

歌ことば「橘」「梅」「桜」における関連対の抽出

ホドシチェク ボル
大阪大学

山元啓史
東京工業大学

要旨

本稿の目的は、歌ことばの辞書を開発するにあたり、従来の「見出し語とその解説」による辞書記述に加え、「見出し語-関連語」形式の関連対の追加を提案することである。関連対によれば、従来の辞書に不足していた語と語を取り持つ関係概念を示すことができるだけでなく、古代語の意味記述の困難さを解消する方法であることを示す。和歌を題材とするネットワーク構造のデータから、見出し語（梅、桜、橘）との関連対となる語の抽出を試みた。R パッケージ linkcomm (Kalinka and Tomancak 2011) の3種の計算方法を用いて行った結果、どの計算方法においてもほぼ同様の抽出ができ、それら語は和歌の文脈において各見出し語の関連対として取り出せたことが確認できた。

キーワード: 和歌, 辞書編集, 八代集, 関連対, 重み, R パッケージ, linkcomm

A Study on the Extraction of Relational Pairs of ‘Orange’, ‘Plum’, and ‘Cherry’ Flowers in Poetic Japanese

Bor Hodošček
Osaka University

Hilofumi Yamamoto
Tokyo Institute of Technology

Abstract

The present paper proposes using pairwise term information such as “flower–spring” between an index term and subordinate terms for the development of a classical poetic Japanese dictionary. This kind of information links terms with their contexts and allows us to infer their true meanings in old Japanese. To extract subordinate terms, we use linkcomm: an R package for the generation of link communities from network data (Kalinka and Tomancak 2011). We conduct an experiment extracting the subordinate terms of three representative terms of flowers, ‘orange’, ‘plum’, and ‘cherry’, using average, mcquitty, and single clustering methods. Results identified minimal differences between extracted subordinate terms of each clustering method, and showed the subordinate terms to be quite natural in the context of classical Japanese poetry.

Keywords: classical Japanese poetry, dictionary compilation, pairwise term information, the Hachidaishū, R package, linkcomm

1 はじめに

歌ことば辞典の開発にあたり、「見出し語-関連語」形式の関連対の記述を行っている。「見出し語」を掲げ、その意味・用法を記述するのが従来の一般的な辞書の仕様であるが、それに関連対を提示することにより語と語の関係概念から、語の使われ方や意味・文脈が推測しやすくなるだけでなく、すでに過去のものとなってしまった古代語の意味記述の助けとなるものと期待している。しかしながら、関連語の選定は、本来の見出し語の選定以上に難易度が高い。そこで、古代語のテキストから生成したネットワーク構造から、見出し語「橘」「梅」「桜」と関連対となる語の抽出を試みる。linkcomm (Kalinka and Tomancak 2011) の3種の計算方法を用いて行った結果、どの計算方法においてもほぼ同様の抽出ができ、それら語は和歌の文脈における各見出し語の関連対として自然であることを確認した。本稿では「橘」(Yamamoto and Hodošček 2016) の考察に「梅」「桜」のと比較を加え、コミュニティ分析の手法を説明し、その結果を述べる。

2 方法

和歌で主役として用いられる「桜」「梅」「橘」の3語(主役級の語: 主役語)について、いわゆるドラマや映画でいう脇役級の語(脇役語)をコミュニティ分析を通して抽出することを試みる。

2.1 ネットワークコミュニティの抽出

主役語は、辞書の見出し語であるので、すでにわかっているものとする。問題は、主役に対する脇役の出現をどう特定するかである。語が作る群を特定することは、主役を中心とする俳優のコミュニティに、どんな役者が集まっているかを明らかにするのに似ている¹。このようなネットワーク中に含まれるコミュニティの特定の手法を辞書に記載すべき脇役語(関連語・関連対)の抽出問題として考えてみる。グラフ理論におけるコミュニティ抽出手法を応用して、もっともらしいコミュニティの中心に位置する語とその周辺の語を主観によらず計算手続きによって、抽出できるならば、それらの情報は辞書の

