Knowledge-rich, computer-assisted composition of Chinese couplets

John Lee, Ying Cheuk Hui and Yin Hei Kong
Halliday Centre for Intelligent Applications of Language Studies, Department of Linguistics and Translation, City University of Hong Kong, Hong Kong

Abstract

Recent research effort in poem composition has focused on the use of automatic language generation to produce a polished poem. A less explored question is how effectively a computer can serve as an interactive assistant to a poet. For this purpose, we built a web application that combines rich linguistic knowledge from classical Chinese philology with statistical natural language processing techniques. The application assists users in composing a 'couplet'—a pair of lines in a traditional Chinese poem—by making suggestions for the next and corresponding characters. A couplet must meet a complicated set of requirements on phonology, syntax, and parallelism, which are challenging for an amateur poet to master. The application checks conformance to these requirements and makes suggestions for characters based on lexical, syntactic, and semantic properties. A distinguishing feature of the application is its extensive use of linguistic knowledge, enabling it to inform users of specific phonological principles in detail, and to explicitly model semantic parallelism, an essential characteristic of Chinese poetry. We evaluate the quality of poems composed solely with characters suggested by the application, and the coverage of its character suggestions.

Correspondence:
John Lee, Halliday Centre for Intelligent Applications of Language Studies, Department of Linguistics and Translation, City University of Hong Kong, Hong Kong.
E-mail:
jsylee@cityu.edu.hk

1 Introduction

Among the various literary genres, poetry enjoys perhaps the most elevated status in the classical Chinese tradition. The Tang dynasty (618–907 AD) is generally considered the golden age of poetry; to this day, poem composition is a popular activity in many aspects of everyday life, both for scholars and laypeople. A poem can serve as a delicate vehicle for expressing political viewpoints, a sophisticated form of communication and entertainment, decoration for temples and other buildings, festive messages in weddings and celebrations, as well as words of remembrance in funerals. Famous poems written by ancient poets are etched onto walls in many sites in China, which have become tourist attractions.

Chinese poems are made up of a sequence of couplets—pairs of adjacent lines, each with an identical number of characters, usually five or seven. Table 1 shows the couplet that forms the first two lines of a well-known poem by Wang Zhihuan, entitled 'Climbing Crane Tower', from the 8th century CE. Following the traditional format, a line is read from top to bottom, and lines are arranged from right to left. In this couplet, both lines consist of five characters.

While it is possible to compose couplets in 'free form', most aspire to the more formal and prestigious 'regulated verse' tradition. This tradition, however, imposes a formidable set of requirements on phonology, syntax, and parallelism. The two lines are expected to be syntactically similar and

doi:10.1093/llc/fqu052 Advance Access published on 10 October 2014

Table 1 A couplet whose lines have five characters each

Second line	First line
黄 *huang* 'yellow'	白 *bai* 'white'
河 *he* 'river'	日 *ri* 'sun'
入 *ru* 'enter'	依 *yi* 'rest'
海 *hai* 'ocean'	山 *shan* 'mountain'
流 *liu* 'flow'	盡 *jin* 'extinguish'
Yellow River enters sea—and flows on.	White sun rests on mountains—and is gone.

The English translations are taken from Cai (2008).

Table 2 A Chinese couplet consists of two lines with the same number of characters

Second line	First line
$w_{2,1}$	$w_{1,1}$
$w_{2,2}$	$w_{1,2}$
$w_{2,3}$	$w_{1,3}$
$w_{2,4}$	$w_{1,4}$
$w_{2,5}$	$w_{1,5}$

The schema above represents a pentasyllabic (five-character) couplet, an example of which is provided in Table 1. Each $w_{i,j}$ is a character.

semantically parallel; furthermore, the characters are to follow specific phonological patterns, within each line and across lines. Amateur poets often struggle to remember and apply these patterns while composing creative poems. Our web application provides an assistive environment for Chinese couplet composition. It checks conformance to phonological requirements, and dynamically suggests characters based on lexical, syntactic, and semantic properties.

A distinguishing feature of our application is its extensive use of linguistic knowledge from classical Chinese philology. In particular, it informs users of specific phonological principles in detail, and explicitly models semantic parallelism, an essential characteristic of Chinese poetry.

The rest of the article is organized as follows. In Section 2, we summarize previous research in poem composition. Section 3 outlines the functions of the application, and Sections 4 and 5 describe how it makes word suggestions and checks phonological rules. After discussing our data set in Section 6, we evaluate our application in Section 7.

Throughout, the notation illustrated in Table 2 will be used. Each character is labeled as $w_{i,j}$, where i refers to the line number (either 1 or 2), and j refers to the position of the character in line i. In a pentasyllabic couplet, such as the one shown in Table 1, $w_{1,1} \ldots w_{1,5}$ form the 'first line', and $w_{2,1} \ldots w_{2,5}$ form the 'second line'. Another common form, the heptasyllabic (seven-character) couplet, has $w_{1,1} \ldots w_{1,7}$ as its first line and $w_{2,1} \ldots w_{2,7}$ as its second. As we shall see, parallelism is an important feature in couplets; we will be considering pairs of characters on the same position of the lines, i.e. $w_{1,k}$ and $w_{2,k}$, where $k = 1 \ldots 5$ for pentasyllabic couplets and $k = 1 \ldots 7$ for heptasyllabic couplets. These pairs will be referred to as 'corresponding characters'.

