

# Does Prosodic-Foot Disyllabicity Hold a Default Status in Mandarin Speech Perception?

Ivan Chow, Ph.D.

University of Toronto, Canada  
i\_chow@sympatico.ca

## Abstract

An acoustic experiment was conducted to investigate the perceptual processing of native Mandarin speakers in syntactic ambiguity resolution<sup>1</sup>. The statistical majority of disyllabic words in modern Chinese [17] and the prosodic theory of Binary Foot Formation in Mandarin [3], [4], [9], [10], [11], [15] point to the stipulation that the disyllabic prosodic feet have a special 'default' status in the prosodic structure of Mandarin. In a situation where neither contextual nor acoustic information is available, given an utterance that can be parsed into disyllabic or tri-syllabic prosodic feet without violating syntactic constituency, native speakers of Mandarin may be biased towards the disyllabic organization. Although experimental results show that Mandarin speakers are more likely to parse the utterance into one of two possible syntactic structures rather than simply recognizing it as ambiguous cf. [7], [8], [13], utterances organized into disyllabic feet are not favoured over ones organized in tri-syllabic feet. *A fortiori*, there is no indication that the quantitative identity of the prosodic feet is taken into account in the perception and parsing of syntactically ambiguous sentences.

## 1. Introduction

In Mandarin (and in modern Chinese), the majority of words are composed of two syllables. According to Wu & Zhu [17], 73.6% of Chinese words are disyllabic, an additional 12% of monosyllabic words can be readily converted into disyllabic words without changing their meaning. Moreover, there is a tendency for borrowed words and newly invented abbreviations to conform to the disyllabic pattern. For example,

- (1) "cai3 se4 dian4 shi4 ji1" (colour television) → "cai3 dian4"  
"kong1 qi4 tiao2 jie3" (air-conditioning) → "kong1 tiao2"  
"luo4 hua1 sheng1" (peanuts) → "hua1 sheng1"  
"pi2" (beer) → "pi2 jiu3"  
"guan1" (Guam) → "guan1 dao3"  
"dian4" (shop) → "dian4 pu4"  
"qi1" (wife) → "qi1 zi"  
"qing2" (clear, sunny) → "qing2 tian1"

In addition, utterance in Mandarin tend to be organized prosodically into disyllabic feet, making the disyllabic foot a

prominent building block for the prosodic structure in Mandarin [3], [4], [9], [10], [11]. Feng [10], [11] proposed a Prosodic Foot Formation Rule that forbids prosodic feet to be formed with less than two syllables. A binary foot consisting of two syllables is thus the smallest prosodic foot structure in Mandarin.

- (2) Prosodic Foot Formation Rule in Mandarin [11: 54]

\* f  
|  
σ

Shih [15] proposed a set of rules governing the formation of prosodic feet in Mandarin, stated as follows:

- (3) Direct Duple Meter Approach
- Foot Construction
    - Immediate Constituency: Link immediate constituents (syllables) into disyllabic feet.
    - Duple Meter: Scanning from left to right, string together unpaired syllables into binary feet, unless they branch to the opposite direction.
  - Super-foot Construction  
Join any leftover monosyllable to a neighbouring binary foot according to the direction of syntactic branching.

According to this set of rules, syllables dominated by the same syntactic constituent are linked together in pairs. Unpaired syllables are linked together into adjacent binary feet according to their direction of branching. In the case where a syntactic constituent dominates a paired prosodic foot plus an adjacent unpaired syllable, a 'super-foot' is formed by merging the unpaired syllable with the paired foot to form a tri-syllabic foot. As a result, constituencies aside, a given sequence of five syllables is organized in one of the two following ways:

- (4) Possible prosodic feet formation of a sequence of five syllables

σ σ σ σ σ → σ σ | σ σ σ or σ σ σ | σ σ

Given the special status of the disyllabic foot in prosodic foot formation and its statistical majority in the repertoire of Chinese words, an interesting question arises in the linguistic processes of speech parsing among Mandarin speakers. Does disyllabicity hold a certain 'default' phonological status in speech perception? In other words, if a given string of syllables can be organized into either disyllabic (binary feet) or tri-syllabic groups (super-feet) without violating syntactic constituency, will the binary foot structure be favoured over the tri-syllabic duper-feet structure?

