

本文章已註冊DOI數位物件識別碼

► Tonal Patterns in Chinese Regulated Verse: A Corpus Study

doi:10.6519/TJL.2005.3(1).1

Taiwan Journal of Linguistics, 3(1), 2005

臺灣語言學期刊, 3(1), 2005

作者/Author : San Duanmu; Nathan Stiennon

頁數/Page : 1-31

出版日期/Publication Date : 2005/06

引用本篇文獻時，請提供DOI資訊，並透過DOI永久網址取得最正確的書目資訊。

To cite this Article, please include the DOI name in your reference data.

請使用本篇文獻DOI永久網址進行連結:

To link to this Article:

[http://dx.doi.org/10.6519/TJL.2005.3\(1\).1](http://dx.doi.org/10.6519/TJL.2005.3(1).1)



DOI Enhanced

DOI是數位物件識別碼 (Digital Object Identifier, DOI) 的簡稱，是這篇文章在網路上的唯一識別碼，用於永久連結及引用該篇文章。

若想得知更多DOI使用資訊，

請參考 <http://doi.airiti.com>

For more information,

Please see: <http://doi.airiti.com>

請往下捲動至下一頁，開始閱讀本篇文獻

PLEASE SCROLL DOWN FOR ARTICLE



**TONAL PATTERNS IN CHINESE REGULATED VERSE:
A CORPUS STUDY***

San Duanmu and Nathan Stiennon

ABSTRACT

It is well known that Chinese regulated verse should meet certain tonal requirements. We offer a corpus study and examine to what extent the requirements are satisfied. The corpus consists of all regulated poems with seven-syllable lines in the anthology of *Three Hundred Tang Poems*. The result shows that only about 32% of the lines and 1% of the poems meet the strict version of the tonal requirements, according to which every syllable must use a designated tone. On the other hand, 95% of the lines and 68% of the poems satisfy the relaxed version of the tonal requirements, according to which the tonal choices for the first, third, and fifth syllables of a line are more flexible. We also discuss a some other points of interest: the Gu Ping constraint, comparisons of our results with those of five-syllable lines (Ripley 1980), and word frequencies.

1. INTRODUCTION

It is often stated that, unlike older verse or folk verse, Chinese regulated verse has special requirements for tonal alternations. The requirements were probably introduced by Shen Yüeh and others in the fifth century in imitation of the alternation between long and short syllables in Sanskrit meter (Mair and Mei 1991). It is rarely shown though how faithfully regulated verse follows the tonal rules. We are aware of only one published corpus study in this regard, which is Ripley

* This work was supported by the Undergraduate Research Opportunity Program at the University of Michigan. The authors contributed to this study equally. Their names are listed in alphabetical order. We would like to thank two anonymous reviewers and Yuchau Hsiao for their helpful comments.

(1980), based on his dissertation (Ripley 1979). However, Ripley's work only considers poems with five syllables per line. In this article we offer a study that considers poems with seven syllables per line.¹

In regulated verse each syllable belongs to one of two tonal categories, either even (E) or oblique (O), which must alternate in certain ways. The E and O tones probably correspond to what were historically long and short vowels, respectively (Zhou 1948). Previous descriptions provide four archetypal seven-syllable lines (Wang 1958), which we label as A through D.

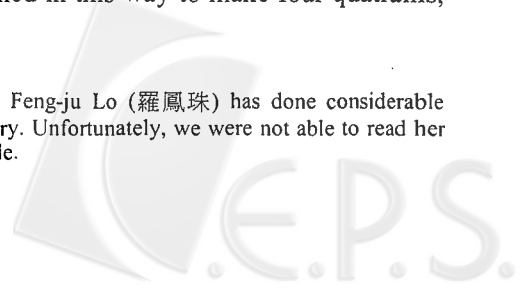
- (1) Four line types
- a. EEOEEEO (A)
 - b. OOEEOOE (B)
 - c. OOEEOOE (C)
 - d. EEOOEE (D)

The lines can form four couplets with mirrored patterns, shown in (2).

- (2) Four couplet types
- a. EEOEEEO (A)
OOEEOOE (B)
 - b. OOEEOOE (B)
EEOEEEO (A)
 - c. OOEEOOE (C)
EEOOEE (D)
 - d. EEOOEE (D)
OOEEOOE (C)

The final syllables of even-numbered lines tend to rhyme. In our corpus, rhyming syllables tend to have E tone. Thus it is usually couplets of type (2a) and (2c) that are encountered. Poems in regulated verse consist of quatrains composed of two differing couplets ending in the same tone. The couplets in (2) can be combined in this way to make four quatrains, shown in (3).

¹ A reviewer points out that Professor Feng-ju Lo (羅鳳珠) has done considerable amount of corpus study on Chinese poetry. Unfortunately, we were not able to read her work or compare it with ours in this article.



- (3) Four quatrain types
- a. EEOOEO (A)
OOEEOE (B)
OOEEEO (C)
EEOOEE (D)
 - b. OOEEOE (B)
EEOOEO (A)
EEOOEE (D)
OOEEEO (C)
 - c. OOEEOE (C)
EEOOEE (D)
EEOOEO (A)
OOEEOE (B)
 - d. EEOOEE (D)
OOEEEO (C)
OOEEOE (B)
EEOOEO (A)

Again, couplets that end in O are rare in our corpus, so the most common quatrains are of type (3a) and (3c). Eight-line poems in regulated verse consist of two quatrains of the same type. Two significant variants must be noted. First, in many poems the pattern of the first line is replaced with the pattern of the last line in the quatrain. This gives rise to the quatrain types in (4).

- (4) Quatrain types with first line rhyming
- a. EEOOEE (D)
OOEEOE (B)
OOEEEO (C)
EEOOEE (D)
 - b. OOEEOE (C)
EEOOEO (A)
EEOOEE (D)
OOEEEO (C)

- c. OOEEOOE (B)
- EEOOOEE (D)
- EEOOEEO (A)
- OOEEOOE (B)
- d. EEOOEEO (A)
- OOEEEOO (C)
- OOEEOOE (B)
- EEOOEEO (A)

Most poems in our corpus employ quatrains of type (3a), (3c), (4a), or (4c). The purpose of the alternative quatrain pattern in (4) is to allow the first line to rhyme with the second and fourth lines (Wang 1958). We will call this pattern “first-line rhyming”. In our corpus, the second quatrain of eight-line poems does not use first-line rhyming.