2 Previous Work

We first sketch the development of poem composition in ancient China (Section 2.1), which informed the design of our application. We then review related research on the related task of the automatic poem generation (Section 2.2), and finally compare our application with other semiautomatic approaches (Section 2.3).

2.1 Background

To provide the background for the design of our system, we summarize the literary history of phonological patterns and semantic parallelism. The interested reader can find more details in Furuta (1994) and Wang (2003).

2.1.1 *Phonological patterns*

Although informal conventions were likely already in place even before, phonological rules were spelled out explicitly for the first time from the third to fifth centuries CE. The four kinds of tones—'level', 'rising', 'falling', and 'entering'—were named; the ways in which word tones can be combined in a couplet were enumerated, most famous by Shen Yue in his 'Eight Prohibitions' *ba bing* 八病 (Tang, 2002). One of the rules, for instance,

Table 3 The twenty semantic categories for nouns, in a taxonomy designed for characterizing semantic parallelism in traditional Chinese poems (Wang, 2003); English translations of the category names are taken from Harvey and Kao (1979)

Celestial	Geographic	Architectural	Instruments	Clothing
Seasonal	Literary references	Calendar coordinates	Products of civilization	Food
Locations	Numbers	Colors	Flora	Fauna
Place names	Personal names	Human relationships	Body parts	Human emotions

When two corresponding characters belong to the same categories, they are considered 'exact match'. When they belong to two different categories that are separated by dotted lines (e.g. 'Celestial' and 'Seasonal'), they are considered 'related match'.

stipulates that the corresponding characters cannot be homophones. Our application enforces this and other phonological requirements (Section 5).

2.1.2 *Semantic parallelism*

A few centuries later, in the Tang dynasty, the so-called 'regulated verse' was formalized as a poetic genre. The scholar Shangguan Yi formulated two sets of rules, called 'six harmonies' and 'eight harmonies' (*liu dui* 六對 and *ba dui* 八對), which characterize ten phonological rules and four others on how characters should correspond with respect to part of speech (POS) and semantics (Tan, 2003). The notion of semantic parallelism was further expounded in the book *Weiwendishige* 魏文帝詩格, which classified nouns into eight semantic categories. Characters occupying the same position on the two lines, $w_{1,i}$ and $w_{2,i}$, were expected to belong to the same category. Following this tradition, a more fine-grained taxonomy, with thirty-seven categories, was fleshed out in the book *Shiyunhebi* 詩韻合璧 (Wang, 2003).

The contemporary scholar Wang Li refined this latter taxonomy into twenty categories, as listed in Table 3. Based on these categories, three degrees of parallelism are defined. The first is 'exact match', where two corresponding characters belong to the same category. For example, *bai* 'white' and *huang* 'yellow' in the couplet in Table 1 are both 'Colors'.

The second is 'related match', where two characters belong to two related categories. As shown in Table 3, of these twenty categories, thirteen pairs of categories are deemed 'related', including, for example, 'Flora' and 'Fauna' and 'Celestial' and 'Geographic'. The couplet in Table 1 also contains a pair of corresponding characters in 'related match': *ri* 'sun' belongs to 'Celestial'; *he* 'river' belongs to 'Geographic', which is related to 'Celestial'. It is also acceptable, if less than ideal, for corresponding characters to belong to a pair of related categories.

The third is 'loose match', where both characters are nouns but belong to neither the same nor related categories. Our application models semantic parallelism with this taxonomy (Section 4).

2.2 Automatic poem generation

Although some poem generation systems rely on grammatical and semantic templates (Oliveira, 2012), most current approaches use a combination of handcrafted rules and statistical language models. The former typically serves as a filter to enforce hard constraints on rhyming or tones, while the latter seeks to optimize fluency and semantic coherence.

The system developed by Greene *et al.* (2010) incorporates a meter model, a word selection model, and a trigram language model. To enhance coherence, the language model is modified on the fly given the words on the previous line.

Two haiku systems adopt different strategies. Wong and Chun (2008) store a set of keyword seeds and a database of sentence fragments. The system selects semantically related fragments based on cosine similarity. Netzer *et al.* (2009) mine relevant expressions according to word association norms. After generating a POS sequence, n-grams from a large English corpus matching the relevant expressions and POS sequences are selected.

Other systems assume not only a seed keyword but a sentence or a semantic representation as the starting point. Both Gervás (2001) and Diaz-Agudo *et al.* (2002) take a prose message in Spanish as input, and then transform it into poetry. As an example of a knowledge-rich system, McGonagall takes into account both semantic information and patterns of meter (Manurung *et al.*, 2012). It proved difficult in the evaluation, however, to produce text which was both semantically coherent and of high quality metrically.

