Interaction of Syntax, Semantics and Pragmatics on Discourse Prosody in Standard Chinese

Xinyi Wen[1, 2], *Yuan Jia*[1], *Aijun Li*[1]

[1] Institute of Linguistics, Chinese Academy of Social Sciences, China
[2] Department of Linguistics, University of Chinese Academy of Social Sciences, China
wenxy1994@126.com, summeryuan_2003@126.com, liaj@cass.org.cn

Abstract

The interface of discourse structure and prosody has been increasingly regarded as an important research area, with abundant studies conducted to account for prosodic features on the discourse level. Still, the key issue of whether discourse structure on different linguistic layers exert intertwined effect on prosodic features remains unsolved, since most existing studies focused merely on a single layer. The present study aims to investigate the interactive effect of syntactic, semantic and pragmatic factors on prosody in Chinese discourse. Specifically, the effects of syntactic function, centering structure and information structure on stress and boundary distribution are examined through log linear analysis. Materials selected from the Annotated Speech Corpus of Chinese Discourse (ASCCD) are annotated following a multi-layered annotation scheme. The results indicate that, apart from the individual effect of each aforementioned factor, interactive effects of centering structure and information structure on both stress and boundary distributions are observed. This finding highlights the significance of an interactive perspective towards prosody studies on the discourse level.

Index Terms: Prosody, Chinese Discourse, Interaction, Stress, Boundary

1. Introduction

The relationship between prosody and factors on other linguistic layers has long been an important research issue. Prosodic structure is believed to constrain syntactic phrasing [1][2]; in turn, syntactic, semantic pragmatic factors would exert influence on prosodic features [3].

When extended from sentence to discourse level, a growing number of studies on the interaction between prosody and other factors have been conducted. For example, prosodic cues of topic changing [4] and disambiguation in reference resolution [5] have been investigated in detail, while the influence of prosodic cues on perceiving discourse functions such as troubles in interaction was also investigated [6]. The above-mentioned studies aimed to figure out the correspondence between prosodic cues and discourse functions with specifically designed materials, and have suggested significant roles of pitch range, intonation contour, pause duration, amplitude, and speaking rate in differentiating various discourse functions. Apart from the studies based on designed materials, corpus studies have also been conducted on natural reading texts, so as to explain prosodic features within discourse theories [7][8][9][10]. The results indicated that, the degrees of stress are significantly influenced by the information structure and centering structure in discourse, while the degrees of boundary are significantly influenced by rhetorical relation, transition status, as well as information structure and centering structure of the following entity. Besides, theories on discourse prosody have also been developed based on large-scale corpus studies [11].

Despite the widely acknowledged interaction between prosody and aforementioned factors on both sentence and discourse levels, the issue of whether these factors combine to affect prosody has not reach a consensus yet [12][13][14]. [15] conducted a preliminary research on the interaction of syntax, semantics and pragmatics on acoustic cues, finding that cues to the three dimensions are orthogonal instead of interactive. Still, the results are restricted on sentence level. Whether interaction exists on discourse level remains to be examined, considering the potential cross-layer interaction suggested in previous studies on discourse prosody.

Against the above background, the present study aims to investigate the interactive effect of syntax, semantics and pragmatics on discourse prosody in standard Chinese. Since the effect of various discourse factors on stress and boundary have been widely investigated, the present study narrows down the scope of prosodic features to the stress degree of each entity and the boundary degree before each entity, so as to offer comparable results to previous studies. For ease of modeling, one factor on each linguistic layer is selected as a representative, as [15] did in their research on the sentence level. On syntactic layer, the syntactic function of each entity is our major concern. On semantic layer, we adopt a famous framework on theorizing the salience of semantic entities realized in each utterance, Centering Theory [16], to define different types of centers and further investigate their effect on prosody. On pragmatic layer, information structure [17] is selected as a representative, considering the solid foundation laid on the relationship between information structure and prosodic features [10][18][19].

