T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	「山吹」をめぐる和歌語彙の空間
Title(English)	Lexical Space of Yamabuki (Japanese Kerria)
著者(和文)	山元啓史
Authors(English)	Hilofumi Yamamoto
出典(和文)	人文科学とコンピュータシンポジウム論文集, Vol. 2011, No. 8, pp. 141-146
Citation(English)	IPSJ Symposium Series, Vol. 2011, No. 8, pp. 141-146
発行日 / Pub. date	2011, 12
権利情報 / Copyright	ここに掲載した著作物の利用に関する注意: 本著作物の著作権は（社）情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof.

Powered by T2R2 (Tokyo Institute Research Repository)

「山吹」をめぐる和歌語彙の空間

山元啓史

東京工業大学大学院社会理工学研究科

要旨

本研究では和歌の共出現パターン（一首に見られる任意２語の組み合わせ）を利用して、語彙ネットワークを描き、和歌用語がどのような集合を形成しているのかを分析している。ある用語を含む和歌を材料に、この語彙ネットワークを描画したところ、通常は、ある用語をハブ（中心となる語）としてにネットワークが描かれるが、時々、見出し語としては扱われにくいような語や一見トピックにあまり深く関連していないと思われるような語が、ハブとして描かれることがある。たとえば、「山吹」をトピックとしてネットワークを描画したところ、「蛙」「井出」「八重」の語が「山吹」に伴うハブとして描かれた。「山吹」と「蛙」「井出」は、歌ことば辞典でともに詠み込まれる用語として解説されているが、「八重」は用語集や辞典では、それ単独としても取り上げられていない。「八重」のように、あまり辞典の見出し語としては取り扱われない語であってもネットワークによる分析において、ハブとして存在し、語と語の関係を表すのに、重要な役割を担う語であることがわかった。

Lexical Space of *Yamabuki* (Japanese Kerria)

Hilofumi Yamamoto

Graduate School of Decision Science and Technology, Tokyo Institute of Technology

Abstract

This project is a lexical study of classical Japanese poetic vocabulary through network analysis based on graph theory. The analysis is based on cooccurrence patterns, defined as any two words appearing in the same poem. When we draw a network model from cooccurrence patterns, we can in general observe a main hub node derived from a topic word. 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 the network models of it, 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. I conclude that a term such as *yahe* can be shown as a hub node which takes a role in connecting 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.

(c) Information Processing Society of Japan

1　はじめに

語と語の関連性を調査するにはどのような方法を用いるとよいのだろうか。和歌や俳句の専門家に「○○はどんな語とともに使われるか」と尋ねてもよいが、その際、専門家の口からは出てきた語は、想起できた語であって、想起されなかった語は仮に重要であったとしても指摘されずに見逃されている可能性がある。従来、大切だと考えられる語（想起できる語）を見出し語として辞書・辞典の編纂が行われてきているが、この方法では想起できない語は、辞書に記載されないことになってしまう。つまり、口頭で指摘できない限り、その事実があったとしても、指摘できないことになってしまう (Wittgenstein, 1953)。このように考えると、辞書・辞典に記述される内容は「編纂者が指摘できる語」という前提で集められていることになる。本論では、見出し語として扱われにくい語、一見トピックに関連していないと思われるような語、単独ではあまり目立たないが語と語の関係に重要な役割を担っている語が、語彙ネットワークの描画により指摘できることを示す。

2　「山吹」をめぐって

人間の意識はトピックに関係の密な語、意味を大きく担っている語に向けられる。トピックとは必ずしも関連していない語、たとえば、抽象的な語、一般的な語は、語と語の関係を取り持つのに重要な役割を担っていても、意識的に注目しない限り、辞書の見出しに扱われることはない。

本稿では「山吹」をトピックとして取り扱う。「山吹」は八代集の中では、表 1 に示す頻度で出現する。「桜」「梅」「橘」のように決して多くはないが、たまに詠まれるというほど少ない語でもない。計量的な分析としては単純に表 1 のように、その頻度を数えて終わりであり、量的特徴も質的特徴も「山吹」という語のみでは何も語れない。そこで、「山吹」をめぐる空間に視野を広げ、そこにどのような質的な特徴があるのかを分析する。