It has also been noted that poets often do not restrict the tones of certain syllables. In particular, there is the so-called “1-3-5 Rule”. A careful examination of the 1-3-5 Rule will be given later. Meanwhile, we will assume a literal interpretation of the rule, which says that the first, third, and fifth syllables in a seven-syllable line are free to be either E or O, but the second, fourth, sixth, and seventh syllables must conform to one of the archetypal patterns (Wang 1958). If the 1-3-5 Rule is used, the eight possible quatrains (four with first-line rhyming and four without) are shown in (5).

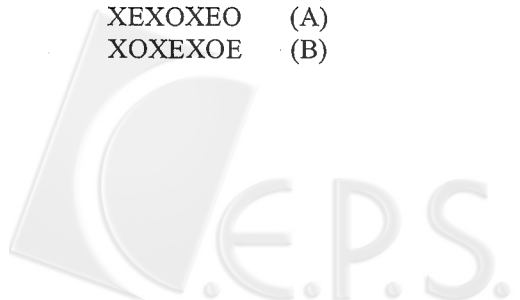
(5) Quatrain types with the 1-3-5 Rule (X is any tone)

Without first-line rhyming

- a. XEXOXEO (A)
- XOXEXO E (B)
- XOXEXOO (C)
- XEXOXEE (D)
- c. XOXEXOO (C)
- XEXOXEE (D)
- XEXOXEO (A)
- XOXEXO E (B)

With first-line rhyming

- b. XEXOXEE (D)
- XOXEXO E (B)
- XOXEXOO (C)
- XEXOXEE (D)
- d. XOXEXO E (B)
- XEXOXEE (D)
- XEXOXEO (A)
- XOXEXO E (B)



- | | | | | | |
|----|--------|-----|----|--------|-----|
| e. | XOXEXO | (B) | f. | XOXEXO | (C) |
| | XEXOEO | (A) | | XEXOEO | (A) |
| | XEXOEE | (D) | | XEXOEE | (D) |
| | XOXEXO | (C) | | XOXEXO | (C) |
| g. | XEXOEE | (D) | h. | XEXOEO | (A) |
| | XOXEXO | (C) | | XOXEXO | (C) |
| | XOXEXO | (B) | | XOXEXO | (B) |
| | XEXOEO | (A) | | XEXOEO | (A) |

The quatrains in the right column exhibit first-line rhyming. Note that in an eight-line poem, first-line rhyming only occurs in the first quatrain. If the first quatrain is from the right column, the second quatrain is the corresponding form in the left column (Wang 1958). Again, in our corpus rhyming syllables usually have E tone, so most quatrains are of type (5a)-(5d).

Previous descriptions also mention an additional constraint called “Gu Ping” (literally ‘lone even’), which prohibits a line with no more than one E tone in the first six syllables (Wang 1958). This rule would take precedence over the 1-3-5 Rule. For example, a line of type OOOEOE violates the Gu Ping rule because there is only one E tone in the first six syllables, even though this line can be considered a valid variation of type B under the 1-3-5 Rule (i.e. XOXEXO). This will be discussed later.

2. METHODS

The corpus for our study is taken from the anthology of *Three Hundred Tang Poems*, which contains poems from many Tang writers. Many online sources have posted the entire anthology, where one can obtain the text. For ease of reference, we follow the edition by Qiu (1976) and number the poems sequentially. There are a total of 320 poems in the anthology, of which 231 are regulated poems. Among the regulated poems 114 have seven syllables per line, with a total of 672 lines. This is shown in (6).

(6) Regulated poems with seven syllables per line

Lines per poem	# of such poems	Total lines	Total syllables
8	54	432	3024
4	60	240	1680
All	114	672	4704

The entire corpus is arrayed in a spreadsheet. Then the tonal category (E or O) for each syllable is determined, in order to determine to what extent the sequence of E and O tones in each poem agrees with the principles outlined in the previous section (see below).

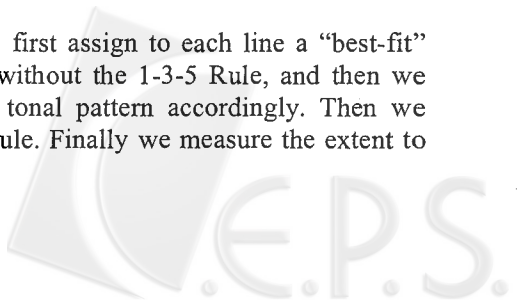
Our initial hypothesis is that every line in the corpus conforms to a line pattern in (1); every couplet conforms to a couplet pattern in (2); every quatrain conforms to a quatrain pattern in (3) or (4); and every eight-line poem is composed of two quatrains with tone patterns that are identical except for possible first-line rhyming. If every line is perfect, the analysis would be straightforward. However, if some lines are not perfect, as often happens, then there would be ambiguities. For example, a line with EOOE~~OO~~ has several possible analyses, shown in (7).

(7) Possible analyses of EOOE~~OO~~

Line type	Analysis	Errors
C without 1-3-5	<u>OOEEOO</u>	three syllables
C with 1-3-5	XOXEXOO	none
B without 1-3-5	<u>OOEEOOE</u>	three syllables
B with 1-3-5	XOXEXOE	one syllable

If we consider the line alone, we might choose the analysis “C with 1-3-5”, which seems to be the best fit (no error syllable). However, if we consider the line within a poem, we might be forced to make another choice; for example, if all other lines in the poem suggest that this line should be “B with 1-3-5”, then we have to say that this line has one error syllable, instead of saying that this line is perfect but all the other lines have errors.

Under such considerations, we first assign to each line a “best-fit” tonal pattern from the list in (1), without the 1-3-5 Rule, and then we assign to each poem a “best fit” tonal pattern accordingly. Then we reassign patterns using the 1-3-5 Rule. Finally we measure the extent to



which the poems agree with these ideal forms, tallying the number of lines, couplets, quatrains, and poems that match each pattern.

2.1. Determining E and O tones

Our corpus contains a total of 4704 characters tokens (1289 different characters). First, all characters were converted into Pinyin, the Roman spelling system, using the Convert function of the Chinese word processor NJStar. Then the tones for each syllable, represented as the digits 1 through 5 (where 5 indicates unstressed toneless syllables), were extracted and placed in a new column. Some examples are shown in (8).

