Template-Free Construction of Poems with Thematic Cohesion and Enjambment

Pablo Gervás
Instituto de Tecnología del Conocimiento
Universidad Complutense de Madrid
Ciudad Universitaria, 28040 Madrid, Spain
pgervas@ucm.es

Abstract

Existing poetry generation systems usually focus on particular features of poetry (such as rhythm, rhyme, metaphor) and specific techniques to achieve them. They often resort to template-base solutions, in which it is not always clear how many of the alleged features of the outputs were already present in the template employed. The present paper considers two specific features – thematic consistency, and enjambment – and presents an ngram based construction method that achieves these features without the use of templates. The construction procedure is not intended to produce poetry of high quality, only to achieve the features considered specifically in its design. A set of metrics is defined to capture these features in quantitative terms, and the metrics are applied to system outputs and to samples of both human and computer generated poetry. The results of these tests are discussed in terms of the danger of to ignoring certain features when designing construction procedures but valuing them during evaluation even if they arise from hard-wiring in the resources or serendipitous emergence, and the fundamental need for poets to develop a personal voice – fundamental for human poets and inconsistent with the application of Turing tests.

1 Introduction

Computer poetry generation has existed for some years now. Yet existing work in this field has very rarely applied existing techniques from natural language generation such as content planning, referring expression generation, lexical choice or surface realization. With the sole exception of Manurung's pioneering work (Manurung, 2003), attempts at computational poetry generation in the past have resorted to more generic artificial intelligence techniques, such as case-based reasoning, evolutionary programming or statistical language modelling rather than traditional natural language generation methods. At a lower level of granularity, these attempts operate more in terms of string manipulation than linguistic representation, and most of the solutions can be seen as template based generation. This is partly due to the properties of poetry, which, in contrast with prose, allows for evocative use of language that need not build complete sentences, but rather can get away with simple phrases concatenated into verse. Whereas this sort of tolerance is acceptable in the early stages of exploration of the field – when a poetry generator that could string simple phrases into verse was better than nothing –, at some point researchers interested in computer poetry generation need to consider the possibility of advancing beyond this.

The present paper considers a subset of the desired features of poetry as a text – thematic consistency, and enjambment – that are a (maybe optional) characteristic of human generated poetry but are often overlooked by computer generated poetry. A set of metrics is defined to capture these features in quantitative terms, and these metrics are tested on samples of both human and computer generated poetry. The results of these tests are discussed in terms of whether the features are indeed optional or whether they can help to distinguish instances

of simpler poetry from more elaborate examples. A new poetry generation system is proposed that specifically addresses some of the new proposed features and produces poems that score reasonably well on the proposed metrics.

2 Previous Work

The goal of this paper involves consideration of a number of poetic concepts establishing the subset of poetry-specific features that are being considered in the present paper (section 2.1), and a subset of the existing poetry generators that consider some of these features (section 2.2).

2.1 Definitions of the Poetic Features Considered

The idea that a given stanza should observe a *thematic unity* has been a classic consideration in traditional disciplines (Korpel and de Moor, 1998). In Arabic poetry, the lack of unity among the verses of a poem is denounced as a severe defect (Moreh, 1988). The formulation employed to justify this as a defect is that "In such poems it is possible to transfer a verse of one poem to another poem of the same meter and rhyme, or to change the order of the verses in the same poem without affecting the meaning or the subject". Consideration of this as a defect may be too strict even for most human poets. But it clearly establishes a criterion that may allow distinction of different degrees of elaboration for computer generated poetry.

Enjambment is a term used in poetry to describe cases where the meaning runs over from one poetic line to the next, without terminal punctuation (Baldick, 2008). Lines without enjambment, in which the syntactic unit (phrase, clause, or sentence) corresponds in length to the line, are called *end-stopped*. Enjambment has been identified as a sign of maturity in Shakespeare's poetry, with his later works distinguished by more frequent use of enjambment (McDonald, 2006). Although the correspondence between metric unit and syntactic units can be considered a positive feature, it seems reasonable to explore the possibility of establishing quantitative measures to identify the use of enjambment as an elaborate feature that requires skill and that many poets have considered an extremely positive feature which helps tie different lines together.

