Lexical Modeling of *Yamabuki* (Japanese Kerria) in Classical Japanese Poetry

Hilofumi Yamamoto*

Tokyo Institute of Technology
University of California, San Diego

Keywords: corpus linguistics, co-occurrence weight, visualization, Japanese literature, network modeling

Abstract

This project is a lexical study of classical Japanese poetic vocabulary through network analysis. The analysis is based on co-occurrence patterns, defined as any two words appearing in a poem. We developed the corpora of classical Japanese poetry based on the eight anthologies compiled under imperial order called the Hachidaishū which were established from ca. 905 to 1205. The co-occurrence weighting, *cw*, allows us to examine the patterns of poetic word constructions through mathematical modeling(Yamamoto, 2006). As a result, we could in general observe a main hub node derived from a topic word. We also encountered other hub nodes which do not indicate topic words nor entry items in a poetic dictionary. We conclude that a term such as *yahe* can be shown as a hub node to connect a topic word with other peripheral words, and plays a supporting role to form a poetic story in the poem even if it is not included in a dictionary for Japanese poetic words.

1 Introduction

This project is a lexical study of classical Japanese poetic vocabulary through network analysis based on graph theory. The analysis is mainly conducted with co-occurrence patterns defined as any two words appearing in a poem.

Many scholars of classical Japanese poetry have tried to explain the constructions of poetic vocabulary relied on their sensitivity and experience. As scholars can only demonstrate constructions that they can consciously point out, those that they are unconscious of will never be demonstrated. Generally, when writing a dictionary, since the writer of dictionary picks up only what it is conscious, important knowledge in the subconscious mind of researchers is not stated. In order to describe what is in the subconscious mind of humans for the development of dictionary, it is important to adopt a computer-assisted description that does not depend on researcher's intuition, sensitivity, consciousness. To this end, we will use co-occurrence weighting methods on corpora of classical Japanese poetry for the selection of lexical items to describe in a dictionary.

2 The flowers of *Yamabuki* (Japanese kerria)

Three words, *yamabuki* (Japanese kerria), *kahazu* (Frog), *Ide* (a proper noun which indicates a placename in Kyoto), are frequently used together in waka and haiku, which are closely related to each other. Not only in poetry, they are used together in Japanese art as well such as in *ukiyoe* (Figure 1 and 2). In the

* yamagen@ila.titech.ac.jp: W1-8 2-12-1 O-okayama, Meguro-ku, Tokyo, 152-8550, Japan

context of Japanese culture, frog and Ide are always used at the same time when *Yamabuki* is used. If this relationship is common sense of Japanese culture, it will be analyzed through poetic texts and three word relation will be drawn by visualization based on graph theory.

Fig. 1 *Yamabuki-to-kahazu* by Hiroshige Utagawa. (`http://www.gekkanbijutsu.co.jp/shop/goods/030761011.htm`)

Fig. 2 *Yamashironokuni Ide no Tamagawa* by Kuniyoshi Utagawa in 1847. (`https://ukiyo-e.org/image/metro/5245-006-01(01)`

3 Methods

We use the Hachidaishū as a material of the present study, which is the eight anthologies compiled by the order of Emperors (ca. 905–1205) and contains about 9,500 poems. We developed the corpora of it and a method of co-occurrence weighting, *cw* (Yamamoto, 2006) which calculates the weight of patterns of any two words appearing in a poem sentence similar to the $tfidf$ method(Spärck Jones, 1972; Robertson, 2004; Manning and Schütze, 1999).

$$
\begin{aligned}
w(t,d) &= (1+\log\ tf(t,d)) \cdot idf(t) \\
cw(t_1,t_2,d) &= (1+\log\ ctf(t_1,t_2,d)) \cdot cidf(t_1,t_2) \\
cidf(t_1,t_2) &= \sqrt{idf(t_1) \cdot idf(t_2)} \\
idf(t) &= \log \frac{N}{df(t)}
\end{aligned}
$$

Where, w is weight, t is a token, N is the number of tokens. The function, idf, is called "inverse document freuency."(Spärck Jones, 1972; Robertson, 2004; Manning and Schütze, 1999) The function cw is called "co-occurrence weight," which allows us to examine the patterns of poetic word constructions through mathematical modeling.

4 Results

As a result, when we draw a network model from co-occurrence patterns, we in general observe a main hub node derived from a topic word. We encountered other hub nodes which do not indicate topic words nor words which we generally see in a poetic dictionary as well. we took *yamabuki* (Japanese kerria) as a topic word and draw its network model. we could observe *kahazu* (frog), *Ide* (place name, proper name) as hub nodes as we expected. Not only those but also *yahe* (eightfold or double flower) could be seen as a hub node. It is never described in any poetic dictionary of classical Japanese.