As a pioneering work on Chinese couplets, the couplet generation tool developed at Microsoft Research Asia accepts as input the first line of a couplet, then leverages phrase-based statistical machine translation techniques to generate the second line (Jiang and Zhou, 2008). About one-fifth of the top-ranked second lines were judged 'acceptable' by humans. This tool has been extended for automatically composing all lines of the regulated verse (He *et al.*, 2012). Another approach is to cast the task as summarization from a large pool of poems based on key terms specified by the user (Yan *et al.*, 2013).

2.3 Computer-assisted poem composition

Research on automatic poem generation has yielded systems that can produce excellent poems, at times approaching the passing of the Turing Test. Yet, it is rare for the output of any automatic couplet generator, however high quality, to satisfy the poet without any further modification—it is unlikely that someone else, let alone a machine, can compose a poem for you that is exactly what you desire. More commonly, poem composition is an iterative process, with the poet working on many drafts before arriving at the final version.

Rather than aiming to generate the poem expected at the end point of this process, computer-assisted poem composition systems support the poet during the process by providing suggestions and feedback. Some systems provide the poet with an initial draft. The Daoxiang poem generator[1] is a typical example. The user is to provide a title and select from a predetermined set of rhyming templates; the system then fills the template with keywords related to the title. The system by Lo (2011) works similarly, but offers not only templates but also a tone dictionary for Chinese characters. In contrast to both systems, ours does not limit the user to any template. The user can freely input desired content, and receive alerts and explanations in case of noncompliance with phonological principles.

More significantly, similar to the concept of the 'reviser' component proposed in Gervás (2013), our system offers dynamic assistance to the poet to improve a working draft. Given the current version of a poem, whether complete or incomplete, our system suggests the next and corresponding characters to be used. Furthermore, it explicitly models semantic parallelism for corresponding characters within a couplet, an essential feature of Chinese poetry, using the taxonomy devised in Wang (2003).

3 System Description

Our application supports the composition of couplets for the two most common forms of regulated verse, i.e. the pentasyllabic couplet, an example of which is shown in Table 1, and the heptasyllabic couplet, where each line has seven characters instead of five. Figure 1 shows the interface for composing a pentasyllabic couplet. There are two columns, each representing a line, with five slots to be filled with characters. At any time, the user can enter a character into any slot. The system then performs two tasks: it detects any violation of phonological rules among existing characters, and gives an explanation; and it provides character suggestions for all other empty slots. We now describe these two tasks.

3.1 Character suggestions

When a character slot is filled, the system offers three candidates for the corresponding and following slots. In Figure 1, the first character in the first line ($w_{1,1}$) is occupied by *ye* 夜 'night'; three suggestions are shown for its corresponding character ($w_{2,1}$). The top-ranked suggestion is *zhao* 朝 'morning', which is semantically related to 'night'.

Fig. 1 The interface for couplet composition. The user has entered the character *ye* 夜 'night' for $w_{1,1}$; as the first choice for $w_{2,1}$, the system proposes *zhao* 朝 'morning', which also belongs to the semantic category 'Seasonal' (see Table 3). In the second position, the two characters entered by the user (i.e. $w_{1,2}$, and $w_{2,2}$) both have the gliding tone, violating the 'alternating tones' rule (see Section 5). These trigger a warning, as shown by their red border, as well as a dialog window that explains the issue. The suggested characters for $w_{1,4}$ are *shan* 山 'hill', *men* 門 'door', and *tai* 苔 'moss'; the first two serve as direct objects to the character at $w_{1,2}$, *zhu* 住 'live', and are therefore more highly ranked

The third character in the first line ($w_{1,3}$) receives suggestions not only for its corresponding character ($w_{2,3}$) but also for the two following it ($w_{1,4}$, $w_{1,5}$). At each empty slot, the user may view up to ten suggestions by clicking on the green 'plus' button.

3.2 Tone information

Word tone is a critical factor in Chinese poetry. As mentioned in Section 2.1, there are four kinds of tones—'level', 'rising', 'falling', and 'entering'. For the purpose of couplet composition, the latter three are often collectively called 'gliding'. Our interface uses color to visually represent tones: blue for the level tone and red for the gliding tone. If a character has the wrong tone, the application flags it by turning the border of the slot to red. The user may then click on the warning dialog on the right side to learn more about the problem. In Figure 1, the user entered gliding-tone characters for both $w_{1,2}$ and $w_{2,2}$, violating one of the tonal rules in Section 5.2. The borders of both slots are thus highlighted.

4 Features

Our application assists the user as he/she edits a poem. Whenever a new character is input, the system dynamically updates its suggestions for both the next character (Section 4.1) and the corresponding character (Section 4.2). This section describes the linguistic features used in performing the search. These candidates will then be filtered by phonological rules (Section 5).

