

【研究ノート】

道央圏の観光地イメージに関する研究

— テキストマイニング手法を用いた観光地イメージ分析試論 —

伊藤 大介

1. はじめに

人間が観光行動をとる際に重要な役割を果たすのは、観光資源の存在を潜在的観光者に伝える広報・宣伝活動である。橋本 (1999:12) は、観光を「(観光者にとっての) 異郷において、よく知られているものを、ほんの少し、一時的な楽しみとして、売買すること」であると定義している。この定義に則って考えるならば、観光資源の広報・宣伝活動は、潜在的観光者に対して観光資源を「よく知られている」状態にする役割を担っていると考えられる。つまり、広報・宣伝により、有形無形の資源は、観光資源となりうるのである (増田 2000)。

このような広報・宣伝により潜在的観光者に伝達される主要な情報は、当該地域の観光資源の各種情報、そして当該地域のイメージである。そのような地域のイメージは、潜在的観光者の目的地決定に重要な影響を与える要因のひとつとされている。Kotler (1993[1996]) は、「まち」イメージは、人々がある「まち」について抱く固定概念、連想、印象の総称であるとして、「まち」マーケティングでは、買い手の選択にイメージが大きな影響を与えることを考慮しなければならないと述べている。

観光地イメージ¹を伝達する媒体には、自治体や旅行業者が発行する観光パンフレット、新聞、雑誌、文学作品、映像などの様々なメディアがある。これら様々な媒体のうち、潜在的観光者がアクセス容易な資料は観光パンフレットである。前田の調査 (前田 1995:110) によれば、観光地に関するパンフレット類 (政府機関・地方自治体および地方観光協会などが編集・配布する案内資料、案内マップなど) は、以下の機能を有している。それは、基本的に美しさ・楽しさを表現することで、その国・地方への来訪を動機づける機能 (興味づけ機能) と便利さ・正確さをもつことによって、その国・地方での旅行を容易にさせる機能 (有用性功能) である。そして、よいイメージづくりに貢献するのは、前者の機能であると前田は述べている。

Gunn (1997:37-38)は、観光地イメージを説明する概念として、目的地の組織イメージ (organic image) と誘導イメージ (induced image) を提示している。前者は、人間が様々なメディアや口コミ情報を通して得てきた知識の蓄積の結果、当該地域に対して抱くイメージである。後者は、目的地が主体となって、観光客を誘引する意図で行われる各種宣伝活動 (各種印刷メディアやTV・ラジオ放送) の過程やパブリシティの結果などで生み出されるイメージのことである。

以上をふまえて、本研究は、北海道・道央圏 79 自治体もしくはそれらの観光協会が発行した観光パンフレットにおいて、自らの地域をどのような言説を用いて紹介しているのかを分析し、その言説がどのような観光地イメージ (つまり、Gunn のいう誘導イメージ) を発信しているのかを明らかにすることが目的である。

2. 先行研究

観光地イメージ研究において、一般に用いられる手法は、アンケート調査である。アンケート調査を用いた観光地イメージ研究は、人間が当該地域に抱くイメージやメディアによって

発信されたイメージがどのように受容されているのかを調査する場合に有効であり、SD 法²が多用される。アンケート調査を用いた研究は、内田 (1986)、内田 (1987)、藤原ほか (1994)、伊藤 (1994)などがある。

しかし、内田の一連の研究 (内田 1997, 2004) は、アンケート調査から地域のイメージを明らかにする場合、研究目的によっては問題のない場合もあるが、アンケート調査が全戸調査ではない以上、必ずしもすべての人々の意見を反映しておらず、そのようなデータに定量分析を行うことの問題点やアンケート回答時の様々な条件の均一性に関する問題点を指摘している。そして上述の問題点を考えた場合、内田は、場所イメージ分析に利用可能なデータの種目を多様にする意味からも、アンケート調査以外の媒体をイメージ分析データとして利用する方法を検討し、観光パンフレットを分析データとして利用している。

内田 (1997) は、観光用パンフレット、ガイドマップ、市勢要覧などの観光客向け資料は、観光地としての公約数的な姿を、コンパクトに、同時に最大限に表現したテキストであるという前提のもと、それらの資料から場所イメージを定量的に求めている。なお、本研究における観光パンフレットの位置づけは、内田のこの前提を踏襲する。

内田 (1997) における対象地域は、新潟、富山、石川、福井、山梨、長野、岐阜、静岡、愛知各県の市制施行 123 市である。観光パンフレットなどの分析対象資料の記事の中で、その市町村の観光地紹介テキストのスペースを 100 としたときの、各トピック (記事) の占有率を計測・数値化し、類似したトピック同士にまとめて、57 項目に分類した。それらの項目に分類後、そのデータに対して主に主成分分析を行って、主成分を抽出することで、当該都市の観光イメージを構成している基本的構造を把握している。結果として、当該分析対象地域の観光パンフレットなどで表現されているイメージの構成要素は、「自然性」・「歴史性」・「都市性」の三次元 (尺度) であることを発見した。そして、その構造を詳細に検討すると、各県ごとの地域特性を反映して地域差があることが判明した。また内田 (2004) は、この研究と同様の分析方法で、中国・四国・九州地方における都市の観光地イメージについて研究を行ったものである。