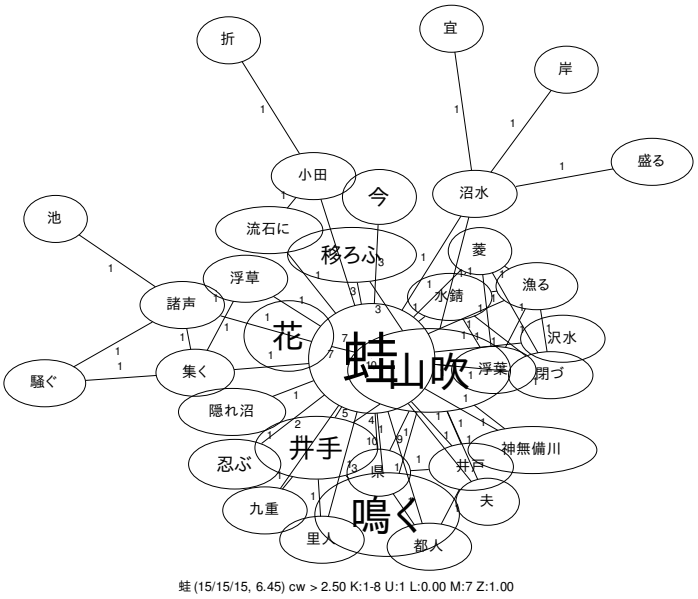


Fig. 3 Graph model of *kahazu* (蛙, frog) before pruning node 蛙.

A minor term *yahe* (eightfold) could be shown as a hub node which plays a major role in connecting a topic word with other peripheral words which support/demonstrate poem stories. These minor words are not seen in poetic term dictionaries.

'*Ide*', a place name in Tsuzuki-gun, Kyoto-fu, is famous for its *kahazu*(Tani, 2006, 25).

In Heian period (795-1185), almost all songs including *kahazu* (flog) are composed with using terms *Ide* and *yamabuki* (a flower, Japanese Kerria) such as in the following song:

> Kahazu naku / Ide no yamabuki / chirinikeri / hana no sakari ni / ahamashi monowo
> (where flogs are crying / flowwers of Japanese kerria in Ide / have fallen / when the flowers fully bloom / I wanted to be and see it)

As we have seen in the explanation of 'Ide', "Ide no Tamagawa" locates at Yamashiro no kuni which is currently called as Ide cho, Kyoto-fu. Kerria originally had been planted in the precincts of Itsutsumidera temple whose owner was Tachibana clan. Because of it, Ide became famous for its kerria.

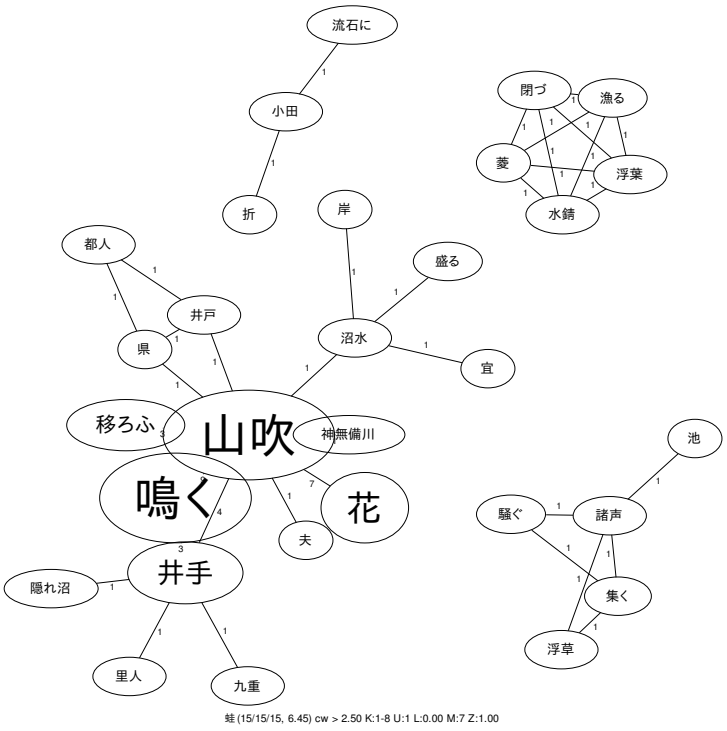


Fig. 4 Graph model of *kahazu* (蛙, frog) after pruning node 蛙.

5 Discussions

The terms *yamabuki*, *kahazu*, and *Ide* are contained in some poetic dictionaries as entry items or collocations. The term *yahe* is, however, not seen in any poetic dictionaries even as a single term.