たとえば、万葉集 1435 には次のような歌がある。

蝦鳴く　甘南備河に　かげ見えて　今か咲
くらむ　山吹の花　（厚見王）

このように「山吹」と「蝦（蛙）」の関係が詠まれている。また、

表 1: 八代集での「山吹」の出現頻度

勅撰集名	頻度
古今集	6
後撰集	4
拾遺集	6
後拾遺集	6
金葉集	5
詞花集	3
千載集	7
新古今集	7
計	44

古池や蛙飛びこむ水の音　芭蕉

は、芭蕉の有名な俳句であるが、これは当初「蛙飛びこむ水の音」が先にできていて、発句の５文字をどうするかという時、芭蕉の弟子の其角が、「山吹や蛙飛びこむ水の音」と詠んだということである。

広重の絵にも「山吹に蛙」という作品があり（図 1)、詩歌だけでなく、広く日本の伝統文化に強い影響を与えていることがわかる。

図 1: 歌川広重「山吹に蛙」の絵: 絵の中に見られる和歌は「春雨のふる日なからも山吹のみのひとつたになく蛙かな—長亭」http://www.gekkanbijutsu.co.jp/shop/goods/030761011.htm。

当然ながら歌枕、歌語、歌ことばの解説書にも「山吹」「蛙」の関係性は詳しく述べられている。「山吹」は「映ろふ」と「移ろふ」が掛けて詠まれたり、「井手」とともに詠まれたりしている (片桐, 1983, pp. 439–40)。

「蛙」は平安時代頃から「蛙」を詠んだ歌のほとんどが「山吹の花」とともに詠まれるようになり、

(c) Information Processing Society of Japan

しかもそのほとんどが「山城の井手（京都府綴喜郡）の山吹」という形で詠まれている (片桐, 1983, pp. 111–2)。

「山吹」「蛙」とともに出てくる「井手」は、「いで〔ゐで〕【井手】京都府南部、綴喜 (つづき) 郡の地名。左大臣橘諸兄（たちばなのもろえ）が別荘を置いた所。西流する玉川は山吹と蛙の名所として知られた。［歌枕］（デジタル大辞林）[1] 」とあり、この「井手の玉川」の前の「井出」だけを詠んで、後の「玉川」を省略することも多く、これらはいずれも地名としての「井手」を指す。ただし、「田の用水として、水の流れをせき止めてためてある所。井堰 (いせき)」という説明もあり（同ページ）、この場合は一般名詞ではあるが、大部分は前者の「山吹と蛙の名所」の「井手」ということになる。

以上のように「山吹」が詠まれる時には「蛙」「井手」「映ろふ」が一緒に詠まれるというのが通説のようである。これに対し、本論では計算手続きによって「山吹」をめぐる語彙の空間を導き出し、上記と一致する関係が見られるか、あるいは上記には見られない重要な関係が指摘できるかということについて検討を行う。

3 方法

国文学研究資料館の二十一代集 CDROM 版より八代集（古今集、後撰集、拾遺集、後拾遺集、金葉集、詞花集、千載集、新古今集）収録の和歌 9503 首を用いる。それぞれの和歌は、古文品詞タグつけシステム `kh` [2]で単位分割し、`t2c` [3]を使って、シソーラスコードに変換し、表記の正規化を行った。それらの和歌の各首に見られる任意２語パターンを用いて、共出現パターンを作成した。

共出現とは「一つの表現域に見出語 i も j も使ってあれば、その i と j とは共出現していると言う (水谷, 1983, p.121)」と定義されている。和歌のテキストの場合、一首に見られる２語の語順や隔たりを問わない組合せを１パターンとする。たとえば、ある歌 s_1 の中に「山吹」と「蛙」が含まれていれば、「山吹–蛙」を１パターンとする。この時、順序（どちらが前か）についても隔たり（２語間は何単語隔てて存在するか）についても問われない。どちらが先にあっても、「山吹」と「蛙」の語の間隔が離れていても、「山吹–蛙」を共出現パターンとする。この「山吹–蛙」のパターンが上記のように２首に見られれば、パターン頻度 C(山吹, 蛙) を２とする。筆者は、このように順序も隔たりも無視し、２語の関係のまったく統語的関係も考慮しないものを「共起」ではなく「共出現」と呼んでいる。