2.2 How Existing Poetry Generators Address the Features under Consideration

Explicit consideration of content as well as form in poetry was a distinguising feature of (Toivanen et al., 2012), developed to generate poetry in Finnish. This approach relied on corpus-based solutions for its generation task, and used separate corpora for form and content. Form was determined by a *grammar corpus* that provided instances of existing poetry that were adapted to create new poems by replacing some of their words with desired content. Content was determined by a *background corpus* from which a word association network for a user provided topic is mined based on word co-occurrences. The network is then used to provide candidate replacements for the words in the template selected from the grammar corpus, which leads to a certain thematic consistency. Because the templates are defined in terms of complete stanzas, the resulting poems do show instances of enjambment (as present in the grammar corpus). The reported version of the system does not consider rhythm or rhyme but mention is made of future work that would do so.

A different approach with potential impact on thematic consistency is the use of mood and sentiment in (Colton et al., 2012), which generated poetry in English. Here, a mood for the day is chosen at the start, then an article from the Guardian newspaper is chosen from which to mine keyphrase that will be combined with a template-based solution for complete stanzas over which rhythm and rhyme controls are imposed. The mood and the newspaper article provide thematic consistency, the template provides syntactic structure that sometimes involves enjambment.

PoeTryMe (Gonçalo Oliveira, 2012) generates poems in Portuguese inspired by a set of seed words by identifying semantic relations that the seed words might be involved in and building verses with spans of text that feature the two words involved in the semantic relation. Because all the verses in a poem are built from the same set of seed words, the resulting poems show a certain thematic consistency. It enforces conformance to a chosen metric. Its construction process is line-based, so it does not in principle allow for one line to syntactically connect to

the following one.

The work of (Veale, 2013) argues that prior poetry generators focus too much on rhyme and too little on having coherent content. To address this problem it uses a rich knowledge base of semantic relations between words mined from the Web. The resulting system produces poems in English that show thematic consistency and apparently insightful use of rhetorical figures such as similes, analogies and metaphor. Due to its specific choice of focus, this approach does not enforce conformance with metrical form either in terms of rhythm or rhyme, and it does not address enjambment.

A recent version of the WASP system (Gervás, 2016) addressed the interplay between theme and metric form in generated poems in Spanish over the conceptual space defined by an ngram language model extracted from a corpus of both poetry and prose texts. Generic guidelines were established to recognise regularity in rhythm and rhyme as valuable, but the system was allowed to explore the construction of stanzas of novel form as determined by the language model. Theme was very broadly stated and established by the choice of texts included in the corpus. The construction procedure is line-based with no mechanisms provided for identifying links between lines, so any enjambment in the resulting poems would be serendipitous.

The importance of evaluating thematic consistency in poetry generation has recently been emphasised by (Gonçalo Oliveira et al., 2017), which present a multilingual system capable of generating in Portuguese, Spanish and English. This work based on the PoeTryMe system evaluates – among other features – the semantic similarity between the generated poems and the seeds used to inspire them using PointWise Mutual Information (Church and Hanks, 1990). It also discusses the difficulties associated with applying metrics on poetic features across outputs in different languages, arising from the need of language-specific resources – lexicons, corpora, semantic knowledge bases ...– to inform any automated evaluation processes.

There has recently been a significant effort to address the task of poetry generation using solutions based on neural networks. Some of these initiatives consider explicitly the issue of thematic consistency.

The work of (Zhang and Lapata, 2014) presents a generator of Chinese classical poetry based on Recurrent Neural Networks. This system operates incrementally generating one line at a time, but at each point considers all previously generated lines as a context.

A different system (Yan, 2016) also uses RNN in an Encoder-Decoder with an iterative polishing schema to generate Chinese quatrains. This refines the poem in several passes by regarding the RNN's hidden state of the last line as the gist of the overall semantic representation of the poem.

In a more elaborate approach (Wang et al., 2016) address the problem in two stages, with an initial stage generating a plan for the poem – also in Chinese –, in which a particular subtopic specified as a chosen keyword is assigned to each line in the poem. The system then generates each line of the poem sequentially using a RNN Encoder-Decoder.

