

Analysis of Prosodic and Rhetorical Structural Influence on Pause Duration in Chinese Reading Texts

Liang Zhang¹, Yuan Jia², Aijun Li³

Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China

liangzhang1025@gmail.com, summeryuan_2003@126.com, liaj@cass.org.cn

Abstract

This paper investigates factors that influence pause duration in Chinese reading texts through examining the stress degree in pre-pausal and post-pausal positions and the rhetorical structure in discourse as a whole. The RSTTool is used in diagramming the rhetorical structures of the texts. The recordings, extracted from the ASCCD corpus, are further analyzed acoustically and statistically by applying Praat and R. The statistical analysis results show that the stress degree in both pre- and post-pausal positions has a significant impact on pause duration. Moreover, the nuclearity in both positions have also been shown to have a remarkable influence. Specifically, the nucleus in pre-pausal and satellite in post-pausal positions can significantly lengthen the pause duration.

1. Introduction

The Rhetorical Structure Theory (RST) (Mann & Thompson, 1988) is a theory of text organization that has led to areas of application beyond discourse analysis and text generation. It has been applied to several linguistic areas, i.e. theoretical linguistics, psycholinguistics, and computational linguistics. RST defines a set of relations to identify the specific relationship that holds between two Elementary Discourse Units (EDUs) of a text. These relations are categorized into two types: mononuclear and multinuclear relations. After the elaborate description and categorization on the relations in large amount of real discourse spans, 30 rhetorical relations were studied in Mann (2005). A novel feature of RST is the concept of nuclearity. As well as presenting the relationship between two text spans, rhetorical relations also convey the information about which span is more central to the writer's purposes. Relations such as background and circumstance are of nucleus-satellite in that the EDUs linked by these relations are distinguished by their centrality: one is called the nucleus (N) and the other is called the satellite (S), with satellite subordinate to the nucleus.

Among various discourse structure theories, RST has been used widely in recent years to diagram the rhetorical hierarchy annotation. The popularity of RST has led to the development of an RST Treebank of manually annotated English texts, which is available for training and testing purposes (Carlson et al. 2003). It consists of 385 Wall Street Journal articles from the Penn Treebank (Marcus et al. 1993) with a total of 176,383 words. Another well-annotated RST corpus is Potsdam Commentary Corpus (PCC), which consists of 172 commentaries from *Mearkische Allgemeine Zeitung*, a German regional daily (Stede, 2004). Some other researchers (Stent, 2000; Taboada, 2004) also tried to apply RST to annotate spoken dialogues in Task dimension.

The development of RST in analyzing Chinese discourse was mainly in the areas of syntactic analysis (Chen, 2008), prosodic analysis (Tseng, 2006), Systemic Functional Grammar (SFG) (Wang & Dong, 1995), and Second Language Acquisition (SLA) (Wang and Xia, 2005). The other application field has been Computer Sciences and Language Processing, aiming principally at auto-annotation of rhetorical structures by training

the model with hundreds of essays or news articles (T'sou et al., 1992; T'sou et al., 1996; Skoufaki, 2009).

Yue (2006) enriched Chinese language resources through building up a Chinese news commentary Treebank, using the RST as the theoretical framework. The corpus, consisting of 400 news texts of about 780,000 characters, has been applied to computation of a priori scores, needed in Chinese summarizers and be used as a platform for training and testing statistics-based discourse parsers. Their annotation efforts have proved, on a fairly large scale, the cross-language transferability of RST and its formalization.

The research on interfacing RST and prosody has attracted much attention. However, little work has been done in Mandarin Chinese discourse. Yang and Yang (2012) examined how rhetorical structures were reflected by boundary prosodic parameters in Mandarin Chinese discourses, through investigating recordings of ten paragraphs of news commentaries, with the prosodic parameters (pause duration, pitch reset, and final lengthening). The results were in line with previous studies (Noordman et al., 1999; Ouden et al., 2009). Nonetheless, no further detailed analysis has been done on stress degree, nuclearity and any other possible influential factors.

The above overview of previous studies shows that RST, which has been adopted in interdisciplinary research, were mainly restricted in the discourse dimension. The empirical study on the interface of prosody and RST, from the perspective of stress degree and nuclearity, is of fundamental importance but is largely under-explored. The work reported in this paper aims to fill this gap. Particularly, this paper is concerned with the duration of rhetorical pause in reading speech of Mandarin. We investigate the effect of the stress degree and nuclearity of preceding and upcoming EDUs, and its rhetorical structure on pause duration at different hierarchy within a RST diagram of the text.

