

SNS の投稿に見る北海道と沖縄の観光イメージに関する考察

A Consideration of Destination Image of Hokkaido and Okinawa in the Comments on the SNS

泉澤 圭亮*

IZUMISAWA, Keisuke

本報告は北海道観光振興機構が行った「北海道観光に関する大都市圏マーケティング調査」において得られた結果が、インターネット上の北海道観光の個人投稿にどのように捉えられているのかについて明らかにするものである。北海道観光と沖縄観光に関する個人投稿をテキストマイニングの手法で解析し、ネットワーク分析を行った。結果、北海道観光の観光イメージには「都市」「交通」「景観」があり、沖縄観光の観光イメージには「海」「景観」があった。

キーワード：テキストマイニング、SNS、観光イメージ

1. 背景

北海道観光振興機構は「北海道観光に関する大都市圏マーケティング調査¹⁾」(以下、「大都市圏マーケティング調査」と記載)において、北海道・沖縄・他アジア地域間の魅力調査を行っている。魅力調査の結果の上位3件は、北海道が「自然風景が美しい」「食べ物が美味しい」「温泉が楽しめる」、沖縄は「海がきれい」「自然風景が美しい」「リゾートとして魅力がある」である。この結果は人びとへ意見を求め、その反応が反映されたものである。人びとの北海道および沖縄に対する「魅力」という概念は、大都市圏マーケティング調査の結果のように「カテゴリー化(categorization)」されているものと考えられる。

カテゴリー化というのは、未分類のものを分類することである。人は未分類の世界をカテゴリー化することによって秩序を形成する。たとえば「動物」「植物」「果物」「乗り物」「履き物」という分類は、人によるカテゴリー化の結果である。そして、カテゴリー化の結果として「概念(concept)」が生まれる²⁾。たとえば「イヌ」「ネコ」「ライオン」「キリン」「ゾウ」は「動物」というカテゴリーを表す概念である。

概念は概念のままに他者に示すことはできないが、それを別のものに置き換えて(=記号化)他者に提示することはできる。概念を記号化したものが言語である。言語の組み合わせによって意味が創出されたものは「コトバ」である。人びとが日常のコミュニケーションの手段として使用しているコトバというのは、発話者の視点(=ものの見方、切り取り方)が反映されていると言えよう。つまり、「自然風景が美しい」「食べ物が美味しい」「温泉が楽しめる」は、人びとの「北海道観光の魅力」というカテゴリーを表す概念が記号化され、最終的にコトバとして示されたものである。さらに、人びとの視点の結果として「自然風景が美しい」という魅力であったとしても、そこには多様な要素が包含されていると考えられる。

*北海商科大学大学院博士後期課程

2. 目的

本報告は、北海道観光と沖縄観光を対象にインターネット上の旅行に関する個人投稿をテキストマイニングの手法で解析を行う。ネットワーク理論を活用し、観光主体の視点から北海道観光と沖縄観光のキーワードを抽出する。その上で、大都市圏マーケティング調査で得られている北海道観光と沖縄観光の魅力の視点が、個人投稿においてどのように捉えられているのかを明らかにすることが目的である。

3. 方法

(1) 分析システム

本報告はテキストマイニングの手法を活用した。表-1 に主要な分析システムの構成を示す。形態素解析エンジンの MeCab がデフォルトで使用する形態素解析用の辞書は IPA 辞書 (IPADIC) である。IPA 辞書の体系は学校文法 (橋本式) が基本である。個人投稿のように文語や口語、新語の混在による多様な文章表現が想定されるテキストを解析の対象とする場合は、その多様性をより形態素解析の結果に反映させやすくするため、異なる辞書との組み合わせが必要であると推察される。そこで、IPA 辞書を含み 3 種類を組み合わせ使用した。

IPA 辞書の他に使用した第 1 の辞書は、インターネット上のフリーの百科事典である「Wikipedia 日本語版」のキーワード一覧^{*1}である。このファイルを MeCab の辞書形式にコンパイルしたものを使用した。この辞書のことを「Wikipedia 辞書」と呼ぶ。第 2 の辞書は、インターネット上の無料の Web 百科事典である「はてなキーワード」のキーワード一覧^{*2}である。Wikipedia 辞書と同様の処理をしたものを使用した。この辞書のことを「はてな辞書」と呼ぶ。