3 Poem Construction Aimed at Thematic Cohesion and Enjambment

From an engineering point of view, the existing work on automated poetry generation tends to select one particular feature of the desired inspiring set of poems and focus on developing a system capable of achieving results that satisfy that particular feature. This is usually done implicitly – with no explicit declaration of a decision to specialise on particular features. This approach allows the reader to imagine that the complete problem of poetry generation has been addressed – which may increase the perceived merit of the solution – but usually leads to disappointment and failed expectations when the outputs are considered. In this paper, the process being proposed focuses on the construction of rhyming poems with a certain degree of thematic cohesion and an ability to join up consecutive verses into syntactically acceptable phrases, which results in enjambment. This does not mean that any other features of poetry are ignored, but it does mean that any that appear in the results do so strictly by serendipity. It also means that the lack of any such additional features in the final results cannot be interpreted as a shortcoming of the system, because it is not designed to achieve those. If and when the engineering challenge of achieving the selected features is solved, the integration with techniques for achiev-

ing other features may be addressed. This is standard procedure in engineering and yet often ignored in the context of automated poetry generation. For the sake of methodological clarity, our contribution starts by defining a set of metrics intended to capture the features that we want our system to exhibit. A method for achieving results with those features is described, followed by a discussion of how well the proposed method fares under the metrics and in comparison with previous work.

3.1 Metrics for the Selected Features

The features that we intend to address are thematic cohesion and enjambment, as defined in section 2.1. The poetry under consideration should also address conformance to a given poetical tradition – which usually includes both rhythm and rhyme according to a classic stanza –, but this is not a feature under study in this paper. Conformance to poetical tradition has already been the goal of several research efforts, and it is not the main focus of this paper. Past efforts in the field have shown that such conformance is achievable algorithmically (see systems reviewed in section 2.2).

The features that we are considering could be measured automatically in different ways. However, the procedures for automating them would very likely be language dependent, as they usually need to rely heavily on linguistic resources. Even procedures based on corpora rather than explicitly declared knowledge are associated with specific corpora in the given language, which makes comparison across languages subjective even if an objective method has been followed (Gonçalo Oliveira et al., 2017). We will consider here metrics that rely on a human judge establishing the extent to which a given poem satisfies a given definition. This is less objective than any automated measure might have been, but it allows a measure of comparison across languages.

Thematic cohesion is a reasonably vague concept that everybody understands intuitively but which is difficult to pin down. For the purposes of this paper we will consider thematic cohesion in terms of co-occurrence within the poem of a number of words which can be considered to be semantically related in some way. To ensure that this broad sense of "related to theme" is captured, we have decided here to rely on a subjective definition of the relation as captured by the intuitions of a human evaluator. The metric is defined as:

$$TC = 10 * RN/TN$$

where TC is *thematic consistency* RN is the number of related nouns and TN is the total number of nouns.

Enjambment is a feature associated directly to the border between one verse and the next. A *line transition* is defined as the border between one line in the poem and the next. A line transition is considered *open* if the line after the transition can be considered a valid syntactic continuation of the line before the transition. These definitions allow the computation of the following metric:

$$EP = 10 * OLT/TLT$$

where EP is *enjambment percentage*, OLT is the number of open line transitions in the poem and $TLTV$ is the total number of line transitions.

3.2 A Generation Procedure Addressing the Features

In order to explore the level of difficulty involved in achieving poems that satisfy the proposed features, we have implemented a poetry generation module that targets these features specifically during construction. The SPAR (Small Poem Automatic Rhymer) system is based on observation of how human poets carry out their task. Namely with a strong base in the set of texts read by the poet before sitting down to write.

3.2.1 The Reference Corpus

SPAR generates based on a corpus of adventure novels that includes (Spanish versions of) *Tarzan of the Apes* by Edgar R. Burroughs, *Sandokan* by Emilio Salgari, *The Jungle Book* and *The Second Jungle Book* by Rudyard Kipling, *Peter Pan* by J.M. Barrie, *Alice in Wonderland* and *Through the Looking Glass* by Lewis Carrol, *The Prince and the Pauper* by Mark Twain, *The Hound of the Baskervilles* and *Study in Scarlet* by Conan Doyle. The choice of texts for the corpus was affected by two main reasons, one historical and one strategical. The historical reason is that the corpus had been compiled

previously to inform a poetry generation exercise in which primary school students were invited to interact with a poetry generator. For this purpose a set of texts considered classical readings for Spanish children had been chosen. The strategical reason is that it was decided that this set of prior readings should not include any poetry, so that the system allow testing of the ability of the system to generate verse inspired by a set of prose texts, and to avoid the risk that any poetic quality appearing in the resulting poems be directly attributed to a loan from poems appearing in the reference texts used as seed.