4.1 Suggesting the next character

Given a newly entered character $w_{i,j}$, the application suggests characters for $w_{i,j+1}$ by considering the following features:

4.1.1 *Character n-gram*

Following most previous poem generation systems, we use an n-gram language model. A character is scored according to the trigram or bigram that it would yield. The language model is trained on the *Complete Tang Poems*, an anthology of all surviving poems from the Tang Dynasty. Repetition of the same character is not allowed, to respect the principle known as *bichongzi* 避重字 (Wang, 2003). In Figure 1, the top-ranked suggestion for $w_{1,4}$ is *shan* 山 'hill', since *qing* 青 'green' is the most popular color for describing hills in our corpus.

4.1.2 *Dependency bigram*

Nonadjacent characters can be syntactically linked; these long-distance relations cannot be detected by character n-grams. Intuitively, if a character has a dependency relation with any of the preceding characters in the treebank, it should be favored. In Figure 1, the three suggested characters for $w_{1,4}$, namely, *shan* 山 'hill', *men* 門 'door', and *tai* 苔 'moss', would all yield frequently occurring bigrams with the preceding character, *qing* 青 'green'. 'Hill' and 'door' are ranked higher than 'moss', because they have been observed in our corpus as direct objects of the verb *zhu* 住 'live', in the slot $w_{1,2}$.

'Moss' has no such grammatical relation with 'live', reflecting the fact that one does not usually live on 'moss'.

4.2 Suggesting the corresponding character

Given the character $w_{1,i}$, the algorithm also needs to suggest candidates for its corresponding character $w_{2,i}$, i.e. the one that occupies the same position on the second line, and vice versa. It considers the following features:

4.2.1 *Horizontal character bigram*

This is similar to the character n-gram feature in the previous section, except that the bigram is 'horizontal', i.e. consisting of $w_{1,i}$ and $w_{2,i}$, for any position i in the line. Furthermore, they must not be the same character or homophones.

4.2.2 *POS parallelism*

The POS tags of $w_{1,i}$ and $w_{2,i}$ must be either identical, according to the tagset of the Penn Chinese Treebank (Xue *et al.*, 2005), or belong to equivalent categories as proposed in Lee and Kong (2012).

4.2.3 *Semantic parallelism*

Corresponding characters ($w_{1,i}$ and $w_{2,i}$) are expected to be related semantically. A candidate character for $w_{2,i}$ is preferred if it belongs to the same semantic category ('Exact match') as $w_{1,i}$ according to the taxonomy in Table 3, or at least to a related category ('Related match'). In Figure 1, this feature helps propel *zhao* 朝 'morning' to the top of the suggestions as the corresponding character *ye* 夜 'night'; both belong to the semantic category 'Seasonal'.

5 Phonological Rules

A ranked list of character candidates, either for the next or the corresponding character, is returned by our application on the basis of the features described above. These candidates are then filtered by the set of phonological rules below.

As mentioned in Section 3.2, a Chinese character may have one of the two main tones, 'level' or 'gliding'. According to the metric of the regulated verse, the tones of the characters need to adhere to various principles. We now summarize them using the pentasyllabic couplet as example, but they can be applied on the heptasyllabic counterpart in a straightforward manner.

5.1 Within a line

5.1.1 *Tone of the last three characters*

The last three characters in the first line ($w_{1,3} \ldots w_{1,5}$) are seen as especially important, due to their effect during a recital (Zhu, 1999). Normally, these three characters cannot be all gliding (三仄); if negative emotions are intended, more words with gliding tone are used toward the end of a line in order to convey a depressed atmosphere (Shi, 2001).

The tones of the last three characters in the second line are also restricted. It is common for a free-verse poem to have level tone for characters at the end of a line, giving it a plain and moderate style (Ho, 2008). To distinguish the regulated verse from free verse, the last three characters ($w_{2,3} \ldots w_{2,5}$) cannot all have level tone (三平).

5.1.2 *Level tones in second line* 孤平

The last character ($w_{2,5}$) must have level tone (孤平). At least two of the preceding characters (from $w_{2,1}$ to $w_{2,4}$) must also have level tone, otherwise the atmosphere formed would be considered too weak (Ho, 2011). But if both $w_{2,4}$ and $w_{2,5}$ are level, then there is no further requirement on the preceding characters.

5.1.3 *Alternating tones*

A line in a regulated verse poetry can be seen as a concatenation of units of words. A unit consists of a pair of words, i.e. $<w_{1,1}, w_{1,2}>$ and $<w_{1,3}, w_{1,4}>$. Normally, the two characters in a pair should have the same tone. Furthermore, adjacent pairs should differ in tone. In practice, the two characters in a pair may not have the same tone; in these cases, the tone of the second character (i.e. $w_{i,2}$ and $w_{i,4}$) is taken to represent that of the pair, as the second one is considered to be more important (Qi, 2002). Hence, we require the tone of $w_{1,2}$

to differ from that of $w_{1,4}$, and likewise for $w_{2,2}$ and $w_{2,4}$.