2. Data

The materials selected in this current study are three reading texts chosen from the Annotated Speech Corpus of Chinese Discourse (ASCCD), which was built by the Phonetic Lab, Institute of Linguistics, Chinese Academy of Social Sciences (CASS). The data were collected from ten Mandarin speakers (5 males and 5 females) in Beijing. The C-ToBI system was used for annotation and four tiers were labeled. In this research, the stress tier (ST) is used, in which each prosodic unit were annotated with one of the four degrees (0: weak; 1: normal; 2: secondary stress; 3: primary stress). The stress degrees of preceding and succeeding rhetorical pauses are extracted for further analysis.

The RSTTool provided by O' Donnell (1997) was used in this study to create the RST diagram. One disadvantage of RST, as mentioned in previous study, is the comparative subjectivity in labeling relations. The diagrams used in this study were double-checked by another researcher in the same field, in order to avoid inaccuracy as much as possible. The resultant RST diagram is showed in Figure 1.

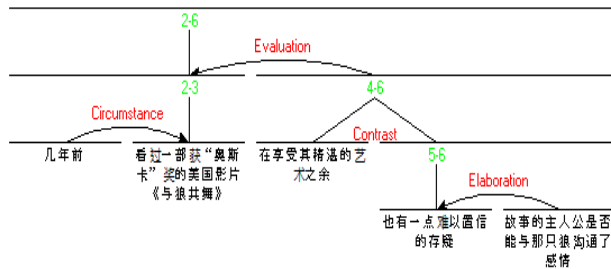


Figure 1. An example of RST diagram

(A few years ago/ I watched an Oscar-winning film: Dances with Wolves/ besides the enjoyment of exquisite art/ I found one point incredible/ whether the hero of the story really communicated with the wolf.)

In all multi-span diagrams, each rhetorical pause is marked with the sum of spans of adjacent EDUs. For instance, “Circumstance” in Figure 1 would be labeled as 4 (2+2, sum of first and second EDU span, with the highest span excluded). Similarly, we get 6 for “Contrast” and 6 for “Elaboration”. For three texts in total, the spans are ranked from 3 to 14. To facilitate statistical analysis, the spans are re-categorized into four hierarchies with 3 consecutive spans in each category.

Pause duration is defined as the silence interval between the ending of one segment and the beginning of the next segment. Based on the rhetorical structures of each discourse, the durations of rhetorical pauses are further annotated and extracted from the corpus.

In the three texts, 156 rhetorical relations are diagramed, and accordingly 1560 pause durations are extracted.

3. Methodology, Results and Analysis

In this section, the study systematically examines three factors that could influence the pause duration: rhetorical hierarchy, stress degree and nuclearity in the preceding and succeeding pause positions. These three factors fall into two categories: prosody and rhetorical structure. Through R for statistical computing, t-test and ANOVA analysis are adopted to investigate the inner-relations and correlations of these variables.

3.1. Rhetorical Relations and Hierarchy

To study the effect of hierarchical position with four levels on pause duration, an ANOVA analysis is conducted. The results show that the depth of the hierarchies significantly affects the duration of the pauses ($p < 0.001$), indicating that pause duration gets longer when the depth of hierarchies increases. This is consistent with the findings from previous studies on Chinese and other languages such as Spanish, English.

The result shows that the schema *Title* takes the longest pause in all, which reflects the widest semantic distance between adjacent segments, while *Summary* takes the second longest. No significant difference is found ($p = 0.166$) between causal and non-causal relationships.

Despite the accordance with the previous results in Yang and Yang (2012), this paper, by involving stress degree and nuclearity, further analyzed their effects on pause duration, and their inner relationship with rhetorical hierarchy.

3.2. Stress Degree

For the stress degree in the pre-pausal position, statistical analysis result (Figure 2) shows a significant difference ($p < 0.001$) between ST1 and ST3. As the stress degree goes higher (from 1 to 3), the pause duration decreases.

The same significant difference is observed in the post-pausal position between ST1 and ST2 ($p < 0.05$) and between ST1 and ST3 ($p < 0.001$). However, with stress degree goes higher, the pause duration increases.