¹たとえば、バーコン数で示されるように主役級の2名の俳優は同じ映画に何度も出るものではない。

記述に役立つはずである。

2.2 材料: 八代集データベースの内容

材料として八代集を用いる。新編国歌大観 CD-ROM 版の二十一代集に相当するデータ(新編国歌大観編集委員会 1996)、国文学研究資料館編集二十一代集データベース(中村他 1999)、新日本古典文学大系本二十一代集に相当する書籍(小島・新井 1989, 片桐 1990, 小町谷 1990, 久保田・平田 1994, 川村他 1989, 片野・松野 1993)、その他、新潮日本古典集成の新古今集(久保田 1979)、ヴァージニア大学日本語テキストイニシアティブ(<http://etext.lib.virginia.edu/japanese/>) 監修、などを参考として、表記を収集し、シソーラスに各々の異なる表記を登録し、正規化を行った(山元 2007)。八代集データベースから、それらに収録のすべての和歌毎に一对ペアを生成し、それらをグラフ処理の基礎データとする。

2.3 グラフデータの抽出

「橘」「梅」「桜」の各データファイルの抽出は、あらかじめ共出現パターン(和歌毎に出現した任意2語の組み合わせパターン)を計算し、パターンの重み、*cw*(山元 2006)については計算せず、すべてのパターンを取り出した上で、R 統計パッケージ *igraph* 形式に変換した(図1)。

```
>head(tachibana)
  V1 V2 V3
1  香 嗅ぐ 2
2  追風 いさよふ 1
3  五月 嗅ぐ 1
4  山郭公 嗅く 1
5  昔 香 5
6  五月 香 2
```

図1: 「橘」のデータファイルの形式; 先頭6行分。

2.4 R パッケージ linkcomm

コミュニティの計算には、linkcomm (Kalinka and Tomancak 2011) を用い、計算手順や結果の表示の流れは、linkcomm の CRAN サイトにある Vignette(Kalinka 2014) に書かれた手順にしたがった。「橘」「梅」「桜」の各ノード・エッジ・リストを八代集データベースより抽出した後、図2の手順で、類

似度計算、デンドログラムの作成、分割密度 (Ahn et al. 2010) の計算を行った²。

```
> install.packages("devtools")
> library(devtools)
> install_github("igraph/rigraph")
> install.packages("linkcomm")
> library("linkcomm")
> tachibana <- read.table("tachibana.txt")
> tachibanadata <- getLinkCommunities(tachibana)
> postscript("output.eps",family="Japan1")
> plot(tachibanadata,type="summary")
> dev.off()
```

図 2: R パッケージ linkcomm による分析手順

3 結果

表 1: 橘のサブクラスタ: 1 番から 10 番までを抽出。average 法、mcquitty 法、single 法のクラスタリングを用いた。括弧内はそれぞれの Maximum partition density を示す。

No.	average (.43)		mcquitty (.43)		single (.38)	
	node	edge	node	edge	node	edge
1	昔	7	昔	7	昔	5
2	匂ふ	6	匂ふ	6	匂ふ	4
3	風	5	今年	4	夢	4
4	夢	5	辺	4	香る	3
5	今年	4	待つ	4	今年	3
6	辺	4	風	4	初む	3
7	待つ	4	夢	4	五月雨	3
8	香る	3	初む	3	折	3
9	五月雨	3	香	3	枕	3
10	初む	3	闇	3	思ひ寝	3

「橘」のネットワークデータから、図 3 のデンドログラム、図 4 のグラフが得られた。コミュニティ発見に際しては、クラスタ分析の 3 手法 (average, mcquitty, single) を用い、方法による違いが、コミュニティの抽出に影響があるかを調べた。「橘」のグラフそのものは、決して大きいものではなく、15 のノードをもつ「昔」が最大ではあるが、それ以外にはサブコミュニティの関係は見つけにくい。図 3 の Partition density は、0 から 1 までの数値で、全体の分割数を示している。

