

テキストマイニングによる都市・景観イメージ分析 - 水・緑環境に着目して -*

An analysis on the image of the city by Text Mining from the viewpoint of water / green environment*

大塚裕子**・森田哲夫***・吉田朗****・小島浩*****・塚田伸也*****

By Hiroko OTSUKA**・Tetsuo MORITA***・Akira YOSHIDA****・Hiroshi KOJIMA*****・Shin-ya TSUKADA*****

1. はじめに

(1) 研究の背景・目的

本研究では、水・緑環境を中心とした都市・景観イメージについて、市民が持つイメージを把握した上で、どのように政策・計画に結び付けていくべきかという検討を行っている。さらに、本稿では都市のイメージ分析の方法論および着眼点を整理するために、テキストマイニング手法を用いて様々な観点からの客観的な分析を行う。

都市のイメージについては、従来から調査研究が進められてきており、都市計画分野の研究者、実務者の主要な課題のひとつであった。一方で、そのイメージを政策や計画にどのように活用するかという実務的な観点による検討は十分にされてこなかった。近年、景観を活用し地域連携やコミュニティ形成を図っていくことを仮説とする萌芽的な研究課題が提起され、今まさに議論することが求められている理由もこれまでの検討の欠如にある。このような研究課題に取り組んでいくためには、1)都市や景観イメージをその地域の市民・住民がどのように共有しているかを明らかにすること、2)都市のイメージが、それまでの政策や計画によって構築された都市そのものへの評価、すなわち政策評価の側面を持つと認識すること、が重要である。

水・緑環境については、良質な都市環境、住環境を確保していくにあたり、近年とくに重要視されている。例えば、国土庁では水環境保全の重要性について広く国民に周知し、水を活かした地域づくりを促進するために「水の郷百選」を認定している。しかし、水・緑環境が都市のイメージに与える影響は、定量的に明らかになっていない。

イメージの定量的分析の研究手法としては、市民・住民のもつ都市や景観のイメージを把握するためにアン

ケート調査やヒアリング調査が行われることが多い。イメージのような主観的な情報を分析するためには、質的データを扱うことになるが、予め設定された選択肢から得られるデータに加え、自由記述データが貴重な情報となる。これまで、自由記述データについては、分析観点の恣意性から活用されてこなかった。近年、自然言語処理技術を用いることにより分析の再現性が確保されるため、分析対象データとして関心が高い。以上を背景に、本研究は、アンケート調査の自由記述データを用い、個人属性・地区特性との関係を考慮し、水・緑環境に着目しながら、地域で共有されている都市・景観イメージを、テキストマイニング手法により分析することが目的である。対象地域は「水を緑と詩のまち」を標榜している群馬県前橋市の利根川左岸地域とした。

(2) 研究の位置づけ

土木計画学、都市計画分野における都市イメージの研究、自然言語処理分野における自由記述データ分析の研究の2つの視点から、本研究の位置づけを整理する。

a. 土木計画学・都市計画分野における位置づけ

都市イメージについては、Kevin Lynch¹⁾が都市の物理的な構成要素を5つのエレメントに分類し分析を行っている。国内では、Lynchの都市のイメージ理論に基づいた三浦ら²⁾の実証的研究、角野³⁾のインタビュー調査による自由回答に基づく都市の主観的なイメージ分析という先駆的研究がある。本研究の着眼点である個人属性と都市イメージの関係について扱っている研究は多く、地区特性との関係分析では、斎藤⁴⁾による研究が代表例である。水・緑環境に着目した研究としては、松浦ら⁵⁾が、河川イメージについて形容詞対の対比較データを用い分析している。

これらの関連研究を踏まえ、本研究の位置づけと特色を2つ整理する。1つめは、本研究は、Lynchによる都市イメージの研究、その後の主観的なデータを用いた研究の系列に位置するという点である。この点に関する分析データと分析手法の特色を述べる。分析データについては、従来の研究が、予め用意された選択肢データや順位・尺度データを用いているのに対し、本研究では、アンケート調査の自由記述データを用いる。これにより、