<sup>1</sup> This acoustic experiment is part of a larger-scale experiment carried out to investigate the prosodic pattern of Mandarin using syntactically ambiguous sentences. Full description of the experiment is found in [5].

## 2. The Acoustic Experiment

### 2.1. The Hypothesis

Using syntactically ambiguous test sequences, an acoustic experiment was conducted to investigate the perceptual patterns of native Mandarin speakers. Research in psycholinguistics indicate that it is more mnemonically costly for listeners to parse a sentence into two possible syntactic structures simultaneously. Consequently, in the case where no acoustic cues are available to help resolving syntactic ambiguity, despite the risk or error, listeners tend to parse the sentence into one of its possible syntactic representations at an early stage rather than recognizing the sentence as ambiguous and delay the decision-making until more information is available [1], [2], [6], [7], [8], [13], [16]. Keeping this in mind, the experiment sets out to investigate whether native speakers of Mandarin are quantitatively biased due to the alleged 'default' status of prosodic foot disyllabicity. If so, subjects are expected to favour the disyllabic alternative over the tri-syllabic one in an environment where no contextual or perceptual cues are available for disambiguation.

#### (5) Hypothesis:

Given the default phonological status and statistical prominence of the disyllabic foot, all else being equal, in a situation where an ambiguous sequence of syllables can be parsed into structures consisting entirely of disyllabic or tri-syllabic feet, listeners are more likely to parse the sequences into disyllabic feet.

### 2.2. Methods

Five test sequences of six syllables were carefully composed for the experiment. With no contextual information and perceptual cues given, each test sequence can be parsed into exactly three disyllabic or two tri-syllabic feet without violating syntactic constituency. The test sentences were designed so that the two alternative structures are equal (or similar) in terms of ease of lexical and semantic retrieval. As a result, preference due to ease of lexical retrieval is considered to be kept to a minimum. Sequences of consecutive tone-3 are avoided in order to eliminate any clue given out by tone-3 Sandhi-triggered melodic changes within a Sandhi domain cf. [3], [4], [9], [15], [18], [20]. The test sequences (printed in bold face) are shown in the example below. Some sequences are embedded in full sentences in order to form a complete line of thought.

#### (6) Test Sequences used in the Acoustic Experiment

- (a) **ta1 cong2 / xiao3bian4 / zhi1dao4** zi4ji3 you3  
tang2niao4bing4  
(He finds out from his urine that he has diabetes.)  
**ta1 cong2xiao3 / bian4 zhi1dao4** zi4ji3 you3  
tang2niao4bing4  
(He has known since he was a kid, that he has diabetes.)
- (b) **san1qian1 / duo1da4 / xue2sheng1**  
(3,000 University of Toronto students.)  
**san1qian1duo1 / da4xue2sheng1**  
(More than 3,000 university students.)
- (c) **hong2se4 / zhi3 na2 / san1 zhang1**  
(Take only 3 sheets of the red ones.)

**hong2se4 zhi3 / na2 san1 zhang1**

(Take 3 sheets of red paper.)

- (d) **gang3shang1 / hui4 zai4 / nei4di4** she4li4  
ban4shi4chu4  
(Merchants from Hong Kong will open offices in Mainland China.)  
**gang3shang1hui4 / zai4 nei4di4** she4li4  
ban4shi4chu4  
(The Hong Kong Trade Association will open offices in Mainland China.)
- (e) **jian4shen1 / cao1lian4 / wo3men2** de si4zhi1  
(Exercises strengthen our limbs.)  
**jian4shen1cao1 / lian4 wo3men2** de si4zhi1  
(Aerobics strengthen our limbs.)

The test sequences were read by eight volunteers<sup>2</sup> (three male and five female, aged 25-50, all native speakers of Mandarin) in a quiet room and recorded with a Sony digital audio recorder. The recording was analyzed digitally with WinPitch for Windows 2000.