クラスタリング手法毎に、サブクラスタに見られる語をエッジ数の多い順に出力したのが、表 1 である。average 法、mcquitty 法、single 法のいずれにおいても、「昔・匂ふ・夢・今年・香(る)」が見ら

²この手順については、“R+igraph: Tutorial on R+igraph-supplementary information”が詳しい。

表 2: 梅のサブクラスタ: 1 番から 10 番までを抽出。

No.	average (.26)		mcquitty (.26)		single (.20)	
	node	edge	node	edge	node	edge
1	折る	11	香	12	香	5
2	匂ふ	9	匂ふ	11	折る	5
3	香	8	折る	10	鶯	5
4	枝	6	枝	8	匂ふ	5
5	文無し	5	鶯	6	植う	5
6	植う	5	雪	6	花笠	4
7	咲く	5	扶す	5	雪	4
8	雪	5	文無し	5	枝	4
9	鶯	5	移す	5	真屋	3
10	主	4	春	5	隙	3

表 3: 桜のサブクラスタ: 1 番から 10 番までを抽出。

No.	average (.47)		mcquitty (.48)		single (.46)	
	node	edge	node	edge	node	edge
1	曙	4	曙	4	高嶺	4
2	高嶺	3	高嶺	3	曙	4
3	春霞	3	春霞	3	止む	3
4	高砂	3	盛る	3	春霞	3
5	尾上	3	高砂	3	吉野	3
6	南	2	尾上	3	南	2
7	山風	2	吉野	3	山風	2
8	紛る	2	面	2	紛る	2
9	捲く	2	柵む	2	捲く	2
10	面	2	堰く	2	面	2

れた。同様に、「梅」においては「香」が、「桜」においては「春霞・高嶺・曙」が共通に見られた。

4 考察

4.1 可視化表現

「橘」の歌にみられる共出現パターンに対してクラスター分析し、類似パターンに整理すると図 4 に示すように可視化できた (スパンサー・サークル方式)。このように可視化技術は通時の変化を要約・分析するのに便利であるが、1) 単語のサイズを一律に決めたために、連語として存在する・しないの区別がなく、それらの語が同じ文に使われたことは確かだが、当時その連語が成立していたかどうかは明確でない、2) 重複したエッジが見られることから、多義語であるにも関わらず、語形が同じであるために、一義的に扱われているなどの問題がある。

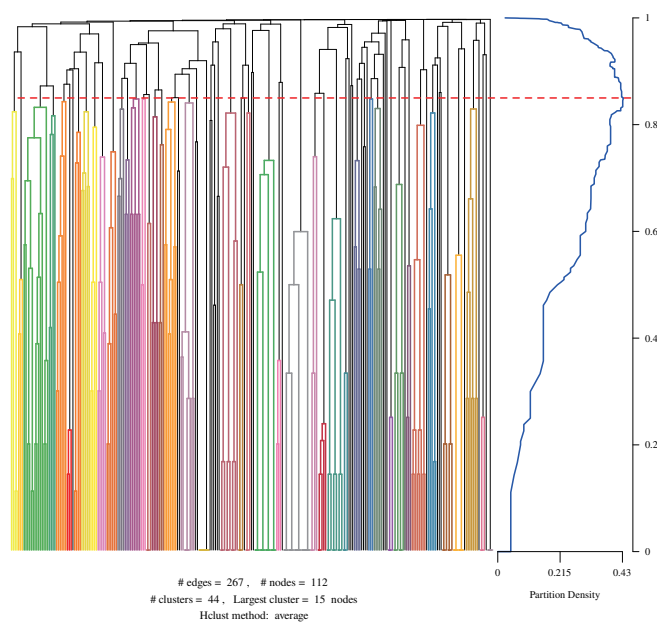


図 3: 橘のデンドログラム: 112 ノード、267 エッジ、44 のクラスタからなり、最も大きいクラスタは 15 ノードからなる。破線はコミュニティの見えやすいところ Partition Density のもっとも深いところ、すなわち見えやすいところに引かれる。