5.2 Across lines

5.2.1 *Alternating tones*

Tones should vary not only within a line but also between the two lines. More precisely, the tone of each character should be different from that of the character located at the same position in the other line. For the same rationale as above, the second character of a unit is seen as tonally more critical. Hence, we require that the tone of $w_{1,2}$ be different from that of its counterpart, $w_{2,2}$, and similarly for $w_{1,4}$ and $w_{2,4}$.

5.2.2 *Rhyming*

The last characters of even lines should rhyme, and their tone should be level. In the context of a couplet, then, the last character of the second line, $w_{2,5}$, is expected to have a level tone. In order to vary the melody of the poem, the last characters of the odd lines usually have a gliding tone (Qi, 2002). Hence, we require the tone of $w_{1,5}$ to be gliding.

6 Data

In this section, we describe the data that were used to derive the features listed in Section 4, and how they were used to train the system.

To suggest the next character, character n-gram counts were obtained from all lines in the *Complete Tang Poems* (Peng, 1960), an anthology of surviving poems from the Tang Dynasty. In total, there are about 39,000 poems, containing 420 K lines and about 2.8 million characters.

To suggest the corresponding character, horizontal character bigram counts were obtained from pairs of corresponding characters from parallel couplets. Unfortunately, couplets in the *Complete Tang Poems* may or may not be parallel, depending on the genre of the poem and the position of the couplet within the poem. The couplets in quatrains and ancient-style poems are not necessarily parallel. In regulated verse, the position of a couplet indicates whether it is expected to be parallel. In the traditional understanding, of the four couplets in an eight-line regulated verse, the second and third couplets should be parallel, whereas the first and fourth couplets should be nonparallel (Kao and Mei, 1971; Qi, 2002). We identified only these parallel couplets to contribute to the bigram counts.

For the Boolean feature ‘semantic parallelism’, we have manually annotated the semantic categories of the characters in a subset of poems from the *Complete Tang Poems*. Specifically, each noun in these poems is assigned its category in the semantic taxonomy presented in Table 3. These annotations cover 78% of all characters in the *Complete Tang Poems*.

For the Boolean feature ‘POS parallelism’ and the dependency bigram counts, POS and dependency information were derived from a treebank of classical Chinese poems (Lee and Kong, 2012).

For each character slot, the system is to suggest *N* characters. It does so by calculating a score for each candidate character, and then retrieving the *N* candidates with the highest scores. More specifically, for each candidate, it constructs a feature vector, say w, using the four n-gram counts (i.e. trigram and bigram, horizontal bigram, and dependency bigram), and two Boolean values (i.e. semantic parallelism and POS parallelism), represented as 0 and 1. Furthermore, the system is also given a real-valued vector, say λ, that represents the weights of the features. The score of the candidate is the dot product $\lambda \cdot w$, in the form of the log-linear model, a widely used framework in various natural language processing tasks (Jurafsky and Martin, 2009).

The weights in λ encode the relative importance of the features. For example, weighing semantic parallelism more heavily than trigrams would promote corresponding characters that are semantically parallel, even though they may result in trigrams that are less frequently attested. A standard procedure would be to optimize the model parameters (feature weights) on a training set, such as the *Complete Tang Poems*, according to the maximum likelihood criterion (Berger *et al.*, 1996), to nudge the user toward the same vocabulary and style in existing poems. We found, however, that users tend to prefer character suggestions that yield semantic parallelism. We therefore manually tuned the weights to give greater emphasis on semantic parallelism in corresponding characters; in future versions of the application, we plan to let the user have the option of adjusting these parameters, so as to stress certain features over others to reflect their personal preference.

For the phonological rules in Section 5, the ancient pronunciations and tones of the characters were harvested from Guangyun 廣韻[2], a Chinese rhyme dictionary (Chen, 1993). This dictionary covers more than 97% of the characters present in the *Complete Tang Poems*. When a character bears multiple pronunciations and tones, the system should ideally choose the one that is correct in the linguistic context. Unfortunately, this disambiguation task would require sophisticated automatic syntactic and semantic analyses. Such analyses are challenging for the state of the art in natural language processing; they are further complicated in our case, because the linguistic context is typically incomplete while the couplet is still being composed. Our system therefore takes a lenient approach: it flags an error only when all possible pronunciations and tones fail to conform to the phonological rules. This approach avoids false alarms, but allows some phonological errors to pass unnoticed in some cases.

7 Evaluation

Automatic poem generation systems are evaluated by the quality of the candidate couplets they produce. The quality can be measured by humans (Jiang and Zhou, 2008), or by comparing the candidates with reference material, typically using the BLEU metric from machine translation (He *et al.*, 2012), or ROUGE from summarization (Yan *et al.*, 2013).