作成された共出現パターンをそのままグラフで描くと重要なパターンもそうでないパターンもすべて含まれてしまうので、何らかのパターンに対する重み付け処理を行う。そこで、テキスト群 (d) において任意の１語 (t) が特徴的であるかを評価する式 *tfidf* (1) (Manning and Schütze, 1999) を利用する。式 (1) は、１語の重要度の評価式であるが、これを任意の２語のパターン (t_1, t_2) がどの程度特徴的であるのかを評価する式 (2) に拡張し、パターンの重み (cw) を計算する[4]。

$$w(t,d) = (1+\log\ tf(t,d)) \cdot idf(t) \quad (1)$$

$$cw(t_1,t_2,d) = (1+\log\ ctf(t_1,t_2,d)) \cdot cidf(t_1,t_2) \quad (2)$$

$$cidf(t_1,t_2) = \sqrt{idf(t_1) \cdot idf(t_2)} \quad (3)$$

ただし、(2) の前半は t_1 と t_2 の２語が共出現した時のテキストの数。(2) の後半 $cidf(t_1,t_2)$ は、(1) の $idf(t)$ を拡張し、２語の idf 値の幾何平均 (3) としたものである。以上で得られた cw 値を相互に比較できるように、標準得点に変換し、正規化を行い、１σ以上の共出現パターンを Graphviz [5]で描く。

4 結果

図 2, 3 は共に「蛙」のモデルである。図２は用語「蛙」に関わるすべてのノードとエッジを削除する前の図である。ほとんどすべてのノードが「蛙」と強く関係しているために、「蛙」以外の語との関係が見えない。そこで、すべてのノードは「蛙」と強い関係があることを前提にノード「蛙」を削除する (図 3)。これをプルーニング（枝の刈り込み）という。こうすることにより、「蛙」以外の語の関係の視界がよくなる。「蛙」から見ると、「山吹」「井手」との関係が見え、それらを中心にさまざまな語が関係しているのがわかる。

[1] `http://kotobank.jp/word/`井手

[2] `kh` は kobun to kana converter の略。2010 年頃から古代語に単位分割するプログラムや辞書は公開されはじめたが、筆者が古代語に関する一連の研究をはじめた当時にはそれらはなかった (山元, 2007b)。

[3] t2c: Token to Code。単位切りした語に分類語彙表準拠のシソーラス体系コードを追加するプログラム (山元, 2009)。

[4] 詳しい説明は山元 (2007a) に記した。

[5] `http://www.graphviz.org/`

(c) Information Processing Society of Japan

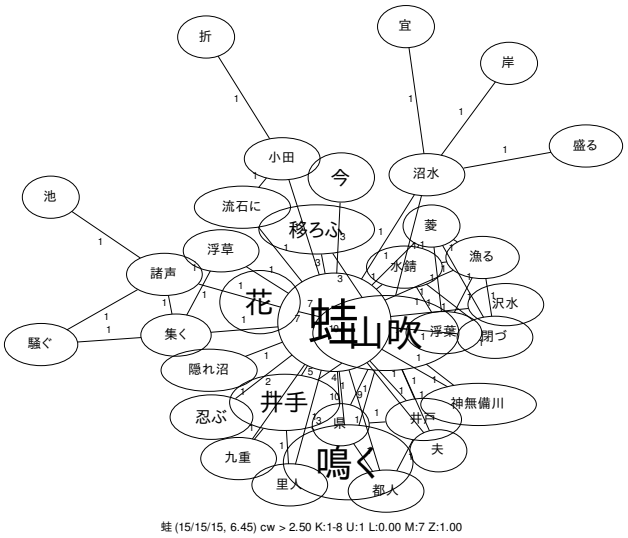

図 2: グラフモデル「蛙」プルーニング前: 古今集から新古今集までに「蛙」は 15 首に出現する。「蛙」自身の *idf* 値は、6.45 で、ネットワークは、共出現ウエイト *cw* は 2.50 以上を描いた。コアノード「蛙」を中心に放射状にエッジが集まっているため、ノード同士の接続の構造がよく見えない（これをスポーク現象と呼ぶ）。

一方、図 4 は「山吹」を用語とする和歌を用いて作られたモデルである。「蛙」の図と同様に、視界をよくするためにノード「山吹」をすでに削除してある。「蛙」の図では、「山吹」「井手」の用語が見られ、それらから枝分かれしているのが見られたが、「山吹」の図では、「蛙」「井手」「八重」の用語が見られ、それらからさらに枝分かれしているのが見られた。

5　考察

「山吹」には、大きく「蛙」「井手」「八重」の３つのノードが連なっていることがわかった。このように「山吹」と「蛙」は強く結びついているのは、文献が指摘するとおりであるが、「山吹」は「八重」とも強く結びついていることがわかった。ところが、この「八重」ということばは歌ことば辞典ではあまり述べられていない。また「八重」と「山吹」の関係についてはまったく指摘されていない。具体的にいえば、片桐 (1983) では「やへむぐら（八重葎）」しか見出しとして扱われていない。確かに八代集データベース (山元, 2009) で「八重」を検索してみると、「八重（やへ）」「八重葎（やへむぐら）」「八重桜（やへざくら）」「八重雲（やへくも）」「八重菊（やへぎく）」は出てくるが、「八重山吹」は出てこない。

