann Wolfgang Goethe-Universität Frankfurt am Main Germany*)

Received Sept. 23, 2010

Revised Oct. 25, 2010

Abstract This study is an investigation of the prosodic encoding of split noun sentences in Chinese Putonghua, for instance, “shu, wo mai le san ben. (Book, I buy ASP three CLAS. ‘I bought three books’)”, in which syntactic fronting highlights the split noun. The question-and-answer paradigm was used to construct contexts where the split noun is either the topic or the focus of the sentence. Acoustic analysis of 280 split sentences read by seven speakers show that the maximum F0 of the base part is higher and the pause after the split noun is shorter in the topic condition than that in the focus condition. But the split noun itself does not differ in either F0 or duration across the two conditions. A perception experiment further shows that the difference in prosody between the two conditions is perceivable, since matched question-and-statement pairs are preferred over unmatched ones.

PACS numbers: 43.70

1 Introduction

Speech communication is a process of exchanging information. It is widely accepted that information structure can be conveyed by means of prosody and syntax.

Syntactic fronting is regularly involved in the highlighting of part of a sentence. The structure is called split sentence or discontinuous construction^[1-3]. An example in Mandarin Chinese is: “shu, wo maile san ben (book, I buy ASP three CLAS, ‘I’ve bought three books’)”. Split structure has been studied in the field of syntax, morphology, semantics and phonetics, especially a cross-linguistic study by Fanselow and Féry^[3] on the split structure of over 100 languages.

Fanselow and Féry^[3] among many other researchers^[1,4-5] have suggested that the splitting of a nominal phrase may convey two different information structures. First, a split noun is the narrow focus of the sentence when the base part, which is the known information, is deaccented;

* This work was supported by the DFG project SFB632, the National Natural Science Foundation of China (60905062) and the “211” Foundation of Minzu University (MY0116).

second, a split noun is the (contrastive) topic while the base part bears a focus. In the latter case, the topic and focus are prosodically marked as “hat-pattern” in languages such as German and Hungarian^[3]. A topic introduces what the remaining sentence is about^[1,7]. For instance, the proposition “wo tingshuo ni maile shu he bi. (I heard that you’ve bought some books and pens)”, may induce “shu, wo maile san ben. (Book, I buy ASP three CLAS)” as a reply. This is equivalent to the English structure, “As for books, I bought three of them” in which shu (book) is the topic of the sentence. Based on syntactic analysis of split constructions, Pan and Hu^[6] have argued that in such construction, the presence of a focus in the base part is an important or even necessary condition, which indicates the connection between the separated head noun and the base part. Besides, split structures can also be used as answers to wh-questions. For instance, in the answer of the question “shenme ni mai le san ben? (What did you buy three?)”, “shu (book)” is the focus of the sentence, i.e. the most prominent part in a sentence^[8–9]. In both contexts, the noun is taken away or ‘split’ from its base position (beside ‘three’) in order to highlight the corresponding information.

It has been found that the fronted (or split) constituent is either realized with a low boundary tone (L%) when the constituent is a focus, or with a high boundary tone (H%) when it is a topic in German^[4] and French^[5]. In Mandarin Chinese, however, the difference in prosodic marking of topic and focus is that there is a large F0 lowering after a focus, but not after a topic^[12]. Wang and Xu^[12] also found that when a sentence initial word is focused, it has a higher maximum F0 and longer duration than when it is the topic of the sentence.

In addition, Fanselow and Féry^[3] have suggested that there is an additional prosodic phrase in a split structure. However, no experimental study was conducted to find out how information structure adjusts the pause duration between the prosodic phrase and the split constituent.

In speech perception, prosody also provides additional information beside syntactic and semantic information^[16]. A number of studies have demonstrated that prosodic variation is perceivable in marked constructions, such as focus and question^[17–18]. This paper investigates whether prosodic encodings, if any, are sensitive cues to distinguish split sentences in a topic context from those in a focus context.

1.2 Research questions

Two general research questions are the subject of this study. These questions were investigated in a production and in a perception study respectively.

(1) What are the differences in prosodic encoding between the split noun as the topic and as the focus of a sentence?