The specific goals of this study are: (1) to verify the individual effects of syntactic function, centering structure and information structure on the distributions of stress and boundary in Chinese reading texts; (2) to examine how the factors on different linguistic layers interact to affect stress and boundary distributions.

2. Methodology

2.1. Data Collection

The materials adopted in this research are 14 reading texts covering different genres, which are selected from Annotated Speech Corpus of Chinese Discourse (ASCCD) [20]. Constructed by the Phonetic Lab, Institute of Linguistics, Chinese Academy of Social Science, the corpus was composed of 18 texts read by 10 Standard Mandarin speakers with equal gender ratio. Recordings were conducted with 16 kHz sampling rate and 16 bit rate in mono channel.

2.2. Data Annotation

To investigate the interaction of discourse structures on prosody across different layers, a multi-layered annotation scheme is adopted in the present study, which is briefed as follows. More details could be found in [21].

Prosodic features are annotated based on Chinese ToBI system (Tones and break indices), including Break Index tier and Stress tier. On the Stress tier, non-accented syllable, the most accented syllable in a prosodic word, the most accented syllable in a second prosodic phrase, and the most accented syllable in a primary prosodic phrase are annotated, denoted as L0, L1, L2 and L3 respectively. On Break Index tier, word boundary, intermediate phrase boundary, and prosodic phrase boundary are annotated, denoted as B1, B2 and B3 respectively.

Discourse features are annotated on syntactic, semantic and pragmatic layers respectively. **On syntactic layer**, we annotate the syntactic functions of each entity. Specifically, subject, object and attribute are annotated, denoted as S, O, A respectively. **On semantic layer**, centering structures are annotated based on Centering Theory, which defines three types of centers as follows: forward-looking centers (hereinafter Cf) are the set of semantic entities in an utterance; back-looking center (hereinafter Cb) is a special member of Cf that links the current utterance to previous discourse; preferred center (hereinafter Cp) is the entity predicted to be the Cb of the following utterance, which is the highest-ranked member in Cf in terms of salience [16]. Since Cf serves as the scope of defining Cp and Cb, only Cp and Cb are annotated in the study, denoted as B and P respectively. Note that Cb and Cp are not mutually exclusive, i.e., an entity could serve as both Cb and Cp in an utterance, and those entities are denotated as BP in this study. **On pragmatic layer**, the information structure of each entity is annotated. Specifically, new information, accessible information and given information are annotated, denoted as 1n, 2a and 3g respectively.

In total, three types of syntactic functions, three types of information structures, and three types of centering structures are involved in the present study to account for the distributions of stress and boundary, as is summarized in Figure 1.

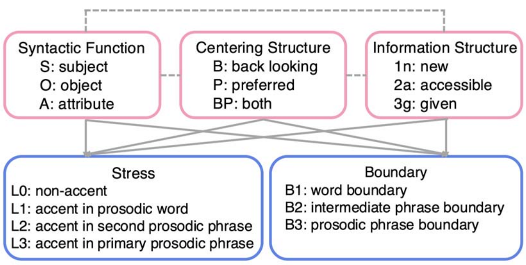

Figure 1: *Symbols of annotation.*

2.3. Data Analysis

The prosodic and discourse information of each annotated entity are extracted with custom-written Praat scripts. Since the extracted parameters are categorical variables instead of continuous acoustic measures, log linear analysis is carried out in the present study to examine the difference between expected frequencies and observed frequencies of each entity. Different from Chi-square tests applied in previous studies involving similar categorical variables, log linear analysis allows for comparisons between more than two variables, thus suitable for investigating the interactive effects of several factors on prosody.

The analysis is conducted using R [22] with the *loglm* function of the MASS package. All variables are entered into the model without a clear distinction between response and explanatory variables. To determine the simplest model that adequately represents the data, a stepwise model selection procedure is conducted with the *step* function, with Akaike's Information Criterion (AIC) taken as the criterion of model evaluation: a smaller value of AIC suggests a better model. After deciding on the suitable model, we refit it to obtain more details such as parameter estimates. The best fit model of stress and boundary will be presented in the next section.