表-1 本報告の主要な分析システムの構成

用途	アプリケーション名	バージョン
基本ソフトウェア	Windows 日本語版	7, 64-bit
統計解析ツール	R	3.0.2, 64-bit
形態素解析エンジン	MeCab	0.996
形態素解析用辞書	IPA 辞書, Wikipedia 辞書, はてな辞書	なし
統合開発環境	RStudio	0.98.501
ネットワーク可視化ツール	Cytoscape	3.1.1, 64-bit

(2) 分析手順

1) 分析対象テキストの取得と処理

本報告が分析対象としたテキストは、すべてインターネット上の旅行クチコミサイト「フォートラベル^{*3}」に掲載されているものである。取得したのは 2014 年の投稿に分類されている「北海道旅行記^{*4}」および「沖縄旅行記^{*5}」である。掲載されている全件を取得した。北海道旅行記は 2,674 件、沖縄旅行記は 2,581 件である。

取得した旅行記のテキストには「本文」のほかに「タイトル」や「旅行の日付」が含まれてい

る。これらは本報告の分析には不要と判断し、本文のみを抽出した。その上で、北海道旅行記のテキストを「北海道原テキスト」、沖縄旅行記のテキストを「沖縄原テキスト」として、それぞれ保存した。

2) テキストクリーニング

テキストマイニングの手法を実施するにあたっては、その下準備として解析対象のテキストにテキストクリーニングと呼ばれる処理を行うことが望ましいとされている³⁾。

北海道原テキストおよび沖縄原テキストを、次の手順でテキストクリーニング処理した。「①半角カタカナを全角カタカナに統一した」「②ローマ字を大文字に統一した」「③改行を削除した」「④空白文字（スペース）を削除した」「⑤URL を削除した」「⑥記号を削除した」。

さらに、フォートラベルの旅行記は、旅行記中に記述された文字列と関連するコンテンツが付与されることがある。「場所の地図」「この施設のクチコミ」に一致する文字列は削除した。

以上の処理を行った北海道原テキストを「北海道分析用テキスト」、沖縄原テキストを「沖縄分析用テキスト」として、それぞれ保存した。

3) ネットワーク分析

語句のネットワークを明らかにするため、北海道分析用テキストおよび沖縄分析用テキストを n-gram 解析した。n=2 とし、bi-gram を取得した。bi-gram にあたり MeCab による形態素解析が実行された。bi-gram の結果から、モノやコトを表すことが可能で主語になり得る品詞として名詞を選び、名詞同士の語句の結合を抽出した。さらに品詞細分類から「固有名詞 - 固有名詞」「固有名詞 - 一般」「一般 - 固有名詞」「一般 - 一般」の組み合わせに分類された bi-gram を抽出した。

以上の結果から、長音符（＝音引き）が含まれる bi-gram を削除した上で、北海道分析用テキストおよび沖縄分析用テキストで頻出上位 1,000 組までを対象に共起ネットワークを作成し、最終的な分析対象とした。北海道分析用テキストの共起ネットワークを「北海道ネットワーク」、沖縄分析用テキストの共起ネットワークを「沖縄ネットワーク」と呼ぶ。

4. 結果

(1) 連結成分

北海道ネットワークおよび沖縄ネットワークの連結成分を抽出した。連結成分には「連結成分数」と「最大連結成分数」の2種類がある。

連結成分数は、ネットワークを構成している語句の連結をひとつずつ数えた総和を示す。「A-B」という2語句の連結も「C-D-E-F」という4語句の連結も、どちらも1として数える。「A-B」「C-D-E」「F-G-H-I」という語句の連結が同一のネットワークにあった場合、連結成分数は3である。最大連結成分数は、ネットワークの中で切断されずに連結している語句の最大数を示す。「A-B」「C-D-E」「F-G-H-I」という語句の連結が同一のネットワークにあった場合、最大連結成分数は「F-G-H-I」で、4である。

表-2 は北海道ネットワークと沖縄ネットワークの連結成分数および最大連結成分数である。連結成分数は北海道のほうが沖縄よりも多い。連結成分数は語句の連結をひとつずつ数えた総和である。北海道のほうが観光主体の視点が多様であることが推察される。最大連結成分数に大差は無いことから、北海道は個別的な視点が大きいものと考えられる。

表-2 北海道ネットワークと沖縄ネットワークの連結成分数および最大連結成分数

ネットワーク名	連結成分数	最大連結成分数
北海道ネットワーク	163	543
沖縄ネットワーク	130	555

(2) ネットワークの中心性

ネットワーク上でどの語句が中心にあるのかを判断するための指標のひとつに、ネットワークの中心性がある。中心性の高い語句は、観光主体の概念においても中心を構成していると考えられる。安田はネットワークの中心性を測る基準を3種類に分類し、①次数中心性、②近接中心性、③媒介中心性を挙げている⁴⁾。