Based on the above results, it is clear that the stress degrees at both preceding and succeeding pause EDUs have a significant influence on pause duration, while the stress degree in the post-pausal position shows a higher influence on each stress degree, as demonstrated in Figure 2. The X-axis is the stress degree in previous EDU, while the boxplot is colored according to stress degree in post-EDU. The post-pausal stress degree, from red to blue, shows an increasing trend without the restriction of stress degree in pre-pausal position. While for pause duration in preceding pause position, there is no visually obvious decrease from normal to primary stress.

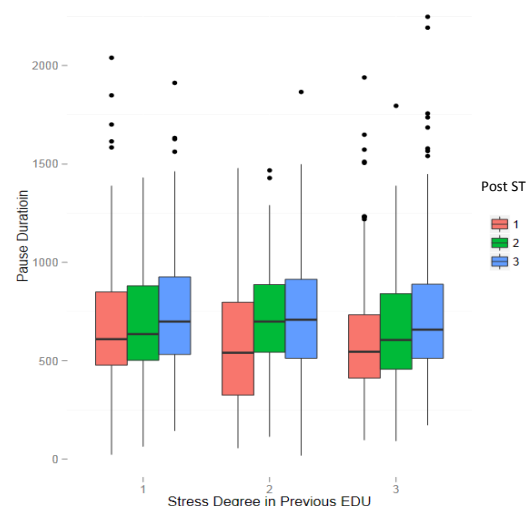


Figure 2. The significant influence of upcoming stress degree on pause duration

In the next step, we examine the distribution of duration in combined stress type in four hierarchies. With reference to the results above, as showed in Figure 3, comparatively speaking, the ST combination “3-1” (up right square) and “1-3” (down left square) generates the shortest and longest duration respectively.

Another noticeable point is that the pause duration and rhetorical hierarchy theory generally but does not necessarily apply to every subordinate category (see circles in Figure 3). This inspires us to make a novel assumption: there is a ranking of the influential factors of pause duration, which leads to the question: what is the ranking of the factors, such as hierarchy,

stress degree, nuclearity, sentence complexity that influence pause duration in reading texts? Further attention should be paid to this interesting point and more experiments need to be done. It is however currently beyond the scope of this paper.

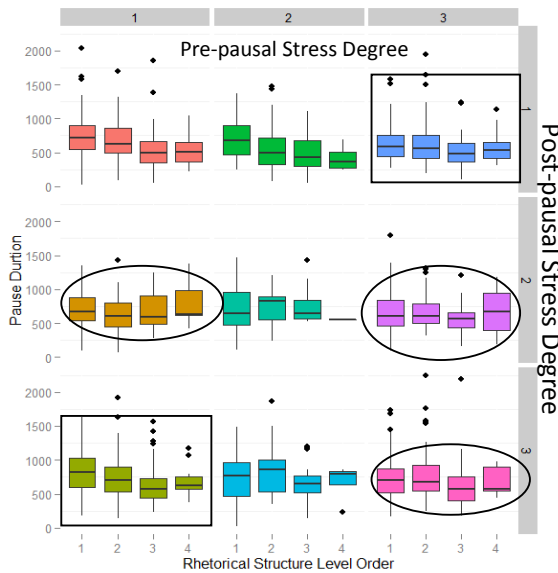


Figure 3. Pause duration in different stress degree combinations in four rhetorical hierarchies

3.3. Nuclearity

This part examines the nuclearity of EDUs in preceding and succeeding rhetorical pause position, aiming at investigating possible influence of nuclearity on pause duration.

3.3.1. Nucleus and Satellite EDU

In the nuclearity of the pre-pausal EDU, statistic result shows a significant difference ($p < 0.001$) between N and S. The result indicates that if the EDU is a nucleus in the rhetorical relation, the duration of its succeeding pause would be significantly longer than that after a satellite EDU.

The same significant difference is found in the post-pausal EDU between N and S ($p < 0.001$). However, the result is opposite to the previous one: if the EDU is a nucleus one in the rhetorical relation, the duration of its preceding pause would be significantly shorter than that before a satellite EDU.

Therefore, when the test came to pause duration in mononuclear relations as a whole, we assume and has confirmed that the influence of nuclearity on duration is neutralized with no significant difference appears in t.test between N-S and S-N combination, even though practically the N-S combination produces slightly longer pause duration than that in the S-N combination.

