Insert a cool title here

Alham Fikri Aji
University of Edinburgh
s1447728@sms.ed.ac.uk

Abstract

ABSTRACT

1. Introduction

..

Automatic poetry generation one challenge in computational creativity. Computational creativity is a field to explore and utilise Artificial Intelligence (AI) to work in creative object domain such as song, performance, poetry, visual design, or other areas [1]. In such creative domain, emotional and intelligence aspect of the author often involved to achieve the

¡¿

2. Background

2.1. Haiku

Haiku is traditional Japanese poetry .

Rules of writing Haiku are clear in Japan. However, problem occurs when Haiku is adopted into different language [2]. The issue comes from different linguistic concepts and designs from Japanese to foreign language. Some Japanese language concepts simply does not exists in English. Cultural difference is also one reason of variety rules in foreign Haiku. Therefore, there is no strict rule of writing Haiku in English [3].

Haiku is three-parts short poem consists of 17 syllables. The syllables are further distributed into each part with ratio of 5, 7, 5. Therefore, the first and the last part have 5 syllables and the second part has 7 syllables. However, converting this rule from Japanese language which has monosyllabic phonetic system into English which has different phonetic system can sometimes be challenging [2]. Thus, 5,7,5 syllables distribution rule is still important in English Haiku, but not completely mandatory [3].

Romaji:

*yuku haru **ya***
tori naki uo no
me wa namida
(Bash, tr. Shirane)

Translation:

spring going–

birds crying and tears
in the eyes of fish

Haiku consists of two different sub-ideas, separated by *kireji*, or "cutting word". In Japanese, *kireji* is used to extend emotional context. It may be used as punctuation. As *kireji* does not exists in English, it is replaced by punctuation mark, such as dash or question mark. In some cases, *kireji* is simply unmarked and understood as implied delay [4]. In example above, *kireji* "ya" is often used to emphasises the previous word. In English, it is translated with long dash.

There is no specific genre in Haiku. Any subjects can be written in Haiku [3]. Haiku poem often can be ¡¿¡¿. This expression can be achieved by choosing words that relate with one of human senses such as visual, hearing, or touch. Moreover, Haiku uses *kigo*, or 'seasonal words' to refer and visualise the poem into one of the seasons in Japan. However, modern Haiku may do not contains any seasonal word. [3]. It is also important to collaborate the five senses perception with correct seasonal theme [5].

The underlined words in

2.2. Automated Poetry Generation

3. Resources

3.1. WordNet

WordNet is huge English lexical database that focuses on relationships between the meaning and the concept of words. WordNet groups nouns, verb, adjectives, and adverbs based on their meaning and usage into set called sysnet. Words in a sysnet shares same concept and can be substituted by each other in most cases. WordNet has about 117.000 sysnets covering broad topics.

Relationship between sysnets are defined by linking them with conceptual relational attribute. This relational

3.2. Standford PoS Tagger

Standford implement

3.3. CMU Pronouncing Dictionary

Carnegie Mellon University Pronouncing Dictionary (CMUPD) provides machine-readable pronunciation dictionary for over 133.000 words. It uses 39 set of phonemes based on ARPA-bet symbol set. Each vowel contains lexical stress pattern information, further categorised into three different levels: 0 for no stress, 1 for primary stress, and 2 for secondary stress[6]. The stress pattern and pronunciation are useful to analyse the beautifulness of a poetry.

Bartlett et al. further improve the CMUPD and add additional syllables information. This syllabified CMUPD provides syllable boundaries that splits between one syllable to another [7]. This information can be used to compute the total syllables of given word which is invaluable resource for fitting the syllable constraints in Haiku. Table .. shows some examples of words and their corresponding pronunciation data.

