



Influence of Contextuality on Prosodic Realization of Information Structure in Chinese Dialogues

Bin Li^{1,2}, Yuan Jia²

¹ University of Chinese Academy of Social Sciences

² Institute of Linguistics, Chinese Academy of Social Sciences

zibolibin@126.com, summeryuan_2003@126.com

Abstract

In this paper, we present a detailed investigation on the influence of contextuality on the prosodic realization of information structure in Chinese dialogues. The materials were selected from the 863 corpus, which contains both isolated sentences and spontaneous dialogues. RefLex was selected as the annotation scheme, which differentiates information structure on the lexical and referential levels. Prosodic data (including duration and pitch range) from 12 groups of spontaneous dialogues were analyzed with the linear mixed effects mode, and each of them consists of 13-22 turns. The isolated sentences corresponding to these dialogues were also analyzed. The analysis results reveal the influence of contextuality. Specifically, the features of prosodic realization of information structure on the lexical and referential levels show a contrary tendency. The statistical analysis indicates that the speakers use duration and pitch ranges as phonetic cues to distinguish information structures on both levels. On the other hand, duration on the referential level is the only phonetic cue affected by contextuality.

Index Terms: contextuality, information structure, prosodic realization, interactional linguistics

1. Introduction

1.1. Previous studies

Throughout much of the twentieth century, modern linguistics suffered from a written language bias. By contrast, interactional linguistics grew out of an interest in spoken language and a desire to study it in its natural habitat: social interaction. Interactional linguists took seriously the observation that “some of the most fundamental features of natural language are shaped in accordance with their home environment in co-present interaction, as adaptations to it, or as part of its very warp and weft.” [1]

Since Couper and Selting [2] formally proposed the concept of interactional linguistics, this line of research has gradually become a new research hotspot in the field of linguistics. The interactional linguistics is mainly concerned with two sorts of questions that implicate a language: (i) what linguistic resources are used to articulate particular conversational structures and fulfill interactional functions? (ii) What interactional function or conversational structure is furthered by particular linguistic forms and ways of using them? [2]

In recent years, the study of interactional linguistics based on Chinese has achieved some fruitful results. Wan [3] found that the minor sentence is the locus of interaction based on the

three factors: next-turn onset, joint utterance completion, and turn unit extension. While, Yue [4] thought the turn-constructive unit with falling pitch or pause in the final syllable, is the most relevant turn-projection unit. Based on TV talk shows, Xie and Fang [5], Chen and Ma [6] studied the prosodic behaviors of conjunctions in Chinese conversations and the prosodic features of turn-takings respectively. In addition, other basic topics of interactional linguistics, such as syntactic choices, formal verification and multimodal studies, were also conducted [7].

All the findings point to the fact that the characteristics of language are influenced by the contexts of social interaction, in which the language is used. Thus the issue of contextuality should be at the heart of international linguistics research. However, in the previous studies, less attention has been paid to the specific realization of prosody during language interactions.

1.2. The present study

To fill the gap, this study will examine the impact of contextuality on the realization of language features. To some extent, the interaction between speakers can be understood as the transmission and sharing of information.

Halliday brought forward the definition of information structure and proposed that information structure is a structure that organizes language into information units [8]. He thought that the information unit is a structure made up of two functions, the new and the given. In an idealized form, each information unit consists of a given element accompanied by a new element [9]. Different from Halliday's dichotomy of information structure, Chafe [10] took the concepts of recoverability of given information and un-recoverability of new information, further into the cognitive field, by establishing the relationship between the information and the consciousness of the speaker and hearer. He added an intermediate state as accessible, which in a sense is new to context, but is cognitively related to what has already appeared. A widely accepted classification scheme [11] also divides information structure into three categories as Chafe did: new, given and accessible. This scheme is also applied in the present research.

Previous studies reveal that prosody conveys discourse-level information, but the extent to which prosodic cues distinguish different kinds of information-structural concepts remains unclear. Iris and Elsi [12] conducted a production study on Mandarin and found that new information is encoded differently from corrective focus: only corrective focus is associated with intensity range expansion. The new-given distinction was reflected only in lengthening and F_0 range expansion. The research by Bi et al. [13] was also based on

Chinese and their results showed no durational and F_0 differences between informational focus and givenness, and no differences in intensity range between informational focus and corrective focus.