(2) Does prosodic variation facilitate perceptual acceptability of split sentences in the relevant topic and focus contexts?

2 Prosodic encoding of split sentences

2.1 Methods

2.1.1 Reading materials

The speakers were asked to read a target sentence in a natural way as an answer to a

preceding question. For each base sentence, contexts of focus and topic were constructed (in line with the terms used in [8–9,13,17]). In the topic context, two objects were mentioned in a question like “Wo tingshuo Baoxin diule shuqian he bijiben, shi ma? (I heard that Baoxin lost bookmarks and notebooks. Is that true?)”, and the answer mentioned one of the objects, for example ‘bookmarks’, which was then the topic in the answer “shuqian, Baoxin diu le san zhang. (As for bookmarks, Baoxin lost three)”. In such a context, “shuqian (bookmarker)” is given and it is the topic. At the same time, “shuqian (bookmarks)” has a contrastive interpretation, ~~because it mentions only one element asked for in the question~~. The second element “bijiben (notebooks)” is not expressed anymore, although it nevertheless triggers a contrast with “shuqian (bookmarks)”. The focus context was introduced simply by a *wh*-question like “What were the three objects that Baoxin lost?”. In the answer, “shuqian (bookmarker)” is then the information focus. In the latter context, the split noun “shuqian, (bookmarker)” is the focus, and it is new. Thus an important difference between the two contexts is that in the topic context the modifier in the base part (“three”) is new (and focused), whereas it is old in the focus context.

All the experimental sentences were in the SVO word order, and they contain 8–11 syllables. In each sentence, the subject was a disyllabic person’s name, the verb was disyllabic or trisyllabic, the object contained a modifier (3–4 syllables) and a disyllabic noun. For each lexical tone, i.e., H (tone 1), R (tone 2), L (tone 3), and F (tone 4), five sentences were constructed with all syllables in a sentence bearing the same tone except for functional words such as “de” and “le”. There were 5 (base sentences) \times 4 (tones: H, R, L, and F) \times 2 (contexts: topic, focus) = 40 sentences in total. The background questions were read by the first author, a 33-year-old female native speaker of Mandarin Chinese. Another 120 filler sentences were used in the experiment, including the sentences with canonical word order and some additional split and canonical sentences. All speakers read the test materials in different randomized orders.

2.1.2 Participants

Seven native Mandarin Chinese speakers from Northern China uttered the sentences, two males and five females, 22–26 years old. They were non-linguistic students at the University of Potsdam. They did not report any hearing or speaking disorder.

2.1.3 Recording

The pre-recorded questions were presented to the participants before they read the target sentences. They were asked to read the sentences at a normal speech rate in a natural way. All the sentences were printed out on paper. The speech was directly digitized onto a computer hard disk in IBM-T40 Laptop with built-in A/D converter of 16 bit at a sampling rate of 22 kHz. Ten sentences were used for practice and were not considered in the experimental results. Total recording time was about 40 minutes.

2.1.4 Speech data annotation

Praat© was used for speech data annotation. F0 was hand checked for every vocal cycle

using Xu's Praat script^[19], which converted the duration of vocal cycles into F0 values automatically. The syllable boundaries were manually labeled. Acoustic parameters i.e., maximum F0, minimum F0, word duration and pause duration, of the subjects, verbs, modifiers and nouns were calculated. The F0 value was converted from Hz to st (semitone) by formula (1), in which the reference F0 is 50 Hz for all speakers.

$$f_{st} = 12 \log_2(f_0/50). \quad (1)$$

2.2 Results

2.2.1 Statistic analysis of F0 variation

Maximum and minimum F0 of all words are graphically displayed in Fig. 1 and Fig. 2 respectively. All values in the two figures are the mean of 140 sentences read by seven speakers. The dash lines are the values of the topic condition, and the solid lines are those of the focus condition. For minimum F0, sentences with all H tones were excluded, since the H tones do not have any point reflecting the lower F0 variation within the pitch range.

