Haiku Generator That Reads Blogs and Illustrates Them with Sounds and Images

Rafal Rzepka and Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University
Sapporo, Japan
{rzepka,araki}@ist.hokudai.ac.jp

Abstract

In this paper we introduce our haiku generator, which, in contrast to other systems, is not restricted to limited classic vocabulary sets and preserves a classic style without becoming too random and abstract because it performs a semantic integrity check using the Internet. Moreover, it is able to analyze blog entry input and, by using nouns and adjectives for web-mining, to stay on topic and still preserve kigo, traditional seasonal words used in Japanese poetry. The haiku generator utilizes grammar templates automatically generated from poems written by Japanese poets and a lexicon of 2,473 kigo words from an online haiku repository. In addition to generating haiku poems, it can output them vocally together with related sound effects and images retrieved from the WWW. Our experiments demonstrate that the proposed system generates high-quality haikus and that using content-related input and multimedia-rich output is effective for increasing users' satisfaction. We have performed impression evaluation experiments and confirmed that our method is especially useful for generating haikus with higher depth and sound-sharpness, which are two very important categories in professional evaluation of Japanese poetry. Next, haikus generated using the proposed method were evaluated by blog authors and blog readers and again, the proposed method outperformed the baseline. We also measured how the presence or absence of multimedia output influenced the evaluation. While using both vocal output and an image achieved higher scores than text alone, there were cases in which some combinations of effects were evaluated higher than all the effects used together. With our original approach to generating poetry, we wish to show the importance of new media and possibilities that are arising from the utilization of the "wisdom of (web-)crowds" in order to achieve higher standards for AI-generated art.

1 Introduction

Since paintings produced by the Aaron system [Cohen, 1995] were shown at the San Francisco Museum of Modern Art in 1979, artificial art has become a widely discussed topic that is not limited to computer enthusiasts. However, the main question remains unanswered – was Aaron, the plotter-equipped art generating program, extending its author's "teachings" and actually creating something, or was it a mere "expert system" with newer capabilities incrementally added to its repertoire? The problem, at least in our opinion, lies in subjective evaluation, which until recently was widely avoided in science and engineering. However, we think that the closer that machines are to humans with their endeavors, the less uncanny uneasiness we will feel with machines that paint [Lindemeier *et al.*, 2013], sing [Fukayama *et al.*, 2010], joke [Dybala *et al.*, 2008], generate stories [y Pérez and Sharples, 2004] or write poetry [Colton *et al.*, 2012]. We believe that, similarly to how the discussions about the meaning of "life" after the discovery of DNA calmed down, the meaning (or rather the breadth of interpretation) of words like "creativity" will also evolve, and one day we will agree that machines can produce art that we enjoy and do not recognize as "artificial".

1.1 Our Approach

In our opinion, one of the factors that make us feel uneasy about artificial art is the unnatural situations in which we encounter it. Normally, artificial art must be specifically sought; robots' paintings are not on the walls around you, computer generated songs are not played on the radio, and nor does your robotic vacuum-cleaner crack an ironic joke when your room is too dirty after a party. To change this situation, we decided to work on programs that can become a lively piece in the so-called Internet of Things that already surrounds, or soon will surround us. Such pieces can be an interactive frame on your wall, TV screen, your car radio or a mirror when you shave or do your make-up. To test our ideas, we created a program that can be changed into a blog module, RSS reader or other software that can follow what you read and enrich the reader's experience or enhance your own blog posts that you write. It can read a blog entry and express impressions about the contents by reading on-spot generated haikus, playing related seasonal sounds and displaying images. In this paper we introduce our system and describe the results of experiments showing that enriching artificial poetry

with multimedia effects can provide a significantly better experience, when not performed in excess.