The emphases in the evaluation of computer-assisted poem composition systems tend to differ, since the utility of these systems depends on the quality of the character suggestions they make. We measured this quality in two studies. First, we asked a human to compose poems with our application, using solely those characters suggested by the system; the quality of these poems was then rated by other human judges (Section 7.1). Second, we estimated how often a poet can expect to find the next or corresponding character he/she wanted within the list of characters suggested by the system (Section 7.2).

7.1 Human evaluation

In this evaluation, we asked a human to compose couplets using our application, under the strict condition that he can only choose from the list of characters suggested by the system. We then asked human judges to rate the quality of these couplets; further, we measured the gap between the quality of these couplets and those drawn from the classics (ceiling), and those automatically composed by the system (floor).

7.1.1 *Experiment setup*

From Table 3, we selected five semantic categories, 'Geographic', 'Human relations', 'Products of civilization', 'Celestial', and 'Body part'. From each of these categories, we randomly chose three characters to serve as $w_{1,1}$, i.e. the first character in the first line. For example, for the category 'Celestial', the three characters *xing* 星 'star', *mu* 霧 'fog', and *feng* 風 'wind' were chosen. The three characters are then used to produce couplets in three different settings:

- ***Automatically generated*** (**'Auto'**). Given the character $w_{1,1}$, the rest of the couplet is constructed one character at a time, always using the top-ranked characters proposed by the system.
- ***Computer-assisted*** (**'Assisted'**). Starting with the character $w_{1,1}$, a human composes a couplet using our application. The human, an amateur poet with a Bachelor's degree in classical Chinese, is constrained to choosing only from the top ten characters suggested by the application—he is not permitted to enter any character himself.
- ***Drawn from corpus*** (**'Gold'**). We retrieved couplets that start with $w_{1,1}$ from the *Complete Tang Poems*. We have avoided the more famous couplets, to minimize the chance of skewing the decisions of our human judges.

With three settings of each of the five categories, we thus produced a total of fifteen couplets for evaluation. For instance, the three couplets produced for the semantic category 'Celestial', starting with the characters 'star', 'fog', and 'wind', respectively, are as follows (English translations are our own):

- **Auto**: 星霜凝碧樹 露雨滿朱枝 'Stars are like frosts on a pleasant tree; and dew like rain on the branches'.
- **Assisted**: 霧裡見秋月 雲邊逢曉風 '[I] gaze at the autumn moon in the fog; [I] feel the morning wind from the edge of the cloud'.
- **Gold**: 風雪軍城外 蒹葭古寺中 'A storm at a garrisoned city; reeds inside an ancient temple'[3].

7.1.2 *Results*

Having verified that these couplets adhered fully to the phonological constraints laid out in Section 5, we asked sixteen human judges to rate these couplets. All judges were native speakers of Chinese. Without knowing the method by which the couplets were created, the judges assigned a score between 1 and 5 to each couplet, with 1 as the lowest quality and 5 the highest.

The average scores for the couplets are shown in Table 4. As can be expected, those drawn from the *Complete Tang Poems* ('Gold') received the highest scores, with an average of 3.6. The average score of those composed by humans using our application ('Assisted') were slightly lower, at 3.3. Again slightly lower was the score received by the automatically generated couplets ('Auto'), at 2.8.

As a subjective evaluation, judges naturally differed in their ratings. The average standard deviation is 0.94 for the 'Assisted' couplets, and 1.1 for both the 'Auto' and 'Gold' couplets. Since some judges may be more generous with ratings than others, the overall average scores mask some variations in the quality of some of the 'Assisted' and 'Auto' couplets. It can be insightful, however, to analyze the relative scores the same judge gave to the couplets. The first column in Table 5 lists the percentage of judges who ranked an 'Auto' couplet higher than a 'Gold' couplet. The 'Auto' couplet in the 'Celestial' category was most highly regarded by the judges. More than half (56%) preferred it to its 'Gold' counterpart. Reflecting the design of our application, even the low-scoring 'Auto' couplets still tend to be highly parallel, in terms of both the POS and meaning of the corresponding characters; the content, however, are not always coherent.

The right column in Table 5 lists the percentage of judges who gave the same or higher score to an 'Assisted' couplet than a 'Gold' couplet. In all but one category ('Products of civilization'), at least half (ranging from 50 to 63%) of the judges did so. This shows that the top-ten list of suggested characters provides ample creative space for a user to compose a high-quality couplet. The computer-assisted compositions in the 'Celestial' and 'Geographic' categories were especially well rated.

7.2 Post hoc evaluation

The evaluation above measures the performance of the application in actual use by an amateur poet;

Table 4 The average score received by the couplets, out of a maximum of five

Method	Average score
Gold	3.6
Assisted	3.3
Auto	2.8

The couplets were created using three different methods: drawn from *The Complete Tang Poems* ('Gold'), written by a human using our application ('Assisted'), and automatically generated by our application ('Auto').