Two factors were statistically tested using repeated measures ANOVA, that is, tone (H, R, L, and F) and context (topic vs. focus). The F and p values of the statistic tests in maximum F0 and minimum F0 of each word are presented in Table 1 and Table 2 respectively. It is of primary relevance for this paper to find out whether the context brings significant differences in the corresponding values. The difference caused by lexical tone is beyond the investigation of this paper and hence will not be covered in detail.

Two observations emerge from Fig. 1 and Fig. 2, with statistical test presented in Table 1 and Table 2. First, the split noun itself does not show any difference in maximum F0 or

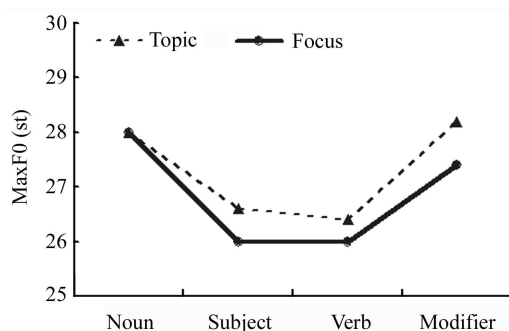


Fig.1 Maximum F0 of all words in split sentences under the focus and topic condition.

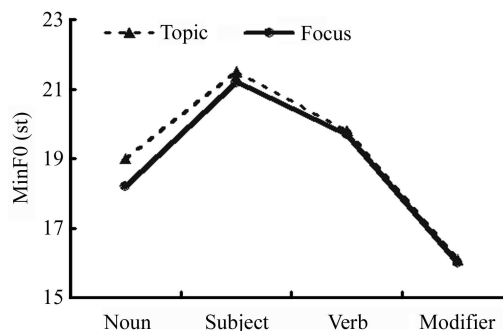


Fig. 2 Minimum F0 of all words in split sentences under the focus and topic condition.

Table 1 Two-factor repeated measures ANOVA in maximum F0 of all words in the split sentences.

	Context $F(1, 12) =$	Tone $F(3, 12) =$	Interaction $F(3, 12) =$
Noun	0.027	15.863***	0.45
Subject	6.597*	14.726***	1.362
Verb	6.25*	22.683***	0.864
Modifier	10.369*	28.268***	2.372

Table 2 Two-factor repeated measures ANOVA in minimum F0 of all words in the split sentences.

	Context $F(1, 12) =$	Tone $F(3, 12) =$	Interaction $F(3, 12) =$
Noun	2.871	23.072***	1.438
Subject	4.955	10.097*	2.056
Verb	0.852	21.836***	0.82
Modifier	0.096	4.432*	2.699

minimum F0 between the two contexts. Second, maximum F0 of the base part is significantly higher in the topic condition than that in the focus condition, but no difference is found in minimum F0. It is worth emphasizing that the modifier in the topic condition has higher F0 than that in the focus condition, because it is new in the former case.

To summarize, F0 of the split noun itself does not differ between the two conditions. However, in the topic condition, the base part has higher F0 than its focus counterpart, which is reflected on maximum F0, but not on minimum F0.

2.2.3 Statistic analysis of durational variation

Figure 3 illustrates mean duration of all words and the pause between the split noun and the base part with topic and focus conditions overlaid in one figure. Similar repeated measures ANOVAs of duration of each word and the pause were carried out, with context and tone as two independent factors. The corresponding F and p values are presented in Table 3.

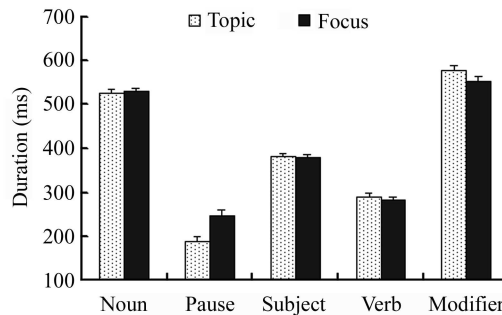


Fig. 3 Duration of all words and the pause after the split noun in split sentences under the focus and topic condition.

Table 3 Two-factor repeated measures ANOVA in duration of all words and the pause after the split noun in the split sentences.