1.2 Real and Artificial Haiku

Seventeen-syllable Japanese miniature poems called *haiku* have fascinated people from both Eastern and Western cultures as a quintessence of "beauty in simplicity", showing that a human being can pass powerful (and beautiful) associations to another person with only several words that are restricted by strict grammatical (three lines of 5-7-5 syllables) and less limited semantic rules (a poem must contain one seasonal word *kigo*). There were and still are many attempts (already in the late fifties [Lutz, 1959]) all over the world to create a computer program that can generate poetry of different kinds. Although this is also the case with haiku, we have noticed that such attempts are missing a very important feature of this particular type of poetry – the brief commentary style of Zen-like simple bliss about what we feel in a specific moment of "now" ([Tosa *et al.*, 2007], [Wong *et al.*, 2008], [Tosa *et al.*, 2009], [Netzer *et al.*, 2009]): the deep connection with the situation we have just witnessed, the feeling we have just experienced. We believe that in the case of art, especially when we talk about haiku, it is not the algorithmic sophistication of the generation process, it is more about feeling that the poet understands a situation and that we can mentally agree with what she/he (or maybe *it*) shares with us. With our system, we want to tackle this problem of emotional unison without overlooking the scientific side of the field and providing a deeper evaluation, which is very often neglected by representatives of the poetry generation subfield of artificial art generation.

2 Proposed Method

The basic approach in our method is to use word occurrences in the WWW and grammar templates extracted from haiku poems created by Japanese poets. The key role is given to related nouns. The whole process of the method is presented in Figure 1.

2.1 WWW Usage

To achieve the above-mentioned feeling of a "given moment", of a situation provoking emotional recollections (usually nostalgic in the case of haiku, because the genre is more about the pensiveness of recollections than conveying vividly distinctive emotions), we concentrated on blogs, which are nowadays a usual means of writing a diary, sharing views and moments online; however, the method could be used with any kind of text, from news articles to entire books. As we work with Japanese language, we decided to utilize Ameba Blog[1], which is the biggest Japanese portal for bloggers, and Yahoo! Japan[2] for searching the related words and their occurrences. Software written in Perl accesses the Web with the text browser w3m[3] for both scrapping contents and reading search hits.

[1]http://ameba.jp
[2]http://yahoo.co.jp
[3]http://w3m.sourceforge.net

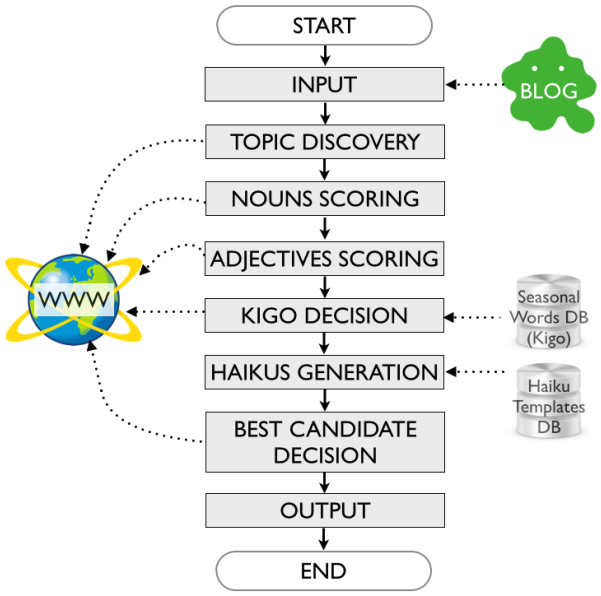

Figure 1: Processing steps of our system.

2.2 Algorithm

A blog site with a title and date information (for season recognition) is used as input. The system checks if nouns are present and whether the whole entry is longer than 50 characters. Nouns are recognized using the standard Japanese morphological analysis tool MeCab[4].

Choosing a Theme

If two or more neighboring nouns are recognized, they are treated as a one-noun phrase. A haiku *theme* is chosen from a list of frequent (in the input blog contents) nouns, and the system tests the importance of one of these against the others by searching the web for their co-occurrences. For example, when one of the theme candidates was "Mt. Moiwa", the other top five frequent nouns from an entry were: "autumn leaves", "wind", "clouds", "mountain" and "observation deck". The acquired hit number of candidate w becomes $N_o(w)$, the number of theme candidate w occurrences within a given entry becomes $C(w)$, and both values are used in the following formula:

$$ThemeScore(w) = N_o(w)^{c(w)} \quad (1)$$

This scoring method allows us to avoid the most common Japanese nouns becoming theme words. The next step is to see if the candidates (starting with ones with the highest $ThemeScore$) are appropriate. The system checks their co-occurrences using a simple co-occurrence probability model (See Figure 2), which is calculated with the following formula:

$$CooccurrenceProb(N_1, N_2) = \frac{Retr(N_1, N_2)}{Retr(N_1)} \quad (2)$$

[4]MeCab: Yet another part-of-speech and morphological analyzer. https://code.google.com/p/mecab/