(8)	Character	Pinyin	Tone	Gloss
	昔	xi2	2	old
	人	ren2	2	person
	已	yi3	3	already
	乘	cheng2	2	ride
	• •	huang2	2	yellow
	鶴	he4	4	crane
	去	qu4	4	depart

Pinyin is based on modern Standard Chinese, which has five tonal categories, which derive from four classical tonal categories. However, Tang poems only distinguish two tonal categories, E (even) and O (oblique). The correspondences among classical tonal categories, the two poetic tonal categories (E and O), and modern tonal categories are shown in (9).

(9)	Classical	Poetic	Modern
	Ping	E	1, 2, 5
	Shang	O	3, 5
	Qu	O	4, 5
	Ru	O	1, 2, 3, 4, 5

The relation between modern tonal categories and the poetic categories E and O is shown in (10).

(10)	Modern	1	2	3	4	5
	Poetic	E, O	E, O	O	O	E, O

All modern tone 3 and tone 4 syllables belong to the poetic category O. However, modern tone 1, tone 2, and tone 5 syllables may come from the classical category Ping, which belongs to the poetic category E, or from the classical category Ru, which belongs to the poetic category O. Thus, all tone 1, tone 2, and tone 5 syllables were manually examined to determine their poetic categories. A summary of modern tone counts is given in (11) and the counts of E and O tones are shown in (12).

(11)	Modern	1	2	3	4	5	Total
	Count	1325	1403	620	1345	11	4704

(12)	Category	E	O	All
	Count	2470	2234	4704
	%	52.5	47.5	100

As can be seen, there are slightly more E tones than O tones.

2.2. Alternate pronunciations and their effect

Some Chinese characters have two or more pronunciations each. For such characters the Convert function in NJStar offers the most common pronunciation. However, less common pronunciations should not be ignored. Some examples are shown in (13), based on the multi-pronunciation table of Tsai (2000).

(13)	Char.	Pronunciations	Tones	Poetic	Gloss
	數	shu4, shu3, shuo4	4, 3, 4	O	number/count
	斜	xie2, xia2	2, 2	E	tilted
	長	chang2, zhang3, zhang4	2, 3, 4	E/O	long/grow
	• •	he2, huo4	2, 4	E/O	and/resonate
	撒	sa1, sa3, sa4	1, 3, 4	E/O	let loose/spread

For some characters, such as 數 and 斜, all the pronunciations belong to the same poetic tonal category. For other characters, such as 長, • • ; and

撒, different pronunciations lead to different poetic tonal categories. Sometimes a different pronunciation has a somewhat different meaning. For example, for the character 長, *chang2* means ‘long’ but *zhang3* means ‘grow’. Similarly, for the character 數, *shu4* means ‘number’, *shu3* means ‘count’, and *shuo4* means ‘frequent’. On the other hand, for the character 斜, both *xie2* and *xia2* mean ‘tilted’. It is possible, sometimes, to determine which pronunciation is the most appropriate for a given poem. On the other hand, not all alternative pronunciations occur in the same dialect. For example, according to *Xiandai Hanyu Cidian* ‘Modern Chinese Dictionary’ (Chinese Academy of Social Sciences 1978), modern Standard Chinese does not have *zhang4* for 長, *xia2* for 斜, or *sa4* for 撒. It is not clear to us how many multiple pronunciations existed for the Tang poets, and therefore we did not try to pick a particular pronunciation among the alternatives for each given line. Instead, we calculated the percentage of characters that have different tone types in order to estimate the upper bound of well-formed line types. The result is shown in (14).

(14)	Types	Tokens
Characters in corpus	1289	4704
Tone can vary (E and O)	65 (5%)	299 (6%)

In our corpus there are 1289 distinct characters, of which 65 (or 5%) can be both E tone and O tone (see Appendix 1 for the list). These characters appear a total of 299 times, which is 6% of the 4704-syllable corpus.

While these characters may affect our judgment on the tonal pattern of a line, their influence on the overall result is not expected to be very large. In particular, for characters that have alternate pronunciations, if we only consider the pronunciation provided by NJStar’s Convert tool, we might misinterpret some of the tones intended by the poet. On the other hand, since the tonal choices made by NJStar represent the most common pronunciation, it is possible that most of them do reflect the intended pronunciation. For those that do not, the error could tilt the result either way: it could either mistake a bad tone for a good tone, or it could mistake a good tone for a bad tone. Thus, the overall result should still be reasonably accurate.

We can also estimate the upper bound on the degree of faithfulness

of the corpus to the canonical tone patterns in (5) by choosing the tone or the rhyme that best fits the line. For example, if NJStar gives a syllable an E tone whereas the line type predicts it to be an O tone, the line would be judged ill-formed on the initial pass, but if the syllable can be E or O according to Tsai (2000), we can choose the O tone instead and the syllable now agrees with the line type. It is possible that this method may sometimes choose a pronunciation that was not intended by the poet, and hence the upper bound is higher than the actual level of best fits. However, since only 299 character tokens (6% of the corpus) have alternate tones, the upper bound would be no more than 6% greater than it would be if the poets' intended pronunciations were known.

2.3. Interpreting line types: without the 1-3-5 Rule

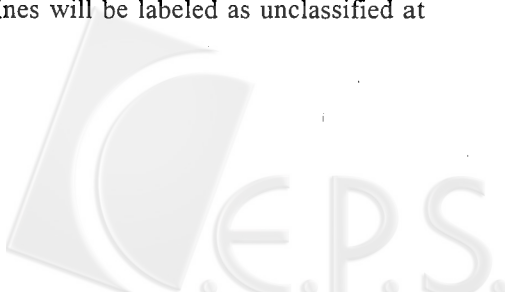
For this task, each line is assigned one of the line types in (1) for an exact match. As will be discussed later, only a third of the lines in the corpus exactly match the tone patterns in (1).

2.4. Interpreting line types: with the 1-3-5 Rule

Next the 1-3-5 Rule is taken into account; each line is assigned one of the line types in (1), while allowing flexibility in the first, third, and fifth syllables. Most lines now fit one of the four line types, but some still do not, such as line 7 in poem 188, shown in (15), where the translation is based on the interpretation of Qiu (1976).