6 Conclusion

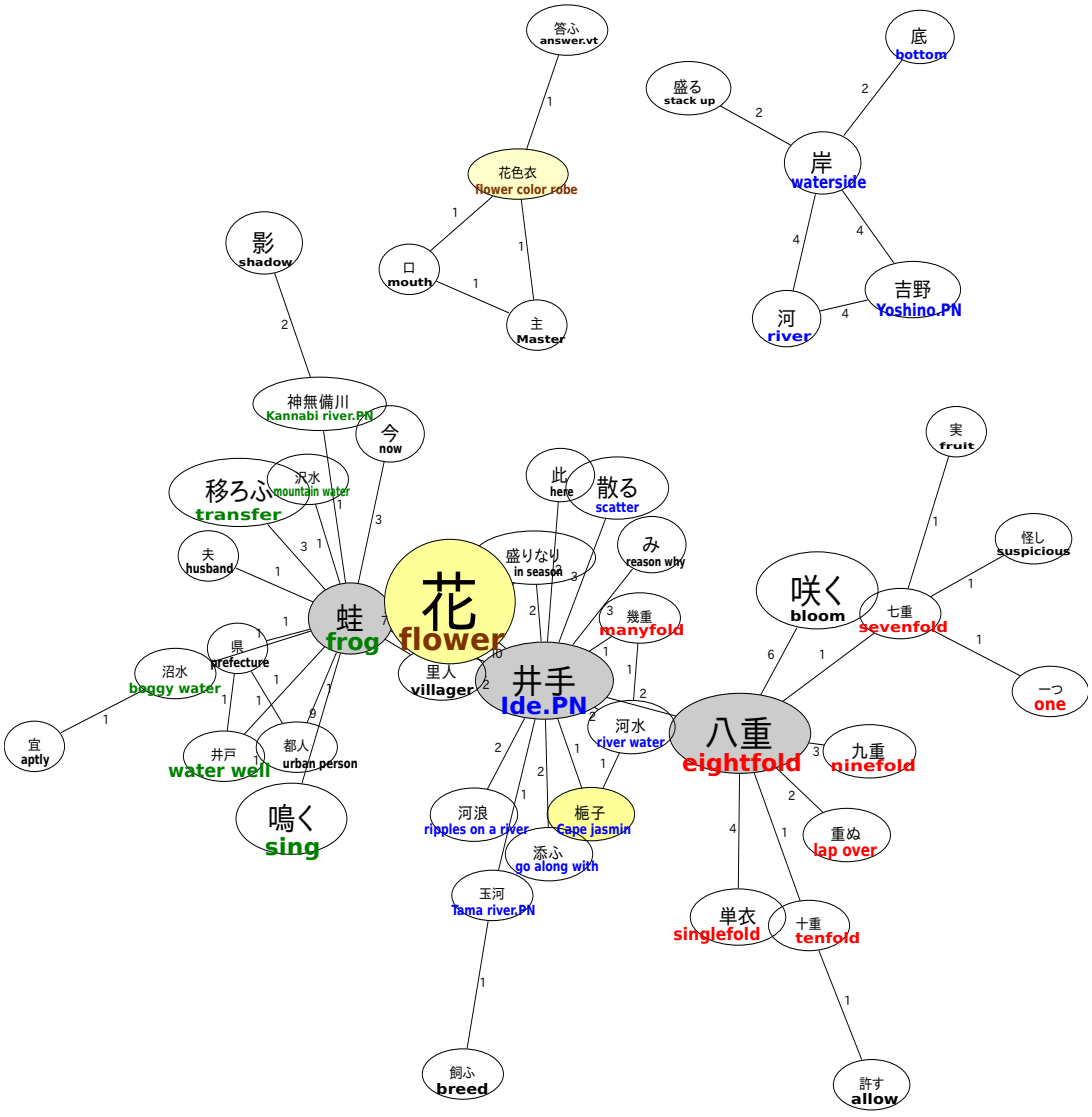
We conclude that a term such as *yahe* can be shown as a hub node which takes an important role to connect a topic word with other peripheral words such as *kokonohe*, *nanahe*, *hitohe*, and plays a supporting role to form a poetic story in the poem even if it is not included in a dictionary.

The finding of this study is that the modeling developed here allows us to 1) discern not only patterns described by experts but also patterns yet undescribed, and 2) identify not only specific or tangible words but also abstract or conceptual words which have a tendency to be left out of dictionaries.

References

Manning, Christopher D. and Hinrich Schütze (1999) *Foundation of statistical natural language processing*, Cambridge, Massachusetts: The MIT press.

Robertson, Stephen (2004) "Understanding inverse document frequency: on theoretical arguments for IDF", *Journal of Documentation*, Vol. 60, pp. 503–520.



山吹 (44/44/44, 5.37) cw > 2.50 K:1-8 U:1 L:0.00 M:7 Z:1.00

Fig. 5 A graph model of *Yamabuki*: a core node, 山吹 *yamabuki*, is pruned. *kahazu* (蛙, frog), *Ide* (井手, place name, proper name), and *yahe* (八重, eightfold or double flower) are observed as hub nodes.

Fig. 6 Variations of Japanese kerria (yamabuki): Single petal (left), white petal (center), and plena petal (right) of *yamabuki*. (`http://mkfarm.blog118.fc2.com/blog-entry-27.html`)

Spärck Jones, Karen (1972) “A Statistical Interpretation of Term Specificity and Its Application in Retrieval”, *Journal of Documentation*, Vol. 28, pp. 11–21.

Tani, Tomoko (2006) *Wakabungaku no kisochishiki*, Kadokawa sensho: Kadokawa Gakugei Shuppan.

Yamamoto, Hilofumi (2006) “Konpyūta niyoru utamakura no bunseki / A Computer Analysis of Place Names in Classical Japanese Poetry”, in *Atti del Terzo Convegno di Linguistica e Didattica Della Lingua Giapponese, Roma 2005*: Associazione Italiana Didattica Lingua Giapponese (AIDLG), pp. 373–382.