In this paper, two kinds of contextuality are considered: the presence of context and the absence of context. We conduct a detailed study on dialogues to investigate (1) whether the contextuality affects the prosodic realization of new, given and accessible information; and (2) if so, how does contextuality affect them. Besides, since the natural dialogues language environment is complex and the data processing is difficult, this study firstly analyzes the dialogues that were recorded based on real corpus.

2. Methodology

2.1. Data collection

The materials adopted in this study were selected from the 863 corpus, which was constructed by the Phonetic Lab, Chinese Academy of Social Science. Dialogues on different topics transferred from natural Chinese dialogues are recorded by two male and two female actors. Besides, the dialogues' order was disrupted, and each actor was asked to read his/her out-of-order part. Thus, we can control presence or absence of the context in our studies. All recordings were conducted with 16 kHz sampling rate and 16bit rate.

In this study, we choose 12 groups of dialogues and each of them consists of 13-22 turns. The total duration is about 10 minutes. The dialogues involve a range of topics, including buying a house, asking for directions, booking a room, etc. There are also 130 isolated sentences corresponding to the dialogues analyzed in this paper.

In the selection of prosodic parameters, we draw on the existing research results. Acoustically, there are some dimensions that are commonly regarded as providing cues about information structure: such as duration, pitch and so on. Regarding pitch, most researchers agree that in Mandarin, information structure is conveyed not by the shapes of the contours of pitch (as is the case in English), but rather by their ranges [12]. Therefore, duration and pitch range are selected as the main phonetic cues in this paper.

2.2. Annotation scheme

Phonetic data were annotated manually with Praat. RefLex annotation system proposed by Riester and Baumann [11] was used to annotate information structure. We also made some adjustments in line with the actual situation of our materials according to Jia's research [14].

The RefLex system is a double-layer annotation scheme, so as to distinguish information structures on lexical and referential levels in detail, corresponding to external expression and internal relationship in dialogues respectively. Based on Chafe's theory [10], the information structure is divided into three categories on both levels: given information for active state, accessible information for semi-active state, and new information for inactive state.

The annotation objects are mainly notional words on the lexical-level, and determiner, nominal and prepositional phrases on the referential level. Besides, according to [11], if an entity occurs earlier than the previous five intonation phrases (given that the prosodic information is available) or clauses (in written texts), then the entity will be reanalyzed,

but the authors also admitted that the choice of a distance of five units is arbitrary to a certain degree. Considering the actual situation of our analyzed materials, we only annotated nouns on the lexical level and we do not conform to the "distance principle" mentioned above. The annotation of information structure of the out-of-order dialogues is consistent with that of the normal dialogues. The annotation tags and examples of annotation are shown in the following tables:

Table 1: Annotation tags of information structure

Lexical level	Referential level
L-NEW	R-NEW
L-ACCESSIBLE	R-BRIDGING
L-GIVEN	R-GIVEN

Table 2: Examples of annotation

Level	Example
Lexical	A:你来的真早。(You came early today.) B:昨天的 作业 (L-NEW)忘了做了,这不,正在赶呢。(I forgot to do the homework yesterday. I am taking the time to write it.) A:哎呀!我把 作业 (L-GIVEN)忘在 宿舍 (L-NEW)里了。(Oh! I forgot my homework in the dormitory .) B:那赶快去拿吧!(Go and get it.)
	A:你来的真早。(You came early today.) B: 昨天的作业 (R-NEW)忘了做了,这不,正在赶呢。(I forgot to do the homework yesterday. I am taking the time to write it.) A:哎呀!我把 作业 (R-BRIDGING)忘在 宿舍 (R-NEW)了。(Oh! I forgot my homework in the dormitory .) B:那赶快去拿吧!(Go and get it.)

2.3. Data analysis

The pitch values and duration were extracted through Praat script and collected into Microsoft Office Excel. The duration of each target entity was extracted according to the onset and offset of the entity. F_0 value was sampled at 10 equidistant measurements of each syllable. In order to eliminate the physiological differences caused by age and gender, the original pitch values are normalized with the LZ-Score formula (Zhu [15]):

$$Z_i^t = \frac{y_i - m_y}{s_y} \quad (1)$$

Within the formula, $y = \log_{10} X_i$, m_y and s_y are the mean and standard deviation of y_i ($i=1, 2 \dots n$), respectively. Pitch range was calculated by subtracting the minimal normalized pitch value from the maximal one of each entity. Similarly, all durations are calculated through the following formula:

$$T_i = \frac{(T_i - T_{min})}{(T_{max} - T_{min})} \quad (2)$$

Where T denotes the duration, and T_i ($i=1, 2 \dots n$) refers to an individual entity's duration. T_{max} and T_{min} are the maximum and minimum of the durations respectively, among entities.