①次数中心性は語句に繋がっている線(=辺)の数が多いほど、次数中心性が高いとする方法である。②近接中心性は任意の語句と他の任意の語句との距離が近いほど近接中心性が高いとする方法である。ネットワーク分析における距離とは、物理的な距離(センチメートルやインチ)で測るのではなく、起点の語句から目的の語句に至る間の語句の経由数のことを示す。③媒介中心性は任意の2語句を結ぶ最短経路上にある語句ほど媒介中心性が高いとする方法である。

本報告では3種類すべての中心性を求めた。北海道ネットワークの中心性の結果は表-3、沖縄ネットワークの中心性の結果は表-4である。

表-3 北海道ネットワークの中心性値と語句

次数中心性		近接中心性		媒介中心性	
語句	値	語句	値	語句	値
ホテル	43	北海道	0.287	ホテル	0.325
北海道	28	ホテル	0.285	北海道	0.293
札幌	28	札幌	0.283	札幌	0.228
バス	25	バス	0.265	小樽	0.159
小樽	22	レストラン	0.264	バス	0.141

表-4 沖縄ネットワークの中心性値と語句

次数中心性		近接中心性		媒介中心性	
語句	値	語句	値	語句	値
ホテル	75	ホテル	0.327	ホテル	0.348
海	62	海	0.317	海	0.304
ビーチ	48	沖縄	0.306	ビーチ	0.252
沖縄	37	ビーチ	0.301	沖縄	0.180
空港	27	空港	0.289	那覇	0.110

(3) 語句の結合関係

抽出した語句の位置関係を北海道ネットワーク図(図-1)、沖縄ネットワーク図(図-2)に記載した。4つの語句はいずれも最大連結成分内に存在しているため、ネットワークは最大連結成分