In the acoustic analysis, prosodic boundaries were found to be signalled rhythmically by pauses and pre-boundary lengthening and melodically by pitch reset. Pre-boundary lengthening refers to the phenomenon where the duration of the last syllable before the prosodic boundary is significantly elongated compared to syllables not in the pre-boundary position. Pitch reset refers to the disruption of the otherwise continuous pitch declination line that runs from the beginning to the end of a sentence or an intonational phrase; the declination line restarts after the boundary at a higher pitch register and continues to decline towards the end of the sentence.

Mandarin is a lexical-tone language in which the shape and register of the pitch contour are integral to lexical retrieval. Unlike atonal languages such as French and English, pitch movements in Mandarin are involved very minimally in the marking of prosodic boundaries (except for pitch reset). For the listening comprehension test, sentences with minimal temporal and melodic boundary cues were selected and digitally altered to eliminate all prosodic boundary cues and to ensure a structurally neutral but lexically comprehensible output. Pauses were eliminated by sensibly deleting silent segments between syllables. Pre-boundary lengthening was neutralized by carefully deleting part of the vowel segment of the syllable in order to render it similar in length to other syllables in non-pre-boundary positions. Melodic adjustments were done only to minimize the difference in pitch register before and after the boundary caused by pitch reset. Impact to the melodic contour shape was kept to a minimum as any drastic variation to the contour shape would result in changes to the meaning of individual words of the sentence. In the extreme case, using monotone would wipe out lexical tone information and render the sentences virtually incomprehensible.

The digitally treated sequences were rendered, checked for unnaturalness, and recorded onto an audio CD for the listening comprehension experiment. For each track, a test sequence is

<sup>2</sup> I am grateful to the staff and graduate students of the Department of East Asian Studies, University of Toronto for their participation in the recording.

played twice with a three-second pause between each repetition. The audio CD was played to 33 volunteers<sup>3</sup> (13 male and 20 female, aged 33-65, all native speakers of Mandarin living in the Greater Toronto area with no reported hearing problems) who had no prior knowledge of the test sequences. An answer sheet was given to each participant. Each question on the answer sheet corresponded to one track (one test sequence) with three answer choices. The subjects were asked to circle only one answer choice they thought corresponded most to the sequence they had perceived. Three answer choices were printed for the question corresponding to each sentence-pair. Answer choices (A) and (B) correspond to the first and second possible structures respectively. (A) corresponds to the disyllabic structure whereas (B) corresponds to the tri-syllabic structure. The structures were printed after the corresponding letters. A third answer choice "(C). Can be either one." was given, meaning that it is not clear whether the audio track corresponds to one structure or the other. According to the design of the experiment, (C) reflects the reality and should be the correct answer—in fact, no prosodic, contextual or lexical information was given whatsoever to prompt listeners to parse the recorded sentences as either of the two possible structures. After explaining the test conditions to the subjects and answering questions concerning the procedure, the audio tracks were played and all answer sheets were collected for analysis afterwards.

### 2.3. Results

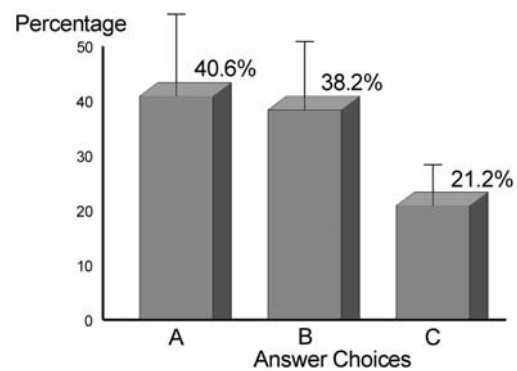
Before looking at the experimental data, we return once more to our predictions. Despite the fact that the answer (C), the correct answer, is given as an option, Nicol & Pickering [13] predict that listeners are more likely going to run the risk of making an incorrect decision; they tend to choose one alternative at an early stage, rather than continuing to parse the two possible structures simultaneously and recognizing the sentence as ambiguous. On the other hand, our hypothesis predicts that, due to the 'default' status of disyllabic prosodic feet in Mandarin, there would be a difference in popularity between the two other alternatives (A & B); the structure composed solely of disyllabic feet (A) should be favoured by native speakers of Mandarin because of their specific native-language experience.