*キーワード：イメージ、水環境、自然言語処理、テキストマイニング

**博(学)、財団法人計量計画研究所 言語・行動研究室

***正会員、博(工)、群馬工業高等専門学校 環境都市工
(〒371-8530 前橋市鳥羽町580, tmorita@cvt.gunma-ct.ac.jp)

****正会員、博(工)、東北芸術工科大学 デザイン工学部

*****工修、財団法人計量計画研究所 東北事務所

*****正会員、博(工)、前橋市都市計画部まちづくり課

分析者が想定していなかった回答を得ることができるとともに、Lynch¹⁾のいう「たくさんの個人のイメージが重なりあった結果としてのパブリック・イメージ」、「それぞれかなりの数の市民たちによってつくられるパブリック・イメージがいくつか集まっている」ものを把握することができると考える。なお、自由記述データを扱うことについては、角野⁴⁾が自由回答を対象に語を抽出しているが、手作業のため、客観性・再現性という点で問題がある。本研究では、単語抽出は形態素解析で行う。分析手法については、従来の研究が因子分析等の多変量解析を用いているのに対し、本研究では、テキストマイニングを用いることにより、語の出現だけではなく、語のつながりを反映した分析が可能となり、従来研究よりも回答者の意識を反映した分析が可能となる。なお、本研究における「都市・景観イメージ」は、都市全体のイメージ、視覚的な景観イメージ、さらには心象風景を含む用語として用いる。

2つめは、研究成果の近年の都市計画課題への対応性に関する位置づけである。本研究は、都市のイメージと個人属性・地区特性との関係に着目する従来研究の流れに属する。個人属性との関係に着目することにより、高齢化等の人口構成変化によるイメージ変化を把握できる。また、地区特性に着目することにより、施策による都市のイメージ変化を把握できる。さらに本研究では、水・緑環境に着目する。河川イメージに関する研究は存在するが、都市全体のイメージの中での水・緑環境に着目するという点で新規性がある。本研究では、水・緑環境のイメージは個人属性により異なる、水・緑環境へのアクセシビリティが高いほど、都市のイメージにおける水・緑環境の占める位置が高いという仮説を考えている。

b. 自然言語処理分野における位置づけ

土木計画分野に限らず、自由記述の意見を行動分析や予測に利用し、マーケティングや政策に活用する試みは多岐にわたっている。なかでも、利活用のための要素技術の研究分野である自然言語処理では、近年、意見分析 (opinion analysis)、評判分析 (sentiment analysis) といった、テキストの感情や評価に関する表現の研究や、表現・構造を手がかりとした自動処理の研究が盛んである⁶⁾。具体的には、web上のレビューサイト、ブログ、アンケート調査時の自由回答等を対象データとして、データが示している書き手や発話者の好き嫌い、肯定・否定、不満、不安・懸念、期待などの感情や評価、およびその評価主体や評価対象、評価の理由などを特定する。

市民参加型計画において収集された自由回答から、市民の要望を自動的に取り出すことを目的に、機械学習手法を用いて意見を要望か否かという観点で分類した研究がある⁷⁾。このようなデータ利用をきっかけに、土木計画学分野、都市計画学分野でも、市民参加型計画にお

いて得られる意見の分類・分析に自然言語処理技術を用いる研究が出てきている⁸⁾⁹⁾。しかし、計画学分野で利活用するためには、意見から何を得たいか、どのように意見が分類され、情報が取り出すことが有効であるのかという出力情報のあり方、処理結果の分析法についてさらに議論が必要な段階である。本研究では、テキストマイニングによって得ることが可能となる様々な観点で分析を行うことにより、都市・景観のイメージ分析の方法論の可能性を探るとともに客観的分析手法の確立を目指す。