の部分のみを抽出した。図中の小さな丸い点は語句を表し、線は語句同士の結合を表している。

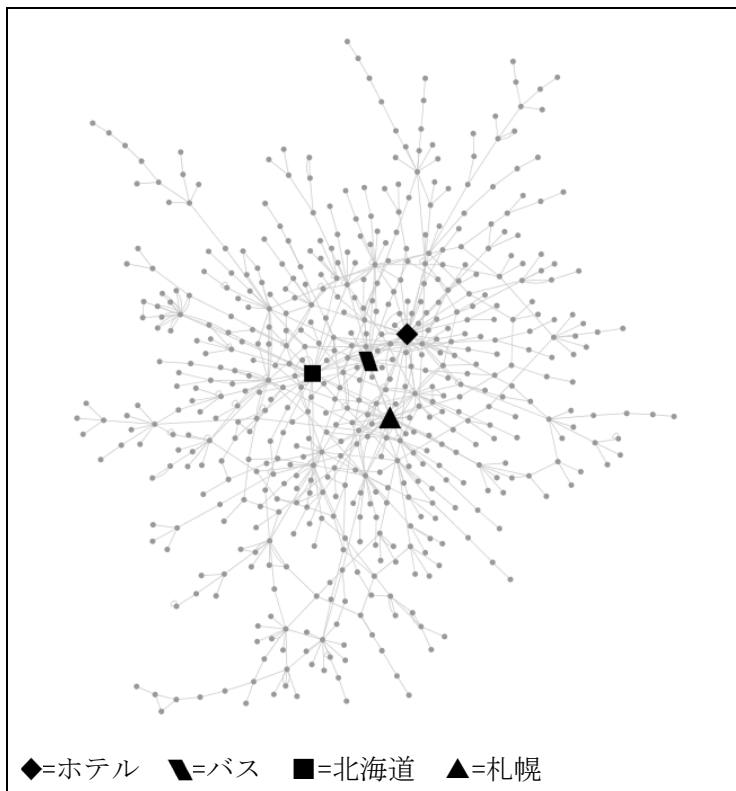


図-1 北海道ネットワーク図

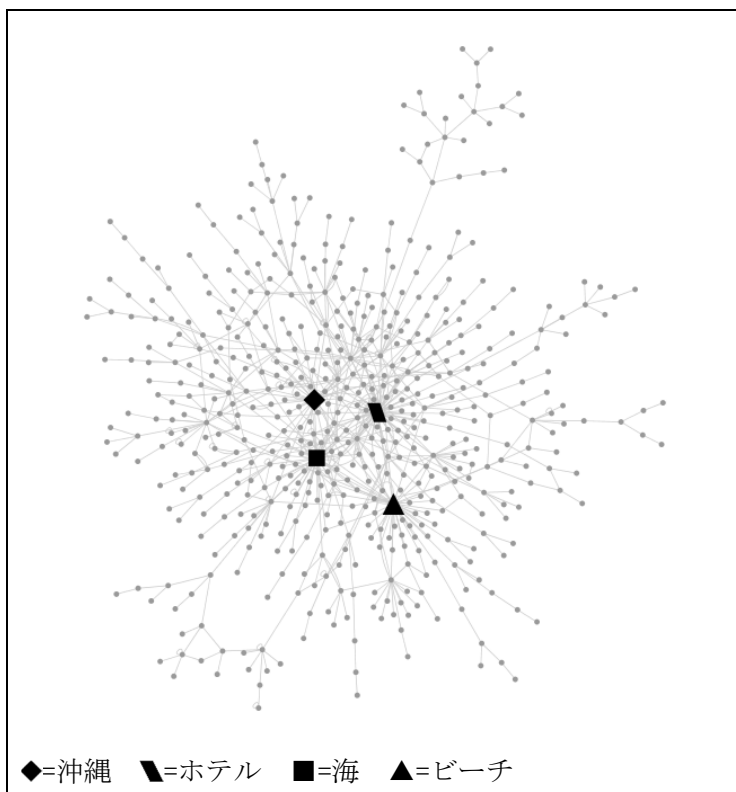


図-2 沖縄ネットワーク図

5. 考察

(1) 北海道旅行記

表-5は北海道ネットワークの中心性値の高い語句と共起関係の語句群である。

表-5 北海道ネットワークの中心性値の高い語句と共起関係の語句群

中心語句	共起関係にある語句群						
北海道	食材	風景	景色	夏	海	花	動物
バス	シャトル	トイレ	車内	車窓	美瑛	函館	空港
札幌	小樽	旭川	函館	釧路	市内	夜景	特急
ホテル	部屋	窓	リゾート	レストラン	温泉	旭川	

1) 北海道

「北海道」と「風景」「景色」「海」「花」があり、これらは景観に関する語句であると考えられる。さらに「食材」があり、これは「食に関する語句」である。これらの語句の結合関係は、大都市圏マーケティング調査にある「自然風景」と「食」というキーワードとの関係を見いだせる。

「風景」や「景色」は小学館の国語大辞典第二版によれば、どちらの語句も意味に大差はなく、「目にうつる自然の様子」である。自然の样子の全体を示しているとともに、「見る」という行為の対象でもある。さらに、「海」と「花」がある。これらも自然に関する語句であると考えられるが、全体を示す「風景」や「景色」に比べ、狭い範囲を示すものである。これは「海」や「花」の存在範囲が狭いということではなく、「風景」や「景色」に「海」や「花」が包含されている可能性があるということである。さらに、「動物」がある。これが野生生物なのか動物園等で飼育されている動物を示しているのかはわからないが、人工物でないことは明らかである。その意味では自然の要素に包含されると言えよう。

「食材」は、これも「風景」や「景色」と同様に全体性がある。北海道の特定の食というよりも、北海道の食という全体性に視点があるものと考えられる。北海道の食がひとつの視点で捉えられ、個別の食への視点が乏しく、北海道の多様な食が独自の魅力を構築できていない可能性が示唆される。

2) バス

「バス」は移動に関する語句である。「トイレ」がある。長距離あるいは長時間の移動が想定されるバスの場合、バスの車内にトイレが設置されている場合や、途中の休憩所でトイレ駐車を設けている場合がある。どちらのケースにしても、短距離には分類されない移動をしているものと考えられる。さらに「車内」「車窓」がある。移動手段という機能に加え、その空間へも視点を向けている可能性がある。

さらに、バスと結合している語句に地域名として「美瑛」「函館」がある。道内でバスが通っている地域はこの2地域だけではないが、美瑛・函館に関する移動手段としてバスが特異的に利用されている可能性がある。