テキスト自体から地域のイメージを分析している研究は、矢部ほか (1995) や遠藤 (2001) の研究がある。

矢部ほかの研究 (1996) は、地域景観イメージを明らかにすることを目的に、全国の小学校校歌歌詞 (2,665 校分) を分析している。この研究におけるテキストデータの分析方法は、歌詞中、地域に存在している事物を具体的に指示している名詞 (句) を景観要素として分類整理を行った後、各学校校歌の景観要素と要素項目出現率を集計している。結果として、校歌には「山」・「川」・「海」の三要素項目の出現率が高いことや各地の地域景観イメージのアイデンティティを表す景観要素が出現することなどを明らかにしている。

遠藤 (2001) は、奈良を研究対象地域として、観光情報誌・パンフレットなどのテキストメディアの言説を分析し、メディアの中で奈良が観光地としてどのようなイメージを付与されているのかを明らかにすると同時に、観光という事象が成立する際、メディアが果たしている役割を検討することを目的としている。この研究におけるテキストデータの分析方法は、テキストメディアに掲載されたキャッチフレーズ 154 件の抽出およびコーディングを行った後、設定された 20 カテゴリに分類するというものである。この結果から、先のテキストメディアで提供されている奈良のイメージで一番反復されているものは、「歴史」というカテゴリであることを明らかにした。その結果を受けて、過去の奈良における観光パンフレットと比較し、イメ

ージは再生産されなければならない、こうした観光地イメージの保持には、テキストメディアは大きな役割を果たすという点では、D. J. ブーアスティンの「疑似イベント」論³の妥当性を認めている。次に、観光客がイメージの受動的享受者に過ぎないというそのブーアスティンの仮説を検証するために、面接調査を行って奈良を訪問する観光客のイメージの読解を分析している。その結果、テキストメディアの言説における奈良のイメージ構造と観光客による奈良のイメージ構造にはズレがあるということが明らかになった。そして、そのズレは、観光客の抱く「観光地」のイメージは、観光メディアにおけるイメージを、観光客が自分たちだけの構造に変換すべく微細な変化を加えている結果であると結論づけている。

3. 本研究の分析手法

前章において述べた観光パンフレットのテキストを分析対象とした観光地（地域）イメージ研究を概観すると、内田（1997）では、テキストを占有率によって数量化し、分析者が設定した項目に分類を行っている。また、遠藤（2001）のテキストデータの数量化の方法は、分析対象のキャッチフレーズが、遠藤の設定した 20 項目に該当しているか否かを基準として数量化を行っている。

そのように分析者が設定した項目に分類を行う場合、分析者の恣意的な項目設定が行われて、分析結果にバイアスが生じる可能性や分析者があらかじめ意図したような分析結果しか得られない可能性がある。

本研究では、上述の先行研究における研究手法はとらず、テキストマイニング（Text Mining）手法を用いて分析を行う。テキストマイニング手法とは、「大量の文書情報の中から、質問の主旨に合致する文書を素早く発見するとともに、文書間の関連性を分析してさまざまにグルーピングし、それらの内容と数量およびその推移を把握することで、新たな知見を得るもの」（石井 2002:14）、また、「テキストデータをさまざまな計量的方法によって分析し、形式化されていない膨大なテキストデータという鉱脈のなかから言葉（キーワード）どうしにみられるパターンや規則性を見つけ、役に立ちそうな知識・情報を取り出そうとする手法・技術である」（藤井ほか 2005:10）と定義されている。

この手法を用いて分析を行う意義は、以下の 2 点にあると考えられる。第一の意義は、観光パンフレットにおけるテキストデータの中から、既存の尺度に依存せず、各自治体が観光パンフレットを用いて訴求するイメージを、テキストマイニング手法を用いて探索的に行うことで、観光地イメージの分析を行う点である。その結果、分析者の主観をある程度排除することが可能となる。先述の先行研究は、トピックを分析者の設定した項目に則って量的データに置換しており、分析者の主観が入り込む余地が多分にあると考えられる。しかし、テキストマイニング手法を用いることによって、そのような主観的項目設定が避けられ、分析者の恣意性を完全に排除することは不可能ではあるものの、その量的データへの置換基準は可視のものとなり、また大量のテキストデータを分析する際にも、その基準を一貫して堅持することが可能となる（太郎丸 2004）。第二の意義は、従来は時間的・労力的に分析が困難であった分量のテキストデータが、テキストマイニング手法を用いることで容易に分析可能となり、先行研究にはみられない観光パンフレットにおけるほぼ全文を分析対象としている点である。