3.2.2 The Poetry Generation Process

SPAR carries out the poetry generation task in five separate stages. First, it builds from the reference corpus a series of models of which words in the available vocabulary appear next to others in the reference texts, and which words rhyme with one another. These models are used to inform later stages. Second, starting from a word provided by the user – which is intended to set the theme for the resulting poem – the system build a set of words related to the seed word. Relation in this context is defined in terms of simple cosine distance in a vector model representation of the reference texts (Salton et al., 1975). This set of words represent the concepts that the system considers might be mentioned in a poem that had the given word as a title. Third, it searches for connections between these words and potential rhyming words. A connection is understood to exist between words if they co-occur within the same window of N words in at least one of the sentence in the set of reference texts. Fourth, by exploring the search space determined by these connections the system builds phrases that might be included in a poem. These phrases are defined as spans of text that either connect target words to one another or a rhyming word to a target word. Each span is built by exploiting an n-gram language model of the reference texts to search for valid sequences of words that connect the desired words. Finally, for a given stanza, it searches for combinations of the resulting verses that satisfy the restrictions on rhyme and can be joined together with a minimum of cohesion. In this context, a minimum of cohesion is understood as having at least one ngram that overlaps the end of the first verse and the beginning of the second one.

3.3 The Resulting Poems

The SPAR system was used to generate a collection of 18 poems in Spanish. The poems were commissioned for the *Festival Poetas* poetry festival, celebrated in Matadero Madrid on 27-29 May 2017. The 18 poems were classical sonnets (14 verses of 11 syllables with rhyme schemes either ABBAABBACDCDCD or ABBAABBACDECDE). The design of the construction process ensures strict enforcement of this form. An example poem is presented in Table 1.

3.4 Applying the Metrics

In terms of the metrics defined in section 3.1, the SPAR system fares reasonably well. The current approach to publication of poetry generation research does not allow for collections of poems so built to be made available widely. This makes it difficult to carry out quantitative comparison between approaches, as only the data made available in each paper can be used. For comparison purposes, the proposed metrics have been applied to the sample poems published for some of the referenced poetry generators. Results for the SPAR collection in comparison with the poems published for some of the systems reviewed in Section 2.2 are presented in Table 2. For reference, results of applying the metrics to two different sets of Spanish poems are also included at the start of the table. These correspond to sets of four sonnets for classical 16th century poets (16C) and for 20th century poet Miguel Hernández (20C). The size of the sample has been selected to match that of available samples for computer generated poems.

As the various systems considered here were not originally design to address the issues on which they are currently being tested, it is important to qualify these numbers with some comments. The output of the system by (Colton et al., 2012) presents two types of poem, one based on a stanza-sized template and another based on loose chaining of independent lines. The results for enjambment in this case are not as meaningful as in other cases, because for one type the enjambment is inherent in the starting template and for the others it is non-existent. The system by (Toivanen et al., 2012) relies on stanza-sized templates for construction, so data on enjamb-

Por una mujer a la maldición.	For a woman to the curse.
De un hombre es un ser que les seguía.	Of a man is a being that followed them.
Miedo por la ley que no comprendía.	Fear for the law he did not understand.
Los celos hacia el mono y la expresión.	Jealousy toward the monkey and its expression.
Miedo por el hombre a una habitación	Fear for the man to a room
y el dios de la ley que no corría	and the god of the law that did not run
por un mono y el pueblo y que podía.	for a monkey and the people and who could.
Amo a esa mujer es su profesión.	I love that woman it is her job.
Quien amo y su mujer en mi carrera	Whom I love and his wife in my race
llega a ser que el hombre de no mostrar.	gets to be the man not to be shown.
Por un mono y la ley y el sonido	For a monkey and the law and the sound
que el dueño y el cachorro no tuviera.	that the owner and the cub would not have.
Sospecha que el cerebro y a juzgar.	She suspects that the brain and to judge.
Profundidades hasta que debido.	Depths until it is owed.

Table 1: Example of sonnet generated by the SPAR system for the seed word "Celos" (Jealousy).

	TC	EP
16C	7.2	8.8
20C	3.7	6.2
SPAR	5.2	4.8
(Colton et al., 2012)	2.9	1.3
(Toivanen et al., 2012)	3.0	2.0
(Veale, 2013)	8.1	0.0
(Gonçalo Oliveira et al., 2017)	3.7	0.0

Table 2: Results for samples of human (16C and 20C) and computer generated poems on the proposed metrics. In each case, average over the available set of poems in the sample is given.

ment refer directly to the chosen set of templates rather than the construction method. The system by (Veale, 2013) focuses explicitly on thematic consistency, and achieves very high scores on that, but has no concern about enjambment. The results on thematic consistency for (Gonçalo Oliveira et al., 2017) are heavily penalised by the fact that the metric only considers nouns, and should not be considered significant, as the poems do show additional indications of consistency in terms of verbs and adjectives.