3) 札幌

「札幌」は地域に関する語句であるとともに、平成26年度の観光入込客数は1,342万人⁵⁾で、

道内トップである。そして「札幌」という語句の中心性が高いということは、北海道を象徴する場所として札幌がイメージされている可能性がある。

「札幌」と直接結合の関係にある語句のうち「市内」「夜景」に関しては、札幌に内包されている要素として言及されたものと考えられる。「夜景」は一般的に夜間の人工光源の景色を指すと考えられ、ここで抽出されている「夜景」も同様であると推察される。ほかの「小樽」「旭川」「函館」「釧路」「特急」に関しては、札幌自体への言及というよりは、札幌に関連性があるものと考えられる。「特急」が抽出されており、これは電車の種類であると考えられる。札幌を經由して「小樽」「旭川」「函館」「釧路」へ電車を移動手段に移動している可能性がある。

そうであるとすれば、「バス」と直接結合の関係にある語句に「車内」や「車窓」が抽出されていた結果を踏まえ、移動の空間および移動に関わる景観というのが北海道観光のイメージを形成している要素のひとつと言える。

4) ホテル

「ホテル」は宿泊施設に関する語句である。観光主体にとっての宿泊施設の部屋というのは、一時的に自らの居住スペースとなり得るものである。「部屋」は、居住スペースについて言及されたものと考えられる。「窓」に関しても部屋の窓のことであると考えられ、窓は部屋の中から外を見ることのできる部分である。ホテルの部屋からの景観性への視点であると考えられる。そして、外（＝屋外）で眺めるというほかに、内（＝室内）からも眺めることのできる対象が北海道には見出されていると考えられる。

「温泉」が「ホテル」と直接結合の関係にある。北海道観光の魅力にも「温泉」があった。しかし、「ホテル」と「温泉」が直接結合の関係にあり、ホテルの媒介中心性値が高いということは、温泉そのものの重要性よりもホテルが有するひとつの機能としての温泉という視点の可能性はある。温泉に入ることが第一の目的というより、目的は別の部分にあり、温泉はその付属物になっている可能性がある。

(2) 沖縄旅行記

表-6 は沖縄ネットワークの中心性値の高い語句と共起関係の語句群である。

表-6 沖縄ネットワークの中心性値の高い語句と共起関係の語句群

中心語句	共起関係にある語句群						
沖縄	魚	石垣	空	風景	那覇空港	空港	
ホテル	部屋	プール	リゾート	レンタカー	空港	ゆいレール	
海	部屋	プール	エメラルド	栈橋	景色	魚	島
ビーチ	シュノーケル	エメラルド	砂浜	西表島	リゾート	プール	宮古島

1) 沖縄

「沖縄」と「風景」「魚」「空」があり、これらは景観に関する語句であると考えられる。これら

の語句の結合関係は、マーケティング調査結果にある「自然風景」というキーワードとの関係を見出すことができる。北海道旅行記と同様に、全体性を示す「風景」、そして個別性を示す「魚」「空」がある。

2) ホテル

「ホテル」は北海道でも抽出されていた。「部屋」と「リゾート」が北海道と沖縄とで共通項として抽出されている。沖縄の観光主体にとっても宿泊施設の部屋というのは、重要な要素であると言えよう。

さらに、「プール」がある。「ホテル」は宿泊施設であり、その基本的機能は泊まることである。そこに「プール」が結合しているということは、ホテルに見出されている視点が宿泊機能に限定されていないことを示していると考えられる。さらに「プール」は「ホテル」のほかに「海」「ビーチ」とも直接結合の関係にあり、「ホテル」「海」「ビーチ」が「プール」を介して相互に連結している。「プール」という要素は、中心をつなぐ役目を見出されており、沖縄観光のイメージを形成している語句と密接な関係にあると考えられる。

3) 海

「海」は沖縄の魅力調査の結果にも表れている語句である。「部屋」が直接結合の関係にある。「部屋」は「ホテル」とも直接結合の関係にあり、ホテルの部屋と海との物理的な距離、たとえば部屋からすぐに海に出ることができること、あるいは部屋から海を眺めることができることを示している可能性がある。「海」と「景色」が直接結合の関係にあることから、部屋からの海の眺めの可能性の方が高い。そうであるとすれば、部屋と海との間の眺望は確保の必要があると言えよう。

さらに「エメラルド」がある。「エメラルド」は「ビーチ」とも直接結合の関係にある。「エメラルド」という語句は、海洋博公園内にある「エメラルドビーチ」およびその名前の由来となった海のエメラルドグリーンという色を表したものであると考えられる。マーケティング調査結果にある「海がきれい」という沖縄の魅力は、「きれい」という部分だけを取り出すならば、ゴミが浮かんでいないことも「きれい」と言える。