	Context $F(1, 12) =$	Tone $F(3, 12) =$	Interaction $F(3, 12) =$
Noun	0.281	5.997**	3.643
Pause	59.631***	0.693	0.693
Subject	0.239	6.678**	0.234
Verb	2.711	46.971***	0.186
Modifier	30.078**	82.964***	0.107

As can be seen in Table 3, the context only affects the pause and the modifier, but not the other words. Looking at Fig. 3, we can see that the pause is longer when the split noun is a focus than when it is a topic (248 ms vs. 189 ms).

In addition, the modifier is about 5% longer when it is new (in the topic condition) than when it is old (in the focus condition) (579 ms vs. 554 ms).

Finally, we need to mention that the duration of the split noun does not differ between the two conditions.

2.2.4 Summary

In the two information structural conditions examined here, i.e., when the split noun is a topic or a focus, the split noun itself does not differ in F0 and duration. The two information structural patterns have a different implementation in the base part: in the topic context the maximum F0 of the base part is higher. As for durational adjustment, the pause after the split noun is shorter and the modifier is longer in the topic context than in the focus context.

3 Speech perception experiment

3.1 Method

3.1.1 Task

A perception experiment was conducted in order to test whether native speakers of Mandarin Chinese could interpret the prosodic differences identified in the production experiment. The task was to estimate the appropriateness of the answers to the questions in terms of intonational variation on a 10-point scale (with 10 being the best). Question-and-answer pairs were played in sequence to the participants. The question was either a focus question or a topic question. The answer was either a canonical sentence or a split sentence in either the topic or focus condition recorded in the speech production experiment. The canonical sentences served as a reference to the split sentences. For each base sentence, there were eight question-and-answer pairs. An example is presented in (4).

(4) An example of the sentences used in the perception experiment.

Topic Question: wo tingshuo Baoxin diule shuqian he bijiben, shi zhen de ma?

(I heard that Baoxin lost bookmarks and notebooks, is that true?)

Focus question: Shenme Baoxin diule san zhang?

(What are the three objects that Baoxin lost?)

Split answer: The split sentences were read either in the topic or in the focus context.

Shuqian, Baoxin diule sanzhang.

(As for bookmarker, Baoxin lost three of them.)

Canonical answer: The canonical sentences were read either in the topic or in the focus context.

Baoxin diule sanzhang shuqian.

(Baoxin lost three bookmarks.)

A female speaker was randomly chosen, and the sentences she had produced in the production experiment were used for the perception experiment. Seven basic sentences were selected for the perception experiment under the condition that the acoustic variation was very close to the averaged results reported in section 2. The participants were allowed to listen to each question-and-answer pair for a maximum of two times, and were then asked to make a judgment. To ensure reliable results, the split sentences were graded three times. In the first series, the split sentences were mixed with the canonical sentences, and in the second and the third series, the split sentences were presented and graded twice in different randomized order. Altogether, the experiment lasted approximately one hour.

3.1.2 Participants:

Twenty-one participants from Beijing Normal University took part in the experiment. They were all native Mandarin Chinese speakers, 20–25 years old and did not report any hearing problems.

3.2 Results:

The values of the acceptability in the four conditions are shown in Fig. 4. This represents the average score of the 21 participants in the three experiment sessions.

Table 4 shows the results of the repeated measures ANOVA on the three factors, i.e., Question (topic vs. focus), Sentence (canonical vs. split) and Answer (read in topic condition vs. focus condition).

Table 4 Two-factor repeated measures ANOVA in duration of all words and the pause after the split noun in the split sentences.

	<i>F</i>	<i>p</i>
Sentence type	29.287***	< 0.001
Question	15.943***	0.001
Answer	32.061***	< 0.001
Sentence type × Question	10.605**	0.004
Sentence type × Answer	13.197**	0.002
Question × Answer	19.513***	< 0.001
Sentence type × Question × Answer	11.023**	0.003

As can be seen in Fig. 4, the three main effects presented in Table 4 can be explained as follows: (1) The main effect on Sentence shows that split sentences are not as well accepted as canonical sentences. (2) The main effect on Question shows that the acceptance is lower when the sentence is an answer to a topic question than to a focus question. (3) The main effect on Answer shows that a sentence in which the split noun is the focus is better accepted than when it is the topic.