なお、本研究は、道央圏の観光地イメージに関する論考であるのと同時に、テキストマイニング手法の観光地イメージ研究への応用の可能性に関する試論でもある。したがって、テキストマイニング手法の分析過程について詳述し、その過程における問題点の洗い出しを行い、今

後の研究への布石としたい。

4. 分析対象地域

北海道庁では、地域特性や可能性を活かした活力ある地域づくり促進を目的に、道南圏、道央圏、道北圏、オホーツク圏、十勝圏、釧路・根室圏の6地域生活経済圏の形成を進めている⁴。

そのうちの1つである道央圏は、2006年1月1日現在、石狩、空知、後志、胆振、日高各支庁、21市48町10村計79自治体から構成されている（図1参照）。本研究において道央圏の観光パンフレットを分析対象として選定した理由は、第一に、第3次北海道長期総合計画において圏域単一の観光施策目標と観光キャンペーンを有しているからである（2006年度は、「楽×楽 HOKKAIDO 道央圏地域観光キャンペーン」（北海道・北海道観光連盟）として6月1日から8月31日まで実施）。第二に、道央圏は、観光入込客数の割合が、北海道全体の半分以上を占め、北海道における観光の一大拠点となっているからである。以上の理由から、道央圏における観光地イメージを測定・把握することが、今後、当該地域の観光施策を進める上で必要性があると考えたのである。ちなみに、道央圏が北海道における一大観光拠点である要因は、圏内に新千歳空港を内包し、道外観光客の出入口になっていること、道内他圏域から札幌圏への観光客が多いこと、札幌市民の近郊への観光客が多いからであり、同時に海、山などに接し、多様な観光資源を有している地域であるからである。



図1 2006年1月1日現在の道央圏各地域および各自治体（北海道総合企画部計画室編 1998:10より作図）

5. 分析対象資料

5.1. 分析対象資料の収集

本研究において分析対象とした観光パンフレットは、筆者が各自自治体に赴き、観光案内所などで入手したもの、および各自自治体に依頼して送付を受けたものを使用した。ただし、ニセコ町、虻田町、壮瞥町、洞爺村の観光パンフレットは、諸般の事情により入手が不可能であった。ニセコ町に関しては、ニセコリゾート観光協会ホームページより、それ以外の3自治体に関しては、西胆振広域観光協議会発行の観光パンフレットから、当該自治体の紹介テキストを抜き出し、分析対象資料として使用した。

5.2. 分析対象資料の選定基準

一般に観光パンフレットと呼ばれるものには、旅行業者が、旅行商品の情報を顧客に紹介するもの、各種交通機関が、利用客増加を図って沿線観光資源を紹介するもの、各市町村が、当該地域の観光資源を紹介するものなどがあり多岐に渡る。

本研究では、誘導イメージ抽出が目的であるので、観光地である各自自治体が発行するもの、つまり道央圏各自自治体もしくはそれらの観光協会などが発行する観光パンフレットを分析対象とする。

本研究で用いた観光パンフレットの選定基準は、①道央圏各自自治体、または観光協会などの公的機関が発行元の観光パンフレットで、②1自治体につき1種類のみ採用する。ただし、1自治体において複数の観光パンフレットが発行されている場合は、(1)最新版、(2)情報量が豊富な資料、(3)記載内容の詳細な資料という三要件のいずれか、もしくは複数を満たす資料を採用する。

分析対象箇所は、上記基準を満たした観光パンフレットのうち、後述の除外箇所の該当項目を除いた観光地紹介の日本語表記テキスト部分である。

分析除外箇所は、外国語表記部分である。しかしアイヌ語表記については、カタカナ表記になっている地名などに関してはデータとして扱ったが、日本語表記にはない特殊表記は、日本語表記に置換可能なものについては対象とした(置換例:「ピッカコタン」⇒「ピリカコタン」)。そのほかの分析対象除外箇所は、ルビ表記部分、誘導イメージ生成に寄与しないと考えられるアクセス情報などの事実列挙データ(観光施設の所在地情報、開催・開館・各種営業・入館料情報、地図・写真およびそれらの説明キャプションのみの情報、観光パンフレットの目次、各種アクセス情報など)である。

6. 分析方法

本研究では、紙媒体である分析対象資料のテキストデータを電子化し、Text Mining for Clementine(以下、「TMC」と記す)を用いて形態素解析⁵を行う。そして、その出力結果を、コレスポネンス分析⁶やWebグラフ⁷によって視覚化し、道央圏各地域における観光地イメージを把握する。