2. 都市・景観イメージのアンケート調査の実施

(1) 対象地域の設定

対象地域は図-1のとおり群馬県前橋市の利根川左岸とした。前橋市は、「水の郷百選」の一つに選ばれており、市も「水と緑と詩のまち」を掲げている。対象地域は、利根川、広瀬川、桃ノ木川、荒砥川、粕川、滝川などの一級河川、敷島公園をはじめとする親水公園や大正用水をはじめとする用水路などの多くの河川が存在するため、水・緑環境が都市・景観のイメージに与える影響が大きいと考えた。前橋市は、「前橋市緑の基本計画」¹⁰⁾により河川沿いへの遊歩道やサイクリングロードの整備、緑化などの水辺環境を軸にしたまちづくりを進めている。したがって、市の政策によって整備された前橋の現状に対するイメージは、市民の政策評価という側面も持つと考える。対象地域は人口約9.9万人（市の約31%）、4.2万世帯（同34%）、面積約27.6km²（同11%）である。



図-1 調査対象地域

（２）アンケート調査の概要

生活の質アンケート調査の概要を表-1に示した。地域内の4,000世帯（抽出率9.5%）を対象に調査を実施し、2,118票を回収した。うち、自由記述欄に記入のあるものを有効票とし1,570票が得られた。

自由記述を促す指示文で「前橋の良い点、好きなところ」と限定したのは、次の2点の理由、1)「水・緑環境」を明示すると、都市・景観の他のイメージとの関連性が出にくい、2)「良い点」を指定せずに「前橋のイメージ」を問うことにより、居住者すなわち当事者としてはネガティブなイメージに焦点が当たりやすくなる、を想定したためである。

表-1 調査方法

調査日	配布：2008年11月中旬～下旬 回収：2008年12月14日（郵送期限）
対象地域	群馬県前橋市利根川左岸 （人口約9.9万人、約4.2万世帯、面積約27.6km ² ）
対象者	対象地域の4,000世帯（抽出率約9.5%）の構成員
調査方法	配布：調査員による戸別配布（ポスティング） 回収：郵送回収
調査内容	1)個人属性（性別、年齢、ふだん利用する交通手段、職業、家族人数、住宅種類） 2)自由記述「前橋の良い点、好きなところ、場所などを、短い言葉や文で、自由に書いてください。」
回収数	世帯：1,293世帯（32.3%） 個人：2,118票回収（うち自由記述欄に記入ある1,570票を有効票とする）

3. 分析方法

（１）データベースの作成

アンケート調査の自由記述文に加え、個人属性、地区

表-2 データベースの作成

個人属性	性別（男性、女性）	
	年齢〔才〕	
	ふだん利用する交通手段（自動車、バス、鉄道、徒歩等）	
	職業（有職、学生・生徒・園児、専業主婦・主夫、無職）	
	家族人数〔人〕	
地区特性	住宅種類（一戸建て、アパート・マンション、社宅・官舎）	
	施設アクセシビリティ	市役所までの自動車所要時間・距離
		最寄り総合病院までの自動車所要時間・距離
		最寄り大規模小売店までの自動車所要時間・距離
		最寄り鉄道駅までの自動車所要時間・距離
		最寄りバス停までの徒歩所要時間・距離
		最寄り幼稚園までの徒歩所要時間・距離
		学区小学校までの徒歩所要時間・距離
		学区中学校までの徒歩・自転車所要時間・距離
		最寄り街区公園までの徒歩所要時間・距離
		最寄り一次遊離所までの徒歩所要時間・距離
		最寄り二次遊離所までの徒歩所要時間・距離
	水辺アクセシビリティ	最寄り水辺までの徒歩所要時間・距離
		利根川までの徒歩所要時間・距離
		広瀬川までの徒歩所要時間・距離
		桃の木までの徒歩所要時間・距離
	モビリティ	幹線道路までの距離
		最寄りバス停の運行頻度/バス停までの距離
	土地利用	地区の人口密度
		市街化区域（用途地域）、市街化区域