We now compare our hypothesis with the experimental results. In total, 165 instances (33 participants x 5 sentence pairs) were collected. The percentage reflects the frequency of the answer choice being chosen.

As predicted, the answer choice (C) received the lowest percentage score of 21.21%. The other two alternatives received almost twice as much as (C), with (A) receiving a slightly but insignificantly higher score than (B). (Figure 7 below).

<sup>3</sup> I would also like to thank the members of the Mandarin Fellowship of the Markham Chinese Presbyterian Church for their participation in the listener comprehension experiment.

(7) Experimental Data



(8) Frequency Percentage for Each Sentence Pair

Sentence Pair	Answer Choice (A) (Disyllabic)	Answer Choice (B) (Tri-syllabic)	Answer Choice (C) (Ambiguous)
(a)	21.2%	48.5%	30.3%
(b)	33.3%	45.5%	21.2%
(c)	45.5%	42.4%	12.1%
(d)	39.4%	36.4%	24.2%
(e)	63.6%	18.2%	18.2%

A more detailed look at the results for each question (figure 8 above) reveals a rather different picture. While the answer patterns of all test sentences consistently show a low percentage score for the answer (C), despite efforts to eliminate bias due to differences in ease of lexical retrieval, sentence pairs (a), (b), and (e) showed a strong preference towards one of the two alternatives while the answer pattern of the other three sequence-pairs showed a somewhat consistent picture with the overall statistics.

### 3. Discussion

First we look at the overall percentage for answer (C) "Can be either one". The significantly lower percentage compared to the two other alternative confirms the prediction put forth in the psycholinguistic research [7], [8], [13]. While no contextual or perceptual information is available in the resolution of syntactic ambiguity, listeners are in fact more likely to parse the sequences as one of the two alternatives rather than simply admitting that there is more than one possible structure and delay their decision-making within the parsing process.

In the overall statistics, much higher than that of (C), the percentages for answers (A) and (B) are almost equal, with an insignificant difference of 1.4%. This indicates that, although listeners are more likely to pick one possible syntactic structure, the disyllabic structure is not favoured over the tri-syllabic one. The quantitative feature of the prosodic organization does not seem to play an important role in the decision-making process of speech parsing. While disyllabic words and binary feet are common in the prosodic organization of utterance in Mandarin, it does not seem to hold a 'default' status in speech perception. Our hypothesis is thus

falsified. Frazier & Rayner's [12] study indicates that, when acoustic, contextual and lexical information is not available, listeners rely mainly on syntactic information to resolve syntactic ambiguity. For example, listeners tend to choose the structure with "the fewest nodes consistent with the well-formedness rules of the language" [12:181]. The fact that the quantitative identity of the prosodic organization does not seem to be a factor in speech perception provides further empirical support to this view. However, whether syntax is the sole source of information in ambiguity resolution is subject to further examination.

Despite efforts in minimizing biases between the two structures caused by reasons other than the quantitative identity of syllabic structures, results in three test sentences showed a strong but inconsistent bias towards one of the two alternatives. Interestingly, the answer scores in sentence pairs (a) and (b) showed bias towards the tri-syllabic structure while the sentence pair (e) showed a strong bias towards the disyllabic alternative. The fact that the disyllabic alternative is not favoured in this bias seems to implicate that the quantitative identity of the syllabic structure is not the cause of the bias in these particular sentence pairs. It is stipulated that differences in ease of lexical retrieval may be one of the many causes of this bias. However, future experimental studies are required to confirm this stipulation. Nonetheless, this observation does not seem to mitigate our experimental findings as there is no indication that the quantitative identity of prosodic foot comes into play in the parsing process in any of the test sentences.