4.2 「橘」と「昔」

「橘」の歌として有名な「五月待つ／花たち花の／香をかげば／昔の人の／袖の香ぞする」(古今 139 番) が表 4 に見られる。「香」は昔の人を思わせる香りであり、「昔」は亡くなった人を偲ぶ想いとともに使われている。この歌以外にも「昔」「匂ふ」「夢」「今年」「香(る)」を使う歌は多く、これらの語と「橘」が示す歌の意味はよく似ている。

4.3 「梅」と「香」

前出の「橘の香」は昔の人の香りであった。「梅の香」は花そのものの「香」、そして「待つ人の香」(古今 34 番、表 5 の 9 番) とあるように、「人の香り」として使われている。「色よりも／かこそあはれと／おもほゆれ／たか袖ふれし／やとの梅そも」(古今 33 番) にも「袖と香」の関係対が見られる。「袖」は「人の袖」で、その人の気配は「袖の香」から感じ取る。古今 32 番から 35 番は「梅の香」と「人」の関係を詠んだ歌が続いている(小島・新井 1989)。この用法「人の香」「花の香」については、歌ことば辞典では「香」の項目を立てて明示的に記載されていない。「梅」にはほぼすべての歌に「香」もしくは「匂ふ」が見られることから、「花」と「香」の

関連としては想像しやすい。しかし「あやなし」や「花笠」は和歌の元となっているエピソードを知らずして、咄嗟に「関連している!」と指摘するのはむずかしいだろう。

4.4 「桜」と「春霞・高嶺・曙」

「桜」と「春」の関連対は「春に桜が咲く」と常識的に理解できる。清少納言『枕草子』冒頭に「春は曙」とあることから、2 語の関連対の存在自体は(有名すぎて)認められても、「では、それはなぜ?」と問われた時に、前提となるエピソードなしには理由は答えにくいだろう。

4.5 間接的な関連対

「橘」「梅」「桜」を通してわかったことは、単純な想像しやすい関連対ではなく、間接的な関連対、つまり、あるエピソードが仲介となってくつついた関連対が存在し、それらは、そのエピソードの認識なしに、即座に指摘することはむずかしい、ということだ。とはいえ、直接的・間接的いずれにおいても、コミュニティ分析を通して関連対を抽出することはできた。

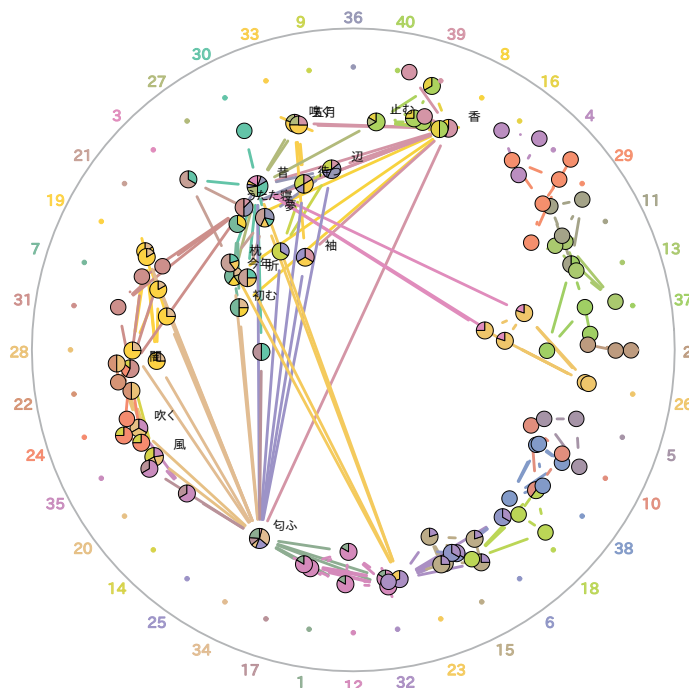


図 4: 橘のネットワーク: この図を見ただけでは、44 のクラスタは見分けにくい。また、核となるノードの語も見えにくい。最大分割深度 (Maximum partition density) は 0.43 であった。