Table 5 The percentage of judges who gave a higher score to an 'Auto' or 'Assisted' couplet than to a couplet drawn from the *Complete Tang Poems* ('Gold')

Semantic category	Auto $\geq$ Gold (%)	Assisted $\geq$ Gold (%)
Geographic	38	63
Human relations	44	50
Products of civilization	31	44
Celestial	56	63
Body part	38	50

here, we gauge the coverage of the character suggestions, by estimating how often a poet would have found what he/she had in mind within the list of suggested characters.

7.2.1 *Experiment setup*

We compiled two sets of couplets, one drawn from poets from the Qing Dynasty, the other from a contemporary Hong Kong poet (Tsoi, 1985). Each set consists of twenty couplets, all composed in regulated verse, split evenly between pentasyllabic and heptasyllabic couplets. We entered each couplet into our application, one character at a time, while keeping count of the number of times the next and/or corresponding character in the couplet appeared as one of the top ten suggestions.

7.2.2 *Results*

As shown in Table 6, if a Qing poet was to use our application, he can expect 27% of the next character to appear among the top ten characters suggested by our application; the proportion for corresponding character is slightly higher, at 33%. The ranks of the suggested characters averaged at 4.7 and 4.9, respectively. The corresponding figures for a contemporary poet are lower.

The lower percentage for the contemporary poet compared to those in the Qing dynasty can be mostly attributed to differences between modern and classical Chinese. The contemporary poet, though composing in the regulated verse style, often adopted modern Chinese word collocations that do not exist in the classical language.

These relatively low recall levels are perhaps not unexpected, given the enormous number of character choices that are possible at any position. Moreover, highly acclaimed poems often utilize unique or even surprising character collocations; semantic parallelism is not always strictly observed in our samples. Data sparseness, of course, is always an issue, and may be ameliorated by incorporating text from other classical Chinese material in our training data. Recall can also be improved by offering more character suggestions. One interesting topic for future research would be to determine if there is an optimal length of the list of suggested characters, and how the length has impact on the speed of composition and the quality of the resulting couplet.

Table 6 Proportion of characters in couplets, composed by Qing and contemporary poets, that are among the top ten suggestions

Time period	Next character (average rank)	Corresponding character (average rank)
Qing	27.0% (4.7)	33.3% (4.9)
Contemporary	22.5% (5.2)	14.2% (5.2)

8 Conclusion

We have described an application that assists users in composing couplets in the regulated verse style in traditional Chinese poetry. It makes suggestions for the next and corresponding characters in the couplet, and alerts users of deviations from phonological rules. In contrast to automatic poem generators, our application offers dynamic assistance to poet during the composition process. Like many of its peers, our application benefits from n-gram statistics derived from collections of poems, in our case the *Complete Tang Poems*. A distinguishing feature of our application is its rich linguistic knowledge from classical Chinese philology, allowing it to provide detailed feedback on phonological patterns, and to explicitly model semantic parallelism.

We measured the quality of poems composed with assistance from our application, and estimated how often a poet can expect to find the next or corresponding character desired within the list of suggested characters. An encouraging finding was that among those couplets composed only with characters suggested by the application, more than half received scores as high as couplets drawn from the *Complete Tang Poems*. This shows that the application can provide a creative space for an amateur poet to compose high-quality couplets.

In future work, we hope to harness fully automatic poem generation systems within our interactive application. For example, during the composition process, the application can repeatedly call a fully automatic system, to obtain suggestions not only for individual character slots but also for the rest of the couplet. Further research is needed to investigate how best to combine the strengths of fully automatic and computer-assisted composition systems.

References

Berger, A., Della Pietra, V. J., and Della Pietra, S. A. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, **22**(1): 39–71.

Cai, Z.-Q. (2008). *How to Read Chinese Poetry*. New York: Columbia University Press.

Chen, P. 陳彭年 (1993). A new revision of Kuang-yün (Sung Edition): A dictionary of rimes for ancient Chinese language with cross-referencing pronunciations 新校互註宋本廣韻. In Yu, N.-W. 余迺永 (ed.), Shanghai: Shanghai Lexicographical Publishing House.

Diaz-Agudo, B., Gervás, P., and Gonzalez-Calero, P. A. (2002). Poetry generation in COLIBRI. *Proc. 6th European Conference on Case Based Reasoning (ECCBR). Aberdeen, UK.*

Furuta, K. (1994). Zhongguo wen xue de dui ju yi shu 中國文學的對句藝術 Translated by Li, M. 李淼. Taipei: Qi ling chu ban she.

Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Prentice Hall.

Gervás, P. (2001). An expert system for the composition of formal Spanish poetry. *Journal of Knowledge-Based Systems*, **14**(3–4): 181–8.

Gervás, P. (2013). Computational modelling of poetry generation. *Proc. Symposium on AI and Poetry, AISB Convention, Exeter, UK.*

Greene, E., Bodrumlu, T., and Knight, K. (2010). *Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation. Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), Cambridge, USA.*

Harvey, B. L. and Kao, K.S.Y. (1979). Text generative modelling of Chinese regulated verse. *Poetics*, **8**: 459–79.