特性に関するデータベースを作成した（表-2）。個人属性データについてはアンケート調査データの回答値をサンプルに割り当て、地区特性データについては、対象世帯の属する町丁・大字別に、生活の質を分析した既存研究¹⁾を参考に、都市施設や水辺へのアクセシビリティ、モビリティ、土地利用を表わす指標を図上計測した。

（２）自由記述欄データの分析方法

本研究の目的を達成するためには、1)自由記述文を客観的・定量的な解析可能なデータに加工すること、2)データを分析仮説に基づき解析すること、が必要である。そこで本研究では、上記2点の課題に対応できるテキストマイニングのためのフリーウェアKH Cooder^{12) 13)}を活用する。KH Cooderは、1)のデータ化にあたり、形態素解析器として『茶筌』¹⁴⁾を用いており、これにより精度の高い客観的な単語抽出を行える。2)については、前橋のイメージを、形態素解析によって得られたキーワードの集合として捉えるだけでなく、キーワード間の関連性やキーワードがどのような文脈で出現しているかといった分析が行える。本研究では、キーワード間すなわち、語と語のつながりのパターンに着目した分析を行う。

語の関連性を自動的に得る方法は、大きく2つある。単語ベクトルを用いたベクトル空間手法と、言語コーパス中での確率を用いる確率手法である。ベクトル空間手法では、独立な事象の確率は足し合わせることができないため、内積を用いる関連度では、不利になる場合がある。そのため、本研究では、どのようなパターンにおいても、算出できる確率手法を採用し、中でも、比較的簡易なJaccard係数を用いることとした。

Jaccard 係数とは、語句 x と y がどれほど共起しているかを示す値である。値が大きくなるほど、共起関係が強いと考える。テキストデータ全ての集合を表-3 のように分割した上で、Jaccard 係数は次式で与えられる。

$$Jaccard(x, y) = \frac{a}{a + b + c} \quad (1)$$

表-3 語句 x と y に対する 2×2 の分割表

	語句 y が現れる	語句 y が現れない
語句 x が現れる	a	b
語句 x が現れない	c	d

3. 語の抽出と個人属性・地区特性との関係

（１）自由記述データからの語の抽出

総サンプル数 1,570 の自由記述データ全文に対して形態素解析を行った結果、上位の頻出語として表-4（次頁）の結果が得られた。総出現数であるため、同一サンプルの中に同じ語が記述されている場合も、出現回数と

してカウントしている。水・緑環境に関する語について着目すると、最上位には、「公園」、「緑」、「敷島」等の緑環境に関する語が出現している。また、「水」、「広瀬川」、「利根川」等の水環境に関する語も多く出現しており、水環境が都市イメージに影響を与えていることがうかがえる。意見に多く現れており、共著者間での協議に基づき、色掛けした水、緑に関する語を前橋の水・緑環境に関するキーワードとして抽出した。

表-4 語の抽出結果（頻出 90 語）

順位	抽出語	抽出数	順位	抽出語	抽出数	順位	抽出語	抽出数
1	公園	422	31	環境	59	61	悪い	28
2	多い	361	32	川	58	62	見る	28
3	緑	313	33	山	56	63	行く	28
4	良い	280	34	病院	56	64	上毛	28
5	敷島	228	35	場所	55	65	台風	28
6	少ない	216	36	都市	53	66	子供	27
7	水	179	37	空気	51	67	中心	27
8	前橋	174	38	施設	50	68	遊歩道	26
9	住む	166	39	生活	47	69	駅	25
10	思う	142	40	地震	47	70	見える	25
11	自然	136	41	赤城山	46	71	広い	25
12	やすい	134	42	車	45	72	特に	25
13	災害	133	43	美しい	44	73	ウオーク	23
14	静か	127	44	交通	42	74	安心	23
15	近い	110	45	人	42	75	温泉	23
16	広瀬川	109	46	整備	42	76	活気	23
17	バラ	92	47	散歩	38	77	四季	23
18	近く	88	48	安い	37	78	身近	23
19	街	85	49	感じる	36	79	東京	23
20	利根川	83	50	天災	35	80	三山	22
21	比較的	79	51	周辺	34	81	市	22
22	恵まれる	73	52	けやき	33	82	水害	22
23	県庁	72	53	物価	33	83	流れる	22
24	きれい	71	54	文化	33	84	スノーバー	21
25	好き	66	55	都会	31	85	出来る	21
26	便利	64	56	公共	30	86	松林	21
27	落ち着く	64	57	不便	30	87	地方	21
28	園	62	58	豊か	30	88	ケヤキ	20
29	町	60	59	たくさん	29	89	充実	20
30	道路	60	60	よい	28	90	地域	20