TMCで自然言語データを分析する際、その対象となるデータは、電子化されている必要がある。したがって、TMCを用いた分析の前段階として、観光パンフレットにおけるテキストデータの電子化を行った。また、この電子化の段階で、TMCでの分析精度向上を目的に、テキストデータに対して、以下の二段階の前処理を行った。

①文章を区切る記号の付与

観光パンフレットにおけるテキストは、文章末に句点などの文章の区切りを示す記号が使用されていないキャッチフレーズや見出し語（ページのタイトルなど）が多く存在する。文章を区切る記号がないテキストデータを、そのまま TMC で分析すると、文章間の区切りが曖昧となり、前後の文章が連結されて認識される。そのように1つの文章として認識されることを防止するために、文字を区切る記号がないテキストに対しては、その文末に句点「。」を付与し、TMC による分析の精度向上を図った。

②各種記号などの処理

TMC では、各種記号（丸括弧「()」、中黒「・」、引用符「“ ”」など）が「主要語」⁸の構成要素として認識・分析される。しかし、観光パンフレット文中で用いられる絵記号や家紋などの記号は認識されない。そこで本研究では、TMC の分析精度向上を目的に、各種記号に対して以下の処理を行った。①丸括弧を用いたふりがな表記は、丸括弧およびその括弧内のふりがなを削除した。また、丸括弧を用いた言い換え語句表記は、同義のものについては丸括弧および括弧内の語句を削除し、単語出現頻度を算出する場合の重複算出を防止した。②並列・同格を表す場合や固有名詞には、中黒が用いられることがある。TMC では、この中黒が2つの単語間に置かれた場合、2つの単語を1つの単語として認識するので、並列・同格の意味で用いられている場合は、読点「、」に置換した（置換例：乗馬・テニス・アーチェリーをはじめ→乗馬、テニス、アーチェリーをはじめ）。③話し言葉によって文章が構成されている観光パンフレットがあるが、その場合も話し言葉特有の表記を、書き言葉に改めた。ただし、話し言葉が、施設名、商品名などの固有名詞として使用されている場合は、後述するデータクリーニングの段階で、単語登録辞書機能を用いて登録処理した。④そのほか、絵記号や家紋、丸囲み文字のような記号類は、TMC では分析が不可能であるので、テキストデータの電子化の段階で入力対象から除外した。

7. 分析

7.1 分析ツール

本研究の分析では、テキストマイニング過程は、Text Mining for Clementine2.0⁹、コレスポンド分析過程は、SPSS10.0J for Windows¹⁰を使用した。

Text Mining for Clementine によって本研究で用いるデータを処理した結果、総文章数は9,493、総主要語数は延べ数 60,745、異なり数 20,564 であった。

7.1.1. 分析対象品詞

本研究では、名詞（相当語句）を対象として分析を行う。仮にどのような「評価」をしているのか知りたい場合は、形容詞や形容動詞に着目して分析を行うべきであるが、本研究では、どのようなトピックを使用して地域のイメージを訴求しているのかを明らかにしたいので、名詞（相当語句）を対象にして分析を行う。ただし、TMC で用いられる品詞体系¹¹のうち、「名詞-非自立-一般」（例：「以外」、「以降」など）、「名詞-接尾-副詞可能」（例：「以降」、「以前」など）、「名詞-副詞可能」（例：「午後」、「金曜」など）といった品詞は、独立した意味を成さない語が多く、分析対象から除外し、「名詞-一般」（いわゆる普通名詞）をコレスポンド分析に使用する。

7.1.2. 独占度による重み付け処理

本研究では、主要語の出現頻度数に着目して、観光地イメージの分析を行う。確かに、表 1 における主要語のうち、「スポーツ」や「テニスコート」といった語は、直接的には観光地のイメージを表しているとはいいがたい。しかし、観光パンフレットにおいて、そのような語が頻繁に用いられているということは、当該地域が自らの地域を「スポーツ」などのレクリエーションを売りとした観光地として訴求したいと考えられる。つまり、出現頻度が多いほど、当該地域にとっては重要性が高く、同時にイメージの喚起力も大きいと考えられる（矢部ほか 1995）。したがって、本研究では主要語の出現頻度に着目して分析を行う。

しかし、主要語の出現頻度の大小を、そのまま観光地イメージの強弱として考えることはできない。なぜなら、出現頻度に依拠した場合、頁数およびテキストデータ量が多い観光パンフレットにおける使用諸単語が出現頻度の上位を占め、頁数およびテキストデータ量の少ない観光パンフレットのテキストデータに含まれる語は除外されて、特定の自治体のイメージだけが抽出されてしまう可能性がある。例えば、本研究で用いたデータにおいて、「三笠」という主要語がある。この主要語は、計 50 個がデータ内に存在し、13,233 個の主要語のうち、出現頻度が高い順にデータを並べると 54 番目に位置しているが、三笠市の市勢要覧にのみ出現している語である¹²。したがって、このような語に対して何らかの処理が必要となる。