In our study, duration and pitch range were submitted for statistical analysis with the linear mixed effects model using R with *lme4* package [16]. The new, accessible and given information, the presence/absence of context, and their

interaction were selected as effects for investigation; speakers was entered as random effect. The interaction effect could reflect the divergence of prosodic realization of information structure under different contextual conditions. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question.

3. Results

This section reports our results of the study from which we can reveal the influence of contextuality on the prosodic realization of information structure in Chinese dialogues. The analyses are conducted on 367 entities on lexical level and 417 entities on referential level, which are annotated and processed as introduced in Section 2. Some statistics are illustrated in Figure 1. Next, the results on lexical and referential levels are reported respectively.

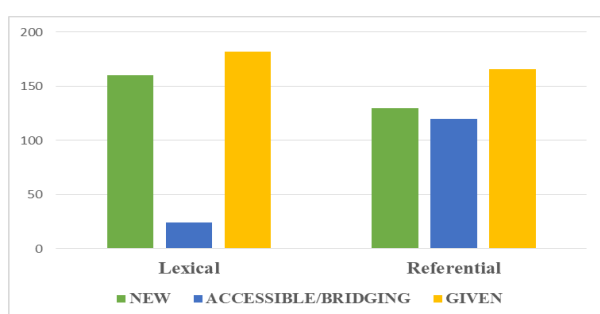


Figure 1: Numbers of different types of information structure

3.1. Lexical Level

3.1.1. Pitch range

Pitch ranges of entities with different information structures on the lexical level are shown in Figure 2. A general pattern can be noticed from the boxplot. In the presence of context, speakers realize new information through a larger pitch range compared with accessible and given. While in the absence of context, the pitch ranges of new and accessible information have been compressed, especially for the accessible ones. For given information, its pitch ranges do not change much, but its pitch register has improved significantly and shows a similar pattern as new information.

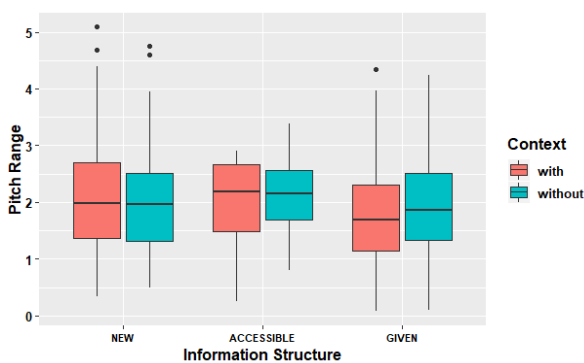


Figure 2: Pitch ranges of entities with different types of information structure on the lexical level.

The statistical analysis results show that, pitch ranges of entities are significantly affected by information structure ($\chi^2(3) = 12.442, p = 0.006$), but are not affected by contextuality ($\chi^2(1) = 0.6195, p = 0.431$). Besides, there is no interaction between the information structure of the investigated entities and the contextuality ($\chi^2(2) = 0.8552, p = 0.652$).

3.1.2. Duration

Durations of entities with different information structures on the lexical level are shown in Figure 3. As the boxplot suggests, contextuality has different effects on new, accessible and given information: the durations of new and accessible are shorter in the absence of context, while given information shows the opposite change, its duration has increased slightly. What's more, the durations of accessible, instead of new ones, are the longest among the three types under the different conditions.

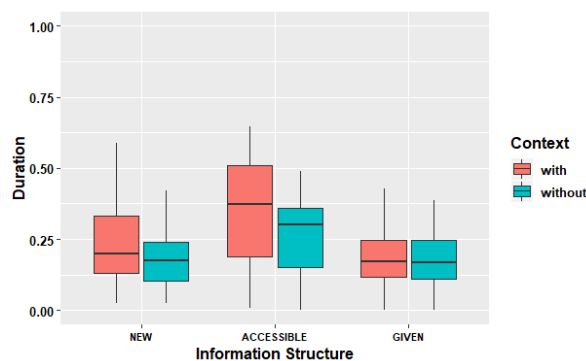


Figure 3: Durations of entities with different types of information structure on the lexical level.