注：緑は緑に関する語、水は水に関する語と著者らが判断した。

（2）個人属性・地区特性との関係

自由記述データから抽出された単語のうち、水環境に関連する単語の出現率を、個人属性別にみると表-3 のようになる。水、広瀬川、利根川等の等の水環境に関する語の出現率は、年齢、職業に影響を受けていることがわかる。一方、性別、住宅種類による差異はみられな

表-5 個人属性別の水関連の語の出現率 **1%有意

個人属性	出現率	χ^2
性別	男性	44.0%
	女性	44.9%
年齢階層	年少（～15 才未満）	27.3%
	生産（15～64 才）	42.0%
	老年（65 才以上）	50.1%
職業	仕事をもっている	42.5%
	学生・生徒・園児	25.6%
	専業主婦・主夫	49.3%
	無職	47.6%
住宅種類	一戸建て	45.1%
	マンション・アパート	40.8%
	社宅・官舎	38.1%

い。地区特性のうちの河川へのアクセシビリティ別にみると（図-2）、アクセシビリティが高くなると水環境に関連する語の出現率が高くなる傾向がうかがえるものの、統計的な検定では、有意な差異は確認できなかった。

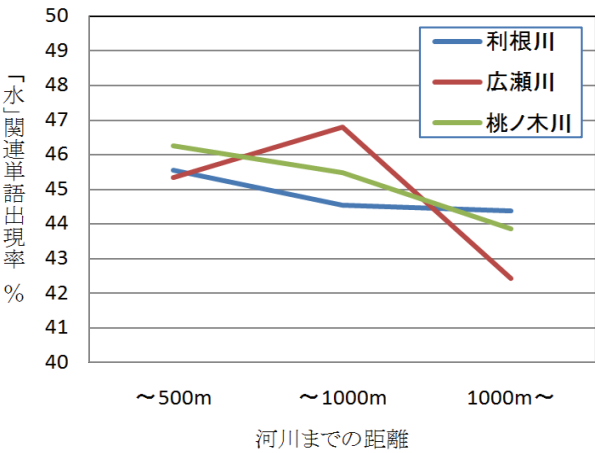


図-2 河川へ距離帯別の水関連の語の出現率

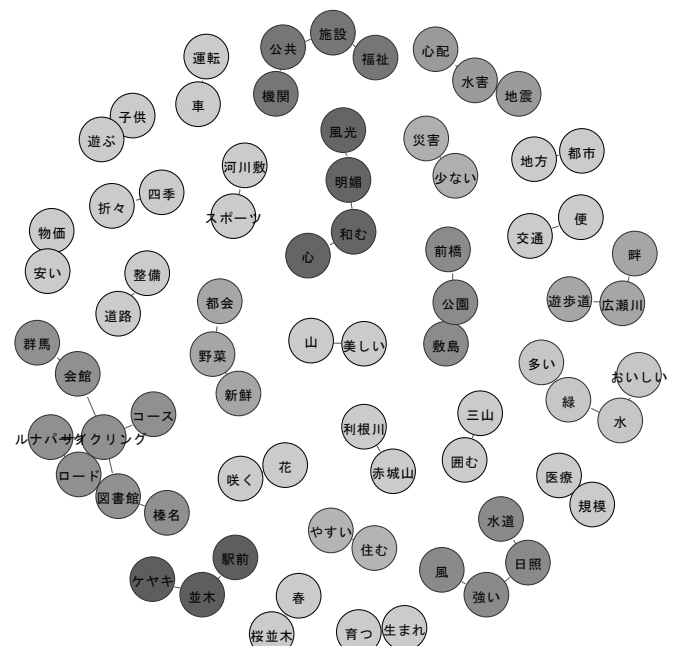
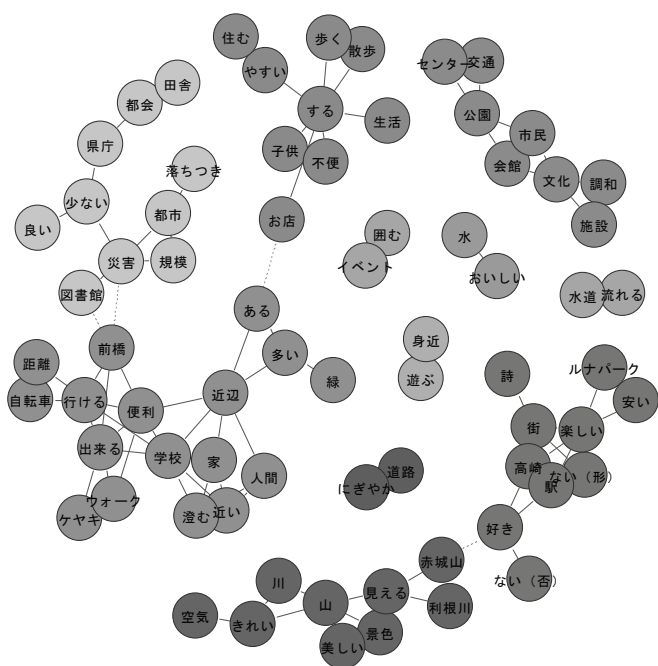
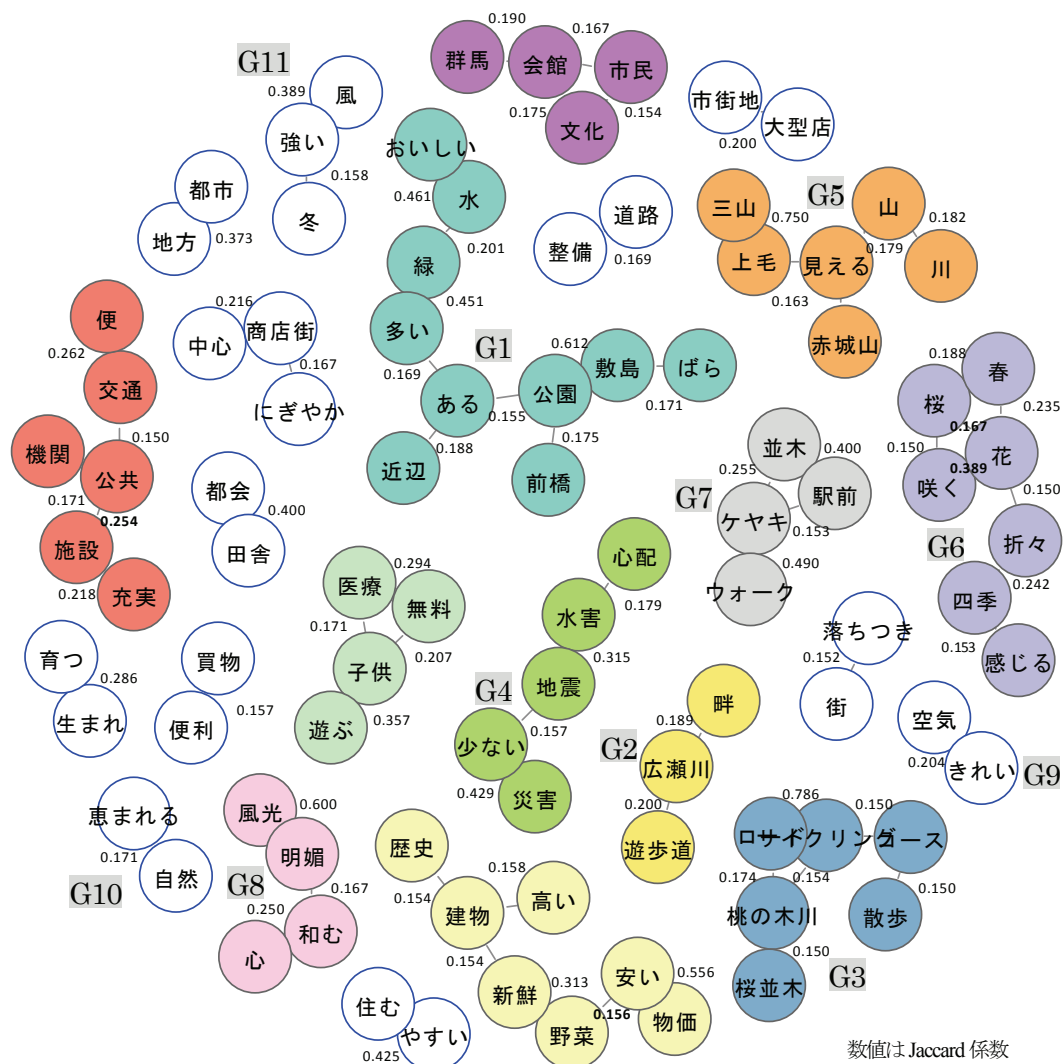
4. 都市・景観イメージ分析