そこで、本研究では、出現頻度数と並行して、独占度 (SPSS Japan Inc 2003:B-3) という指標を用いる。独占度とは、ある主要語が、各自治体において、どの程度独占して使用されているのかを測定する指標である。独占度は、1 が最大値で、例えば、道央圏全自治体のうち、ある単語 A が 1 自治体のみで使用されている場合の数値は、 $1/1=1$ をとる。ある単語 B が、全自治体のうち、3 自治体で使用されていた場合は、 $1/3=0.333\dots$ という数値をとる。つまり、ある単語が、ある自治体で集中的に使用されている場合は、独占度の数値は限りなく 1 に近づくのである。この独占度に依拠して主要語「三笠」をみると、独占度は 1 となり、独占度が低い順にデータの並べ替えを行うと、7514 番目に位置する。

テキストマイニング分析を行う場合、得られるキーワードの数が膨大となる。しかも、本研究で用いたデータにおいてもいえることであるが、それらの大半のキーワードは出現頻度が 1 回であり、テキストマイニング分析を行って得られたキーワードすべてに対してコレスポネンス分析を行うことは現実的ではない（上田ほか 2005:214）。したがって、本研究においては、独占度が 0.05 という数値を用いて、コレスポネンス分析を行う主要語を選定した。

なお、この 0.05 という数値には、理論的裏づけはなく、主要語が出現頻度上位 40 位程度までの数値を採用した

主要語	出現頻度	独占度
自然	294	0.021
歴史	139	0.036
緑	129	0.027
人	129	0.030
心	109	0.030
花	106	0.041
夏	102	0.036
冬	90	0.030
人気	89	0.027
温泉	75	0.045
姿	74	0.045
秋	72	0.030
山	68	0.040
コース	67	0.036
大地	67	0.033
風	65	0.041
春	65	0.035
イベント	61	0.033
森	61	0.042
味	60	0.035
川	53	0.049
魅力	50	0.037
大自然	49	0.037
全国	43	0.049
香り	43	0.045
季節	42	0.039
四季	38	0.050
手	38	0.048
子ども	37	0.045
文化	37	0.049
四季折々	36	0.043
資料	33	0.049
野菜	31	0.047
催し	29	0.046
汗	29	0.042
スポーツ	29	0.042
テニスコート	25	0.043

表 1 独占度 0.05 以上の「名詞一般」の主要語（出現頻度数の降順）

(表 1 参照)。この数値に関しては、独占度という指標自体の妥当性も含め、今後、関連研究などを参照し、理論的検討を加える必要があると考えている。

7.2. データクリーニング

7.2.1. 単語登録辞書機能を用いたデータクリーニング

1 語と認識されなければならないのにも関わらず、分割されて出力される場合(例「ししゃも祭り」→「ししゃも」と「祭り」の 2 形態素に分割)は、「ししゃも祭り」を 1 語の名詞として定義・登録することで、分析者の意図する主要語が出力されるようになる。

なお、観光パンフレットに含まれる標語、俳句、名言は、それら 1 文で 1 つの意味をなしており、形態素解析することで、その 1 文が持つ本来の意味が失われる可能性がある。したがって、そのような場合は、1 単語として登録処理を行った。そのほか、地名・施設名などの固有名詞もこの機能を用いて処理を行った。

7.2.2. 単語置換辞書機能を用いたデータクリーニング

単語置換辞書機能は、単語 A と単語 B という単語は同義のため、単語 B がデータ内に存在する場合は、単語 A として置換するために使用する機能である。本研究では、略称を正式名称に置換する場合に使用した。

このほか、テキストデータの電子化の際に、完全に排除できなかった記号類や係り先主要語を持たない主要語に対しても、Web グラフに出力する過程を考慮して、破棄設定を行った。

8. 分析結果

形態素解析を経た道央圏各自治体の観光パンフレットにおけるテキストデータの主要語を、道央圏内の地域別に分類し、コレスポネンダ分析した結果が、図 2 のイメージ・ポジショニングマップである。本章では、このイメージ・ポジショニングマップとともに、主要語と係り先主要語の関係性が視覚的に把握可能な Web グラフも、イメージ・ポジショニングマップの考察を深めるために使用した¹³。

原点付近に存在する主要語は、道央圏各地域において等しく使用されている主要語であり、道央圏全体の観光地イメージを代表するイメージでもある。原点付近には、「テニスコート」、「自然」といった語があり、道央圏全体では、レクリエーション施設や自然的観光資源を地域の特色として訴求している自治体が多いと考えられる。