3. Results

3.1. Stress Distribution

To begin with, the frequency of entities with different stress degrees and discourse structures is presented in Figure 2 to offer a panorama of the data. The abscissa represents the stress degree, the ordinate represents the frequency of entities with different features, and the colors differentiate syntactic functions. It is obvious from the plot that, entities with different syntactic functions, information structures and centering structures differ in their stress degrees, which is further confirmed in the results of log linear analysis.

For the data of stress distribution, the best fit model (Goodness-of-fit test statistics: AIC = 87.78, Likelihood ratio chi square = 11.779, df = 16, p = 0.759) involves the following factors: centering structure, syntactic functions, information structure, stress degree, centering structure and syntactic functions, centering structure and information structure, syntactic functions and information structure, centering structure and stress degree, syntactic functions and stress degree, information structure and stress degree, centering structure, syntactic functions and information structure, centering structure, information structure and stress degree. Since our research purpose is to find out the effect of other factors on stress degrees, the interactions involving stress in the model are our major concerns. Specifically, a two-way interaction indicates a significant effect of a feature on stress distribution, while a three-way interaction indicates a significant interactive effect of two features on stress distribution.

Consistent with the results in previous studies, all of the three investigated factors exert an influence on stress distribution. The parameters of the model are listed in Table 1.

Table 1: *Parameters of modelling.*

Factor	Df	Deviance	p
Syntactic structure	2	15.46	0.000^{***}
Information structure	2	14.66	0.001^{***}
Centering structure	2	26.23	0.000^{***}

Apart from the above-mentioned two-way interactions indicating individual influence of each layer on stress distribution, a three-way interaction of information structure,

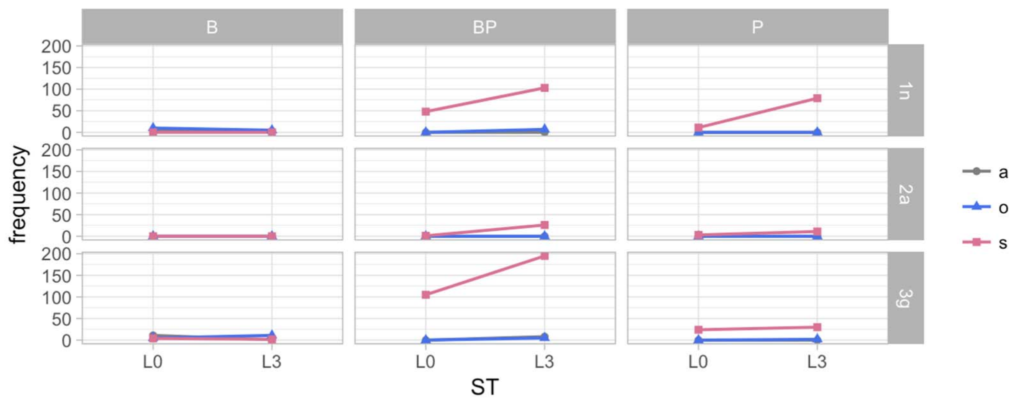

Figure 2: *Frequency of entities with different stress degrees and discourse structures.*

centering structure and stress is also observed, suggesting that the effect of information structure on stress is conditioned on centering structure. To further illustrate the interaction, a mosaic plot (Figure 3) is presented, which is a graphical method to visualize data from two or more qualitative variables. Counts of entities with different stress degrees are divided into their relative proportions across centering structure and information structure; in other words, the area of each cell is proportional to the frequency of the entity it represents. Besides, the colour of each cell demonstrates the residuals obtained from the refitting procedure: positive residuals are shaded blue, suggesting a significantly higher observed frequency than the expected frequency; negative residuals are shaded red, suggesting a lower one.

