The rhythmic constraint on prosodic boundaries in Mandarin Chinese based on corpora of silent reading and speech perception

Wei Lai [1], *Jiahong Yuan*[1], *Ya Li*[2], *Xiaoying Xu*[3], *Mark Liberman*[1]

[1] University of Pennsylvania, Philadelphia, United States
[2] Chinese Academy of Science, Beijing, China
[3] Beijing Normal University, Beijing, China

weilai@sas.upenn.edu, jiahong.yuan@gmail.com, yli@nlpr.ia.ac.cn, xuxiaoying2000@bnu.edu.cn, markyliberman@gmail.com

Abstract

This study investigated the interaction between rhythmic and syntactic constraints on prosodic phrases in Mandarin Chinese. A set of 4000 sentences was annotated twice, once based on silent reading by 130 students assigned 500 sentences each, and a second time by speech perception based on a recording by one professional speaker. In both types of annotation, the general pattern of phrasing was consistent, with short "rhythmic phrases" behaving differently from longer "intonational phrases". The probability of a rhythmic-phrase boundary between two words increased with the total length of those two words, and was also influenced by the nature of the syntactic boundary between them. The resulting rhythmic phrases were mainly 2-5 syllables long, independent of the length of the sentence. In contrast, the length of intonational phrases was not stable, and was heavily affected by sentence length. Intonational-phrase boundaries were also found to be affected by higher-level syntactic features, such as the depth of syntactic tree and the number of IP nodes. However, these syntactic influences on intonational phrases were weakened in long sentences (>20 syllable) and also in short sentences (<10 syllable), where the length effect played the main role.

Index Terms: prosody, rhythmic constraint, syntax-prosody mapping, implicit prosody, Mandarin Chinese

1. Introduction

Although a hierarchy of prosodic units appears to exist in all languages, the nature of this hierarchy remains uncertain. Prosodic phrasing can be defined from different perspectives [1-3], and is influenced by multiple factors [4, 5], including syntactic structure [6-9], syntactic relations such as c-command [6], the structural distance between two words [7], branchingness [8], and the alignment of a word's left or right edge with the left or right edge of a syntactic constituent [9].

However, theories of syntax-prosody mapping generally fail to take into account the rhythmic/prosodic weight, which is most straightforwardly reflected as phrase length. For example, in Mandarin, a subject noun that is short is more likely to group with the fol*lowing predicate to form a prosodic unit than a long one (e.*g., *中国表示/* "China claims/" vs *中华人民共和国/表示/* "People's Republic of China/ claims/"). Similar phenomena are reported for various syntactic components in other languages such as Japanese [10] and Korean [11].

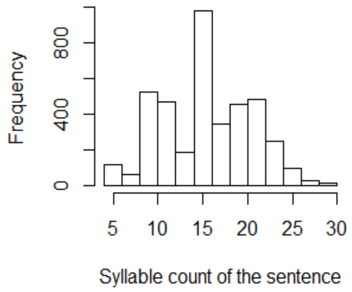
The idea that prosodic phrasing is affected by phrase length establishes a sort of rhythmic constraint on prosodic phrases. [12] defines the rhythmic patterns derived by the postlexical prosodic unit as "macro" rhythmic units, in contrast with "micro" rhythmic units related to foot, syllable, or mora. Like "micro" rhythmic units, these "macro" rhythmic units (such as accentual phrases, or APs) also follow the constraint that prosodic units at the same level tend to have a similar size. Thus in Japanese, most APs have 4-5 syllables [13], and in Korean and French, most APs have 3-4 syllables or 1.2 content words [14]. And other studies suggest that the construction of still larger phonological phrases can be constrained by the size of the domain [15, 16]. Phrase length has also been claimed to influence the attachment of relative clauses in a disambiguation context, in both behavioral [17, 18] and neurolinguistic [19] studies.

This study explores rhythmic constraints on prosodic phrasing in Mandarin. Different levels of prosodic phrases as well as different terminologies have been proposed by previous prosody models of Mandarin Chinese [20-22]. The one that is most popular in China, which was also adopted for the construction of corpora in the early stage of our study, proposes a hierarchy of "prosodic word - prosodic phrase - intonational phrase" [22]. In this paper, we use the term "rhythmic phrase" rather than "prosodic phrase", to better describe the rhythmic nature of this prosodic unit.