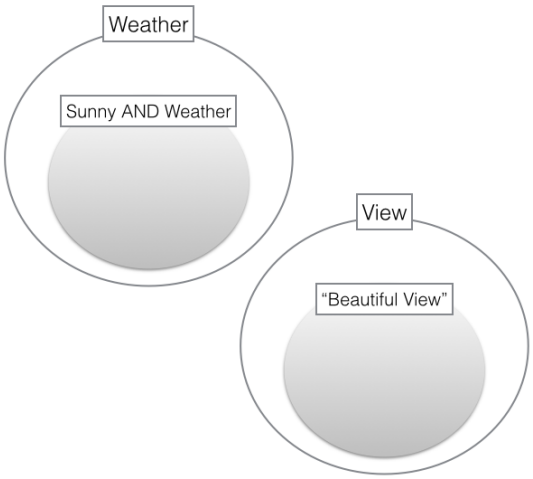

Figure 2: Model of cooccurrence probability in the case of nouns AND queries (left) and adjectives with nouns as exact match queries (left).

where $Retr(N_1, N_2)$ is the number of search hits for a query containing both N_1 and N_2 nouns. This probability is used later to score nouns. The values of the top five nouns become N_1 (e.g. "autumn leaves") and N_2 is the candidate's (e.g. "Mt. Moiwa") value. After repeating this process for all five nouns and summing up the probabilities, the candidate with the highest value is chosen.

Scoring Nouns for Generation
Using co-occurrence probability as explained above, the system calculates words (nouns and adjectives) that are related to both the chosen theme word w, and the title t of the blog. The following method is used, and the closer the co-occurrence of noun N_1 to the title and theme, the stronger relationship they have. If the value approaches 1 and the N_1 score exceeds 1, it changes into a reverse-ratio score.

$$NounScore(N_1) = \frac{CooccurrenceProb(N_1, w)}{CooccurrenceProb(N_1, t)} \quad (3)$$

Acquiring Appropriate Adjectives
Our system also acquires an appropriate adjective $Adj(N_1)$ for every chosen noun by using 20 web search engine snippets[5] (the same approach to co-occurrences is used; see Figure 2). A snippet is parsed and adjectives are extracted. First, the retrievals are validated using the following method:

$$AdjV(N_1, Adj(N_1)) = \frac{Retr(Adj(N_1), N_1)}{Retr(N_1)} \quad (4)$$

The retrievals are then scored with:

$$AdjS(N_1, Adj(N_1)) = AdjV(N_1, Adj(N_1)) \times C(N_1, Adj(N_1)) \quad (5)$$

[5]We have experimented with different numbers and lengths of snippets to ensure that the same pairs are retrieved with the fewest number of searches.

Table 1: Evaluation of adjective ranks.

Rank	*Score*
First place	4.66
Second place	3.45
Third place	4.17
Fourth place	3.33
Fifth place	4.25

where $C(N_1, Adj(N_1)$ is the number of snippet occurrences for $Adj(N_1)$.

In order to confirm whether the items retrieved as the top adjectives are the most related to the blog, we performed a small-scale experiment that showed that the system's first choice (first rank adjective) were preferred by the majority of six human subjects, who read the blogs and evaluated five adjectives from the system's ranking module (the extracted nouns were randomly reordered to preserve fairness). However, as shown in Table 1, values did not decrease constantly, which suggests that we need to test different ranking methods.

Acquiring Appropriate Seasonal Words
Similarly, our system chooses the most appropriate seasonal words kigo from a kigo-database[6] using the following formula:

$$CooccurrenceProb(N_k, w) = \frac{Retr(N_k, w)}{Retr(N_k)} \quad (6)$$

where $Retr(N_k, w)$ is the number of hits for a query containing both kigo and the theme word, and $Retr(N_k)$ is simply the number of kigo web occurrences. The kigo-database consists of 2,473 seasonal words divided into four seasonal categories: spring (601), summer (802), autumn (525) and winter (545).

