

古今集データベースによる歌語の視覚化

Visualisation of the construction of poetic vocabulary using the database of the Kokinshū

山元啓史

Hilofumi Yamamoto

オーストラリア国立大学

The Australian National University

Japan Centre, Faculty of Asian Studies, BPB, ANU, Canberra ACT, 0200, AUSTRALIA

キーワード：歌枕 地名 和歌 モデル 視覚化 共出現パターン

Keyword: *Utamakura*, Place name, classical Japanese poetry, modelling, visualisation, co-occurrence pattern

概要

本研究は、ひとつの和歌あるいはその現代語訳に共出現する任意の2語のパターンの頻度と重みづけを利用し、和歌語彙の構造を視覚化した。材料は古今集の和歌、1,000首およびその現代語訳10種を用いた。視覚化は任意の2語それぞれの重みづけと、それら2語が各和歌に共出現した時の頻度を用いて、2語の結びつきの重要度を計算し、重要な共出現パターンからグラフで表現した。その結果、分析対象とした歌語について、和歌データだけを用いたネットワークは共出現頻度も少なく、2語の共出現関係も図からは推論しにくい。現代語訳データもあわせて表示すると、注釈で述べられている関係を観察することができた。本研究は、主に語誌の研究として発展してきた現代語訳をもとに、共出現パターンを分析することにより、古今集の語彙の体系化を試みた。

Using mathematical analyses, this study examines the connotations of classical Japanese poetic vocabulary, especially those of *utamakura*, place names, with reference to the *Kokinwakashū* (in short, *Kokinshū*), the anthology compiled by imperial order in the tenth century (ca. 905). Using the frequency and weight of co-occurrence patterns between any two terms appearing in the translations of poems, I developed the structures of co-occurrence patterns of poetic vocabulary. I used 1,000 poems in the *Kokinshū* (OP, original poems; except for *chōka*, *sedōka*, and other different styles of poems) and the ten contemporary Japanese translations (CT). The models delineated by both OP and CT data could indicate the similar relationships between the terms with those which are described in the annotations. Since the co-occurrence patterns are indicated based on the significance of the two terms appearing in the ten CT, I could observe the strength of the relationships between the terms and the associations of terms in sub-network clusters. I could also identify the poems of the *Kokinshū* by retracing through the network models.

1 はじめに

本研究は、歌枕(和歌に見られる地名)を視覚化によって分析する。材料は、古今和歌集(905年、以下古今集あるいはKKS)とその現代語訳を用い、和歌あるいはその現代語訳文に共に出現する任意の2語の頻度と重みづけによって、すべての共出現パターンとその重みを計算し、重要なパターンを抽出し、視覚化した。「立田」と「吉野」に焦点をあて、視覚化を通して、同じ奈良県近隣に位置するふたつの地名の特徴を述べる。

和歌を鑑賞するためには、和歌に表記されている語や文法の理解だけでなく、それらに明記されていない含意の理解が不可欠である。含意は作品の内容、文脈、背景、受け手によって決定される意味であり、表記されている語から推論するためには、その世界の常識を用いる必要がある。和歌の世界の常識は一般的に辞典、注釈書などに詳しく記述されているが、歌語の体系的な研究を行う上では、いくつか問題がある。たとえば、注釈の著者によって、その記述の多少や観点の差が見られること、どんな語と共に用いられるかという記述（共起情報）があまり見られないこと、などである¹。本研究では、注釈を直接用いるのではなく、注釈に添えられた現代語訳を用いて、注釈と同様の説明が可能かどうかを試みる²。

歌枕とは広くは和歌に用いられた語のことであり、狭くは和歌に用いられた地名のことである。しかしながら、それは「単なる地名ではなく、特定の人事的観念が結合した地名」（片桐, 1970, 257）といわれている。つまり、歌枕は単にその土地の名前を示すだけでなく、景色、風物、季節などその土地に関する観念を持っているのである。本稿では「立田」「吉野」を対象語として取りあげ、それぞれが地名であるだけでなく、どのような「特定の人事的観念」が歌の中で示されているのか、について現代語訳文に共出現する2語の関係から分析する。「立田」「吉野」は、歌によく用いられる地名で、ともに奈良県に位置する土地である。古今集では、「立田」は、13の和歌で、「吉野」は、22の和歌で、使われており、頻度で第1、2位を占めている。