4 Discussion

The application of the metrics to computer and human generated poems gives rise to some insights.

Thematic consistency is very difficult to evaluate. Simple perusal of the various poems gives a human reader a very solid intuition of whether a particular theme is being pursued, but this intuition is extremely difficult to quantify. Approaches that rely on automated means for extracting word associations from statistical analysis of corpora – such as (Toivanen et al., 2012) or the SPAR system itself– sometimes come up with word associations for which the rationale is very difficult to follow. This makes them score less well under human evaluation for consistency than they should, as they generally have followed strict construction procedures to achieve significant presence in their output of the desired words. System based on knowledge bases capturing semantic relations between words – such as (Colton et al., 2012), (Veale, 2013) and (Gonçalo Oliveira et al., 2017) – fare irregularly, with (Veale, 2013) – which focus specifically on thematic cohesion – achieving the highest score.

The proposed metrics relies exclusively on nouns, and should be extended to consider other types of words.

In comparison with the results provided for human generated poems, it might seem that modern poetry departs from the degree of thematic consistency shown by earlier poems. The problem that has been observed during application of the metrics is that the use of figurative language can significantly cloud the issue of consistency. Where the poet is working on one or more metaphors to illustrate his theme, a literal understanding of "is related to" will undermine his score even where a human reader will find obvious connections. In this sense, the set of poems used to represent 20th century poetry make heavier use of metaphorical associations. Further

work should address the role of metaphor in the application of metrics of this type.

The size of the samples is also problematic, both with respect to this particular measure and in general. For this type of quantitative metrics, application to a much larger sample would be desirable. This suggests that some means should be found in the field to associate with particular publications data sets of the resulting poems, so that this kind of empirical testing might be applied.

Concerning enjambment, the results on the metrics illustrate that the feature has been generally ignored by poetry generation systems in the past. It is also clear that the comparison between human and computer generated poems shows a significant gap with respect to this feature. Researchers working on poetry generation would do well to address this aspect specifically in future work.

Finally, the metrics as applied to the different systems considered here show that the use of templates, in spite of working considerably well regarding the quality of the outputs, is actually obscuring the fact that many significant issues underlying the task of poetry generation are being side-stepped. Enjambment as considered here is a case in point, but there may be multiple others in similar circumstances. In view of this, we advocate for a progressive transition from template-based solutions to more elaborate techniques for generating text. This may involve discovery of new methods of text generation, but it may also be achieved by more informed consideration of existing natural language generation techniques in the cause of poetry generation.

With respect to the SPAR system itself, a number of issues require comments.

Because the search spaces involved are so large, each of the stages described in section 3.2.2 can take between one and three hours of computing time. With smaller search spaces, the system might finish in shorter times, but the probabilities of finding valid combinations decrease in proportion. The density of correct verses that can be generated from a given (non-poetry) corpus is very low. This is what makes poetry generation so difficult. For these reasons, this particular approach to the automatic generation of poetry is not yet in a position to be used interactively.

The nature of the corpus – a set of adventure novels popular among young adults – has a strong influence in the results that can easily be perceived by anyone reading the poems. In contrast with poetry originating from other sources, the poems include frequent references to bears and wolves – Baloo and the Seeonee pack from the Jungle Book –, monkeys and lions – from Tarzan of the Apes –, a small bottle – as used by Alice – or to children's bedtime – the Darling children in Peter Pan. This peculiarity of the generated poems may make it less likely for readers to find connections between the poems and their own personal feelings, but it helps create an illusion of a joint general background and, in some sense, a particular voice for the automated poet.

Human poets dedicate a significant amount of energy to find a personal *voice*. This implies being able to produce poems that are significantly different from any others that had been produced before, and which can be attributed to that poet by someone familiar with their prior work. For a human poet, to have a part of their work declared indistinguishable from that of their peers, or – even worse – indistinguishable from the classics would be a radical sign of failure. This is an important issue for computer generated poetry, related to the expectation of originality traditionally associated with creativity. This is an important argument against the recent trend in the consideration of potentially creative outputs generated by computers to apply Turing test style evaluations, where success is associated with machine results being indistinguishable from human efforts. In the field of poetry, results indistinguishable from prior efforts are a sign of failure, not success.