2.3 Haiku Text Generation

To create grammatical templates, we have randomly chosen 312 from 20,000 poems found on the Internet, among which most (about 17,000) were written by modern (postwar) Japanese amateur poets, and the remaining 3,000 are masterpieces carefully selected by specialists. We did not choose only from the masterpieces because they often use archaic grammar, and we did not want to be biased towards any particular trends in order to reach the widest user tastes possible. From the randomly chosen poems we extracted the seven most common grammar patterns for the obligatory three lines of haiku which could deal with different syllabic lengths of theme words. It must be underlined here that in the case of haiku poems, the wording, not the form variations play the most important role, and for that reason we limited the number of templates (although this feature can be easily extended). For example, for four-syllable theme words the pattern was:

[6]http://www.haiku-data.jp/

- Theme word + possessive particle “no” (first line)
- Three-syllable kigo + possessive particle “no” + three-syllable noun (second line)
- Five-syllable noun or noun phrase (third line)

When, for example, “Mt. Moiwa” *mo-i-wa ya-ma*, five syllables) was the theme word, a slightly different template (with adjective) was used and the following haiku was generated:

Mount Moiwa
is this the view of yellow
gingko leafs?

When a haiku candidate is generated, every line is automatically searched for on the WWW. If the exact match query does not generate any results, the next template is used and another candidate is tested. With this simple step, it is rare to produce lexically or even semantically erroneous lines.

2.4 Adding Voice, Sounds and Images

For reading poems out loud we used a popular program for Japanese speech generation, $AquesTalk_2$[7], and inserted one-second pauses between the lines. We also prepared 10 sound effects such as birds, cicadas or chimes, which were associated by searching their titles with nouns using the Yahoo! Japan search engine. For instance, if the word “summer” was used in a haiku and “cicada” had the highest occurrence, the sound of the insect was chosen to illustrate the poem (the user presses a button to play it). For the incorporation of images, the system simply retrieved the first image from a Yahoo! Japan image search when a kigo from the given haiku was queried. In the default setup an image is displayed automatically, while the options for reading out loud and sound effects are hidden under buttons, and the user can decide whether he or she wants to listen to them.

3 Evaluation Experiments and Results

First, we evaluated the haiku itself using the Semantic Differential (SD) [Osgood *et al.*, 1957] scale with Varimax [Kaiser, 1958]. This method was used for evaluating human-created poetry in the past [Luber, 1976] and is still being enhanced to measure poetic values [Vala, 2011]. We preferred this impression-based and aesthetic sensibility-centered approach to ROUGE measures for document summarization performance evaluation, which concentrate on importance, relevance and coherence [Yan *et al.*, 2013] or BLEU-based evaluation using poetry experts [Zhang and Lapata, 2014].

The pairs of adjectives we used are shown together with the evaluation results in Figure 3. The haiku poems generated with the proposed method were shown in random order to the subjects (23 bloggers in their twenties, 13 males and 10 females) together with haiku produced using random words from an input blog, which the subjects chose freely. The proposed method achieved 4.51 points on average on a 7-point scale (1 worst, 7 best) while the baseline (haiku generated using input blog words randomly inserted into templates) was 3.68, and a statistically significant difference was confirmed ($p < 0.05$ in t-test). Normalized Varimax results are shown in Table 2, and the averages in Table 3.