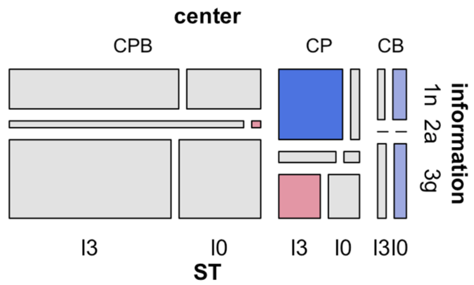

Figure 3: *Mosaic plot of the three-way interaction between centering structure, information structure and stress degree.*

As is indicated in Figure 3, the effect of information structure varies with different types of center. For Cbp, no significant effect is observed. For Cb, the observed frequency of new and given entities without stress is significantly higher than the expected. For Cp, the observed frequency of given entities with L3 stress is significantly lower than expected frequency, while the observed frequency of new entities with L3 stress is significantly higher than expected frequency. In other words, an obvious tendency that the percentage of higher-degree stress decreases as the givenness of the entities increases is observed only in Cp.

3.2. Boundary Distribution

The frequency of entities with different boundary degrees and discourse structures is presented in Figure 4. The abscissa represents the boundary degree, the ordinate represents the frequency of entities with different features, and the colors of lines differentiate syntactic functions. The plot indicates that entities with different syntactic functions, information structures and centering structures differ in their degrees of boundary, which is further confirmed in the results of log linear analysis.

For the data of boundary distribution, the best fit model (Goodness-of-fit test statistics: AIC = 117.6, Likelihood ratio chi square = 3.595, df = 24, p = 0.999) includes the following factors: centering structure, syntactic functions, information structure, boundary degree, centering structure and syntactic functions, centering structure and information structure, syntactic functions and information structure, centering structure and boundary degree, syntactic functions and boundary degree, information structure and boundary degree, centering structure, syntactic functions and information structure, centering structure, information structure and boundary degree. As is introduced in 3.1, interactions involving boundary in the model are our major concerns.

The two-way interactions involving boundary suggest that, all of the three investigated factors exert an influence on boundary distribution. The parameters of the model are listed in Table 2.

Table 2: *Parameters of modelling.*

Factor	Df	Deviance	p
Syntactic structure	4	46.86	0.000^{***}
Information structure	4	11.25	0.024^{*}
Centering structure	4	115.86	0.000^{***}

Similar to the results of stress, a three-way interaction involving boundary is observed. The effect of information structure on boundary degree is conditioned on centering structure, which is further illustrated by Figure 5. From the mosaic plot, the following observations are made. For Cbp, the observed frequency of new entities after B1 is significantly lower than the expected frequency, while the observed frequency of given entities after B3 is significantly higher. For Cp, the observed frequency of given entities after B1 and B2 is significantly lower, the observed frequency of accessible entities after B3 is significantly higher, while the observed frequency of new entities after B1 is significantly higher than the expected frequency. For Cb, the observed frequency of new and given entities after B1 is significantly higher, while the observed frequency of new and given entities after B3 is significantly lower. That is to say, an obvious tendency that the percentage of higher-degree boundary increases as the