（1）全体イメージの分析

都市・景観イメージを分析するために、つながりの強い語を結び視覚的に表現する共起ネットワークを用いる。共起関係の強さの尺度として、前述の Jaccard 係数を適用した。全サンプルを用い、Jaccard 係数が 0.15 の共起ネットワークを示したものが図-3（次頁）である。表-4 に示された水・緑環境のキーワード群について示す。

- G1：大規模公園である敷島公園・前橋公園、敷島浄水、を代表とする水・緑環境のイメージである。
- G2：前橋中心部の代表的景観である広瀬川と河畔の遊歩道のイメージである。
- G3：桃の木川と河畔の桜並木、サイクリングロード、そこでの散歩のイメージである。
- G4：地震、水害などの災害に関係するイメージである。
- G5：上毛三山、赤城山を主とする群馬、前橋の代表的なイメージである。
- G6：前橋の四季折々の花々のイメージである。
- G7：前橋駅から大規模ショッピングセンターへと続くケヤキ並木のイメージである。
- G8：心和む風光明媚な前橋のイメージである。
- G9：きれいな空気のイメージである。
- G10：自然に恵まれたイメージである。
- G11：冬の風の強いイメージである。

以上のように、利根川左岸の住民が抱いている前橋の都市・景観イメージを把握することができ、水・緑環境の位置づけが大きいことが明らかになった。



(2) 個人属性別の都市・景観イメージ

水関連の語の出現率との関連が確認できた個人属性について都市・景観イメージを分析する。

図4(前頁)の学生(サンプル数 78)の共起ネットワークをみると、水・緑環境に関連する語は、全サンプルの場合と比べまとまっており、赤城山、利根川の景観、山、川に関する群となっており、水・緑環境に関連する語の共起は少なく、自宅・学校周辺の身近な環境、都市施設に関連する共起がみられる。

次に、図5(前頁)の65歳以上(サンプル数 639)の共起ネットワークでは、全サンプルの語の群と類似しており、敷島公園・前橋公園の群、広瀬川と河畔の群、桜並木の群、災害に関する群がみられる。全サンプルの場合と異なるのは、医療施設、福祉施設、公共施設に関する語の群が発生している点である。

以上より、都市・景観イメージは、個人属性により異なる点と共通する部分が存在することを明らかにした。

5. おわりに

(1) 本研究のまとめ

市民の視点から前橋の都市・景観のイメージを把握することができ、個人属性、地区特性により異なる傾向のあること、すなわち、「たくさんの個人のイメージが重なりあった結果としてのイメージ」、ならびに、「それぞれかなりの数の市民たちによってつくられるイメージ」を把握することができた。さらに、水・緑環境がイメージの一部分を構成していることを明らかになった。