He, J., Zhou, M., and Jiang, L. (2012). *Generating Chinese Classical Poems with Statistical Machine Translation Models. Proc. Conference on Artificial Intelligence (AAAI), Toronto, Canada.*

Ho, M. W. 何文匯 (2008). Jintishi gelv qian shuo 近體詩格律淺說 [A basic introduction to the metric of Chinese regulated verse]. In Man Wui, H.O. (ed.), 何文匯. *Yue du 粵讀 [Cantonese Pronunciation]*. Hong Kong: Mingpao Group, pp. 167-200.

Ho, M. W. 何文匯 (2011). Jintishi "gu ping" za shuo近體詩"孤平"雜說 [A broad discussion of "gu ping" in Chinese regulated verse]. In Man Wui, H.O. 何文匯 (ed.), *Wenhui wen xuan* 文匯文選 *[Selected Works of Man Wui]*. Hong Kong: Commerical Press, pp. 80–109.

Jiang, L. and Zhou, M. (2008). *Generating Chinese Couplets using Statistical MT Approach. Proc. 22nd International Conference on Computational Linguistics (COLING), Manchester, UK.*

Kao, Y.-K. and Mei, T.-L. (1971). Syntax, diction and imagery in T'ang poetry. *Harvard Journal of Asiatic Studies*, **31**: 49–136.

Lee, J. and Kong, Y. H. (2012). *A Dependency Treebank of Classical Chinese Poems, Proc. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Montreal, Canada.*

Lo, F.-J. 羅鳳珠 (2011). 以語言知識庫爲基礎的智慧型作詩填詞輔助系統[Empowering classical poetry and lyrics composition: An intelligence model based on language knowledge database]. 教學科技與媒體 [Research of Educational Communications and Technology] (95):36–52.

Manurung, R., Ritchie, G., and Thompson, H. (2012). Using genetic algorithms to create meaningful poetic Text. *Journal of Experimental and Theoretical Artificial Intelligence*, **24**(1): 43–64.

Netzer, Y., Gabay, D., Goldberg, Y., and Elhadad, M. (2009). *Gaiku: Generating Haiku with Word Associations Norms. Proc. NAACL HLT Workshop on Computational Approaches to Linguistic Creativity*. Boulder, Colorado.

Oliveira, H. G. (2012). Poetryme: A versatile platform for poetry generation. *Computational Creativity, Concept Invention, and General Intelligence*, **1**: 21.

Peng, D. 彭定求 (1960). *The Complete Tang Poems (Quantangshi)* 全唐詩. Beijing: Zhonghua shuju.

Qi, G. 啓功 (2002). *Shi wen shenglu lun gao 詩文聲律論稿 [A Sketch on the Discussion of Metric of Chinese Poems and Proses]*. Beijing: Zhonghua shuju.

Shi, X. 施向東 (2001). *Shi ci gelv chujie 詩詞格律初階 [A Basic Introduction to the Metric of Shi Poetry and Ci Poetry]*. Tianjin: Tianjin University Press.

Tan, R. 譚汝爲 (2003). *Shige xiuci jufa yu jianshang 詩歌修辭句法與鑑賞 [The Rhetoric and Appreciation of Chinese Poems]*. Macau: Associação da Ciência Linguística de Macau.

Tang, Z. 唐作藩 (2002). *Yinyunxue jiaocheng 音韻學教程 [Chinese Phonology Textbook]*. Beijing: Peking University Press.

Tsoi, L. K. 蔡聯歡 (1985). *Cai Renqiu shi ci xuan 蔡紉秋詩詞選 [Selected Poems by Tsoi Lan Kwan]*. Hong Kong: Wonderful Printing.

Wang, L. 王力 (2003). *Hanyu shiluxue* 漢語詩律學 *[The Metric of Chinese Poems]*. Hong Kong: Zhonghua shuju.

Wong, M. T. and Chun, A.H.W. (2008). Automatic Haiku Generation using VSM, *Proc. 7th WSEAS International Conference on Applied Computer and Applied Computational Science (ACACOS), Hangzhou, China*.

Xue, N., Xia, F., Chiou, F.-D., and Palmer, M. (2005). The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. *Natural Language Engineering*, **11**(2): 207–38.

Yan, R., Jiang, H., Lapata, M., Lin, S.-D., Lv, X., and Li, X. (2013). *i, Poet: Automatic Chinese Poetry Composition through a Generative Summarization Framework under Constrained Optimization. Proc. 23rd International Conference on Artificial Intelligence (IJCAI), Beijing, China*.

Zhu, C. 朱承平 (1999). *Shi ci gelu jiaocheng 詩詞格律教程 [A Course of Shi Poetry and Ci Poetry Metric]*. Guangzhou: Jinandaxue chubanshe.

Notes

1 Accessible at http://www.poeming.com

2 Accessible at http://w3.uch.edu.tw/luo4/

3 This couplet was drawn from《歲暮江寺住》in the *Complete Tang Poems*, by Qi Ji 齊己.