#### 4. Conclusion

The acoustic experiment discussed in this paper investigated whether Mandarin speakers are quantitatively biased towards disyllabic prosodic feet in speech perception. Due to its statistical majority within the repertoire of words in Chinese and its status as the 'default' quantitative feature of prosodic feet in Mandarin, it was stipulated that Mandarin speakers would favour a disyllabic structure in a syntactic ambiguity situation where no contextual or perceptual information is available to help with the disambiguation. However, our experimental data falsified this stipulation. Given an ambiguous sequence that can be parsed into disyllabic or tri-syllabic groups without violating syntactic constituency, experimental data indicate that, native speakers of Mandarin are just as likely to choose one structure or the other, regardless of their prosodic make-up. In other words, the quantitative aspect of the prosodic organization does not seem to be taken into account in the process of speech parsing. On the other hand, consistent with psycholinguistic research, despite the risk of error, native speakers of Mandarin tend to parse the sentence into one possible structure rather than simply recognizing the utterance as ambiguous and try to parse more than one structure simultaneously.

#### 5. References

- [1] Beach, C., 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30, 627-643.
- [2] Bresnan, J., 1971. Sentence stress and syntactic transformations. *Language*, 47, 257-281.
- [3] Chao, Y-R., 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- [4] Chen, M., 2000. *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge, UK: Cambridge University Press.
- [5] Chow, I., 2003. *Prosodic Structures in French and Mandarin*. Ph.D. Thesis. University of Toronto.
- [6] Cooper, W.E.; Paccia-Cooper, J., 1980. *Syntax and Speech*. Cambridge, Mass.: Harvard University Press.
- [7] Crain, S.; Fodor, J.D., 1985. How can grammars help parsers? D.R. Dowty, L. Karttunen, & A. Zwicky (eds.), *Natural language parsing*. New York: Cambridge University Press.
- [8] Crain, S. and Steedman, M. J., 1985. On not being led up the garden path: The use of context by the psychological parser. D.R. Dowty, L. Karttunen, & A. Zwicky (eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge, England: Cambridge University Press.
- [9] Duanmu, S., 2000. *The Phonology of Standard Chinese*. Oxford University Press.
- [10] Feng, S., 1995. *Prosodic structure and Prosodically constrained syntax in Chinese*. Ph.D. dissertation, University of Pennsylvania.
- [11] Feng, S., 2001. *Yun Lü Ci Yü Ke Xue Li Lun De Kou Jian (The Construction of Prosodic Word and Scientific Theories)*. Shi Jie Han Yü Jiao Xue (Chinese Pedagogy in the World), 55:1, 53-64.
- [12] Frazier, L.; Rayner, K., 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- [13] Nicol, J.; Pickering, M., 1993. Processing Syntactically Ambiguous Sentences: Evidence from Semantic Priming. *Journal of Psycholinguistic Research (JpsyR)*. New York, NY. 22:2, 207-237.
- [14] Shen, X. S., 1992. A Pilot Study on the Relation between the Temporal and Syntactic Structures in Mandarin. *Journal of International Phonetic Association*, (1992) 22:1/2, 35-43.
- [15] Shih, C., 1986. *The prosodic domain of Tone Sandhi in Chinese*. Ph.D. thesis, University of California- San Diego.
- [16] Speer, S.R.; Kjelgaard, M. M.; Dobroth, K. M., 1996. The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25, 247-268.
- [17] Wu, J.; Zhu, H., 2000. *Han Yu Jie Lü Xue (Chinese Prosody)*. Yu Wen Chu Ban She, Beijing, China. <http://www.ywcb.com>.
- [18] Xu, Y., 1997. Contextual Tonal Variations in Mandarin. *Journal-of-Phonetics (Jphon)*. Ann Arbor, MI. January, 1997. 25:1, 61-83.
- [19] Yip, P., 1992. Grammatical Identity of a Chinese Word. *Journal of the Chinese Language Teachers Association*, 28:3, 53-73.
- [20] Zhang, N., 1997. The Avoidance of the Third Tone Sandhi in Mandarin Chinese. *Journal-of-East-Asian-Linguistics (JEAL)*, Irvine, CA. October, 1997. 6:4, 293-338.