本研究は従来の注釈・語誌の研究を基盤とし、計算手続きにより、従来手作業では難しかった歌語の体系的な構造、いわゆる語彙の構造を明らかにするものである。

2 材料

古今集（略 OP; original poems）、1,111 首のうち、長歌・旋頭歌などの雑体歌をのぞく 1,000 首、およびその現代語訳 10 種（略 CT; contemporary translations）を利用した。現代語訳 10 種は、表 1 にある注釈書に収録されているものを利用した。底本はすべて定家本である。各著者による現代語訳の方法は、注釈書の序文や凡例において明らかに述べられていないものもあるが、だいたいにおいて逐語訳を基本に字句を補ったり、語順を任意に変更しているようである。表 2 に現代語訳の例を示す。

表 1: 古今集の注釈書 10 種の一覧

著者 (年)	ページ数	底本	現代語訳の方法
金子 (1933)	1105	定家本 (貞応二年)	逐語訳
窪田 (1960)	1449	定家本	逐語訳
松田 (1968)	1998	定家本 (梅沢本)	不詳
小沢 (1971)	544	定家本 (貞応二年)	語順・語法を変えている
竹岡 (1976)	2278	定家本 (伊達本)	逐語訳
奥村 (1978)	434	定家本 (貞応二年)	作者の作意を尊重
久曾神 (1979)	1260	定家本 (伊達本)	字句を補っている
小町谷 (1982)	407	定家本 (貞応二年)	不詳
小島・荒井 (1989)	483	定家本 (詰訓本)	不詳
片桐 (1998)	3022	定家本 (嘉禄二年自筆書写本)	逐語訳

¹ 語の体系的な研究がなぜ重要であるかについては、田島 (2004, 19) に譲る。共起あるいは共出現パターン (co-occurrence patterns) といわれる。水谷 (1979) はおそらく公になっている最初の共出現パターンを利用した和歌の計量研究であろう。

² 注釈の利点は和歌に記述されていない語や表現が記述されていることである。欠点はその歌に直接関係のない関連情報があったり、他の歌で既に出た語の説明が省略されたり、簡略化されたりしていることである。

表 2: 現代語訳の例 KKS 298

兼覧王	立田姫手向くる神のあればこそ秋の木の葉の幣と散るらめ
金子 (1933)	立田姫は秋の神だが、それすら暮れて行かれる折には、お手向けなさる道の神様があるからこそ、このように、御自身お染めなされた秋の木の葉が、手向けの幣のように散るのであろう。
窪田 (1960)	竜田姫は、手向けをするべき神があるので、そのつかさどる秋の木の葉が、幣のように散るのであろうよ。
松田 (1968)	竜田姫には、旅中、供え物をささげる道祖神があるからこそ、秋の紅葉した葉が、幣のように散っているのだらう。
小沢 (1971)	もはや秋の終りで龍田姫が帰り道にお着きになった。山の紅葉が色とりどりの幣となって散っているのは、姫が道の神様にそれを供えていらっしゃるのだな。
竹岡 (1976)	竜田姫が、旅にあつて手向けの神があるので、それであるように秋の木の葉が、幣として散っているのだらう。
奥村 (1978)	秋も終りに近づき、秋の女神の龍田姫がお帰りになる。姫が道中の安全を祈って手向けをなさる神があるので、紅葉が幣のように散っている。
久曾神 (1979)	秋も終りに近づき竜田姫がお帰りになる際に、旅の安全を祈って手向けられる神があるので、秋の紅葉が幣となって散っているのだらう。
小町谷 (1982)	秋の末近くなって帰り道についた龍田姫が、道中の無事を願って手向けをする神があるので、秋の木の葉が幣となって散っているのだらう。
小島・新井 (1989)	竜田姫が手向けをする神さまがおられるからこそ、秋の木の葉のもみじが手向けものとして散るのだらう。
片桐 (1998)	秋をつかさどる龍田姫が旅立ちにあたって手向けをする神があるので、あのように、秋の木の葉が幣を撒くように散っているのだらう。