(15) 已 忍 伶-傳 十 年 事 OOEEEO
already endure alone ten year thing
'I already endured it alone for ten years'

Under the 1-3-5 Rule, this line is XOXEXEO, which does not match any of the types in (1). If the sixth syllable were O, the line would match type (1c), which is XOXEXOO. Such lines will be labeled as unclassified at this stage.



2.5. Interpreting poem types: with the 1-3-5 Rule

Next, each poem is assigned a tone pattern using the quatrain types in (5). This assignment is made so as to most closely match the sequence of line types assigned in the previous step, which takes the 1-3-5- rule into consideration. As mentioned above, the two quatrains of an eight-line poem are either of the same type or else differ by first-line rhyming in the first quatrain. Consider poem 173 in (16) and the analysis of its best-fitting pattern in (17).

- (16) 朝 聞 游-子 唱 離 歌• • EEEOOEE
dawn hear traveler sing depart song
'At dawn I hear your parting song'
- 昨 夜 微 霜 初 度 河。 OOEEOE
last night mild frost first cross river
'Last night a first frost came over the river'
- 鴻-雁 不 堪 愁-裡 聽• • EOOEEO
goose not bear grief-in hear
'In grief, I can't bear to hear wild geese cry'
- 雲 山 況 是 客 中 過。 EEOOOEO
cloud mountain yet be you through pass
'Yet the cloud mountains are where you have to pass'
- 關 城 樹 色 催 寒 近• • EEOOEEO
Gate City tree color hurry cold near
'In Gate City, tree colors herald cold winter'
- 御-苑 砧 聲 向 • • 多。 OOEEOEO
palace laundry sound near evening quicken
'Near the palace, laundry sounds quicken in the evening'

莫 見 長-安 行-樂 處•• OOEEOO
 not see Chang-An happy place
 'Don't think Chang-An is just a happy place'

空 令 ••月 易 蹉-跎。 EOOOEE
 vain let time easy waste
 'It is easy to let times go by in vain'

(17)

Poem Types

Line	Poem 173	No L1 rhyming				With L1 rhyming				
			5a	5c	5e	5g	5b	5d	5f	5h
1	XEXOXEE	D	A*	C*	B*	D	D	B*	C*	A*
2	XOXEXOE	B	B	D*	A*	C*	B	D*	A*	C*
3	XOXEXOO	C	C	A*	D*	B*	C	A*	D*	B*
4	XEXOXEO	A	D*	B*	C*	A	D*	B*	C*	A
5	XEXOXEO	A	A	C*	B*	D*	A	C*	B*	D*
6	XOXEXOE	B	B	D*	A*	C*	B	D*	A*	C*
7	XOXEXOO	C	C	A*	D*	B*	C	A*	D*	B*
8	XOXOXEE	-	D*	B*	C*	A*	D*	B*	C*	A*
Poem type score:			5	0	0	2	6	0	0	1

The text of poem 173 and the tonal category of each syllable are shown in (16). By examining the second, fourth, sixth, and seventh syllables, a line type is assigned to each line, shown in the third column in (17). For instance, the second, fourth, sixth, and seventh syllables of line 1 have tones E, O, E, and E, respectively. This identifies the line with (1d), so it is marked as type D. The eighth line, however, has tones O, O, E, and E for its second, fourth, sixth, and seventh syllables and this does not match any line in (1). Thus line 8 is left without a line type.²

² A reviewer points out that the second character in line 8 of poem 173, namely 令 'let', is thought to have E tone by Qiu (1976), instead of O tone as we marked it. The character indeed has two pronunciations, one with E and one with O. According to *Xiandai Hanyu Cidian* (1978), the meaning 'let' should have the E tone, as we marked. However, it is possible that both tones were available to the poet and that E was a proper choice at that time. If so, (5b) would still be the type we assign to poem 173 and line 8 would be without error. This does not affect our discussion of poem 173, which is used as an example to demonstrate how we assign line types and poem types.



The rest of (17) is devoted to assigning to each quatrain in the poem a quatrain type from (5). The sequence of line types in column 3 is DBCA ABC-. This sequence does not admit a perfect match with any of the patterns in (5), but our purpose here is to find the pattern that most closely matches the poem. There are eight possibilities for the tone pattern of the first quatrain. As mentioned above, the second quatrain is assumed to resemble the first except without first-line rhyming. These eight possible patterns are represented in columns 4-11. Discrepancies between these patterns and the assigned line types in column 3 are marked with asterisks. Pattern (5a), for instance, expects the line types of each quatrain to be A, B, C, and D, in that order. Lines 2, 3, 5, 6, and 7 of the poem match this pattern, but lines 1, 4, and 8 do not. The last row of (17) indicates the number of lines in which a pattern agrees with the assigned line types. The poem type with the highest score is (5b), which disagrees with the actual poem in only the fourth and eighth lines. Poem type (5b) is thus assigned to poem 173.

All eight-line poems in our corpus yield a unique best-fit pattern by this method. However, a few four-line poems are ambiguous. One example is poem 299, shown in (18).

(18)	Line	Poem 299		5a	5b	5c	5d	5e	5f	5g	5h
	1	XEXOXEE	D	A*	D	C*	B*	B*	C*	D	A*
	2	XOXEXOO	C	B*	B*	D*	D*	A*	A*	C	C
	3	XOXEXEO	-	C*	C*	A*	A*	D*	D*	B*	B*
	4	XEXOXEE	D	D	D	B*	B*	C*	C*	A*	A*
	Poem type score:			1	2	0	0	0	0	2	1

Line 3 does not fit any of the four line types, and the remaining three lines do not fit any of the quatrain types in (5). Again columns 4-11 show the possible quatrain patterns, with asterisks indicating disagreement with the assigned line types in column 3. As indicated by the scores in the last row, the two best candidate quatrain types each agree with the assigned line types in only two lines. In such cases the poem is assigned the quatrain type which minimizes the number of mismatched syllable tones. This is shown in (19), where the mismatched tones are underlined.