The interactions can be explained as follows (see Fig. 4). (1) The interaction between Question and Sentence indicates that canonical sentences are better accepted than split sentences.

(2) The interaction between Question and Answer indicates the matched question-and-answer is rated higher than an unmatched pair to a focus question, except in the case of a topic question. (3) The interaction between Sentence and Answer indicates the acceptance is higher for matched question-and-answer pairs than unmatched pairs in canonical sentences, but these differences are small in split sentences. (4) The overall interaction among Sentence, Question and Answer implies that the judgment is affected by all three factors.

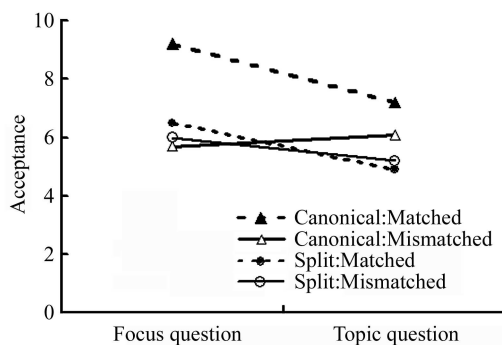


Fig. 4 Acceptance degree of sentences with split and canonical order, showing the conditions of whether the question is a topic or a focus question and if the answer is matched or mismatched.

A simple-effect test shows that the matched question-and-answer has a higher rate than an unmatched pair to a focus question, which holds true in both canonical (9.2 vs. 6.1, $F(1, 46) = 28.055$, $p < 0.001$) and split sentences (6.5 vs. 6, $F(1, 46) = 6.753$, $p = 0.017$). However, in a topic question, the matched pair gets a lower rate than an unmatched pair in canonical sentences (5.7 vs. 6.1, $F(1, 46) = 4.449$, $p = 0.048$) and the difference does not reach a significant level in split sentences (5.2 vs. 4.9, $F(1, 20) = 1.968$, n.s.).

To summarize, the perception results show that matched question-and-answer pairs are preferred in the focus context. It implies that the listeners are aware of the acoustic difference of the split sentences between the topic and focus condition. However, it may be hypothesized that the listeners are not familiar with the topic context (see below for an elaboration). If this is the case, the matched pair does not show any preference.

4 General discussion

This study provides experimental evidence that both syntax and prosody are active in marking information structure. In addition, the corresponding prosodic variation is perceptibly meaningful.

4.1 The prosodic encoding of split sentences

One of the most important results of the production experiment is that maximum F0 of the base part is higher in topic context than in focus context (see Fig. 1). The possible explanation is that the split noun exhibits post-focus lowering in the focus condition, even though there is a

long pause after the focus. In other words, the domain of a focus is the whole sentence, whereas that of a topic is only on itself. This is consistent with the findings of Wang and Xu^[12] that a topic does not lower F0 of the following words. It is worth mentioning that the post-focus lowering in split sentences is not as big as that in canonical sentences as found in Xu^[13] and Wang and Xu^[12]. The reason may lie in the opposite effect by pitch-reset at the beginning of the base part (see Fig. 2). In line with the parallel encoding mechanism as proposed in PENTA intonational model^[24–25], our results show that when both post-focus lowering and pitch reset at prosodic boundaries are applied, the two effects are added, which causes a much less lowering effect after the focus.

The prosody of the two contexts differs in the modifier as well, as it is realized with higher F0 and longer duration in the topic condition than in the focus condition. The modifier is new and focused in the topic context since it is not mentioned in the question, hence it is accented; however, it is given in the focus context, hence deaccented. This result is in accordance with Pan and Hu^[6].

As for the split noun, it does not differ acoustically between the topic and focus conditions. The split noun is prominent in both cases: either it is a contrastive topic (as mentioned in the introduction) or it is a focus. The results show that the two realizations are identical.

Minimum F0 in the base part (see Fig. 2) does not differ between topic and focus contexts. It is in agreement with Shen^[26] and Wang^[14,20] that minimum F0 is mostly used on indicating prosodic phrase structure, whereas the encoding of topic and focus does not apply in minimum F0.