2. Corpus

A text corpus and a speech corpus were built on the same set of 4000 sentences that had been collected from newspapers. Two levels of prosodic units – prosodic/rhythmic phrases (RP) and intonational phrases (IP) – were annotated twice, once based on silent reading and a second time based on speech perception. For the silent reading task, 130 undergraduate students from Beijing Normal University participated to earn credit. Each of the students was assigned 500 sentences, such that each sentence was annotated by 12-17 students. The students were given a twenty-minute lecture on the general concept of prosodic phrases before they began to parse. For the speech perception task, the 4000 sentences were produced by a professional female speaker and recorded in a sound booth in the Chinese Academy of Science. Then, a professional female annotator at the Chinese Academy of Science listened to the audio files of these sentences and annotated the phrase boundaries according to her perception. The distribution of sentence lengths in terms of syllable count is shown in the following histogram. In total, there are 14699

RP and 5468 IP annotated via speech perception, and 209195 RP and 93621 IP annotated via silent reading.



The Stanford Parser, version 3.2.0 [23], was used to parse these sentences and to provide information about word segmentation, POS tagging, phrase structure, and dependency relations for each of the sentence.

3. Experiment

The goal of our study is to answer the following questions: a) What does the distribution of phrase lengths look like for the two kinds of phrases, and how are they influenced by the length of the sentence? b) How does the factor of length interact with syntax in affecting prosodic boundaries? The experiment presents three analyses to address these questions.

3.1. Overall distribution of phrase lengths

The first analysis aims to gain a general sense of the phrase lengths. We extracted the length of RP, IP, and the length of the sentence they came from for both text based and speech based annotation.

3.2. Interaction between constraints for RP

The second analysis aims to investigate how rhythmic phrase boundaries are affected by phrase length and by syntax. According to an edge-based approach, the presence of a boundary should be largely determined by the syntactic structure, and the locations of an adjacent boundary should not matter. But according to a length-constraint approach, the location of a preceding boundary will influence the following boundary.

In order to evaluate the syntax-rhythm interaction on the RP level without making any assumption about this issue, we focused on the rhythmic phrase boundary between the initial two words of a sentence with silent-reading annotation. For each sentence, we extracted the total length of the two words, the presence or absence of rhythmic boundaries between them, and the type of XP constituent in both left-edge-based (the second word) and right-edge-based (the first word) ways.

3.3. Interaction between constraints for IP

The third analysis aims to investigate the interaction between syntactic constraints and rhythmic constraints on intonational phrase boundaries. Since intonational phrases have been reported to be associated with syntactic features that come from the upper levels of the syntactic tree [24], we extracted three upper-level syntactic features for the sake of this analysis: the depth of the tree, the number of secondary nodes in the tree, and the number of Inflectional Phrase (InfP) nodes in the tree. The number of InfP nodes and the number of secondary nodes are used to evaluate the amount of syntactic splits higher on a tree, and the depth is used to evaluate the degree of syntactic nesting of a tree.

4. Results

Results from the above three analyses are reported in the following subsections.

4.1. Overall distribution of the length of RP and IP

Figure 1 shows the density plot (bandwidth = 0.4) of the rhythmic phrase lengths and intonational phrase lengths in terms of syllable count, as annotated via silent reading and via speech perception. Figure 2 shows the lengths of RP and IP for sentences of different lengths.