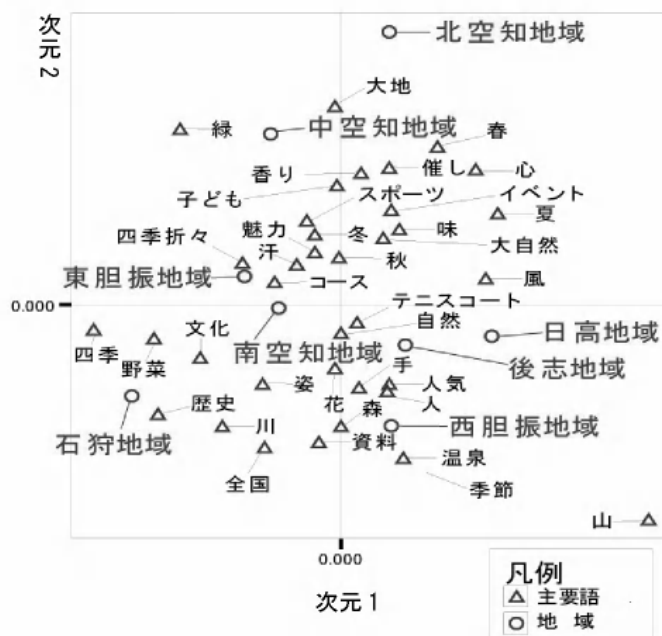


図 2 道央圏地域別の観光地イメージ・ポジショニングマップ

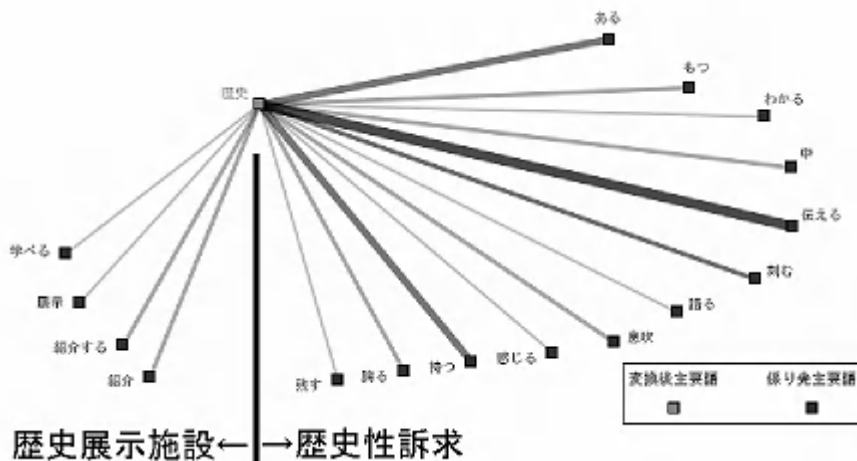


図3 主要語「歴史」とその修飾先を示したWeb グラフ

石狩地域には、「歴史」、「文化」といった語が至近に存在していることから、歴史のおよび文化的観光資源が強く訴求されていることがわかる。主要語「歴史」とその係り先主要語の関係を、Web グラフを用いて視覚化すると、当該地域の歴史性を訴求している場合と、歴史展示施設（郷土資料館など）を紹介している場合に分けられる（図3参照）。

西胆振地域は、「温泉」というイメージが強いが、この背景には、西胆振地域は、全国でも有数の温泉街である登別温泉や洞爺湖温泉を域内に内包していることがある。

図の上半分にみられる特徴としてあげられるのが、「スポーツ」、「汗」、「コース」といったレクリエーション関連の語である。これらのイメージが強い東胆振・南空知・中空知地域などは、都市機能が集積した札幌に近接した地域であり、都市部からレクリエーションが目的の入込客を対象とした観光施設が整備されていることが背景として考えられる。

北空知・中空知地域は、至近に配置される主要語の数が少なく、位置的にもポジショニングマップの外縁に位置しているが、その理由としては、これら地域は、産業基盤が稲作を中心とした農業地帯や観光地として後発の旧産炭地であるため、観光地としてイメージをあまり発していないと考えられる。

9. まとめ

本章では、前章で得られた道央圏各地域の観光地イメージの結果についてまとめ、さらに本研究の問題点について言及する。

まず、分析で得られた道央圏の観光地イメージについて整理する。

第一に、自然やレクリエーションといったイメージが、道央圏全体で等しく訴えられていることがわかった。

第二に、温泉・文化施設などの自然的・文化的観光資源が比較的多く存在する石狩、南空知、日高、後志、西胆振、東胆振に対して、中空知および北空知は、観光地として強いイメージがみられず、その要因には、当該地域が農業地帯や旧産炭地であるためであると考えられる。しかし、それら地域では、スポーツやイベントのイメージ訴求が強く、レクリエーション観光を全面に出した観光客誘致が観光パンフレットにおいて行われていることがわかった。