As for durational adjustment due to information structure, the split sentences in the topic context show shorter pause after the split noun than those in the focus context. In the topic context, the question is like “Wo tingshuo Baoyin diu le shuqian he bijiben, shi ma? (I heard Baoxin lost bookmarks and notebooks. Is that true?)”, then the modifier (‘three’ in the example) is what the listener expects to know about the noun (shuqian ‘bookmark’). Without hearing the modifier, the addresser would consider that the answer is not complete yet, so that the base part is required to be closely connected with the split noun. In the focus contexts, instead, the answer with only the noun already provides enough information for the question like “shenme Baoxin diu le san zhang? (What were the three objects Baoxin lost?)”. The base part is simply a repetition of the question so that it can be separated from the noun. This differs from a language as English or German, which not only assign a different shape in the tonal contour of a topic and a focus, but which produce a clear phrasing between the fronted part and the base part only in the case of the topic condition. Only in this case the base part contains a second prominent constituent, a focused one. When the fronted part is focused, the base part is given and deaccented. It is then not phrased separately from the fronted part (see in Ref. [3]).

In summary, focus highlights the most prominent information in a sentence, and the domain of a focus is generally the whole sentence. In a split sentence, when the split noun is the focus, it lowers F0 of the base part even though there is a long pause after the focused word. Topic, on the other hand, is to draw the listener’s attention to the initial position of an utterance^[27].

It is realized with raised F0 at the initial position and the remaining intonational contour is left intact.

4.2 The perception of split sentences

The perception experiment confirms that the different prosodic encodings found in the production experiment are perceivable by the listeners. In the focus context, the listeners preferred matched question-and-answer pairs. It indicates that an answer to a focus question is preferred when realized with a longer pause after the split noun and lower maximum F0 in the base part. More experiments are required to examine how the two factors work together to influence the perceptive judgment.

Another finding of the perception experiment is that the acceptance of a topic context is lower than that of a focus question. It could be that the participants had problems understanding the question when designed to introduce a sentence with a certain word as a topic. The relationship between focus and F0 variation is well established. However, the prosodic cues of topic are more complicated. In addition, the domain of a topic is not easy to detect. A topic can be either just the word or the whole sentence. Another reason that listeners did not like topic context might be that the answer was only partial. For a topic question like “I heard you bought books and pencils” a full answer could be ‘Books, I bought three. Pencils, I bought two.’

To sum up, our study give some elements of the production and perception of topics in Mandarin Chinese, especially in the context of discontinuous nominal phrases, but more work on this kind of structure remains to be done, especially in topic contexts.

5 Conclusions

It has been shown that syntactic fronting is effective in highlighting information (see for instance the work of Pan and Hu^[6]). This article demonstrates that prosody also plays a role for the same purpose in speech communication. Topic and focus are both prominent in a sentence; however, they are encoded differently in prosody, and this difference is perceivable. This indicates that topic and focus are different in terms of communicative function. More specifically, focus highlights the focused constituent from the rest of a sentence, whereas topic draws the listeners’ attention to the beginning of an utterance.

The conclusions are as follows:

(1) The F0 value and duration of the split noun itself does not differ between the two conditions.

(2) The base part bears a focus in the topic context, and is given in the focus context. As a result, in the topic condition, the base part has higher F0 than its focus counterpart, which is reflected in maximum F0, but not in minimum F0. In addition, the pause between the base part and the split noun is shorter in the topic context than in the focus context.

(3) Listeners are aware of the acoustic difference between the split sentences in the topic and the focus condition. They prefer matched question-and-answer pairs in the focus context.

However, the matched pair does not show any preference in the topic context. The reason might be that the context eliciting the topic sentences are difficult to interpret.

6 Acknowledgements

We are particularly grateful to Yi Xu and Shinan LÜ for their constructive comments. We would also like to thank Lingling Su for the work of speech annotation, Torsten Roloff for the data-format-converter program, Prof. Hua Shu at Beijing Normal University for providing the facilities to run the perception experiment, and Stuart Howe for English-proof reading. All remaining errors are ours.