(19)	Line	Poem 299		(5b)		(5g)
	1	XEXOXEE	D	XEXOXEE	D	XEXOXEE D
	2	XOXEXOO	C	XOXEX <u>OE</u>	B*	XOXEXOO C
	3	XOXEXEO	-	XOXEX <u>OO</u>	C*	XOXEX <u>OE</u> B*
	4	XEXOXEE	D	XEXOXEE	D	XEXOX <u>EO</u> A*
	Syllable Mismatches:			2		3

In this case, poem type (5b) is chosen because it differs from poem 299 in just two syllables, whereas poem type (5g) differs from poem 299 in three syllables.

A reviewer suggests, following Qiu (1976: 383), that there was a reason for the sixth syllable in line 3 of poem 299 to have a tonal violation. The intended pattern is XOXEXOO (line type C), as our analysis predicts. If the poet uses O for the fifth syllable, which is allowed by the 1-2-3 rule, the line would be XOXEOOO, which contains a trisyllabic OOO, which is thought to be disfavored. To avoid OOO, the poet need to use E for the sixth syllable, which gives XOXEOEO (the actual line is EOEEEOE). Such consideration, i.e. the choice of one tone influences the choice of another, is called • ‘counter-balance’ within a line. Given this, one might consider separating two cases of tonal violations: those that serve the purpose of counter-balance, and those that do not. It is worth noting though that OOO occurs quite frequently. In our corpus, it occurs in 138 lines, which is 20.5% of the 672 lines. In any case, our algorithm counts tonal violations regardless of the motivation. As we will report below, over 90% of the lines have no violations under the 1-3-5 rule. This means that over 90% of the lines do not use syllables 2-4-6 for counter-balance. If we treat lines with counter-balances as good lines, the total number of good lines would be a few percentages higher.

2.6. Comparison to findings from pentasyllabic regulated verse

The ideal tonal patterns for heptasyllabic regulated verse are similar to those for pentasyllabic regulated verse (Wang 1958). The pentasyllabic tone patterns, shown in (20), resemble the last five syllables of each heptasyllabic pattern in (1).

- (20) Line types of pentasyllabic regulated verse
- a. OOEEO (A)
 - b. EEOOE (B)
 - c. EEEEO (C)
 - d. OOOEE (D)

Because of the similarities in the tonal requirements, it is natural to compare results from heptasyllabic poetry to those from pentasyllabic poetry, such as that provided by Ripley (1980). The pentasyllabic analogue to the 1-3-5 Rule is the 1-3 Rule, which regards the tone of the first and third syllables of a line as variable. We will use the line type labels A, B, C and D to refer to both the seven-syllable line types in (1) and the corresponding five-syllable line types in (20).

3. RESULTS

We offer the results of best fit for lines, couplets, and poems, and compare our results with those of Ripley (1980). We also discuss the issue of Gu Ping, the 1-3-5 Rule, and word frequencies.

3.1. Lines

When lines are judged against the strict patterns in (1), without the 1-3-5- rule, the results are abysmal. The same is true for Ripley's (1980) study. A comparison is shown in (21) and (22), where "no match" means the line is not a perfect A, B, C, or D.

(21) Best fit of poems, without the 1-3-5 Rule

	Present corpus		Ripley (1980)	
Perfect	1	0.9%	2	0.5%
Imperfect	113	99.1%	462	99.5%
All	114	100%	464	100%

(22) Best fit of line types, without the 1-3-5 Rule

Line type	Present corpus		Ripley (1980)	
A	51	7.6%	480	12.9%
B	83	12.4%	744	20.0%
C	23	3.4%	249	6.7%
D	59	8.8%	385	10.4%
No match	456	67.9%	1854	49.9%
Total	672	100%	3712	100%

It is quite clear that the patterns in (1) are too strict for the general practice of Regulated Verse. In our corpus, only one poem (less than 1%), the four-line poem 277, match the strict tonal requirements. Similarly, in Ripley's corpus, only 2 poems (also less than 1%) match the strict tonal requirements.

With regard to line types, only a third of the lines in our corpus exhibit the tone patterns in (1); the presence of alternate pronunciations (at about 6%) would not have changed the result very much. In Ripley's corpus, 50% of the lines match the ideal forms. The lower rate of matches in our corpus is probably due to the length of the lines. Consider a specific example. Suppose most poets follow the 1-3-5 Rule. A heptasyllabic line type A is then XEXOXEO, where X can be either E or O. For each X, there is a 50% chance that it is E and a 50% chance that it is O. The chance for XEXOXEO to be exactly EEOEEEO is then 12.5% ($50\% \times 50\% \times 50\%$). In contrast, the chance for the corresponding pentasyllabic line type XOXEO to be exactly OOEEO is 50% ($50\% \times 50\%$). In other words, there is a greater chance for a shorter line to match a strict pattern by chance.

Next, we takes the 1-3-5 Rule into consideration, whereby a line needs only to match the patterns in (1) in syllables two, four, six and seven. Now around 90% of the lines in our corpus match the patterns (without regard to alternate pronunciations). Similarly, pentasyllabic lines in Ripley's corpus also show great improvement, with a match rate of 93%. This is shown in (23).



(23) Line type matches, with respect to 1-3-5 Rule

Line type	Our corpus		Ripley's corpus	
	Line count		Line count	
A	117	17.4%	815	22.0%
B	199	29.6%	937	25.2%
C	100	14.9%	713	19.2%
D	194	28.9%	994	26.8%
No match	62	9.2%	253	6.8%
Total	672	100%	3712	100%

Thus, both corpora show that the 1-3-5 Rule reflects a general practice in composing regulated verse by Tang poets.

It is worth noting that in both corpora there are more type B and type D lines (which end in E syllables) than type A and type C lines (which end in O syllables). This is especially so in our corpus, as shown in (24).

(24) Ratio between lines of type B and D and lines of type A and C

	B&D	A&C	Ratio
Ripley's corpus	52%	41%	5:4
Our corpus	59%	32%	2:1

The reason for the higher percentage of B and D lines is that (a) they end in an E syllable, which is favored for rhyming lines, and (b) there are more rhyming lines than non-rhyming lines because of first-line rhyming. Specifically, if only even lines rhyme (which is mostly the case), the ratios between rhyming and non-rhyming lines are as in (25).

