

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

Hilofumi Yamamoto*

Tokyo Institute of Technology
University of California, San Diego
yamagen@ryu.titech.ac.jp

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 based on graph theory. The analysis is based on 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 based on their intuition and experience. As scholars can only demonstrate constructions that they can consciously point out, those that they are unconscious of will never be demonstrated. When we develop a dictionary of poetic vocabulary using only our intuitive knowledge, the description will lack important lexical constructions. In order to conduct more exact and unbiased descriptions, it is necessary to use computer-assisted descriptions of poetic word constructions using co-occurrence weighting methods on corpora of classical Japanese poetry.

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. We also developed a method of co-occurrence weighting (Yamamoto, 2006) which calculates the weight of patterns of any two words appearing in a poem sentence similar to the *tf-idf* method (Spärck Jones, 1972; Robertson, 2004; Manning and Schütze, 1999). The CW allows us to examine the patterns of poetic word constructions through mathematical models.

As a result, when we draw a network model from co-occurrence patterns, we can in general observe a main hub node derived from a topic word. Additionally, we also encounter other hub nodes which do not indicate topic words nor entry items in a poetic dictionary. For instance, when we take *yamabuki* (Japanese kerria) as a topic word and draw its network model, we will observe *kahazu* (frog), *Ide* (place name, proper name), and *yahe* (eightfold or double flower) as hub nodes. 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. We conclude that a term such

*TokyoTech: 2-12-1 O-okayama, Meguro-ku, Tokyo, 152-8550, Japan

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.

1 Introduction

This project is a lexical study of classical Japanese poetic vocabulary through network analysis based on graph theory. The analysis is based on 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 based on their intuition and experience. As scholars can only demonstrate constructions that they can consciously point out, those that they are unconscious of will never be demonstrated. When we develop a dictionary of poetic vocabulary using only our intuitive knowledge, the description will lack important lexical constructions. In order to conduct more exact and unbiased descriptions, it is necessary to use computer-assisted descriptions of poetic word constructions using co-occurrence weighting methods on corpora of classical Japanese poetry.

The relationship between *Kerria*, *Kahazu*, and *Ide* is shown in Figure 1 ¹

2 Methods

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. We also developed a method of co-occurrence weighting (Yamamoto, 2006) which calculates the weight of patterns of any two words appearing in a poem sentence similar to the *tf-idf* method (Spärck Jones, 1972; Robertson, 2004; Manning and Schütze, 1999). The CW allows us to examine the patterns of poetic word constructions through mathematical models.

3 Results

As a result, when we draw a network model from co-occurrence patterns, we can in general observe a main hub node derived from a topic word. Additionally, we also encounter other hub nodes which do not indicate topic words nor entry items in a poetic dictionary. For instance, when we take *yamabuki* (Japanese *kerria*) as a topic word and draw its network model, we will observe *kahazu* (frog), *Ide* (place name, proper name), and *yahe* (eightfold or double flower) as hub nodes.

能因はうやうやく懐から錦の小袋を出します。その中に入っていたのは、なんと鮑屑！普通ならゴミとして捨てられるものでしょう。しかし、それは、長柄

¹<http://www.gekkanbijutsu.co.jp/shop/goods/030761011.htm>.



Figure 1: The picture of “Yamabuki To Kahazu” (kerria and frog) by Hiroshige Utagawa.

の橋（淀川の支流長柄川にかかる橋、有名な歌枕）を造ったときの鮑屑だというのです。しかし、相手もさる者、懐中から包み紙を取り出し、ひからびた蛙を披露します。これもまた有名な「井手の蛙」（井手は京都府綴喜郡井手町、山吹と蛙で有名な歌枕）だと言うのです。鮑屑とひからびた蛙を秘宝として携帯し、見せびらかしあう姿は、一般の人から見れば、奇人変人以外の何者でもないでしょう。しかし、これがまさに数奇の真骨頂なのです。（和歌文学の基礎知識、谷知子 p.25）

（参考）古今和歌集仮名序相当部分やまと歌は、人の心を種として、万の言の葉とぞ成れりける。世の中にある人、事・業しげきものなれば、心に思ふ事を、見るもの聞くものに付けて、言ひ出せるなり。花に鳴く鶯、水に住む蛙の声を聞けば、生きとし生けるもの、いづれか歌を詠まざりける。力をも入れずして天地を動かし、日に見えぬ鬼神をもあはれと思はせ、男女の仲をも和らげ、猛き武人の心をも慰むるは歌なり。