古語辞典において「八重」は「やへ【八重】(1) 幾重にもたくさん重なっていること。『臣の子の—の紐解く』紀歌謡 127 (2) 八枚重なること。『紅梅の御衣をぞ—にも過ぎて幾つとなく奉りたる上に』〈栄花菩花〉」とあり (大野他, 1990, p. 1351)、「山吹」をとりたてて、形容するためのことばではないような説明である。実は「山吹」の花には２種類あって、一重のものと八重のものがあるのだが、その解説は「山吹」の歌ことば関連の辞典には見られない。

一方、モデル上では「蛙」「井手」「八重」の３つのハブ（それぞれの語が放射状に他の語と関連付けられているノード）が示されている。単純に辞典作りの感覚から言えば、「蛙」「井手」は、直示的定義[6]ができるのに対し、「八重」は、具体的事物を伴った上で、その何かが「八重」であるというように事物の様子を述べる語である。それだけでは、直接的に具体物を示す語でないことは明らかだ。しかし、この図のように明示的に「八重」がハブとして示されると「八重山吹」が和歌で詠まれる「山吹」であることは明らかである。そのことを歌ことばの専門家が知らないはずはないと思われるが、歌ことば辞典

[6]直示的定義（ちょくじてきていぎ、英:ostensive definition）とは、実例を示して言葉の意味を伝える方法。定義法のひとつ。

(c) Information Processing Society of Japan

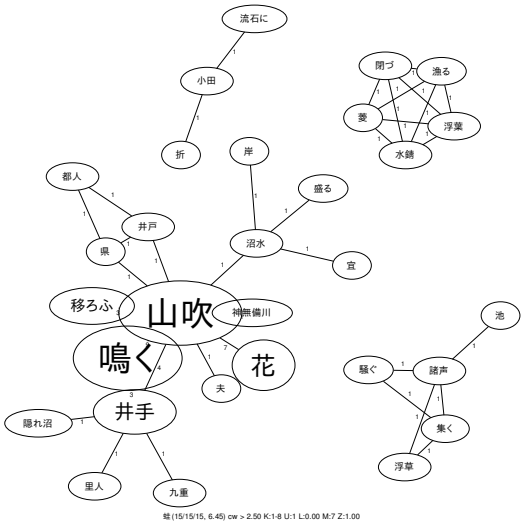

図 3: グラフモデル「蛙」プルーニング後: 古今集から新古今集までに「蛙」は 15 首に出現する。「蛙」自身の *idf* 値は、6.45 で、ネットワークも共出現ウエイト *cw* は 2.50 以上を描いた。このようにコアノード「蛙」を刈り取り（プルーニング）、視界をよくすると数多くのエッジで見えなかった語と語の関係がよく見えるようになる。

で説明されていないのは、常識過ぎる常識だからなのだろうか。理由は何であれ、「八重」と「山吹」の関係が指摘されなかったことから、ここでは一般的に、和歌のネットワークでハブとして示される用語は、将来の辞書作りにおいて、見出し語として検討すべき候補と考えてよかろう。いいかえれば、辞典などの見出し語検討に至っては、執筆者の経験や知識のみで行うのではなく、可視化モデルのような外在的基準を利用して、見出し語を検討することにより、従来よりも緻密な知識が記述できるものと考えられる。

6 おわりに

本稿は「山吹」とともに詠まれる語彙の検討を文献と共出現パターンによる可視化の両者によって行った。歌ことばを解説する辞典類では「山吹」は「八重」の関連性についてあまり述べられていない。一方、共出現パターンによる可視化モデルでは、「山吹」に対して、「蛙」「井手」「八重」の３つのハブノードが見られ、「山吹」は「八重」の関係が示された。可視化モデルのような外在的基準を利用して、辞典などの見出し語の検討を行えば、従来よりさらに緻密な知識が記述できると考える。

参考文献

片桐洋一 (1983) 『歌枕歌ことば辞典』，第 35 巻，角川小辞典，角川書店，東京.

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

水谷静夫 (1983) 『語彙』，第 2 巻，朝倉日本語新講座，朝倉書店，第 1 版.

大野晋・佐竹昭広・前田金五郎（編） (1990) 『岩波古語辞典』，岩波書店，第〔補訂版〕版.