[7]http://www.a-quest.com

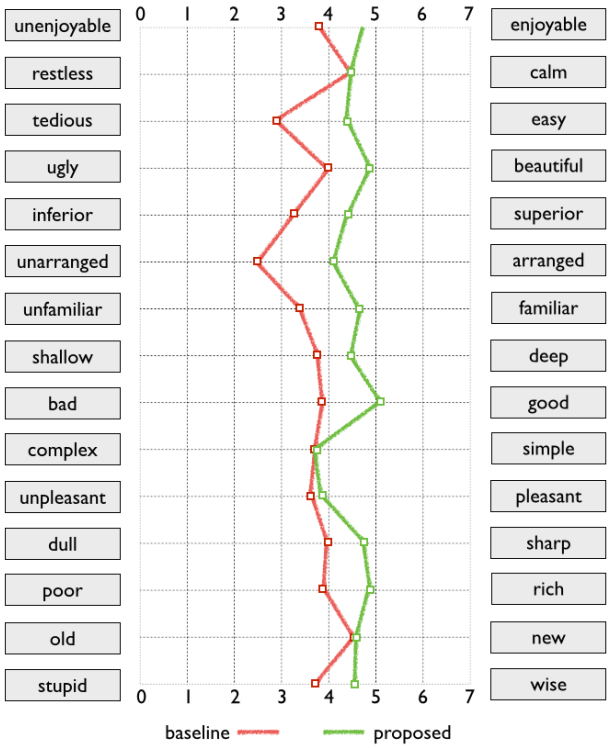

Figure 3: Comparison of impressions for each pair of adjectives (SD).

3.1 Blog Author vs. Non-Author Evaluation

In the next experimental phase, we employed criteria that are used by Japanese professional critics of traditional poetry. We also needed to verify whether being an author of a blog causes differences in haiku evaluation. This time we used 34 people from different backgrounds and ages, half of whom were writing blogs regularly and half with no experience of posting a blog entry (topics were not limited). The bloggers input their own blog entries, and their evaluation was compared with third-person evaluation by the non-authors. We used a common scale from 0 (“I strongly disagree”) to 7 (“I strongly agree”) and asked the following questions, where the first three criteria are commonly used in professional haiku rating [Tsuji, 2001]:

- Did you have seasonal associations expressed in the input blog?
- Did you feel a poetic lingering specific to haiku?
- Did you imagine the scene/scenery?
- Was the sound rhythm appropriate?
- Do you think it was a high-quality haiku?
- Did the poem related to the contents of the input blog?
- Would you like to use the system on your blog?

The results are shown in Table 4 for blog authors and in Table 5 for non-authors.

Table 2: Factors from Normalized Varimax.

	Factor 1	*Factor 2*	*Factor 3*
Enjoyable	**0.634**	0.371	0.374
Calm	0.157	**0.639**	-0.182
Fun	**0.565**	**0.641**	-0.0013
Beautiful	0.312	**0.817**	0.141
Superior	**0.606**	**0.561**	0.071
Arranged	0.427	0.318	0.480
Familiar	**0.875**	-0.018	-0.056
Deep	0.196	**0.504**	0.191
Good	**0.834**	0.346	0.134
Simple	-0.152	10.011	**0.563**
Sharp	0.496	0.288	0.209
Rich	**0.721**	0.441	-0.302
Pleasant	0.188	-0.027	**0.859**
New	0.488	**-0.551**	-0.001
Wise	**0.894**	0.246	0.039
Proportion	0.317	0.204	0.111

Table 3: Averages for Normalized Varimax.

	Satisfaction	*Depth*	*Articulateness*
Baseline	-0.349 (q=0.815)	0.296 (q=0.836)	-0.203 (q=0.785)
Proposed	-0.381 (q=1.071)	0.322 (q=1.059)	0.222 (q=1.234)
t-test	$p < 0.10$	-	-

The proposed system outperformed the baseline in all categories; however, not every difference was statistically significant. Blog authors evaluated the baseline system at 2.99 and the proposed system at 4.55, and the non-authors at 4.37 and 5.05 respectively, showing that blog writers are satisfied with our systems output, but they are harsher in their critique when compared to the non-authors. The results may suggest that the proposed method does not fully agree with the emotions the authors want to convey with their blogs, and needs further development.

3.2 Evaluating Multimedia Effects

As the final stage of our experiments, we conducted a comparison between plain text and multimedia-enriched haiku poems. This time we used five randomly chosen blogs from the Ameba portal and showed the system output to 33 subjects (16-25 years old; 19 males, 14 females) divided into three groups of 11 members for different combinations of media. One evaluation scenario had the output poems displayed simply as plain text, and the second scenario had the addition of haiku read by the previously-mentioned voice synthesis software, with sound effects played and a related image displayed. We set a fixed number of blogs in order to avoid generating new poems when a user enters a different blog as input; therefore, in this experiment the question about relatedness to the blog content is omitted. This decision was also made in order to concentrate on the effectiveness of multimedia and to simplify the new type of experiment. As shown in Table 6, the multimedia-enriched system outperformed the text-only baseline, especially in the professional criteria of "Seasonal Impression" (recollections of the annual seasons) and "Lingering Sensation" (poetic depth).