3 方法

ある和歌あるいはその現代語訳に共出現する任意の2語のパターンの頻度と重みづけを利用し、和歌語彙の共出現パターンを視覚化した。和歌とその現代語訳はすべて β 単位で分割した³。ただし、地名に限っては、「立田山」を「立田」と「山」のように短い単位に分けた。さらに、分類語彙表(中野, 1994)に準じたコードをそれぞれの語につけ、そのコードによって、和歌中の語あるいは現代語訳中の語が表記が異なっても対応できるようにした⁴。視覚化には共出現パターンの重要度を計算し、重要な共出現パターンから順にネットワークで表現した。

共出現パターンの重要度 (cw ; co-occurrence weight) は、(1) のように定義できる。

$$cw(t_1, t_2, d) = (1 + \log \text{ctf}(t_1, t_2, d)) \cdot \sqrt{\text{idf}(t_1) \cdot \text{idf}(t_2)} \quad (1)$$

ただし、前半の $1 + \log \text{ctf}(t_1, t_2, d)$ は、ある条件で抽出された資料の数 d で、 t_1 と t_2 の2つの用語が共に出現した時の資料の数 (co-occurrent term frequency) を対数変換し、1を加えたものである。また、後半の $\sqrt{\text{idf}(t_1) \cdot \text{idf}(t_2)}$ は、 t_1 と t_2 の2つの用語の idf 値の幾何平均である⁵。 idf (inverse document frequency) は式 (2) で定義できる (Manning and Schütze, 1999, 543)。

$$\text{idf}(t) = \log \frac{N}{df(t)} \quad (2)$$

ただし、 N はすべての資料の数、ここではすべての和歌あるいは現代語訳文の数である。 $df(t)$ (document frequency) は、語 t の出現する資料の数、ここでは当該の和歌あるいは現代語訳文の数である。 $\text{idf}(t)$ は、

³ 語の調査単位については、田中 (1971)、宮島 (1994) を参照されたい。宮島 (1994, 113) は、日本語の語彙調査で一番困ることとして、『「単語」という単位が確立していないこと』と述べている。

⁴ 分類語彙表に準じた語にコードを付け、そのコードで作品中の意味成分を調査する提案や試みは、田島 (1995) や西端 (1996) などに見られる。

⁵ 式 (1) は、 $w = \text{tfidf}$ (Rocchio, 1971; Spärck Jones, 1972) を任意2語の共出現の重みを計算するために拡張したものである。頻度、 tf の代わりに共出現頻度、 ctf を、 idf の代わりに2つの idf の幾何平均を用いた。

N と $df(t)$ の比を対数変換したものである (Spärck Jones, 1972; Robertson, 2004)。

視覚化の手順としては、対象語「立田」「吉野」を含む和歌をまず検索し、その和歌あるいはその和歌に対応する現代語訳文に共出現する任意の2語の頻度と2語それぞれの idf 値を使って cw を計算した。このように計算された2語の cw 値の高いパターンから順にグラフで描いた。描画には、Graphviz (Ellson et al., 2005) を用いた。

4 結果

本稿では、現代語訳を用いた分析について詳しく述べる。地名、人名のような固有名詞は、さまざまな文章に出てくるのではなく、特定のある限られたの文章に出てくる性質をもつため、結果的に idf 値の高い用語となる。「立田 ($idf=8.52$)」「吉野 ($idf=6.08$)」のように idf の高い語 (5-10) を調査の対象としてネットワークを描いた場合、対象語を中心に放射線状の図形 (spoke effect) が出るということがわかっている (Yamamoto, 2005)。しかし、対象語との共出現パターンを除くと放射線状の図形は取り除かれ、放射線状の図形に蔽われていた構造が見える⁶。本稿では、対象語との共出現パターンを削除した図、つまり、「立田」「吉野」との共出現パターンを除いた図を結果として示す。それぞれのネットワークモデルは語を示すノードとそのつながりを示す枝からなり、ノードの大きさは相対的な頻度を、枝に添えられた数字は2つのノードの共出現頻度を示す。