しかし、沖縄の「海がきれい」という魅力には、色彩という要素が影響を及ぼしていると言えよう。また、「魚」という語句も抽出されており、それが「海」と直接結合の関係にあるのだから、海にいる魚のことを表しているものと考えられる。魚が見える海の透明度も沖縄にとっては重要な要素であると考えられる。

4) ビーチ

「ビーチ」は「海岸」を意味する語句で、全体性が含意される。直接結合の関係にある語に「シュノーケル」があり、これは「体験」に関する語句である。シュノーケルを用いたスポーツのことはシュノーケリングと呼ばれる。おもに浅瀬で行われ、海中での呼吸が可能であるため遊泳できる。海を活用したスポーツであるのだから、海的环境と密接な関係があると言えよう。さらに「砂浜」があり、「ビーチ」に含まれる「砂浜」という部分を示した語句であると考えられる。

6. まとめ

本報告は、北海道観光振興機構の「北海道観光に関する大都市圏マーケティング調査」で得ら

れている北海道観光の「自然風景」「食」「温泉」という魅力、および、沖縄観光の魅力を構成している「海」「自然風景」「リゾート」という魅力が、観光主体のイメージの中でどのように捉えられているのかを明らかにすることを目的とした。そのため、インターネット上の旅行クチコミサイトに投稿されている北海道観光と沖縄観光を対象とした個人旅行記を、テキストマイニングの手法で解析し、ネットワーク分析を行った。

結果、北海道観光の中心の観光イメージは「都市」「交通」「景観」であった。大都市圏マーケティング調査で挙げられている北海道観光の魅力と比較すると、「景観」という点で類似性が指摘できる。

他方、「食」と「温泉」は中心の観光イメージの構成要素として認知されていた。「食」や「温泉」は中心の観光イメージに付属している機能というイメージが強いものと考えられる。したがって、「食」や「温泉」を北海道観光の中心の観光イメージに位置づけるのならば、それらを北海道観光の包含できる観光資源へと変化させることが必要である。

沖縄観光の中心の観光イメージは「海」「景観」であった。共起ネットワーク上で媒介中心性値の高い語句が相互に連結していることから、沖縄観光の観光イメージは複合的な要素の連携で成立していることが示唆される。このことから大都市圏マーケティング調査における「海」「自然風景」という魅力と一致が見られる。そして、このような全体性の結果として、「リゾート」という視点が創出されているものと考えられる。

沖縄に見出されているイメージは「海」を中心とした統一感にあると推察される。海洋資源を保全していくことが必要であると指摘できる。

補注

*1 Wikipedia 日本語版のキーワード一覧は次の URL から取得できる。

(<https://dumps.wikimedia.org/jawiki/latest/>、2015年5月29日取得)。

*2 はてなキーワードのキーワード一覧は次の URL から取得できる。

(<http://developer.hatena.ne.jp/ja/documents/keyword/misc/catalog/>、2015年5月29日取得)。

*3 フォートラベル、(<http://4travel.jp>、2015年6月9日取得)。

*4 フォートラベル「北海道旅行記 (ブログ)」、(http://4travel.jp/dm_travelogue_list-todofuken-hokkaido.html、2015年6月9日取得)。

*5 フォートラベル「沖縄旅行記 (ブログ)」、(http://4travel.jp/dm_travelogue_list-todofuken-okinawa.html、2015年6月9日取得)。

参考文献

- 1) 北海道観光振興機構『北海道観光に関する大都市圏マーケティング調査概要版(2009年3月)』。
- 2) F.ウンゲラー・H.S.シュミット(1998)『認知言語学入門』池上嘉彦・坪井栄治郎・西村義樹・塩谷英一郎・野村益寛・本多啓訳、大修館書店、pp. 2-4。
- 3) 石田基弘・金明哲(編著)(2012)『コーパスとテキストマイニング』、共立出版、p. 2。
- 4) 安田雪(1997)『ネットワーク分析 何が行為を決定するか』、新曜社、p. 83。
- 5) 北海道経済部観光局『北海道観光入込客数調査報告書』、

(http://www.pref.hokkaido.lg.jp/kz/kkd/toukei/H26_irikomi_honbun.pdf、2016年7月29日取得)。

付記：本報告は2015年10月10日・11日に北海道北見市において開催された「北海道地域観光学会 第2回全国大会」で発表したものを再構成したものである。

(2016年7月31日受理)