3.3.2. Nuclearity and Rhetorical Relations

A closer examination is performed on the components of rhetorical relations in both combinations. Among the 345 tokens of N-S relations, nearly one third are *Evaluation* relations, which,

unlike other relations, carry particular emotions (Hou, 2012). Moreover, from the discourse analysis point of view, *Evaluation* indicates the change of footing registration of the writer, with the purpose of jumping out of the narration and gaining interaction and solidarity with readers. The change in registration were clearly obtained by the reader, and thus in the recordings, they may produce a longer pause to indicate the register change.

With the purpose of further analysis of the distribution of mononuclear relation in four hierarchies, Figure 4 is plotted out with the nuclearity combination and pause duration.

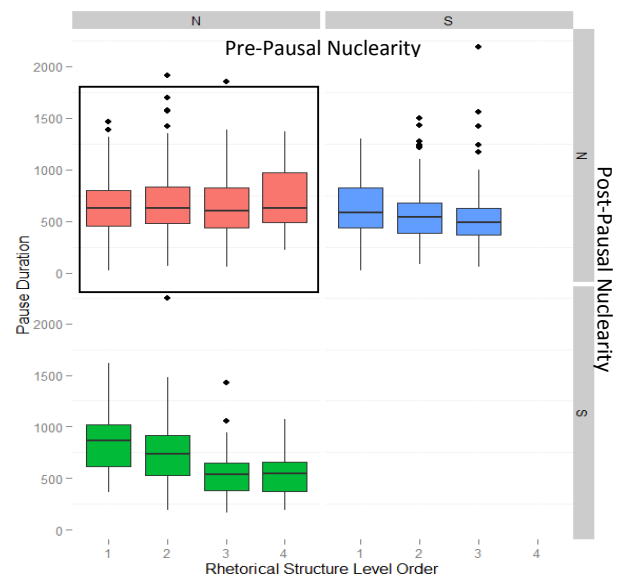


Figure 4. Pause duration in different nuclearity combinations in four rhetorical hierarchies

Two points are notable in the boxplot above.

1. The neutralization of pre-and post-pausal nuclearity influence could be resulted in by the uneven distribution of S-N combination in four levels.

2. The pause duration in multinuclear relations (in the square) does not follow the hierarchy pattern show a decrease with the increase of hierarchy. There was no significant difference at all between each rhetorical hierarchy. This leads back to the assumption that mentioned in the stress degree section. Is the duration of multinuclear relations not restricted by the hierarchy theory? Could nuclearity rank first in the restraints? These questions definitely require more attentions and further detailed experiments.

4. Discussion

The study, with the adoption of RST and stress degree on three reading texts may suffer from several biases in the following aspects. First, it is a comparatively small corpus, which may result in skewness of the data. To be specific, there is no S-N combination in forth rhetorical hierarchy. Secondly, though double-check was applied during the process of labeling rhetorical relations, there could still be few mistakes since the

inevitability of subjectivity. Third, the accuracy and consistency of stress degree annotation, which was done based on pure perception of experienced linguistics.

The results also show a gendered variability, in which the male speakers tend to have longer pause duration than the female speakers in every dimension.

5. Conclusion

Through RST labeling on Chinese reading texts and statistical analysis on stress degree and nuclearity in pre- and post-pausal position, the present study explores the influential factors of pause duration. The following observations can be inferred based on the results: stress degree in both pre- and post-pausal positions has significant influence on pause duration, in which “1-3” indicates longest duration while “3-1” the shortest. Nuclearity in both positions separately showed remarkable effect on pause duration: nucleus in pre-pausal and satellite in post-pausal position significantly lengthened pause duration.

This is a pioneer research between RST and prosody with the parameter hierarchy, stress degree combination and nuclearity, which not only jumped out of the hierarchy restriction, but also casts new light on the interface research.

Further experiments are worth investigation given the assumed sequence of influential factors of pause duration, such as stress degree combination, multinuclear relation, and rhetorical hierarchy. Larger corpus is needed for a more thorough research. It may also involve logical speech discourses, such as presentations or story-telling, which are more complicated in topics and less structured in construction.

6. Acknowledgements

This research is supported by National Program on Key Basic Research Project (973 Program) under Grant 2013 CB329301, as well as the Innovation Program of Chinese Academy of Social Sciences “Key Laboratory of Phonetics and Speech Science”.