4.1 立田

図1は、「立田」のネットワークモデルである。「立田」のモデルは、大きく2つに分かれ、秋をイメージする「紅葉」を中心とするモデルと、「沖つ白波」の歌 KKS 994 を中心とするモデルである。

神様の山である立田山に向かって、色あざやかな紅葉が川一面に流れて行く情景を幣 (神にささげる布織物の贈り物) に見立てた歌が、ネットワークから読みとれる。このように紅葉の流れゆくさまを幣に見立てた歌は、KKS 298 (表2) や KKS 300 に見られる。

KKS 300 神奈備の山を過ぎ行く秋なれば立田川にぞ幣は手向くる —清原深養父

ネットワークの中にさらにネットワークが見られる。「神代」を中心とするサブネットワークには、KKS 294 の歌が見える。

KKS 294 千早振る神代も聞かず立田川唐紅に水括るとは —在原業平

上記から「立田」は秋を紅葉で彩る山であることがわかるが、そこには「神奈備」「神代」という語が含まれていることからわかるように、「神」が存在していることが認められる。一方、「沖つ白波」のモデルは KKS 994 が基となっている。KKS 994 は伊勢物語 (23段) や大和物語 (149段) にも出てくる有名な歌である。

KKS 994 風吹けば沖つ白波立田山夜半にや君が一人越ゆらむ —よみ人知らず

「立田山」は想う人が越える山である。「白波」は「波が立つ」という意味関係に由来する「立田」の縁語である。この場合の「立田」は紅葉に代表される秋の「立田」ではなく、河内国から大和国への重要な交通路としての「立田」である (片桐, 1983, 251)。このことはモデルでは、図1のように2つのネットワークに分かれて描画されている。

⁶ ゆえにモデル中には「立田」「吉野」の地名が見られないことに注意。対象語を含む歌が抽出された後、その歌に対応する現代語訳を用いて共出現パターンを作成しているため、すべての語は対象語と高い cw 値で結びついている。