Table 4: Blog authors evaluation.

	Baseline	*Proposed*	*t-test*
Seasonal Impression	2.28	5.00	**
Lingering Sensation	2.82	4.12	*
Scene Imagination	3.24	4.41	+
Sound Rhythm	3.88	5.24	+
Haiku Quality	2.53	4.24	**
Relatedness to Blog	2.18	3.71	**
Willingness to Use	3.47	5.18	**
			**$p < 0.01$
			*$p < 0.05$
			+$p < 0.10$

Table 5: Non-authors evaluation.

	Baseline	*Proposed*	*t-test*
Seasonal Impression	3.76	5.23	+
Lingering Sensation	4.29	5.13	
Scene Imagination	4.76	4.80	
Sound Rhythm	4.76	6.18	+
Haiku Quality	3.41	3.88	
Relatedness to Blog	5.29	5.59	
Willingness to Use	3.47	5.18	
			+$p < 0.10$

To determine the effect of combining different types of multimedia, we again divided 33 subjects into three groups and asked them to separately evaluate voice, sound and images. It appears that combining all effects together is probably not the best solution, as different sets are stronger and weaker depending on the criteria (see Table 7). For example, the two above-mentioned professional criteria achieve higher scores only in the case of combining text with sound effects, which may suggest that this pair influences our imagination more than when, for example, the text is read by a synthetic voice.

4 Conclusions and Future Work

In this paper we introduced our haiku generating system, which can illustrate any blog not only with written poetry, but also with sounds and images. We performed experiments on these various outputs to see how the generated haiku were evaluated by lay people, both blog authors and blog readers. Our proposed system outperformed baseline systems in both a text-only comparison and comparison between text vs. text with multimedia. This dominance was also preserved in different types of evaluation, and was especially visible when

Table 6: Evaluation comparison between poetry in textual and multimedia forms.

	Only text	*With effect*	*t-test*
Seasonal Impression	4.77	5.38	*
Lingering Sensation	3.97	4.66	*
Scene Imagination	3.80	4.66	*
Rhythm	5.32	5.66	
Haiku Quality	3.61	4.22	*
Willingness to Use	4.90	5.31	**
			*p $<$ 0.05
			**p $<$ 0.10

Table 7: Differences between combinations of effects.

	All effects	*Voice*	*Sound*	*Image*
Seasonal Impression	5.38	4.82	**5.64**	5.45
Lingering Sensation	4.66	4.45	**4.73**	4.55
Scene Imagination	4.69	4.45	4.64	**5.00**
Rhythm	5.66	5.64	**5.91**	5.36
Haiku Quality	4.22	4.18	4.00	**4.36**
Willingness to Use	5.31	**5.64**	5.37	4.73

subjects were asked about haiku characteristics that have been valued by Japanese people for centuries. We discovered that bloggers like the system output but are more critical than readers, which suggests that the generated haiku were often not conveying associations that the bloggers would like to be expressed. However, higher scores in the third-person evaluation by non-authors indicated that blog readers would be more entertained if a blog had a haiku-generating module with the capability of reading poems, playing sounds and displaying season-related images which, using an actual example of a picture of a camisole (Figure 4) as a symbol of a summer word, can automatically change every time the search engine ranking is updated. As the next step in this research, we plan to make a more interactive version by devising a module with learning capabilities to experiment with crowd-based, half human, half machine-based art collaboration. We believe that our new approach targeted at Internet-savvy young generations of Japanese has the potential to entertain them and enrich their reading experience, while at the same time producing high-quality poetry that can be respected also by their parents and grandparents, who have much higher knowledge and wider experience in both reading classic haiku and creating their own. We also plan to experiment with different combinations of semantic differential evaluation in order to enlarge the still relatively small gap between random baseline poems and artificially created ones, although we realize that human beings treat randomness differently when labeled as "poetry". We believe that our work could spark an interesting multidisciplinary discussion on this topic.