References

- [1] Carlson, L., D. Marcu, & M.E. Okurowski (2003), Building a discourse-tagged corpus in the framework of rhetorical structure theory, in van Kuppevelt, J. and R. Smith, editors, *Current Directions in Discourse and Dialogue*, Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 85–112.
- [2] Chen, Liping. (2008), Research on Rhetorical Structure Theory and Sentence group. *Journal of Suzhou University (Philosophy & Social Science)*. Jul. No. 4.
- [3] Dai, Weihua. & Xue, Yan. (2004). Rhetorical Genre Studies and Rhetorical Structure Theory. *Foreign Language Education*. Vol.25. No. 3. pp. 35-39.
- [4] Mann, W. & Thompson . (1988). S. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243-281
- [5] Mann, W. (2005), RST Web Site, now <http://www.sfu.ca/rst>
- [6] Marcus, P.M., M.A. Marcinkiewicz, and B. Santorini (1993), Building a large annotated corpus of English: the Penn Treebank, *Computational Linguistics*19 (2), pp. 313–330.
- [7] Noordman, L., Dassen, I., Swerts, M., Terken, J. (1999). Prosodic markers of text structure. In: van Hoek, K., Kibrik, A., Noordman, L. (Eds.), *Discourse Studies in Cognitive Linguistics, Selected Papers 5th International Cognitive Linguistics Conference*. John Benjamins, Amsterdam, pp. 133–148.
- [8] O' Donnell, Michael. (1997). RST-Tool: An RST analysis tool. In Proceedings of the 6th European Workshop on Natural Language Generation, Gerhard-Mercator University, Duisburg, Germany.
- [9] Ouden, J.N. den, Noordman, L. & Terken, J.M.B. (2002). The Prosodic Realization of Organizational Features of Texts. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 Conference, Aix-en-Provence, France, (pp. 543-546)*. Aix-en-Provence, France: Laboratoire Parole et Langage.
- [10] Rao, Rajiv. (2010), Final Lengthening and Pause Duration in Three Dialects of Spanish. *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA, USA.
- [11] Skoufaki, Sophia. (2009). An Exploratory Application of Rhetorical Structure Theory to Detect Coherence Errors in L2 English Writing: Possible Implications for Automated Writing Evaluation Software. *Computational Linguistics and Chinese Language Processing*. Vol. 14, No.2, pp. 181-204
- [12] Stede, Manfred. (2004). The Potsdam Commentary Corpus. In *Proceedings of the ACL 2004 Workshop 'Discourse Annotation'*, Barcelona. 2004
- [13] Taboada, Maite. (2004). Rhetorical Relations in Dialogue: A contrastive study . In C L. Moder & A. Marinovic-Zic (Eds.), *Discourse Across Languages and Cultures*. Amsterdam and Philadelphia: John Benjamins.
- [14] T'sou, Benjamin, Ho, H., Lai, B., Lun, C. & Lin, H. (1992). A Knowledge-based Machine-aided System for Chinese Text Abstraction. In *Proceedings of International Conference on Computational Linguistics*, France, pp. 1039-1042.
- [15] T'sou, Benjamin. , Lai, Tom., Chan, Samuel., Gao, Weijun. & Zhan, Xuegang (2000). Enhancement of Chinese Discourse Marker Tagger with C4.5. In Proceedings of the Second Chinese Language Processing Workshop. Hong Kong. 38-45.
- [16] Wang, Wei & Dong, Jiping. (1995) Rhetorical Structure Theory and Systemic Functional Grammar. Shandong Foreign Language Teaching. Vol.2.
- [17] Wang , Yanping, & Xia, Zhen. (2005). A Contrastive Study of Rhetorical Structure in Argumentative Texts: An Empirical Study in English-Chinese contrastive Rhetoric. *Journal of Henan University (Social Sciences)*. Vol. 45. No. 5.
- [18] Yang, Xiaohong & Yang, Yufang. (2012). Prosodic Realization of Rhetorical Structure in Chinese Discourse. *IEEE Transactions on Audio, Speech , and Language Processing*. Vol 20. No.4
- [19] Yue, Ming. (2006). Annotation and Analysis of Chinese Financial News Commentaries in Terms of Rhetorical Structure. Ph.D. Thesis. Communication University of China. China
- [20] Zvonik, Elena. & Cummins, Fred. (2002). Pause Duration and Variability in Reading Texts. In *preceeding of 7th International Conference on Spoken Language Processing, Interspeech*, Denver, Colorado, USA.
- [21] Tseng, Chiuyu. (2006). Higher Level Organization and Discourse Prosody. *The Second International Symposium on Tonal Aspects of Language*. pp.23-34