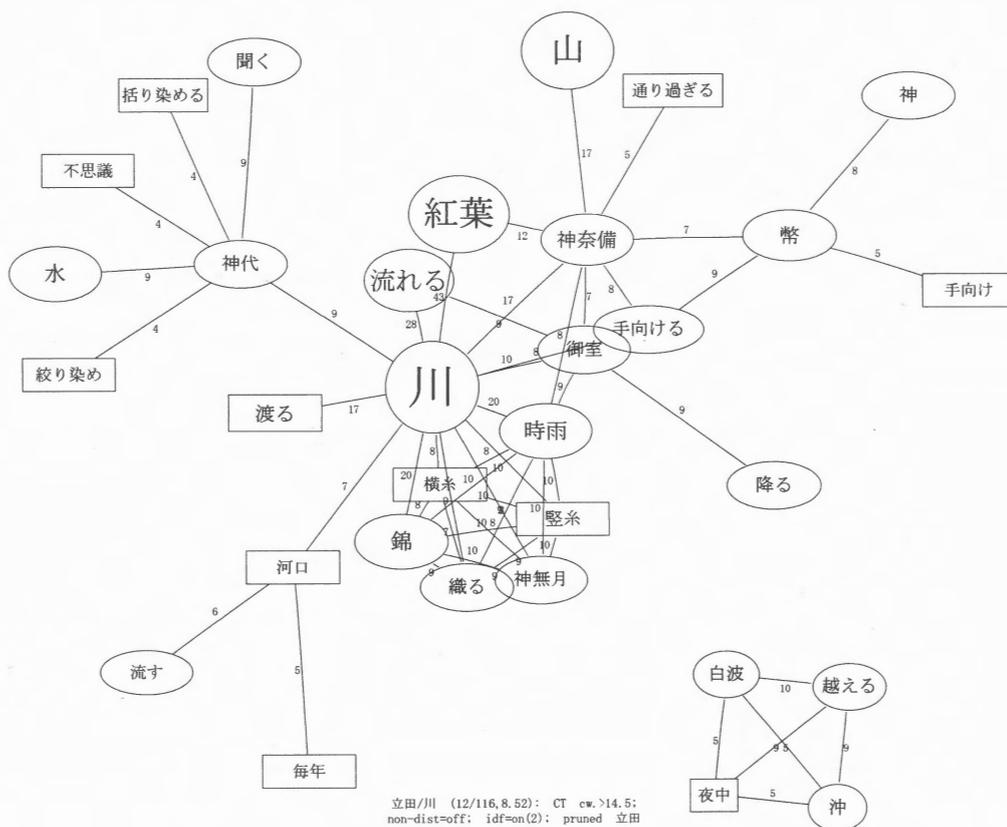


図 1: ネットワークモデル「立田」(川/白波): 立田のネットワークは、大きく2つの部分に分かれ、その「川」の部分と「白波(右下)」の部分を描画。「立田」はOPでは12回、CTでは116回出現する。「立田」のidf値は8.52。14.5cw以上を対象に描画。「川」は、28ノード(OP18ノード、CT10ノード)、49の枝からなっている。「白波」は、4ノード(OP3ノード、CT1ノード)、6の枝からなっている。各ノードの大きさは相対的頻度を示している。楕円中の語はOPにも出現する語で、矩形中の語はCTのみに出現する語。

4.2 吉野

図2は、吉野川を中心とするモデルで、図3は、吉野山を中心とするモデルである。

「吉野川」のモデルは、「烈しい」「流れ」「妹山」「男女」という語からもわかるように、男女間の心理の激しい変化が、「吉野川」の流れとして表現されている。

KKS 471 吉野川岩波高く行く水の早くぞ人を思ひ初めて —紀貫之

「妹山」は「妹背山」ともいい、「妹背」とは「愛し合う女と男。夫婦」(広辞苑)の意味である。「吉野山」のモデルには「妹山」を中心とするサブネットワークが見える。「妹山」に流れ落ちる川とその谷の狭間を「男女」「夫婦」の「仲」「間」に見立てている歌KKS 828がモデルから読みとれる。