(25) Ratio between rhyming- and non-rhyming lines (B&D : A&C)

	With first-line rhyming	No first-line rhyming
4-line poems	3:1	2:2 (1:1)
8-line poems	5:3	4:4 (1:1)

In Ripley's corpus, all poems have eight lines each (3712 lines for 464 poems). If some poems have first-line rhyming and some do not, the overall ratio should fall between 5:3 and 1:1. In our corpus, there are 60 four-line poems and 54 eight-line poems. The overall ratio, therefore, should fall between 3:1 (for four-line poems with first-line rhyming) and

1:1 (for eight-line or four-line poems without first-line rhyming). In both corpora, the actual ratio falls between the high bound and the low bound, as shown in (26).

(26) Ratio between rhyming- and non-rhyming lines (B&D : A&C)

	High bound	Actual	Low bound
Ripley's corpus	5:3	5:4	1:1
Our corpus	3:1	2:1	1:1

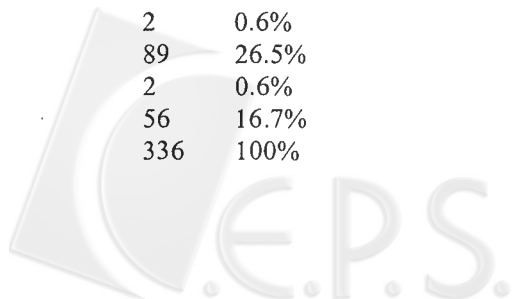
This shows how first-line rhyming skews the distribution of line types in favor of B and D.

3.2. Couplets

We refer to couplet types by the lines which make them up. For example, an AB couplet is an A line followed by a B line. We expect it to occur in the beginning of an ABCD quatrain or at the end of a CDAB quatrain. Couplet types AC, BD, CA and DB are found only with first-line rhyming in the beginning of the poems of (4). For example, a BD couplet is expected to occur at the beginning of a BDAB quatrain, where the first and fourth lines are of the same type, but no ideal poem form should end with a BD couplet. The summary in (28) gives the couplet counts in the present corpus that fall into each category, without regard to multiple pronunciations.

(27) Couplet type counts in our corpus

Type	Without 1-3-5 Rule		With 1-3-5 Rule	
AC	0	0%	2	0.6%
BD	1	0.3%	38	11.3%
CA	0	0%	1	0.3%
DB	7	2.1%	41	12.2%
AB	22	6.5%	105	31.3%
BA	0	0%	2	0.6%
CD	9	2.7%	89	26.5%
DC	0	0%	2	0.6%
No match	297	88.4%	56	16.7%
All	336	100%	336	100%



Without the 1-3-5 Rule, just about 10% of the 336 couplets are consistent with the possible quatrains. In contrast, with the 1-3-5 Rule more than 80% of the couplets are well-formed.

In Ripley's corpus, there are 1217 couplets that are made of that he calls "model" or "near model" lines (Ripley 1989: 132). In addition, there are four line types, with a total of 346 lines, that that he considers to be not as good but acceptable (Ripley 1989: 133, Figure 6); these are lines that follow the 1-3-5 Rule but have three O tones or three E tones in a row (OOOEO, EEEEOE, EEOOO, and OOEEO). If these lines are all accepted (as their counter-parts are in our corpus), they add up to 346 more couplets, which give the total of 1536. In (28) we compare couplet matches in both corpora.

(28) Couplet counts, with the 1-3-5 Rule

	Our corpus		Ripley's corpus	
Match	280	83.3%	1536	82.8%
No match	56	16.7%	320	17.2%
All	336	100%	1856	100%

Again, the two corpora show very similar results.

3.3. Four-line poems

As noted earlier, without the 1-3-5- rule, only one poem (poem number 277) matches a quatrain type perfectly. With the 1-3-5 Rule, 62% of four-line poems match a quatrain type in syllables two, four, six and seven. The result in (29) indicates the distribution of four-line poems among the eight possible quatrain types listed in (5). Here "poor lines" refers to lines which do not agree with their assigned best-fit types and FLR is the abbreviation for first line rhyming.

(29)	Poem type	Number of poor lines					Total
		No FLR	0	1	2	3	
	A	0	5	0	0	0	5
	B	0	0	0	0	0	0
	C	2	0	0	0	0	2
	D	0	0	0	0	0	0
	With FLR						
	A	14	7	4	0	0	25
	B	0	0	0	0	0	0
	C	21	6	1	0	0	28
	D	0	0	0	0	0	0
	Total	37	18	5	0	0	60
	%	62	30	8	0	0	100

As can be seen, most non-perfect poems deviate from the model pattern by one line. These numbers represent the upper bound in considering alternate pronunciations and may be up to 6% too high.

3.4. Eight-line poems

Eight-line poems have a greater chance of a line disagreeing with that in the model poem type. As shown in (30), about 40% of eight-line poems match the quatrain forms with respect to the 1-3-5 Rule.

(30)	Poem type	Number of poor lines								Total	
		No FLR	0	1	2	3	4	5	6		7
	A	4	1	1	1	0	0	0	0	0	7
	B	0	0	0	0	0	0	0	0	0	0
	C	0	4	2	0	0	0	0	0	0	6
	D	0	0	0	0	0	0	0	0	0	0



With FLR										
A	7	4	9	3	0	0	0	0	0	23
B	0	0	0	0	0	0	0	0	0	0
C	9	4	4	1	0	0	0	0	0	18
D	0	0	0	0	0	0	0	0	0	0
Total	20	13	16	5	0	0	0	0	0	54
%	37	24	30	10	0	0	0	0	0	100

No non-perfect poem deviates from the model pattern by more than three lines. Again, these numbers represent the upper bound in considering alternate pronunciations and may be up to 6% too high.

3.5. Faithfulness of lines to best fit

In section 3.1 we discussed line type matches when each line was considered by itself. We can also consider line type matches after each poem was assigned a poem type. In the latter case, the type of a line is determined by the expected line sequencing of the poem. The two methods may yield different results, with the second method having slightly fewer good lines. The result of the second method is shown in (31). *These numbers reflect the upper bound, because alternate pronunciations have been considered.*

(31) Line type matches based on poem types (with 1-3-5 Rule)

Line type	Number of mismatched syllables					All
	0	1	2	3	4	
A	110	7	2	1	0	120
B	190	17	2	5	0	214
C	95	21	3	3	0	122
D	189	20	0	7	0	216
Total	584	65	7	16	0	672
	87%	10%	1%	2%	0.0%	100.0%

87% of the lines agree with the poem types we have assigned. This result is quite close to that in section 3.1 (about 91%), where line types were

judged without reference to poem types. Also, it can be seen that most imperfect lines have just one mismatched syllable under the 1-3-5 Rule.