Durations of entities are significantly affected by both information structure ($\chi^2(2) = 9.8126, p = 0.007$) and contextuality ($\chi^2(1) = 4.6111, p = 0.0318$). Similar to results on pitch ranges, there is no interaction between the information structure of the investigated entities and the contextuality ($\chi^2(2) = 1.6456, p = 0.439$).

To sum up, information structure significantly affects both pitch ranges and durations, while contextuality affects durations alone, and no interaction effects can be noticed. That is, on the lexical level, duration and pitch range are the phonetic cues to distinguish information structures, but the presence or absence of context does not affect the prosodic realizations of information structures.

3.2. Referential Level

3.2.1. Pitch range

Pitch ranges of entities with different information structures on the referential level are presented in Figure 4. Contrary to the performance on the lexical level in the absence of context, the pitch ranges of new and accessible information became larger. The range of the given information has not changed and its pitch register has been reduced. Another difference compared to the lexical level is that the pitch ranges of new are more compressed than given ones in the presence of context.

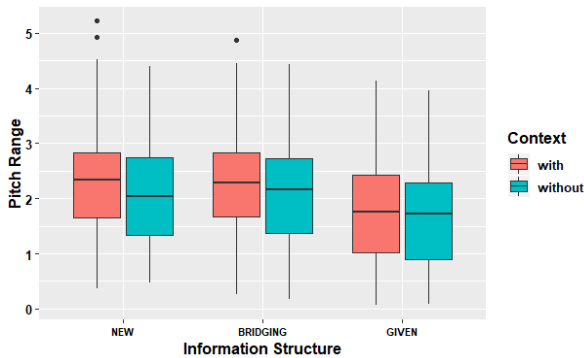


Figure 4: Pitch ranges of entities with different types of information structure on the referential level.

Similar to the statistical results on the lexical level, pitch ranges of entities on referential level are also significantly affected by information structure ($\chi^2(2) = 27.038, p = 0.000$), but there is no significant difference between pitch ranges of entities uttered in different conditions ($\chi^2(1) = 2.5992, p = 0.107$) and no interaction effects noticed ($\chi^2(2) = 0.7132, p = 0.7$).

3.2.2. Duration

Durations of entities with different information structures on the referential level are presented in Figure 5. From the boxplot we can know that the performance of durations on referential level are also contrary to the lexical level. Except for the accessible information, the durations of new and given are much longer in the absence of contexts. The durations of new information are the longest in the out-of-order dialogues, while accessible ones have the longest under the condition of different contextuality.

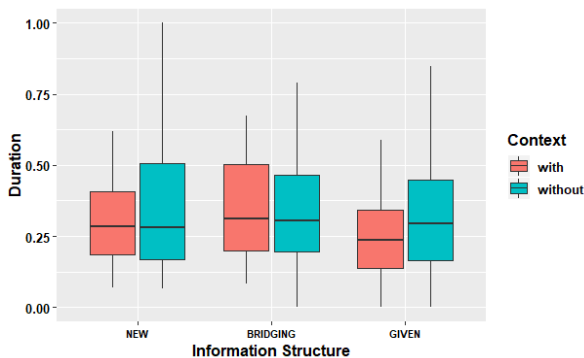


Figure 5: Durations of entities with different types of information structure on the referential level.

The statistical results of durations on the referential level present a different pattern from those on the lexical level. Durations of entities on referential level are significantly affected by both information structure ($\chi^2(2) = 9.9961, p = 0.0068$) and contextuality ($\chi^2(1) = 6.9748, p = 0.008$). An interaction effect between the two factors is also spotted ($\chi^2(2) = 7.317, p = 0.026$).

To sum up, the results of pitch range on the referential level present a similar pattern to those on the lexical level. While durations are affected by both information structures, contextuality and their interaction. The results indicate that the duration is the only cue affected by contextuality in the

prosodic realization of information structures on referential level.

4. Discussion and Conclusions

This section offers further discussions of the results in Section 3, and answer the research questions put forward in Section 1 on that basis.

The statistical results tell us the contextuality only affects the durations of entities with different information structures on referential level. Although it is not statistically significant, we can also see from the boxplot that on the lexical level, in the absence of context, the speaker shows a tendency to compress the pitch range and duration, while on the referential level, this tendency is in the opposite direction.