個人属性・地区特性の影響については、本研究においては、個人属性の方の影響が大きいという結果になった。個人属性によりイメージが異なることは、高齢化等の社会状況変化によりイメージが変化すること、地区特性により異なることは、都市イメージによる地域連携の可能性を示唆していると考えられる。

(2) 今後の研究課題

今後の研究課題を3点整理する。1つめは、イメージ構造の階層性を分析することであり、例えば、加齢によるイメージの形成過程を把握することである。2つめは、地域の広がりイメージの関係を把握し、都市・景観イメージによる地域連携の可能性を検討することである。3つめ、都市施策による都市・環境のイメージ変化を分析することである。

謝辞：本研究を遂行するにあたり、群馬工業高等専門学校環境都市工学科の長塩彩夏氏、専攻科環境工学専攻の青木清剛氏の協力を得た。本研究は、2009年度科学研究費補助金・基盤研究(C)(課題番号 20560499)の助成を受けた。また、本分析には、KH Cooder¹³⁾(著作権者、樋口耕一氏)を使用した。ここに記し、感謝の意を表す。

参考文献

- 1) Kevin Lynch: THE IMAGE OF THE CITY, The Massachusetts Institute of Technology and the President Fellows of Harvard College, 1960. ケビン・リンチ, 丹下健三・富田玲子訳: 都市のイメージ, 岩波書店, 1968.
- 2) 三浦周治・海宝弘和・大石幹也: 都市のイメージ理論の実証的検討とその関連 - ケース・スタディ 横浜市関内地区および周辺地域 -, 日本都市計画学会都市計画論文集, No.14, pp.409-414, 1979.
- 3) 角野幸博: 地域イメージの構成要素に関する研究 - 大阪府南北地域を事例に -, 日本都市計画学会都市計画論文集, No.16, pp.373-378, 1981.
- 4) 斉藤和夫・石崎裕幸・田村亨・梶谷有三: 都市のイメージ構造と地域特性の関係に関する分析, 土木学会土木計画学研究・論文集 Vol.14, pp.467-474, 1997.
- 5) 松浦茂樹・島谷幸宏: 都市の河川イメージの評価と河川環境整備計画, 土木学会土木計画学研究・論文集 Vol.4, pp.205-212, 1986.
- 6) 大塚裕子・乾孝司・奥村学: 意見分析エンジン計算言語学と社会学の接点 -, コロナ社, 2007.
- 7) 大塚裕子・内山将夫・井佐原均: 自由回答アンケートにおける要求意図判定基準, 自然言語処理, vol.11 No.2, pp.21-66, 2004.
- 8) 福田大輔・庭田美穂・屋井鉄雄: 疑問型表現自由回答データを用いた社会資本整備に対する市民の関心の抽出方法に関する基礎的研究, 土木学会土木計画学研究・論文集 Vol.24, pp.139-148, 2007.
- 9) 鄭蝦榮・羽鳥剛史・小林潔司・白松俊: ファセット学習モデルを用いた公的討議のプロトコル分析, 土木学会土木計画学研究発表会・講演集 Vol.36, 2007.
- 10) 前橋市: 前橋市緑の基本計画, 2008.
- 11) 森田哲夫・吉田朗・小島浩・馬場剛・樋野誠一: 都市環境に関わる諸施策を評価するモデルシステムの提案, 土木学会論文集 D vol.64 No.3, pp.457-472, 2008.
- 12) 樋口耕一: テキスト型データの計量的分析 - 2つのアプローチの峻別と統合 -, 理論と方法, 19(1), pp.101-115, 2004.
- 13) KH Cooder, <http://khc.sourceforge.net/>, 2010.5.7 (閲覧).
- 14) ChaSen「茶筌」形態素解析器, <http://chasen-legacy.sourceforge.jp/>, 2010.5.7 (閲覧).