The poems generated by the SPAR system cannot be confused with poems generated by a human. There is a clear tendency in them towards the surreal, an occasional warping of the rules of grammar to achieve metric correction, and a fixation with wild animals that arises from its background readings. That is in a way, the voice of the system. Maybe a relatively inmature voice at this stage, but clearly personal, different from what came before and recognisable once a number of poems have been read. To devote efforts to eliminate the small quirks that constitute at this stage the voice of the system would be detrimental to its perception as a poet with no significant advancement in terms of having modelled significant human abilities.

Possible improvements would be getting the system to become aware of more features of poetic texts to take into account – such as metaphor or aliteration – and to start operating with more elaborate definitions of purpose or intended message.

5 Conclusions

The use of templates in poetry generation leads to output poems of considerable quality, but clouds the actual capability of the systems in question to emulate fundamental abilities of human poets. Work in this field should progress away from the use of templates and make better use of existing natural language generation techniques.

Thematic consistency is very difficult to identify even for human judges, and it is therefore extremely difficult to automate. Any attempt to do so would need to find a solution for figurative use of language and the role of metaphorical connections in poetry.

Enjambment is a relevant and popular feature of human poetry that has not been addressed by poetry generation systems in the past. Metrics on enjambment can currently act as discriminators for human vs. computer generated poetry.

The use of Turing test evaluations for poetry generation is inconsistent with the basic tennets that define success and failure for human poets. Further effort should be made to evaluate computer generated poetry in ways that allow the attribution of quality independently of the ability to distinguish it from human poetry.

Acknowledgments

This project has been partially supported by project IDiLyCo (MINECO/FEDER TIN2015-66655-R), funded by the Spanish Ministry of Economy and the European Regional Development Fund.

References

C. Baldick. 2008. *The Oxford Dictionary of Literary Terms*. Oxford Paperbacks. Oxford University Press.

K. W Church and P. Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29, mar.

S. Colton, J. Goodwin, and T. Veale. 2012. FullFACE poetry generation. In *Proc. of 3rd International Conference on Computational Creativity*, ICCC 2012, pages 95–102.

P. Gervás. 2016. Constrained creation of poetic forms during theme-driven exploration of a domain defined by an n-gram model. *Connection Science*.

H. Gonçalo Oliveira, R. Hervás, A. Díaz, and P. Gervás. 2017. Multilingual extension and evaluation of a poetry generator. *Natural Language Engineering*, page 1–39.

H. Gonçalo Oliveira. 2012. PoeTryMe: a versatile platform for poetry generation. In *Proc. of the ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence*.

M.C.A. Korpel and J.C. de Moor. 1998. *The Structure of Classical Hebrew Poetry: Isaiah 40-55*. Oudtestamentische Studiën. Brill.

H. M. Manurung. 2003. *An evolutionary algorithm approach to poetry generation*. Ph.D. thesis, University of Edimburgh, Edimburgh, UK.

R. McDonald. 2006. *Shakespeare's Late Style*. Cambridge University Press.

S. Moreh. 1988. *Studies in Modern Arabic Prose and Poetry*. E.J. Brill.

G. Salton, A. Wong, and C. S. Yang. 1975. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620, November.

J. M. Toivanen, H. Toivonen, A. Valitutti, and O. Gross. 2012. Corpus-based generation of content and form in poetry. In *Proc. of 3rd International Conference on Computational Creativity*, ICCC 2012, pages 175–179.

T. Veale. 2013. Less rhyme, more reason: Knowledge-based poetry generation with feeling, insight and wit. In *Proc. of 4th International Conference on Computational Creativity*, ICCC 2013, pages 152–159.

Zhe Wang, Wei He, Hua Wu, Haiyang Wu, Wei Li, Haifeng Wang, and Enhong Chen. 2016. Chinese poetry generation with planning based neural network. In *Proceedings of 26th International Conference on Computational Linguistics*, COLING 2016, pages 1051–1060. ACL.

Rui Yan. 2016. I, poet: Automatic poetry composition through recurrent neural networks with iterative polishing schema. In *Proc. of the 25th International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 2238–2244. AAAI Press.

Xingxing Zhang and M. Lapata. 2014. Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2014, pages 670–680. ACL.