In Iris and Elsi's research [12], they found new information structures have longer duration and larger pitch range compared with givens. In our view, in the absence of context, the prosodic realization of given information should converge with new ones. From this perspective, the change on the referential level accords with our expectations to a larger degree. According to the results we assume that instead of the specific things, listeners and speakers are more sensitive to their defining components in communication. In this paper, the speaker presents a strategy to make his intentions clearer by increasing the length of time. Certainly, further research is needed to examine our findings in more detail.

The statistical analysis also indicates that the speakers use duration and pitch range as the phonetic cues to distinguish information structures on both level. There are two interesting things that deserve our attention in the presence of context. One is the durations of accessible are the longest on both level. We speculate this may be caused by the difference in the length of the annotated entities. The lexical form of accessible entities might be more variable than new and given information. In addition, the pitch ranges of given information are larger than new ones on referential, which needs further research to explain it.

This study still has limitations in various aspects. Future work will focus on the prosodic realization of information structure of natural dialogues, and more phonetic cues, e.g. intensity and formant need to be analyzed. Besides, other influencing factors, such as emotion, the contents of dialogues etc. should be taken into consideration.

5. Acknowledgements

This work is supported by the National Key R&D Program of China (No. 2017YFE0111900), Key Project of National Social Science foundation (16AYY016) and the Research Foundation of Beijing Municipal Science and Technology Commission (No. Z181100008918002).

6. References

- [1] E. Couper-Kuhlen and M. Selting, *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press, 2018.
- [2] E. Couper-Kuhlen and M. Selting, *Studies in Interactional Linguistics*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 2001.
- [3] Q. Wan. "The Minor Sentence as a Locus of Grammar and Interaction in Chinese." In M. Fang and X. L. Cao., *Interactional Linguistics and Chinese Language Studies, the second volume*, 2018, pp.16-32.

- [4] Y. Yue, "Turn-Constructional Unit as an Important Research Topic in Interactional Linguistics," *Contemporary Linguistics*, vol. 19, no. 2, pp. 246–271, 2017.
- [5] X. Y. Xie and M. Fang, "Prosodic Behaviors of Discourse Markers in Chinese Conversation," In M. Fang, *Interactional Linguistics and Chinese Language Studies, the first volume*, 2016, pp. 225-245.
- [6] Y. D. Chen and R. F. Ma, "Prosodic Features of Turn-takings: A Case Study Based on Talk Show Data," In M. Fang, *Interactional Linguistics and Chinese Language Studies, the first volume*, 2016, pp.246-269.
- [7] M. Fang, X.Y. Li and X.Y. Xie, "Interactional Linguistics and Chinese Studies in Interactional Perspectives," *Language Teaching and Linguistic Studies*, no.3, pp.1-16, 2018.
- [8] M. A. K. Halliday, "Notes on Transitivity and Theme in English," *Journal of Linguistics*, vol.3, no.2, pp.199-244, 1967.
- [9] M. A. K. Halliday. *An Introduction on Functional Grammar*. London: Edward Arnold, 2000.
- [10] W. Chafe, *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press, 1994.
- [11] A. Riester and S. Baumann, "Information Structure Annotation and Secondary accents," in *Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena*, pp. 127, 2011.
- [12] C. O. Iris and K. Elsi, "Prosody and Information Structure in a Tone Language: an Investigation of Mandarin Chinese," *Language, Cognition and Neuroscience*, vol. 30, nos. 1-2, pp.57-72, 2015.
- [13] Y. Bi, L. Y. Ganushchak, A. E. Konopka, G. Ren, X. Sui and Y. Chen, "Prosodic Encoding of Information Structure in Mandarin Chinese : Evidence from Picture Description Task," in *Proceedings of Speech Prosody*, 2016.
- [14] Y. Jia. *The Construction of Multi-layered Annotation Scheme of Chinese Discourse and the Interface Research of Prosody*. Beijing: China Social Sciences Publishing House, 2018.
- [15] X. Zhu. "Normalization of Pitch---How to Deal with the Random Differences of Tones", *Language Sciences*, no.2, pp.3-19, 2004.
- [16] W. Bodo, "A Very Basic Tutorial for Performing Linear Mixed Effects Analyses (Tutorial 2)," 2013.