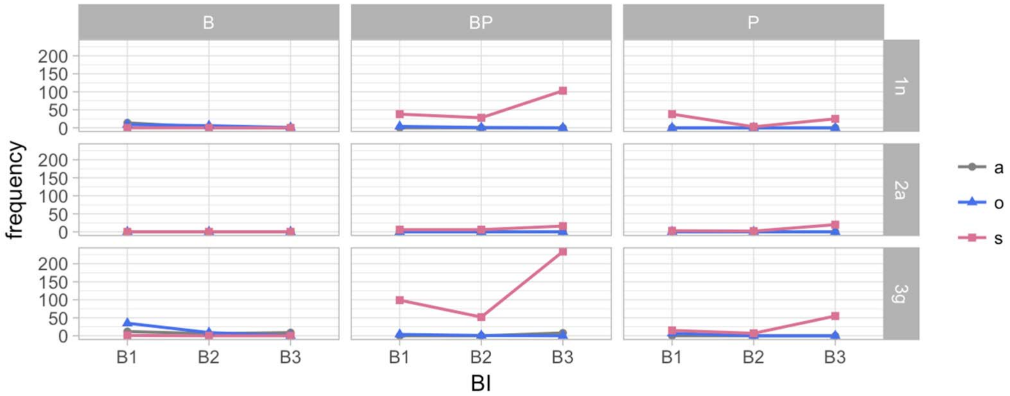

Figure 4: *Frequency of entities with different boundary degrees and discourse structures.*

givenness of the entities increases is observed only in Cp.

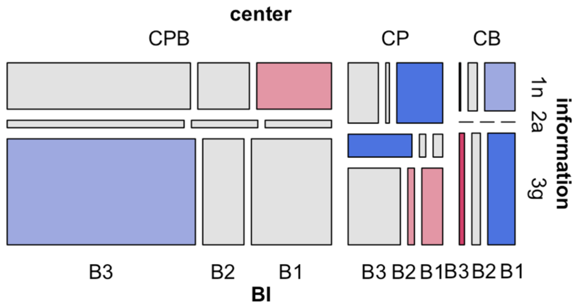

Figure 5: *Mosaic plot of the three-way interaction between centering structure, information structure and boundary degree.*

To sum up, apart from the individual effect of syntactic function, centering structure and information structure, the interactive effects of centering structure and information structure on both stress and boundary distributions are also observed.

4. Discussion and Conclusion

The present study investigates the interaction of syntax, semantics and pragmatics on discourse prosody in standard Chinese. Syntactic function, centering structure and information structure are selected as the representative factors on each linguistic level. This section answers the research questions put forward in Section 1, and offers further explanations to the results presented in the previous section.

In terms of the individual effect of factors on each linguistic layer, all of the factors investigated exert some effect on the distribution of stress and boundary, which conforms to the results in previous studies [7][8][10]. Instead of simple repetitions of previous studies, the analyses of each linguistic layer's effect on prosody are conducted on the same reading materials and annotated on the same entities, hence allowing for a straightforward comparison between the effect of different layers.

As for the interactive effect across different linguistic layers on prosody, interactions of centering structure and information structure are found to influence both stress and boundary distributions. Comparing the results in Section 3 with previous results of information structure's effect on prosody alone [10], we notice that typical effects of information structure on stress and boundary occur only in Cp: for stress distribution, the frequently observed pattern that new information tends to be accented while given information tends to be de-accented is obvious only in Cp; for boundary distribution, readers' tendency to pause before given information is obvious only in Cp. In Cb and Cbp, instead, those effects are less significant, which is possibly due to the nature of Cb. Cb refers to the center semantically linked to previous discourse, hence it is more familiar to the speaker; even if the center is lexically new, the effect of information structure might be neutralized by the semantic link. As for Cbp, its similar pattern to Cb suggests that, the nature of Cb might prevail when the nature of Cp and Cb co-exist in an entity. Without a thorough interactive study across different linguistic layers, the observations above might have been ignored, which highlights the significance of an interactive perspective towards prosody studies on the discourse level.

It is also worth noting that, the present results contradict Wagner & McAuliffe's finding that factors cross linguistic layers affect prosodic cues orthogonally [15], which might reveal the difference between the sentence and discourse levels. However, the contradictory results in the two studies could also arise from different prosodic cues adopted in analyses, as well as different factors selected as representatives for each layer. Although the representatives are established on solid research foundation, the one-factor-for-one-layer method is merely an expedient for modelling, and insufficient to represent the linguistic data on each layer systematically. Therefore, future studies could take acoustic cues into consideration, and should also try to accommodate various factors on each layer in modelling, so as to achieve more comprehensive results.