次に、本研究においてテキストマイニング分析処理を行った過程で浮上した問題点について述べ、その解決策を示す。

レスポンス分析に用いた主要語に「四季」および「四季折々」という語がある。これらは、当該観光地が、観光客を通年受入可能な観光資源を有していることを訴求している語である。このような語が用いられる理由は、観光需要の季節的変動は、観光関連企業の収益を低下させるため、オフ・シーズン対策が必要となるからであり、通年受入可能な観光施設やオフ・シーズンのイベントによって、自らの地域が四季を通じて楽しめる観光地であると訴求するのはどこの地域においても行われているためである。いわば「四季」や「四季折々」といった語は、観光パンフレットにおける常套語句であり、このような語が、当該地域の観光地イメージを訴求する語であるのかは疑問の余地が残る。したがって、今後は、そのような語を排除して分析を行った結果と比較し、どのように処理を行うことが妥当であるかを検証したい。

分析に用いた主要語に、「自然」と「大自然」、または「イベント」と「催し」という同義語があった。それらの語については、本来ならば TMC の単語置換辞書機能で処理すべきだが、原文を尊重し、今回は別々の主要語として扱った。今後、分析結果に大きな影響を与えない範囲で、同義語を同一の単語として扱う基準を設定し、この問題点を解決していきたい。

本研究は、観光パンフレットにおけるテキストの意味的側面に焦点を絞った研究であり、形態的側面（文字のフォントなど）は分析対象ではなかった。今後は、形態的側面に加え、パンフレットの大半を占める写真・イラストなども分析対象にする予定である。具体的には、写真・イラストは、先行研究（日高ほか（1998）、姫野ほか（2003）など）で行われている数量化手法を用いて、テキストデータと数値データを融合して扱い、データマイニング手法を用いて分析したい。また、テキストの形態的側面に関しても、今後、重み付け処理の手法などを考案していきたいと考えている。

謝辞

本論文の執筆にあたり、多くの助言・ご指導を賜り、各種分析ソフトなどの使用に関して便宜を図って下さった北海道大学大学院文学研究科・佐々木亨助教授には深く感謝いたします。また、ご多忙の中、観光パンフレット送付に協力して頂いた北海道内各自治体の観光関係部署の方々にも深く感謝いたします。

註

- 1) 場所のイメージに関する研究は、人類学、社会学、地理学、都市工学などの多様な学問領域において行われている。また場所のイメージは、対象・目的もしくは学問領域の違いによって、地域景観イメージ、まちイメージ、地理的イメージ、場所イメージ、地域イメージなど様々な呼称が存在するが、本研究では、観光パンフレットを分析対象としているので、観光地イメージという呼称を使用する。
- 2) SD 法 (semantic differential) とは、相対する意味を持つ形容詞の対を直線状の両端に配置し、回答者に点数を付けさせる方法で、イメージや感性などを問う際に用いられる質問方法である。
- 3) 疑似イベントとは、メディアに報道されることを期待して仕組まれたり、メディアの情報をアリの的に追認したりする、現実に対する関係が曖昧な出来事をさしている。ブーアスティンによれば、観光は疑似イベントの典型的一例であり、人びとはガイドブックや広告などの複製メディアによって表現された事象とおなじような経験をしたいと願うようになる。そして、現実によってイメージを確かめるのではなく、イメージによって現実を確かめるために旅行するようになるという（長谷編 1997:2）。
- 4) 北海道総合企画部計画室編（1998）を参照した。
- 5) 形態素解析とは、文章を、意味を持つ最小の言語単位である形態素に分かち書きするとともに、各形態素の品詞などを決定する機能である。

- 6) コレスポネンズ分析 (correspondence analysis)とは、クロス集計表を元データにして、行の要素と列の要素を多次元空間にビジュアルに表現する手法である。本研究では、行要素には地域名、列要素には主要語を用いて分析を行った。
- 7) 項目間の関連性の強さを、線の細太で視覚化できるグラフであり、本研究では「主要語」とその修飾先の語である「係り先主要語」との関係性把握に用いる。
- 8) TMCによる形態素解析を経ることによって分かち書きされた文の文節のうち、「て」、「に」、「を」、「は」などの機能語を除いたそれ自体が意味を持つ語が主要語である。このほか形態素解析処理によって、「主要語末尾品詞 (主要語の品詞情報)」、「係り先主要語 (主要語の修飾先)」などが構造化されたデータとして出力される。
- 9) Text Mining for Clementine は、データマイニングソフト Clementine のプラグインツールである。使用した Clementine のバージョンは、9.0 である。
- 10) SPSS でコレスポネンズ分析を行うためのオプションとして、SPSS 10J Categories を使用した。正規化の方法は対称的正規化である。コレスポネンズ分析を行う前段階として、本研究で用いたデータのクロス集計表にカイ 2 乗検定を適用した結果、有意と判定された。
- 11) TMC における形態素解析処理には、奈良先端科学技術大学院大学情報科学研究科自然言語処理講座が開発した形態素解析ツール「茶筌」が用いられているが、その「茶筌」は、情報処理振興事業協会 (IPA)が設定した IPA 品詞体系に基づいて形態素解析を行う。
- 12) 主要語「三笠」の品詞は、「名詞-固有名詞」である。本研究では、「名詞-一般」を分析対象としているが、ここでは、出現頻度と独占度を各々単独に依拠した場合の数値的な違いが、特に際立って確認できる「三笠」を例として示した。また 13,233 個の主要語は、係り先主要語を排除した後の異なり数であり、同時に「名詞-一般」以外の品詞も含んでいる数である。
- 13) このほか、イナーシャの寄与率 (各次元がもとの情報の何割を説明しているか示す数値) や行および列プロフィール (クロス集計における行または列の合計を 100%とした 2 元表) を参照している。