Figure 4: An example of an image illustrating summer: "camisole".

Acknowledgments

We would like to thank Mr. Takuya Emori, our former student who was the main contributor for the presented work but never had a chance to share his research with a wider audience.

References

[Cohen, 1995] Harold Cohen. The further exploits of Aaron, painter. *Stanford Humanities Review*, 4(2):141–158, 1995.

[Colton *et al.*, 2012] Simon Colton, Jacob Goodwin, and Tony Veale. Full face poetry generation. In *Proceedings of the Third International Conference on Computational Creativity*, pages 95–102, 2012.

[Dybala *et al.*, 2008] Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. Humor prevails! – implementing a joke generator into a conversational system. In *AI 2008: Advances in Artificial Intelligence*, pages 214–225. Springer Berlin Heidelberg, 2008.

[Fukayama *et al.*, 2010] Satoru Fukayama, Kei Nakatsuma, Shinji Sako, Takuya Nishimoto, and Shigeki Sagayama. Automatic song composition from the lyrics exploiting prosody of the Japanese language. In *Proc. 7th Sound and Music Computing Conference (SMC)*, pages 299–302, 2010.

[Kaiser, 1958] Henry F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.

[Lindemeier *et al.*, 2013] Thomas Lindemeier, Sören Pirk, and Oliver Deussen. Image stylization with a painting machine using semantic hints. *Computers & Graphics*, 37(5):293–301, 2013.

[Luber, 1976] Raymond F. Luber. Evaluation of poetic mood with the semantic differential. *Psychological Reports*, 39(2):499–502, 2015/04/15 1976.

[Lutz, 1959] Theo Lutz. Stochastische texte. *Augenblick 4(1)*, 1959.

[Netzer *et al.*, 2009] Yael Netzer, David Gabay, Yoav Goldberg, and Michael Elhadad. Gaiku: Generating haiku with word associations norms. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, CALC '09, pages 32–39, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[Osgood *et al.*, 1957] C.E. Osgood, G.J. Suci, and P.H. Tannenbaum. *The Measurement of Meaning*. Illini books. University of Illinois Press, 1957.

[Tosa *et al.*, 2007] Naoko Tosa, Hideto Obara, Michihiko Minoh, and Seigow Matsuoka. Hitch haiku. In *Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts*, DIMEA '07, pages 6–7, New York, NY, USA, 2007. ACM.

[Tosa *et al.*, 2009] Naoko Tosa, Hideto Obara, and Michihiko Minoh. Hitch haiku: An interactive supporting system for composing haiku poem. In Scott M. Stevens and Shirley J. Saldamarco, editors, *Entertainment Computing - ICEC 2008*, volume 5309 of *Lecture Notes in Computer Science*, pages 209–216. Springer Berlin Heidelberg, 2009.

[Tsuji, 2001] Momoko Tsuji. *Hajimete-no haiku zukuri (Creating your first haiku, in Japanese)*. Nihon Bungeisha, 2001.

[Vala, 2011] Jaroslav Vala. Exact research on the reception of poetry. *Procedia - Social and Behavioral Sciences*, 29(0):520 – 527, 2011. The 2nd International Conference on Education and Educational Psychology 2011.

[Wong *et al.*, 2008] Martin Tsan Wong, Andy Hon Wai Chun, Qing Li, SY Chen, and Anping Xu. Automatic haiku generation using VSM. In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*. World Scientific and Engineering Academy and Society, 2008.

[y Pérez and Sharples, 2004] Rafael Pérez y Pérez and Mike Sharples. Three computer-based models of storytelling: BRUTUS, MINSTREL and MEXICA. *Knowledge-based systems*, 17(1):15–29, 2004.

[Yan *et al.*, 2013] Rui Yan, Han Jiang, Mirella Lapata, Shou-De Lin, Xueqiang Lv, and Xiaoming Li. I, Poet: Automatic Chinese poetry composition through a generative summarization framework under constrained optimization. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, pages 2197–2203. AAAI Press, 2013.

[Zhang and Lapata, 2014] Xingxing Zhang and Mirella Lapata. Chinese poetry generation with recurrent neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 670–680. Association for Computational Linguistics, 2014.