Wittgenstein, Ludwig (1953) *Philosophical investigations*, Oxford: Blackwell.

山元啓史 (2007a) 「ネットワークによる歌ことばのモデリング」,『語彙研究』，第 2007 巻，第 5 号，21–32 頁.

——— (2007b) 「和歌のための品詞タグづけシステム」,『日本語の研究』，第 3 巻，第 3 号，33–39 頁.

——— (2009) 「分類コードつき八代集用語のシソーラス」,『日本語の研究』，第 5 巻，第 1 号，46–52 頁.

(c) Information Processing Society of Japan

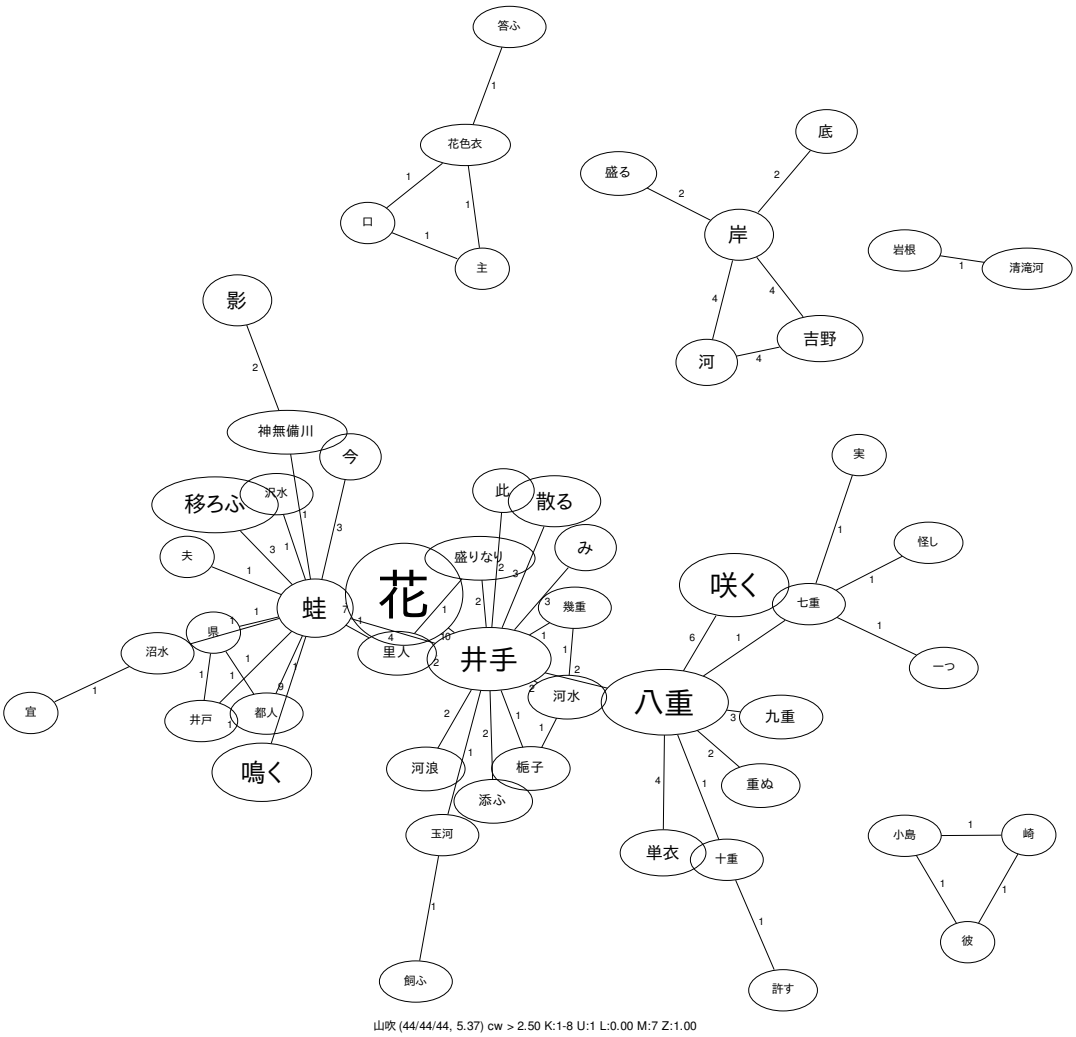

図 4: グラフモデル「山吹」: コアノードの「山吹」はプルーニング済み。「蛙」「井出」「八重」の３つの用語のハブ（そのノードを中心に放射状に他の用語と結びついている様子）が見える。

(c) Information Processing Society of Japan