The experiments and the first draft were conducted at the University of Potsdam, supported by the DFG project SFB632: Information Structure (D2), The writing was completed at the Minzu University of China, supported by the National Natural Science Foundation of China (60905062) and the “211” Foundation of Minzu University (MY0116) to the first author.

References

- [1] Büring D. Topic. In: Bosch P, van der Sandt R eds. *Focus: Linguistic, Cognitive, and Computational Perspectives*, Cambridge Press, 1999: 142–165
- [2] Fanselow G, Cavar D. Distributed deletion. In: Artemis Alexiadou ed. *Theoretical Approaches to Universals*, Amsterdam-Benjamins. 2002: 65–107
- [3] Fanselow G, Féry C. Prosodic and morphosyntactic aspects of discontinuous noun phrases: a comparative perspective. In: C. Féry, van de Vijver eds. *Discontinuous Noun Phrase*, Cambridge University, 2009: 213–237
- [4] Féry C. The prosodic basis of topicalization. In Schwabe K, Winkler S eds. *On Information Structure, Meaning and Form*, Amsterdam- Benjamins, 2006: 73–92
- [5] Doetjes J, Delais-Roussarie E, Sleeman P. The prosody of left detached constituents in French. In: *the Proceeding of Speech Prosody*, Japan, 2002: 247–250
- [6] Pan H H, Hu J H. Head noun movement, focus, and topicalization in Mandarin. In: Xu, Jie, Donghong Ji, Kim Teng Lau eds. *Syntax and Semantics*, Singapore, 2003: 119–156
- [7] Lambrecht K. *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge Press, 1994: 27–35
- [8] Halliday M A K. Notes on transitivity and theme in English. *Journal of Linguistics*, 1967; **3**: 199–244
- [9] Bolinger D L. Accent is predictable: if you’re a mind reader. *Language*, 1972; **48**: 633–644
- [10] Eady S J, Cooper W E. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, 1986; **80**: 402–415
- [11] Féry C, Kügler F. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, 2008; **36**: 680–703
- [12] Wang B, Xu Y. Prosodic encoding of topic and focus in Mandarin. In: *Proceedings of Speech Prosody 2006*, Dresden, Germany: 1013–1016
- [13] Xu Y. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 1999; **27**: 55–105
- [14] WANG B, LÜ S N, YANG Yufang. The pitch movement of stressed syllable in spoken Chinese, *Acta Acustica* (In Chinese). 2002; **27**(3): 234–240

-
- [15] Chen Y Y, Gussenhoven C. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 2008; **36**: 724–746
- [16] Hruska C, Alter K. Prosody in dialogues and single sentences: how prosody can influence speech perception. In: Anita Steube ed. *Information Structure: Theoretical and Empirical Aspects*, Walter de Gruyter-Berlin-New York Press, 2004: 211–226
- [17] Liu F, Xu Y. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 2005; **62**: 70–87
- [18] Chen S W, Wang B, Xu Y. Closely related languages, different ways of realizing focus. *Interspeech 2009*, Brighton, UK, 2009: 1007–1010
- [19] Xu Y. TimeNormalizeF0.praat, <http://www.phon.ucl.ac.uk/home/yi/tools.html>, 2005–2008
- [20] WANG B, YANG Y F, LÜ S N. Acoustic-phonetic analysis on the boundaries of prosodic hierarchical structure of Chinese sentences, *Acta Acustica (In Chinese)*, 2004; **29**(1): 29–36
- [21] Ladd D R. *Intonational phonology*. Cambridge Studies in Linguistics 79. Cambridge: Cambridge University Press, 1996: 160–167
- [22] Cooper W E, Eady S J, Mueller P R. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, 1985; **77**: 2142–2156
- [23] Xu Y. Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 2005; **46**: 220–251
- [24] Xu Y, Xu C X. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 2005; **33**: 159–197
- [25] Prom-on S, Xu Y, Thipakorn B. Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America*, 2009; **125**: 405–424
- [26] Shen J. Type of intonation in Mandarin Chinese. *Fang Yan*. 1994(3): 221–228
- [27] Umeda N. “F0 declination” is situation dependent. *Journal of Phonetics*, 1982; **10**: 279–290