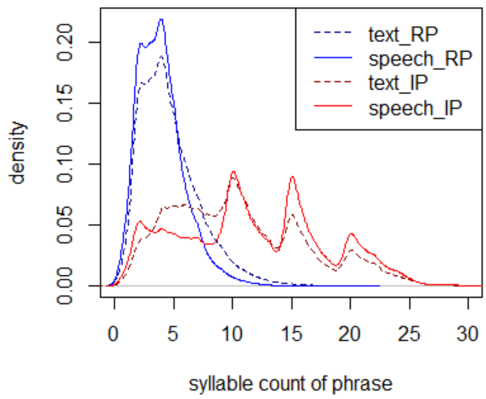


Figure 1: *Overall density of length of RP and IP in terms of syllable count based on text and speech*

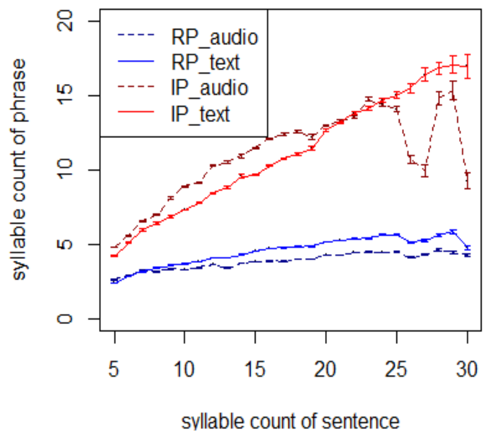


Figure 2: *Relationship between phrase length (y-axis) and sentence length (x-axis)*

As shown in Figure 1, the distributions of length for RP and IP are very different. Rhythmic phrases are mainly 2-5 syllables long. A clear peak can be observed at the length of 2-4 syllables and the density drops rapidly after 4. This pattern appears in both silent reading and speech perception results. In contrast, intonational phrases vary widely from as short as 2 syllables to as long as more than 20 syllables. The distributions of IP length for silent reading and for speech perception are also similar. Three peaks are shown respectively at the values of 10, 15 and 21, and it is difficult to determine which is the preferred length for IP. However, if we compare this pattern to the distribution of sentence length in Section 2, it is quite clear that the IP length parallels the sentence length in peaking at 10, 15 and 21.

The two levels of phrases also differ in how they are influenced by the length of the sentence. In Figure 2, as the sentence length increases from 5 syllables to 30 syllables, the averaged IP length accordingly increases from 5 syllables to 18 syllables in an almost linearly way for both speech perception and silent reading. By contrast, the averaged RP length remains quite steady as 2.5 to 5 syllables long as the sentence becomes longer, albeit a slight increase.

4.2. Interaction between length and syntax for RP

Figure 3 displays the relationship between the probability of the RP boundary between the first two words and the total length of the two words. We can see that they are highly correlated: the probability rises from 0.1 to around 0.7 as the total length of the two words increases from 2 syllables to 8 syllables. This relationship remains stable under all three conditions where the syntactic constituents are balanced by the left edge (the second word), by the right edge (the first word) and by both sides.

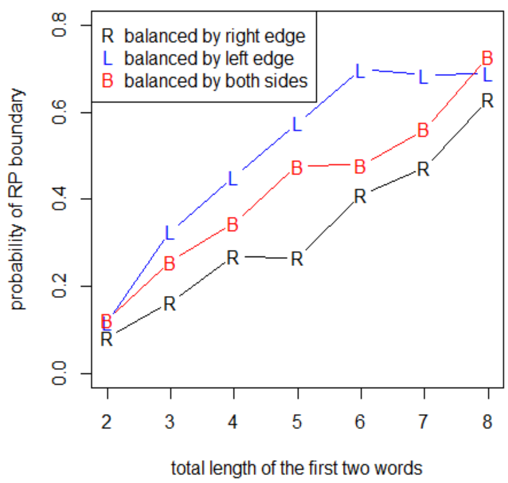


Figure 3: *Influence of the total length of the first two words on the probability of the rhythmic boundary between them with balanced syntactic context*

The fact that the length of the domain heavily influences the probability of a prosodic boundary within the domain does not mean that syntax is irrelevant to prosodic boundaries. Figure 4 shows how length influences the probability of a prosodic boundary for two specific types of XP constituents – adverb phrases (ADVP) and quantifier phrases (QP), based on the left edge and the right edge of the constituents. Here length influences the probability of prosodic boundaries just as it did in Figure 3, but the syntactic juncture type also has a strong influence, which modifies the slope of the relationship between length and boundary probability. Thus ADVP shows a greater juncture power than QP does, since the probability of a prosodic boundary increases more rapidly at the edge of an ADVP than at the edge of a QP.