平安時代に入ると、「かはづ」をよんだ歌のほとんどが「かはづなく井手の山吹散りにけり花のさかりにあはましものを」（古今集・春下・読人不知）「音に聞く井手の山吹見つれどもかはづの声はかはらざりけり」（貫之集）のように山吹の花とともによまれるようになり、しかもそのほとんどが「山城の井手（京都府綴喜郡）の山吹」という形でよまれているのである。山城の井手は『伊勢物語』に「山城の井手の玉水...」とあるように清流で知られていたから、ここに住んでいる「かはづ」も、やはり河鹿であろう。普通の蛙であれば名所になるはずもない。鴨長明の『無名抄』には、「世の人の思ひてはべるは、ただ蛙をば皆かはづといふぞと思へり。それも違ひはべらねどかへるも、かはづと申す蛙は、ほかにはさらに侍らず。ただ井手の川にのみ侍るなり。をどあり色黒きやうにて、いと大きに

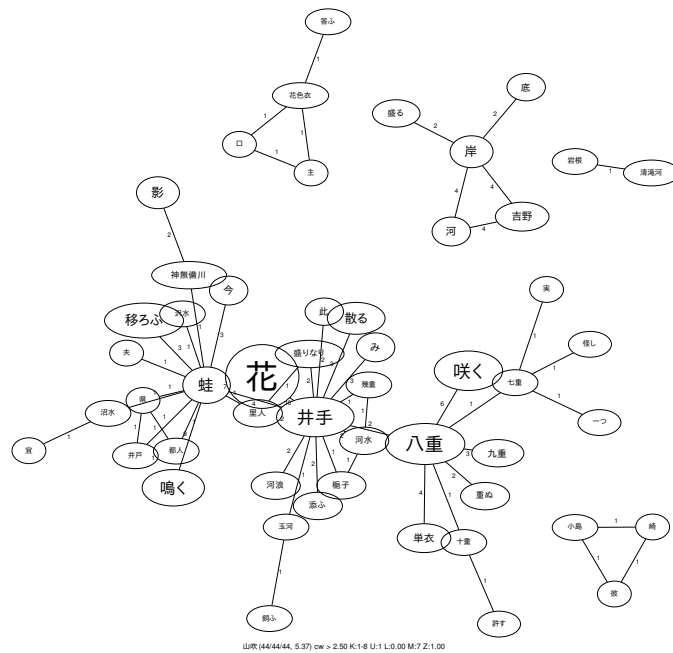


Figure 2: Graph model of *Yamabuki*: a core node, 山吹 *yamabuki*, is pruned. *kahazu* (蛙, frog), *Ide* (井手, place name, proper name), and *yahe* (八重, eightfold or double flower) are observed as hub nodes.

もあらず。世の常の蛙のやうにあらはに跳り歩くことも、いとせず。常に水にのみ住みて、夜ふくるほどにかれが鳴きたるは、いみじく心澄み、物哀れなる声になむはべる」とある。なお、『伊勢物語』百八段の「よひごとにかはづのあまた鳴く田には水こそまされ雨は降らねど」は「田の蛙」であり河鹿ではなからう。蛙を「かはづ」と呼ぶことも、やはりこのようにあったのである。(歌ことば辞典、片桐洋一、p.112)

「井手の玉川」は「井手」の項にも述べたように山城国、今の京都府綜昏郡井手町。相楽郡和束の山中から発してこの井手を通り木津川にそそぐ川が玉川である。「かはづなく井出の山吹散りにけり花のさかりにあはましものを」(古今集・春下・読人不知)の影響が強く、「駒とめてなほ水かはむ山吹の花の露そふ井手の玉川」(新古今集・春下・俊成)のほか、「井手」の語がなくても「玉川の岸の山吹影見えて色なる浪に蛙鳴くなり」(後鳥羽院集)など「山吹」や「蛙」がよまれていれば、この「井手の玉川」である。また『大和物語』一六九段に見える三輪大社へ使いに行った男がこの地で見たかわいい幼女に思いを寄せ、大きくなったら迎えに来ると言って帯を取り替えたという話によった「とりかへし井手の下帯ゆきめぐりあふせうれしき玉川の水」(長秋詠藻)の「玉川」もこの「井手の玉川」である。(歌ことば辞典、片桐洋一、p.263)

4 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.

5 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.
- Spärck Jones, Karen (1972) “A Statistical Interpretation of Term Specificity and Its Application in Retrieval”, *Journal of Documentation*, Vol. 28, pp. 11–21.
- 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.