表 4: 「橘」のサブクラスタから抽出された「昔」に関わる歌

No.	歌番号	作者	歌
1.	10139	読人不知	さ月まつ／花たち花の／かをかけは／【むかし】の人の／袖のかそする
2.	40215	高遠	【昔】をは／花たちはなの／なかりせは／なにゝつけてか／思ひいてまし
3.	60070	花山院	宿近く／はな橘は／ほりうへし／【むかし】をしのふ／妻となりけり
4.	70175	公衡	おりしもあれ／花たちはなの／かほるかな／【昔】を見つける／夢の枕に
5.	80238	俊成	誰か又／花橘に／おもひいてん／我も【むかし】の／人となりなは
6.	80240	式子内親王	かへりこぬ／【むかし】を今と／思寝の／夢の枕に／匂ふたちはな
7.	80241	忠良	橘の／花ちる軒の／しのふ草／【むかし】をかけて／露そこほるゝ
8.	80244	読人不知	郭公／花たちはなの／かをとめて／なくは【むかし】の／人や恋しき
9.	80245	俊成女	橘の／匂ふあたりの／うたゝねは／夢も【むかし】の／袖の香そする
10.	80246	家隆	今年より／花さきそむる／橘の／いかて【昔】の／香ににほふらん
11.	81953	寂然	吹風に／花橘や／にほふらん／【むかし】おほゆる／けふの庭かな

和歌というきわめて短い文脈において、主役を2人(2語)も出演させては、焦点がぼけて舞台監督としては失敗である。その意味で2大主役級である「桜」と「梅」が同時に出現するはずもなく、「似たことば」という類似対は計算結果には顕れにくい。クラスター分析は類似距離の計算によるものであるが、言語の世界ならぬ計算上の世界では、関連対と類似対は区別されないため、語の意味としての両者の弁別はむずかしいものとする。

以上、抽出した関連対を和歌の実例と付きあわせて見る限り、サブクラスタに見られる語は脇役語と考えると良いだろう。ただし、いくつかの脇役語までが有効かは不明である。

5 おわりに

本稿では、歌ことば辞典の開発にあたり、従来の「見出し語とその解説」による記述に加え、「見出し語-関連語」の形式による記述を提案し、「橘」「梅」「桜」のコミュニティ分析によって、関連対の抽出を行い、それらは実際の和歌において関連した語であることが確認できた。

参考文献

Ahn, Yong-Yeol, James P. Bagrow, and Sune Lehmann (2010) "Link Communities Reveal Multi-scale Complexity in Networks", *Nature*,

表 5: 「梅」のサブクラスタから抽出された「文無し」「花笠」「植う」に関わる歌

No.	歌番号	作者	歌
1.	10041	躬恒	春のよの／やみは【あやなし】／梅花／色こそ見えね／かやはかくるゝ
2.	20029	読人不知	心もて／おるかは【あやな】／梅花／かをとめてたに／とふ人のなき
3.	30016	躬恒	香をとめて／誰おらさらむ／梅の花／【あやなし】かすみ／たちなかくしそ
4.	30031	能宣	匂ひをは／かせにそふとも／むめのはな／色さへ【あやな】／あたにちらすな
5.	40051	能宣	梅花／にほふあたりの／夕暮は／【あやなく】人に／あやめられつゝ
6.	11081	読人不知	あをやきを／かたいとによりて／鶯の／ぬふてふかさは／梅の【花かさ】
7.	20032	読人不知	春雨の／ふらは野山に／ましりなむ／梅の【花かさ】／ありといふ也
8.	50662	読人不知	むめの【花かさ】／きたるみのむし／雨よりは／風ふくなどや／おもふらん
9.	10034	読人不知	やとちかく／梅花【うへ】し／あちきなく／まつ人のかに／あやまたれけり
10.	31008	広庭	いにし年／ねこして【うへ】し／我宿の／わか木の梅は／花さきにけり
11.	40062	嘉言	わか宿に／【うへ】ぬはかりそ／梅花／あるしなりとも／かはかりそみん
12.	70612	忠教	ほり【植し】／若木の梅に／咲花は／年もかきらぬ／匂ひなりけり
13.	81443	忠平	をそくとく／つみにさきぬる／梅の花／たか【うへ】をきし／種にか有らん