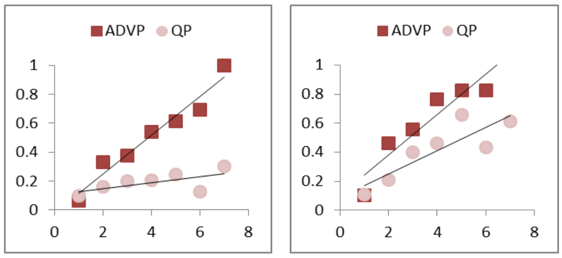


Figure 4: *The probability of a RP boundary (y-axis) in the first two words influenced by the total length of the two words (x-axis) for ADVP and QP junctures based on the right edge (left graph) and the left edge (right graph)*

The k and R^2 values for the fitted linear regression lines with total length as dependent factor and probability as independent factor for different types of XP are shown in Table 1, as well as the averaged probability for each XP type.

Table 1. *Mean, k and R^2 of the regression lines with total length as dependent factor and probability as independent factor for different syntactic constituents*

XP	Right-edge based			Left-edge based		
	Mean	*k*	R^2	*Mean*	*k*	R^2
ADJP	0.23	0.09	0.64	0.42	0.10	0.88
ADVP	0.52	0.13	0.95	0.59	0.14	0.87
DP	0.20	0.09	0.77	0.48	0.15	0.86
NP	0.40	0.08	0.97	0.33	0.08	0.95
P/PP	0.41	0.07	0.58	0.67	0.16	0.83
QP	0.19	0.02	0.42	0.41	0.08	0.74
VP	0.07	0.02	0.16	0.53	0.13	0.97

4.3. Interaction between length and syntax for IP

Figure 5 shows the relationship between the complexity of syntax (number of InfP nodes, secondary nodes and syntactic depth) and the number of IP boundaries within a sentence.

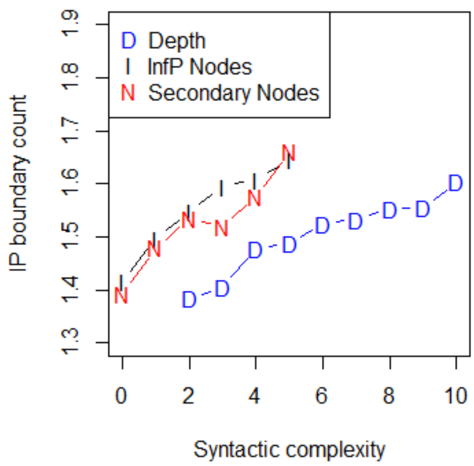
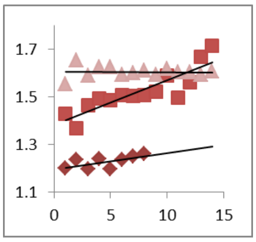


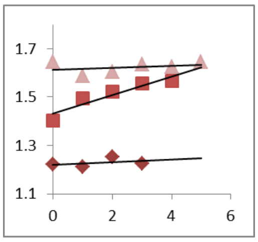
Figure 5: *Relationship between the count of intonational phrase boundaries within a sentence and the syntactic complexity (syntactic depth, number of IP nodes and number of secondary nodes in the tree) of the sentence*

Although each sentence normally only has 1 to 2 IP boundaries, we can still observe a very clear tendency that the number of IP boundaries in a sentence tend to increase as the complexity of the sentence becomes higher.

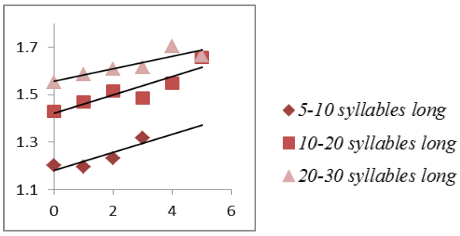
However, the trend in Figure 5 is not a pure reflection of the syntactic influence. The factor of length is also mediating in this trend, e.g., the longer the sentence, the deeper its syntactic tree, and the more InfP nodes it is likely to have. In order to tease apart these two influences, we divide all the sentences into three groups according to their lengths (short: 5-10 syllables, middle: 10-20 syllables, long: 20-30 syllables) and replicate the experiment again. The result is shown in Figure 6.