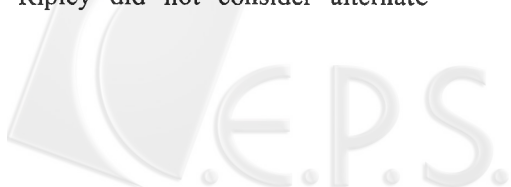
3.6. “Gu Ping”

Previous descriptions of pentasyllabic and heptasyllabic Chinese regulated verse mention a prohibited tonal pattern called “Gu Ping”, or ‘lone Even’ (Wang 1958). Literally, Gu Ping refers to a line that contains only one E tone, not counting the last syllable. In heptasyllabic verse, Gu Ping can arise when the 1-3-5 Rule is applied to XOXEXO to create OOEOOE. Gu Ping can also arise through a violation of an even-numbered syllable, such as EOOOOE, which comes from XOXEXO with a violation of the fourth syllable. In our corpus no Gu Ping line was found. The rarity of Gu Ping lines supports the long-held idea that Gu Ping lines are consciously avoided in regulated verse.

In his pentasyllabic corpus, Ripley (1980: 133-135) found a total of 62 Gu Ping lines. They are shown in (32), along with the corresponding strict patterns. Underlined syllables indicate changes allowed by the 1-3-5 Rule.

(32)	Line	Strict	Count
	<u>O</u> E O O	EEOOE	2
	OO <u>O</u> E O	OOEEO	47
	<u>O</u> E O OO	EEEOO	13
	Total		62

The number of Gu Ping lines in Ripley’s corpus was surprisingly high, even though his corpus has over five times as many lines as ours. Ripley argues that not all Gu Ping lines are equally bad. Instead, he suggests that only OEEOE should be banned strictly, whereas OOEO and OEEOO are acceptable (under certain circumstances). In any case, it remains a question why Gu Ping lines do not show up in our corpus at all. There are several possibilities. First, our corpus has longer lines, and longer lines have less chance of having just one E tone. Second, our corpus has better poems (probably the best 300 from all Tang poems judged by the compiler). Third, Ripley did not consider alternate pronunciations.



3.7. The 1-3-5 Rule

We discussed earlier that the literal interpretation of the 1-3-5- rule is that the first, third, and fifth syllables in a heptasyllabic line can take any tone. However, as is often noted, the literal interpretation is a simplification (Wang 1958, Qiu 1976, Ripley 1980). This can be seen in the data (33), which shows, for each line type (A-D), the rate of tone change at each syllable position (S1-S7) from the strict pattern in (1).

(33) Rate of tone change at each syllable position (%)

Type	S1	S2	S3	S4	S5	S6	S7	Lines
A	35.8	1.7	29.2	2.5	7.5	1.7	0.8	120
B	43.0	3.7	2.8	3.3	15.4	2.8	0.5	214
C	36.9	4.9	48.4	3.3	16.4	13.9	0.0	122
D	45.4	3.7	58.8	4.2	2.8	3.7	0.5	216
All	41.4	3.6	33.8	3.4	10.1	4.9	0.4	672

If S1, S3, and S5 are equally free, one would expect the rates of change to be similar in those positions. However, there are fewer changes for the fifth syllable (10.1% overall) than for the first and third. In addition, different line types also differ in their patterns of change. For example, the third syllable of type B has far fewer changes than the third syllable in other line types. Evidently, a more accurate characterization of the 1-3-5 is needed.

One such characterization was given by the Qing scholar Wang Shizhen (王士禛, 1634-1711, cited in Qiu 1976: 256-258). According to him, the first syllable is free, but the third and fifth are conditional. In particular, the third and fifth syllables can change their tones as long as it does not destroy an E-tone pair. For example, OO can change to EO, but EE should not change to OE (unless the next tone is E). This is shown in (34), where the free positions are underlined. The predicted free positions are compared with the actual results, repeated in (35).

- (34) Type Pattern (free positions underlined)
 A EEQOEE
 B QOEEQOE
 C QOEEEOO
 D EEQOOEE

(35) Rate of tone change at each syllable position (%)

Type	S1	S2	S3	S4	S5	S6	S7	Lines
A	<u>35.8</u>	1.7	<u>29.2</u>	2.5	7.5	1.7	0.8	120
B	<u>43.0</u>	3.7	2.8	3.3	<u>15.4</u>	2.8	0.5	214
C	<u>36.9</u>	4.9	<u>48.4</u>	3.3	16.4	13.9	0.0	122
D	<u>45.4</u>	3.7	<u>58.8</u>	4.2	2.8	3.7	0.5	216
All	41.4	3.6	33.8	3.4	10.1	4.9	0.4	672

The prediction works fairly well in that the free positions mostly have higher rates of tone change than other positions. On the other hand, some differences among the free positions remain to be explained. For example, in line B the fifth position has much fewer changes than the first position; indeed, the fifth position in line B has fewer changes than the fifth position of line C, although the latter is supposed to be non-free. Similarly, in lines C and D the first position has fewer violations than the third position, whereas in line A it is the reverse.

Let us now consider the case in Ripley's corpus. The expected positions that allow change are shown with underline in (36), where each line type corresponds to the last five syllables of a heptasyllabic line. The actual results are shown in (37). The total number for each line type is based on those that Ripley discussed (he only discussed 3459 of the 3712 lines in his corpus), and the percentages are based on the numbers in his Figure 11 (Ripley 1980: 140).

- (36) Type Pattern (free positions underlined)
 A QOEEO
 B EEQOE
 C EEEOO
 D QOOEE



(37) Rate of tone change at each syllable position (%)

Type	S1	S2	S3	S4	S5	Lines
A	<u>37.1</u>	0.7	12.6	4.5	0.0	815
B	3.6	0.3	<u>20.5</u>	0.1	0.0	937
C	<u>53.0</u>	0.4	45.7	28.1	0.0	713
D	<u>59.3</u>	0.0	4.1	0.7	0.0	994
All	37.7	0.3	19.1	7.1	0.0	3459

The predicted free positions (underlined) still have high rates of changes, although as in our corpus the rates are not always even. However, the rates of two positions are rather unexpected: the third and fourth positions of line C are not free, yet they have very high rates of change.