Word	Pronunciation
Haiku	HH AY1 - K UW0
Poetry	P OW1 - AH0 - T R IY0
Thesis	TH IY1 - S AH0 S0

Table 1: Examples of Syllabified CMU Pronunciation Dictionary

4. Project Specification

In this project, we aimed to build automated Haiku generator. This Haiku generator will takes input of one or some topic keywords from user, and randomly generates Haiku that related with the keyword.This keyword is not mandatory, topic will be randomly chosen in case when user does not provide the input.

The generated Haiku will follow English Haiku rules as follow:

1. It consist of three lines with 5,7,5 syllables distribution
2. Using *kireji* is not necessary
3. Using seasonal words is preferable, but not required.
4. Does not consider about rhymes

However, as there is no strict rule of writing English Haiku, we allow user to customize the output Haiku structure. Therefore, we put the rules above as default rule and user may change it preferences as desired.

This project will be implemented in Java, thus we expect that it should run in most standard computers. A Java applet of the project will be developed and further deployed on web platform.

5. Methodology

The brief concept of implementing Haiku generator is by learning the grammatical pattern and rules of existing Haiku poems and build a new grammar skeleton based on it. The generator will select words to fill-in the grammar skeleton based on statistical approach. Other resources such as CMUPD and WordNet will be supporting the decision making of words selection.

5.1. PoS Tag Extraction

Learning grammar structure is done by applying Part of Speech (PoS) tagger algorithm for every collected Haiku poems. PoS tagger gives tag for each word in the input sentence. Therefore, in this phase we aim to acquire grammar tag pattern of the We are only interested in the grammar tag, therefore we can remove the Haiku after tagging. One simple illustration of this process is shown below:

The reason of choosing this method over using grammar tree or context free grammar is because sometimes poetry does not follow standard grammar rules. By using formal grammar rule, we might achieve less poetic result. By learning from existing Haiku, we hope that we are able to capture the grammatical pattern used in Haiku.

Haiku	Extracted PoS Tag Information
Three strokes of lightning	CD NNS IN NN
One hit mountain frightening	CD NN NN JJ
Dark clouds thunder loud	JJ NNS NN RB

Pattern 1	Pattern 2	Pattern 3
CD NNS IN NN	DT NNS IN NNP	JJ JJ NN
CD NN NN JJ	JJ NN CC JJ NNS	VBG IN DT NN IN NN
JJ NNS NN RB	VBG DT NN	NN IN

Generated Skeleton
CD NN NN JJ
JJ NNS NN RB

5.2. Creating Haiku: Grammar Skeleton

Grammar skeleton is sequence of tags that will be used to create a Haiku. The generated Haiku should follow the grammar tag defined by the skeleton. In this phase, grammar skeleton is created by combining several tag patterns in previous part.

The/ leaves/NNS of/IN Autumn/NNP lovely/JJ gold/NN and/CC brown/JJ colors/NNS painting/VBG the/DT landscape/NN

Illustration of creating a grammar skeleton based on tag data is shown above.

5.3. Creating Haiku: Word Filling

5.4. Creating Haiku: Poetic Scoring

6. Evaluation

References

[1] Simon Colton, Geraint A Wiggins, et al. Computational creativity: the final frontier? In *ECAI*, pages 21–26, 2012.

[2] Kei Grieg Toyomasu. Haiku for people. URL `http://www.toyomasu.com/haiku/`.

[3] Haiku in english. URL `https://www.ryukoku.ac.jp/haiku/haiku_en.html`.

[4] Ashley Capes. An introduction to haiku - form and structure: Kireji and kigo. URL `http://ashleycapes.com/2013/08/26/an-introduction-to-haiku-form-structure-kireji-kigo/`.

[5] Michael Dylan Welch. Becoming a haiku poet. URL `http://www.haikuworld.org/begin/mdwelch.apr2003.html`.

[6] Carnegie Mellon University. The cmu pronouncing dictionary. URL `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

[7] Susan Bartlett, Grzegorz Kondrak, and Colin Cherry. On the syllabification of phonemes. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 308–316. Association for Computational Linguistics, 2009.