(a) Depth of the tree



(b) Count of InfP nodes



(c) Count of secondary nodes

Figure 6: *Relationship between syntactic complexity and number of IP boundaries for sentences of different lengths*

As shown in Figure 6, the sensitivity of intonational phrase boundaries to syntactic complexity varies as the sentence length differs. Depth of the tree and counts of InfP nodes do not seem to have a clear effect on the IP boundaries of short sentences and long sentences. Only the sentences of medium length are sensitive to these features in terms of IP boundaries. The number of secondary nodes seem to have an effect on sentences of all three levels of length. But in general, if a sentence is too long or too short, the grouping or splitting of intonational phrases can be more a direct result of sentence length and has less to do with the syntactic complexity of the sentence. The k and R^2 values for the fitted regression lines in Figure 6 are reported in Table 2.

Table 2. *k and R^2 for fitted lines with syntactic complexity as dependent factor and the number of IP boundaries as independent factors for sentences of different lengths*

Features	Short		Middle		Long	
	k	R^2	k	R^2	k	R^2
Depth	0.001	0.01	0.019	0.76	0.007	0.44
IP nodes	0.005	0.16	0.040	0.88	0.004	0.10
Secondary nodes	0.038	0.79	0.040	0.82	0.025	0.80

5. Discussion

In the literature, much more attention has been paid to the role of syntax in determining prosodic phrasing than to the role of rhythm. Despite several investigations of the effects of length on accentual phrases, sandhi application and clause attachment, our knowledge of rhythmic constraint is generally still limited to the broad idea that when a sentence is longer, it tends to be divided into more prosodic units.

One important finding of this study is that one of the prosodic units investigated, the so-called “rhythmic phrase”, is almost always about 2-5 syllables long. Interestingly, this unit is about the same length of an AP in Japanese and Korean, although there is no simple analogue to the AP in Mandarin Chinese. But this finding opens up the possibility that some version of this 'macro' rhythmic unit might be universally available across languages.

Another prosodic unit investigated in our paper, the intonational phrase, does not show clear preference towards a certain range of length, at least as counted in syllables. This observation is consistent with the claim that “the rhythmic nature of the unit would be weakened as the size of the postlexical prosodic unit increase, from prosodic word to intonation” [12]. However, we did find the number of "rhythmic phrases" in intonational phrases was generally again about 2 to 4 or 5.

Finally, it is worth noting that the distributions of phrase lengths based on silent reading and speech perception were quite similar in our study. Previous studies have also found that implicit prosody computed during silent reading is similar to overt prosody in planned reading [17] in terms of disambiguation, but prosody in unplanned reading is a another story [25]. The difference between implicit prosody and overt prosody elicited in different ways needs to be examined by future research.

6. Conclusions

This study explores the interaction between rhythmic constraints and syntactic constraints on rhythmic phrases and intonational phrases in Mandarin Chinese. We used two types of prosodic annotation for the same set of 4000 sentences, one based on silent reading and a second one based on speech perception.

We found that rhythmic phrases are normally 2-5 syllables long, in both implicit prosody and overt prosody, and this range of lengths is hardly at all influenced by the length of the sentence. By contrast, the length of intonational phrases is strongly influenced by sentence length, in both implicit prosody and overt prosody. We find that the probability of a rhythmic-phrase boundary between two words is influenced by both the total length of the two words and the type of syntactic juncture between them. For intonational-phrase boundaries, the length-syntax interaction results in an alternation of main effect on IP boundaries from length to syntax then to length again when the sentence is short, medium, and long.

7. Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (NSFC) (No.61305003), the Beijing Normal University (11-30, 10-12-02), the Fundamental Research Funds for Central University (2010105565004GK), and the State Language Commission 12th Five-Year language research programme (YB125-41).

8. References

[1] E. O. Selkirk, *The phrase phonology of English and French*. Routledge, 2015.