KKS 828 流れては妹背の山の中に落つる吉野川の縦しや世の中⁷ —よみ人知らず

一方「吉野山」のモデルは、「雪」「白雪」「山奥」「有明の月/里」「隠れ所」などの用語で構成されている。

⁷ 図3には「たとえ」が見られる。これは「縦しや」の現代語訳であろう。「縦しや」は「吉野」の掛詞である。

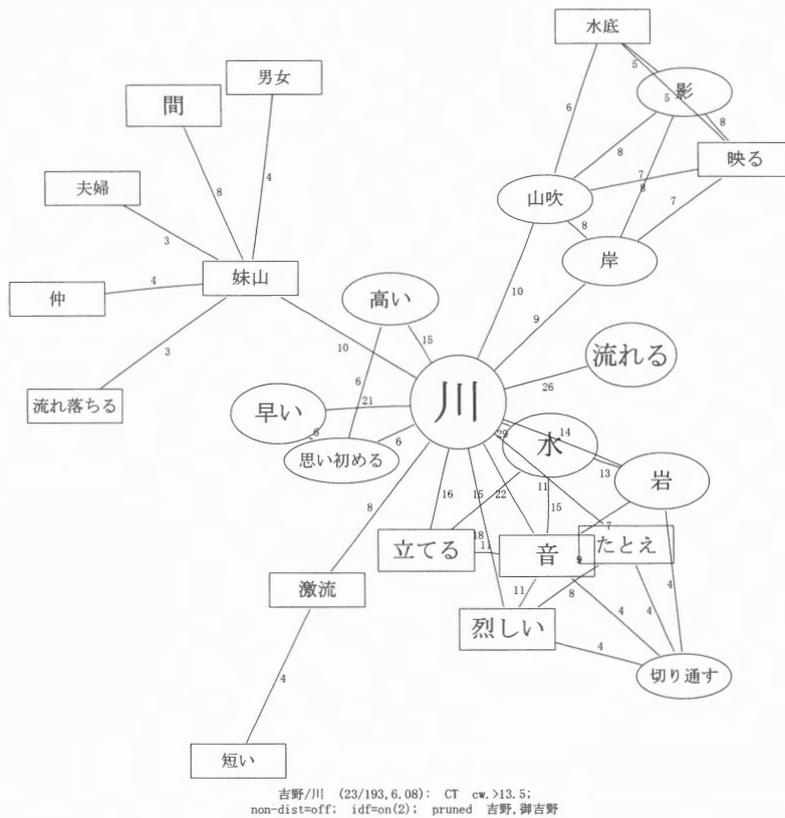


図 2: ネットワークモデル「吉野川」: 「吉野」は OP では 23 回、CT では 193 回出現する。「吉野」の *idf* 値は 6.08。13.5 *cw* 以上を対象に描画。このモデルは 25 ノード (OP 11 ノード、CT 14 ノード) 43 の枝からなっている。各ノードの大きさは相対的頻度を示している。楕円中の語は OP にも出現する語で、矩形中の語は CT のみに出現する語。

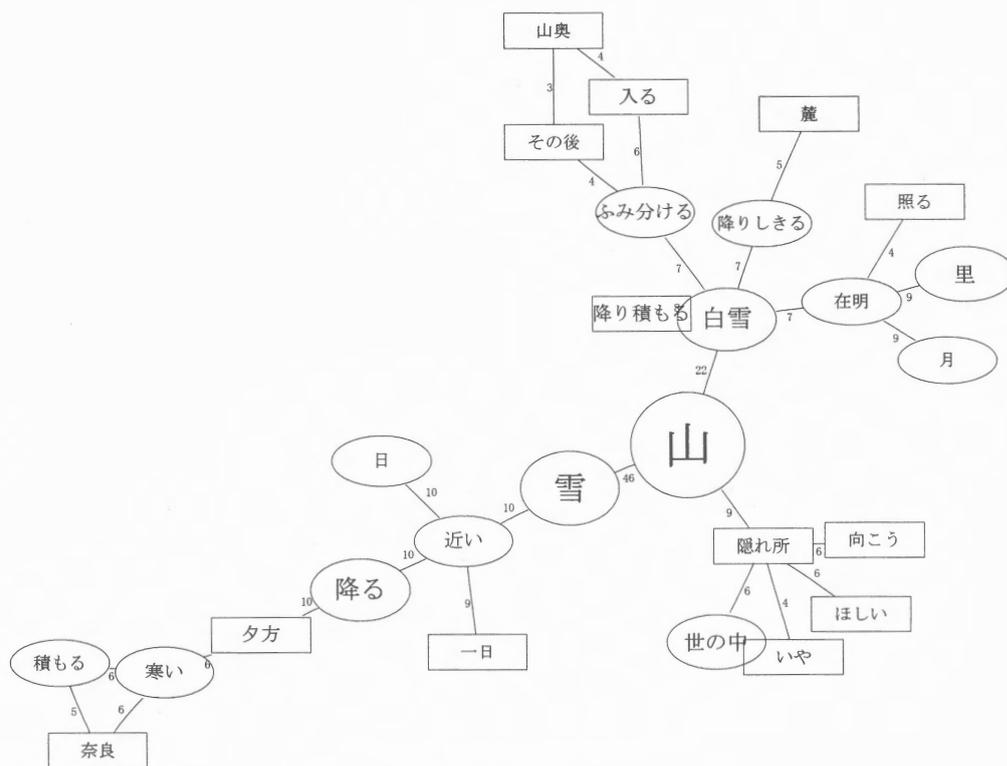
KKS 332 朝朗け有明の月と見るまでに吉野の里に降れる白雪 一坂上是則

「隠れ所」を中心とするサブネットワークが見られ、「吉野山」が世の中 (人間社会) を嫌った人の隠遁の地であることがわかる。このサブネットワークを構成する語はほとんどが現代語訳文にのみ出現する語であることから、「隠れ所」「隠遁」という意味は和歌そのものからは、