参考・引用文献

Boorstin, Daniel J

1963[1964] The Image: or What Happened to the American Dream. Harmondsworth (『幻影の時代』星野郁美 後藤和彦訳:創元新社)

遠藤英樹

2001 「観光という「イメージの織物」—奈良をイメージとした考察—『社会学評論』52(1):133-146

藤井和美 小杉考司 李政元

2005 『副詞・心理・看護のテキストマイニング入門』中央法規出版

藤原篤 福永太郎 川崎清

1994 「京都の都市イメージにおける伝統的様子の想起要因」『日本建築学計画系論文集』462:107-116

Gunn, A clare

1997 Vacationscape: developing tourist areas. Taylor & Francis

長谷政弘編

1997『観光学辞典』同文館出版

橋本和也

1999 『観光人類学の戦略—文化の売り方・売られ方—』世界思想社

日高圭一郎 伊藤解子 嶋心治 佐谷宣昭 坂井猛 萩島哲

1998 「観光画像情報からみた観光資源に関する一考察—北九州市を事例として—」『日本建築学計画系論文集』512:213-220

姫野由香 佐藤誠治 小林祐司 金チョン希

2003 「観光資源が写された景観画像の構図解析手法」『日本建築学計画系論文集』569:139-145

北海道経済部観光振興課

2004 『北海道観光の概況』北海道経済部観光振興課

北海道総合企画部計画室編

1998 『第3次北海道長期総合計画基本計画』北海道庁.

石井哲

2002 『テキストマイニング活用法』リックテレコム

伊藤悟

1994 「北陸地方における都市のイメージとその地域的背景」『人文地理』46(4):353-371

Kotler, Philip, Donald H Haider and Irving Rein

1993[1996] *MARKETING PLACES*. Free Press. (『地域のマーケティング』井関利明監訳 前田正子 千野博 井関俊幸訳:東洋経済新報社)

前田勇

1995 『観光とサービスの心理学 観光行動学序説』学文社

2003 『21世紀の観光学 展望と課題』学文社

増田辰良

2000 『文化経済学ライブラリー④観光の文化経済学』芙蓉書房出版

那須川哲也

2006 『テキストマイニングを使う技術／作る技術 基本技術と適用事例から導く本質と活用法』東京電機大学出版局

日本マーケティング・リサーチ協会編

2004 『マーケティング・リサーチ用語辞典—改訂新版—』同友館

大隅昇

2000 「調査における自由回答データの解析 - InfoMiner による探索的テキスト型データ解析 -」『統計数理』48(2):339-376

SPSS Japan Inc

2003 『Text Mining for Clementine トレーニング』SPSS

太郎丸博

2004 「特集 非定型データ分析の可能性」『理論と方法』19(2):131-133

内田治

2006 『すぐわかる SPSS によるアンケートのコレスポネンダ分析』東京出版

内田順文

1986 「都市の『風格』について—都市のイメージによる都市の評価の試み—」地理学評論 59(5):391-405

1987 「地名・場所・場所イメージ—場所イメージの記号化に関する試論—」『人文地理』39(5):391-405

1997 「中部地方における都市のイメージについて—観光パンフレットを用いた場所イメージの定量的分析の試み—」『国土館大学文学部人文学会紀要』31:134-147

2004 「中国・四国・九州地方における都市の観光イメージについて—観光パンフレットを用いた場所イメージの定量的分析の試み—」『国土館大学地理学報告』13:1-16

上田隆徳 黒岩祥太 戸谷圭子 豊田裕貴

2005 『テキストマイニングによるマーケティング調査』講談社サイエンティフィック

矢部恒彦 北原理雄 徳山郁義

1995 「小学校校歌に謳われた全国の地域景観イメージに関する研究」『日本建築学会計画系論文集』472:111-122

(いとう・だいすけ／北海道大学大学院)