5. Acknowledgements

This research was supported by National Key R&D Program of China (2017YFE0111900), Key Project 'Interaction of Grammar, Semantics and Prosody' of National Social Science foundation (16AYY016), and Innovation Program of Chinese Academy of Social Sciences.

6. References

[1] D. R. Ladd, "Intonational phrasing: the case for recursive prosodic structure," *Phonol. Yearb.*, vol. 3, pp. 311–340, 1986.

[2] E. Selkirk, "The interaction of constraints on prosodic phrasing," in *Prosody: Theory and experiment*. Dordrecht: Springer, 2000.

[3] H. Truckenbrodt, “Phonological Phrases: Their Relation to Syntax, Focus, and Prominence,” Ph.D. dissertation, MIT, 1995.
[4] J. Hirschberg and C. H. Nakatani, “A prosodic analysis of discourse segments in direction-giving monologues,” *Proc. 34th Annu. Meet. Assoc. Comput. Linguist*., pp. 286–293, 1996.
[5] J. C. Tyler, “Discourse prosody in production and perception,” Ph.D. dissertation, The University of Michigan, 2012.
[6] F. Roberts, A. L. Francis, and M. Morgan, “The interaction of inter-turn silence with prosodic cues in listener perceptions of ‘trouble’ in conversation,” *Speech Commun*., vol. 48, no. 9, pp. 1079–1093, 2006.
[7] Y. Pang, Y. Jia, A. Li, D. Song, and R. He, “Influence of Dependency Parsing on the Prosody of Chinese Discourse,” in *Proceedings of speech prosody*, 2016.
[8] Y. Dong, Y. Jia, and A. Li, “Prosodic Features of Chinese Reading Texts within the Framework of Centering Theory,” in *Proceedings of NCMMSC*, 2015.
[9] L. Zhang, Y. Jia, and A. Li, “Analysis of Prosodic and Rhetorical Structural Influence on Pause Duration in Chinese Reading Texts,” in *Proceedings of Speech Prosody*, 2014.
[10] Y. Jia and Z. Yang, “An interface research on information structure and prosody in Chinese reading texts,” in *Proceedings of Oriental COCOSDA*, 2017.
[11] C. Tseng, “Higher level organization and discourse prosody,” in *Proceedings of Tonal Aspects of Languages*, 2016.
[12] H. Fujisaki, “Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations,” *KTH Department of Speech, Music and Hearing Quarterly Status Report*, 1981.
[13] J. Pierrehumbert, “The phonology and phonetics of English intonation,” Ph.D. dissertation, MIT, 1980.
[14] E. Selkirk, *Phonology and Syntax. The relation between sound and structure*. Cambridge, MA: MIT Press, 1984.
[15] M. Wagner and M. McAuliffe, “Three Dimensions of Sentence Prosody and their (Non-)Interactions,” in *Proceedings of Interspeech*, 3196-3200, 2017.
[16] B. J. Grosz, A. K. Joshi, and S. Weinstein, “Centering: a framework for modeling the local coherence of discourse,” *Comput. Linguist*., vol. 21, no. 2, pp. 203–225, 1995.
[17] M. A. K. Halliday, “Notes on transitivity and theme in English,” *J. Linguist*., no. 3, pp. 199–244, 1967.
[18] K. Lambrecht, *Information structure and sentence form*. Cambridge: Cambridge University Press, 1994.
[19] 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.
[20] A. Li, X. Chen, G. Sun, W. Hua, and Z. Yin, “Speech corpus of Chinese discourse and the phonetic research,” in *Proceedings of ISCSLP*, 2000.
[21] Y. Jia and A. Li, “A Linguistic Annotation Scheme of Chinese Discourse Structures and Study of Prosodic Interactions,” in *Proceedings of ISCSLP*, 2016.
[22] R Core Team, “R: a language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing,” 2013.

[This paper was published at ISCSLP 2018]