3.8. Word frequencies

We thought it would be useful to see what words are used more frequently than others in poetry. Therefore, we calculated the frequencies of all the characters in our corpus. The results are given in Appendix 2. The discussion here is unrelated to prosody or phonology, as the reviewers rightly pointed out. Nevertheless, we thought it is linguistically relevant and might be of interest for some readers.

The most frequent word is • •‘not’, which occurs 40 times. The top ranking of ‘not’ may seem a surprise, but considering the fact that there are 672 lines, 40 represents just 6% of them, which does not seem to be a high percentage for lines with a negation.

The second most frequent word is • •‘person/man’, which occurs 39 times. This word can not only refer to the generic ‘person/man’, but also function like the English suffix ‘-er’ in a compound, such as • • •‘boat-person (boater)’ and • • •‘travel-person (traveler)’. Therefore, its high frequency is quite understandable. Next consider some heavenly objects, shown in (38).

(38) Frequencies of some heavenly objects

Word	• •	• •	• •	• •	• •
Frequency	35	31	27	21	6
Gloss	‘moon’	‘sun’	‘cloud’	‘rain’	‘star’

It may seem natural that clouds and rain often stir emotions in the poet. It is also interesting to see that stars do so much less. However, one might not expect the moon and the sun to have similar effects. A careful look shows that • • can mean 'moon' or 'month', and • • can mean 'sun' or 'day'. The frequencies of the meanings are given in (39).

(39)	• •		• •	
	30	'moon'	12	'sun'
	5	'month'	19	'day'

We now see that 'moon' is nearly three times as likely to be used in a poem as 'sun'. The psychological effect of the moon is probably not unique to Chinese culture, since we also find words like 'moonstruck' and 'lunatic' in English. Similarly, consider the seasons, shown in (40).

(40)	Word	• •	• •	• •	冬
	Frequency	33	18	1	0
	Gloss	'spring'	'autumn'	'summer'	'winter'

Poets seem to be sensitive to 'spring' a lot more than to 'autumn'. In addition, they rarely refer to 'summer' or 'winter'. Finally, consider times of the day, shown in (41).

(41)	Word	• •	晚/(黃)昏	晨	午
	Frequency	32	6+2	1	0
	Gloss	'night'	'evening/dusk'	'morning'	'noon'

Poets seem to be most sensitive to 'night' by far. The words for 'evening' and 'dusk' are also mentioned but not a lot. In contrast, 'morning' and 'noon' are rarely mentioned.

The above examination of word frequencies is rather superficial. Nevertheless, we found it interesting and hope it would be to the general audience.



4. CONCLUDING REMARKS

We hope to have shown that a corpus study offers explicit ways to test theories about the prosodic requirements in poetry. In particular, we have examined the tonal requirements in regulated heptasyllabic verse (114 poems, 672 lines). Our results confirm the traditional view that some version of the “1-3-5 Rule” must be assumed, according to which the first, third, and fifth syllables of a line are free to use any tone. Our results also confirm additional limitations on the 1-3-5 Rule, as previously suggested by Wang Shizhen (王士禛, 1634-1711, cited in Qiu 1976: 256-258), in that the first syllable is indeed tonally free but the third and fifth are subject to the condition that an even-tone sequence EE should not be changed into OE. As is expected, tonal violations do occur, even in non-free positions (the second, fourth, and sixth syllables), but only at low rates (below 5% for each position). Our results are generally compatible with those of pentasyllabic lines (Ripley 1980), although ours seem to be more consistent in two respects. First, Ripley’s corpus has a number of Gu Ping lines (section 3.6) whereas ours has none. Second, Ripley’s corpus shows higher rates of tonal violations in non-free positions (section 3.7). We have also provided information on word frequencies, which enables one to examine which words are favored by poets.

REFERENCES

- Mair, Victor H. & Tsu-Lin Mei. 1991. The Sanskrit origins of recent style prosody. *Harvard Journal of Asiatic Studies* 51, No. 2. (Dec.): 375-470.
- Qiu, Xieyou. 1976. *Xin yi Tang shi san bai shou* [New translation of 300 Tang poems]. Taipei: Sanmin Shuju.
- Ripley, Stephen. 1979. *A statistical study of tone patterns in Tang regulated verse*. Toronto: University of Toronto dissertation.
- Ripley, Stephen. 1980. Some findings on tone patterns in Tang regulated verse. *Journal of Chinese Linguistics* 8.1: 126-48.
- Tsai, Chih-Hao. 2000. Mandarin Syllable Frequency Counts for Chinese Characters. <http://www.geocities.com/hao520/>.
- Wang, Li. 1958. *Hanyu Shilü Xue* [Chinese Versification]. Shanghai: Xin Zhishi Chubanshe.

Duanmu, San; Stiennon, Nathan

Xiandai Hanyu Cidian [Modern Chinese Dictionary]. 1978. Compiled by the Institute of Linguistics, Chinese Academy of Social Sciences. Beijing: Shangwu Yinshuguan.
Zhou, Fagao. 1948. Shuo ping ze [On even and oblique tones]. *Bulletin of the Institute of History and Philology, Academia Sinica* 13. 153-162. Reprinted in Fagao Zhou 1968 *Zhongguo yuyanxue lunwen ji* [Papers on Chinese linguistics], 105-114. Hong Kong: Chongji Publishers.

*San Duanmu and Nathan Stiennon
University of Michigan*

*Correspondence address:
San Duanmu
Department of Linguistics
University of Michigan
Ann Arbor, MI 48109-1285
duanmu@umich.edu*

APPENDIX 1: CHARACTERS WITH ALTERNATE TONAL CATEGORIES

The list is based on alternate pronunciations given in Tsai (2000). There are 65 characters in all. They occur for a total of 299 times.

• ;
• • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ;
• • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ; • • • • • ;
• • • • •

APPENDIX 2: WORD FREQUENCIES

Each paragraph starts with a frequency number and, in parentheses, the number of words at this frequency.

- 40 (1) • •
- 39 (1) • •
- 38 (1) • •
- 35 (1) • •