KKS 950 御吉野の山のあなたに宿もがな世の憂き時の隠れ家にせむ 一よみ人知らず

のように「隠れ家」という語から読みとれる。片桐 (1983, 434) によると、奈良時代までは実際に吉野山に赴くことがあまりなかったがゆえに、山岳信仰の地・隠遁の地としての吉野山であったが、そのような神秘的なイメージは雪をいただく山々の姿とマッチして、吉野山といえば雪がよまれるというようになった、ということである。モデル中の「雪」「山奥」「隠れ所」の語が、結果的にこのようにネットワークで表現されている理由は、上記の説明でわかる。

ところで「吉野の桜」は有名ではあるが、「桜」はこのモデルには見られない。片桐 (1983, 436) によると、吉野山は、平安時代中期までは、雪の名所であるが、桜の名所としての吉野山のイメージはそれほど強くはなく、吉野山と桜の関係が決定的なものになるのは、新古今集を代表とする数々の歌を詠んだ西行の時代、といわれる。古今集のデータにも「吉野」と「桜」を詠んだ歌があるにもかかわらず、このモデ



吉野/山 (23/193, 6.08): CT cw.>13.5;
non-dist=off; idf=on(2); pruned 吉野, 御吉野

図 3: ネットワークモデル「吉野山」: 「吉野」は OP では 23 回、CT では 193 回出現する。「吉野」の *idf* 値は 6.08。13.5 *cw* 以上を対象に描画。このモデルは 27 ノード (OP 14 ノード、CT 13 ノード) 28 の枝からなっている。各ノードの大きさは相対的頻度を示している。楕円中の語は OP にも出現する語で、矩形中の語は CT のみに出現する語。

ルに「桜」が含まれていないのは、共出現パターンの重みづけ (*cw*) がうまく機能しているものと思われる。しかし、それをより確実に結論づけるなら、新古今和歌集 (1201 年) を分析し、そこに「桜」が「吉野山」のモデルに含まれるかどうかを検証するとよいだろう。

5 考察

「立田」と「吉野」を対象語として分析を行った。ここまで、それぞれ歌語としての意味・用法をモデルから推察し、注釈と比較しそのモデルの妥当性を検討してきた。モデルが正しいという前提で、改めて「立田」と「吉野」を比べてみると「立田」は神の山、紅葉/秋の山であるのに対して、「吉野」は男女の山、人間心理の激しさを象徴する川、世の中に疲れた人が隠遁する山、と見てとれる。つまり、「立田」の神とは対照的に「吉野」は世の中 (人間世界) の歌枕として目に映る。同じ奈良県に近隣する 2 つの歌枕であっても、モデルから考察すると非常に異なる観念が含まれていることがわかる。

本稿では詳細に述べなかったが、分析対象とした歌語について、和歌データだけでも視覚化してみた。和歌データだけをを用いたネットワークは共出現頻度も少なく、2 語の共出現パターンも図からは推論しにくいものであった。しかし、現代語訳データもあわせて視覚化すると、注釈で述べられている関係

がモデルから読みとることができた。共出現パターンは重要度によって表示されるため、10種の現代語訳に共通して示されているもの、示されていないもの、結びつきの強弱も観察できた。ネットワークの中にはさらにサブネットワークが見られ、そのサブネットワークを構成する語群が、どの和歌に起因するものなのかも特定できた。また、同じ対象語であっても意味・用法の異なる場合には、別の群として描画されていることがわかった。