表 6: 「桜」のサブクラスタから抽出された「高嶺(1,2)」「曙(2,3)」「春霞(4,5,6)」に関わる歌

No.	歌番号	作者	歌
1.	50036	公実	白雲と／をちの【高ねの】／見えつるは／心まとはす／さくらなりけり
2.	80133	後鳥羽院	みよし野の／【たかね】のさくら／散にけり／嵐もしろき／春の【あけほの】
3.	80114	俊成	又やみん／かた野のみのゝ／桜かり／花の雪ちる／春の【あけほの】
4.	10058	貫之	たれしかも／とめておりつる／【春霞】／たちかくすらん／山の桜を
5.	30042	元輔	【春かすみ】／たちなへたてそ／花さかり／みてたにあかぬ／山のさくらを
6.	80115	成仲	ちりちらす／おほつかなきは／【春霞】／立田の山の／さくら成けり

Vol. 466, pp. 761–764.

Kalinka, Alex T. and Pavel Tomancak (2011) “linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type.”, *Bioinformatics*, Vol. 27, No. 14, pp. 2011–2012.

Kalinka, Alex T. (2014) “linkcomm: Tools for Generating, Visualizing, and Analysing Link Communities in Networks”, Technical report, CRAN Project. Download: Vignette, A step-by-step guide to using linkcomm.

片桐洋一 (1990) 『後撰和歌集』, 新日本古典文学大系, 岩波書店, 東京.

片野達郎・松野陽一 (1993) 『千載和歌集』, 新日本古典文学大系, 岩波書店, 東京.

川村晃生・柏木由夫・工藤重矩 (1989) 『金葉和歌集、詞花和歌集』, 新日本古典文学大系, 岩波書店, 東京.

小島憲之・新井栄蔵 (1989) 『古今和歌集』, 新日本古典文学大系, 岩波書店, 東京.

小町谷照彦 (1990) 『拾遺和歌集』, 新日本古典文学大系, 岩波書店, 東京.

久保田淳・平田喜信 (1994) 『後拾遺和歌集』, 新日本古典文学大系, 岩波書店, 東京.

久保田淳 (1979) 『新古今和歌集』, 新潮日本古典集成, 新潮社, 東京.

中村康夫・立川美彦・杉田まゆ子 (1999) 『国文学研究資料館データベース古典コレクション『二十一代集』(正保版本) CD-ROM』, 岩波書店, 東京.

新編国歌大観編集委員会 (編) (1996) 『CDROM 版新編国歌大観』, 角川書店.

Yamamoto, Hilofumi and Bor Hodošček (2016) “Development of the dictionary of poetic Japanese description”, in *Digital Scholarship in History and the Humanities, the 6th conference of the Japanese Association for Digital Humanities, Japanese Association for Digital Humanities*, pp. 44–46.

山元啓史 (2006) 「歌ことばの可視化とコノテーションの抽出—グラフによる共出現パターンの作り方—」, 『じんもんこん 2006, 人文科学とコンピュータシンポジウム』, 第 2006 巻, 第 17 号, 21–28.

—— (2007) 「モデリングによる歌ことばの変遷と分析—八代集・歌ことばシソーラスの開発—」, 『じんもんこん 2007, 人文科学とコンピュータシンポジウム』, 第 2007 巻, 第 15 号, 163–170.