[2] B. Hayes, "The phonology of rhythm in English." *Linguistic Inquiry* 15, no. 1, pp. 33-74. 1984.

[3] A. Arvaniti and M. Baltazani, "Intonational analysis and prosodic annotation of Greek spoken corpora." *Prosodic typology: The phonology of intonation and phrasing*, pp. 84-117. 2005

[4] A. Cutler, D. Dahan, and W. V. Donselaar, "Prosody in the comprehension of spoken language: A literature review." *Language and speech* 40, no. 2, pp. 141-201. 1997.

[5] S. Shattuck-Hufnagel and A. E. Turk, "A prosody tutorial for investigators of auditory sentence processing." *Journal of psycholinguistic research* 25, no. 2, pp. 193-247. 1996.

[6] E. M. Kaisse and P. A. Shaw, "On the theory of Lexical Phonology." *Phonology yearbook* 2, no. 1, pp. 30. 1985.

[7] J. Rotenberg, "The syntax of phonology." PhD diss., Massachusetts Institute of Technology, 1978.

[8] D. J. Napoli and M. Nespor. "The syntax of word-initial consonant gemination in Italian." *Language*, pp. 812-841. 1979.

[9] E. O. Selkirk, Phonology and syntax: the relationship between sound and structure. MIT press, 1986.

[10] Y. Hirose, "Recycling prosodic boundaries." Journal of psycholinguistic research 32, no. 2, pp. 167-195. 2003.

[11] S. Jun, "The effect of phrase length and speech rate on prosodic phrasing." In *Proceedings of the XVth International Congress of Phonetic Sciences*, pp. 483-486. 2003.

[12] S. Jun, ed. *Prosodic typology II: the phonology of intonation and phrasing*. Vol. 2. Oxford University Press, 2014.

[13] M. Ueyama, "Speech rate effects on phrasing in English and Japanese." *Unpublished manuscript, University of California, Los Angeles* (1998).

[14] S. Jun and C. Fougeron. "Realizations of accentual phrase in French intonation." *Probus* 14, no. 1, pp. 147-172. 2002.

[15] M. Nespor and I. Vogel. *Prosodic phonology*. Vol. 28. Walter de Gruyter, 2007.

[16] E. O. Selkirk, "The interaction of constraints on prosodic phrasing." In*Prosody: Theory and experiment*, pp. 231-261. Springer Netherlands, 2000.

[17] N. Lovrić, D. Bradley and J. D. Fodor. "RC attachment in Croatian with and without preposition." In *Poster presented at the AMLaP Conference, Leiden*. 2000.

[18] D. Quinn, H. Abdelghany, and J. D. Fodor. "More evidence of implicit prosody in silent reading: French, English and Arabic relative clauses." In *Poster presented at the 13th Annual CUNY Conference, La Jolla, CA*. 2000.

[19] H. Hwang and K. Steinhauer. "Phrase length matters: The interplay between implicit prosody and syntax in Korean “garden path” sentences." *Journal of Cognitive Neuroscience* 23, no. 11, pp. 3555-3575. 2011.

[20] C. Shih, "Mandarin third tone sandhi and prosodic structure." *Linguistic Models* 20, pp. 81-124.1997.

[21] S. Peng, M. Chan, C. Tseng, T. Huang, O. J. Lee, and M. E. Beckman. "Towards a Pan-Mandarin system for prosodic transcription." Prosodic typology: The phonology of intonation and phrasing. pp. 230-270. 2005.

[22] S. Feng, “*Hanyu de yunlv, cifa yu jufa.*” [The prosody, morphology and syntax of Mandarin]. Peking University press, 1997.

[23] http://nlp.stanford.edu//software/lex-parser.shtml

[24] H. Che, Y. Li, J. Tao, and Z. Wen. "Investigating Effect of Rich Syntactic Features on Mandarin Prosodic Boundaries Prediction." Journal of Signal Processing Systems 82, no. 2, pp. 263-271.2016.

[25] S. Jun, "The implicit prosody hypothesis and overt prosody in English." Language and Cognitive Processes 25, no. 7-9, pp. 1201-1233. 2010.