6 おわりに

本研究では、主に語誌の研究として発展してきた現代語訳をもとに、古今集の語彙の体系化を試みた。共出現パターンを作成し、共出現パターンの重要度 (*cw*) にしたがって、パターンを描画し、視覚化を試みた。視覚化されたモデルを通して、語の意味・用法の違いを確認することができた。

参考文献

- Ellson, J., Gansner, E., Koren, Y., Koutsofios, E., Mocenigo, J., North, S., Woodhull, G., Dobkin, D., Alexiev, V., Lilly, B., Scheerder, J., G., D. R., and Low, G. Graphviz — graph visualization software. <http://www.graphviz.org/> 2005; The page cited on 4th Feb. 2005.
- 金子 元臣 1933. 古今和歌集評釈: 昭和新版 東京: 明治書院 .
- 片桐 洋一 1970. 歌枕の成立 —古今集表現研究の一部として— [古今和歌集 小町谷照彦, 日本文学研究資料刊行会 編] 第1版 pp. 256-266 東京: 有精堂, [1976] 「国語と国文学」(1970) 昭和45年4月号.
- 片桐 洋一 1983. 歌枕歌ことば辞典 角川小辞典, 第35巻 東京: 角川書店 .
- 片桐 洋一 1998. 古今和歌集全評釈 上・中・下 全3巻 東京: 講談社 .
- 小島 憲之, 荒井 栄蔵 1989. 古今和歌集 新日本古典文学体系, 第5巻 東京: 岩波書店 .
- 小町谷 照彦 1982. 現代語訳対照 古今和歌集 東京: 旺文社 .
- 窪田 空穂 1960. 古今和歌集評釈 (上中下) 東京: 東京堂出版 .
- 久曾神 昇 1979. 古今和歌集 全訳注 (1)-(5) 講談社学術文庫 .
- Manning, C. D. and Schütze, H. 1999. *Foundation of statistical natural language processing* Cambridge, Massachusetts: The MIT press .
- 松田 武夫 1968. 新釈古今和歌集 (上下) 東京: 風間書房 .
- 宮島 達夫 1994. 語彙論研究 東京: 麦書房 .
- 水谷 静夫 1979. 用語による梅・桜の歌の弁別. 計量国語学 12, 1-13.
- 中野 洋 1994. 分類語彙表/フロッピー版 国立国語研究所言語処理データ集, 第5巻 東京: 大日本図書 .
- 西端 幸雄 1996. 「歌物語」3作品の使用語彙の比較 [「歌物語」語彙の数量的分析と研究] 第1版 pp. 3-18, 文部省科学研究費: 重点領域研究 「人文科学とコンピュータ」研究成果報告書.
- 奥村 恒哉 1978. 古今和歌集 東京: 新潮社 .
- 小沢 正夫 1971. 古今和歌集 第13版 日本古典文学全集, 第7巻 東京: 小学館 .
- Robertson, S. 2004. Understanding inverse document frequency: on theoretical arguments for idf. *Journal of Documentation* 60, 503-520.
- Rocchio, J. J. 1971. The SMART Retrieval System: Experiments in Automatic Document Processing in *Relevance feedback in information retrieval* Salton, T. G. (ed.) 1 edn. pp. 313-323 Englewood Cliff, NJ: Prentice-Hall .
- Spärck Jones, K. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28, 11-21.
- 田島 統堂 1995. 語彙と単語. 日本語学 14, 4-11.
- 田島 統堂 2004. 比較語彙論 [語彙研究の課題 田島 統堂 編] 第1版 pp. 1-20 大阪: 和泉書院 .
- 竹岡 正夫 1976. 古今和歌集全評釈 (上下) 東京: 右文書院 .
- 田中 章夫 1971. 新聞語彙調査の同音語と同形語 [電子計算機による国語研究 III] 第1版 国立国語研究所報告, 第39巻 pp. 121-145 東京: 秀英出版 .
- Yamamoto, H. 2005. A mathematical analysis of the connotations of Classical Japanese poetic vocabulary. Ph.D dissertation, The graduate school of the Australian National University.