

---

# ブログテキストマイニングによる 海外観光都市に関する消費者ニーズの探索的調査 ：モナコ公国を事例に

加藤 淳 一

---

## 要約

本研究は、海外の観光都市としてモナコ公国に注目して加藤・石川（2011）の手順（以下、KIP と略す）により市場セグメンテーションを行い、それらを捉える主成分軸を抽出することで消費者ニーズを解明する。

本研究により期待される貢献は大きく2つある。1つ目に本研究によりモナコ公国に対する日本人の消費者ニーズを明らかにすることで、モナコ公国はより一層の観光客を集める上で重要な消費者ニーズを把握する。

2つ目にテーマパークの経営者がハウステンボスと同様にヨーロッパ的リゾート地を競争相手としているならば、消費者が競争相手にどのようなニーズをもつのかを把握する。

第1点目の問題意識について、日本の消費者は心理的・精神的な癒しを求めており、自分の身の回りの出来事だけでなく政治・経済の出来事に関心を持ち、モナコだけでなく韓国にも関心を持っていると考えられる。モナコの観光担当者はこれらの点を考慮に入れることでより一層の集客が望める。

第2点目の問題意識について、日本の消費者は、おいしい食事、家族で過ごす時間、穏やかな天候にモナコの魅力を感じていると考えられた。テーマパーク経営者がヨーロッパ的なりゾート地を競争相手としているならば、これらの点を考慮に入れることで競争相手以上の魅力のあるテーマパークを作れる可能性が考えられる。

キーワード：マーケティング、ブログ、テキストマイニング、観光

## 1. 導入

日本国政府は2010年に観光を含む7つの分野からなる新成長戦略を閣議決定した。これに関連して観光庁は、「訪日外国人3000万人プログラム」を示した。さらに東日本大震災後、その影響を考慮した日本再生戦略が閣議決定された。その中でも、同様に観光市場を重要な成長の要と位置づけている。

このように観光市場に注目が集まる中で、千野（2011）は国内観光市場の予測と雇用への影響を

分析している。その結果、国内観光市場を維持・拡大し雇用も創出させるには、訪日外国人を3000万人にして日本の国内旅行回数を増加させる必要があると主張している。

このような観光市場への注目は国内だけでなく海外においても同様である。また研究方法も多様なアプローチがとられている。千野（2011）のような国内市場を対象にした経済学的研究が行われてきた。その一方で、観光市場を切り開くキーワードを探索的に探り、そのキーワードをもとにした観光市場に関するマーケティング研究が蓄積されている。本研究は多面的な観光市場創造に関連した研究の中で、ブログのテキストをデータとして消費者ニーズの解明を目指す。

マーケティング既存研究をレビューすれば、次のような研究で観光地に関連したブログテキストを分析した研究が蓄積され始めている。Pan et al. (2007) は、観光地についてのブログの記述が肯定的な内容かあるいは否定的な内容かの分類により、観光地の強みあるいは弱みを明らかにした。

Volo (2010)は、Carson (2008) やPan et al. (2007)などの先行研究をレビューして、先行研究は経験の概念を観光地の提供物へと単純化し操作化しており、ブログの読者や潜在的な旅行者への影響を研究していないと主張した。Volo (2010)の研究はブログデータにより観光地での消費者の経験を明らかにした。

三田村等 (2008)は旅行に関連したブログの分析だけでなく、旅行カテゴリーと一般カテゴリーという二つの比較により旅行に関連したブログの特徴を明らかにした。三田村等 (2008)は、ブログのテキストをデータとしてマイニングにより収集したブログに著者等が観光キーワードと呼ぶ単語の出現頻度を調査している

Kato (2012) は、これら先行研究をレビューした上でハウステンボスを対象にした探索的調査により市場セグメンテーションを行い、主成分分析によりそれら市場セグメンテーションを構成している消費者グループを捉える主成分軸を抽出した。その結果、海外リゾート軸が消費者ニーズを表す主成分軸の一つとして抽出された。

海外リゾート軸は、航空、東京、千葉、成田、モナコ、リゾート、モンテカルロ、ホテルなどのような単語によって特徴付けられていた。ここからハウステンボスの競争相手として、海外リゾート地やアジア近隣のヨーロッパ的リゾート地が考えられた。

確かにハウステンボスはオランダの街並みを再現することに力を注いだテーマパークであり（上之郷, 1992）、海外のリゾート地が競争相手としてあがってくることもうなずける。特に海外リゾート軸を特徴付けた単語として現れたモナコは海外リゾート地として知られている。

そこで本研究は、海外の観光都市としてKato (2012)の単語としてもあがっていたモナコ公国に注目して探索的な方法により市場セグメンテーションを行い、それらを捉える主成分軸を抽出することで消費者ニーズを解明する。

本研究により期待される貢献は大きく2つある。1つ目に本研究によりモナコ公国に対する日本人の消費者ニーズを明らかにすることで、モナコ公国はより一層の観光客を集める上で重要な消費者ニーズを把握できる。

2つ目にテーマパークの経営者がハウステンボスと同様にヨーロッパ的リゾート地を競争相手としているならば、消費者が競争相手にどのようなニーズをもつのかを把握できる。これは同様の競

争相手を持つテーマパークの経営者にとって経営戦略立案の一助となると考えられる。

本研究は探索的な手法として加藤・石川（2011）の手順（以下、KIPと略す）を採用する。多くの他の研究が工学研究をマーケティングデータに応用する中で、この手法はマーケティング既存研究に依拠した分析手順を提案した点に特徴がある。節を改めて分析方法の説明を行う。

## 2. 方法

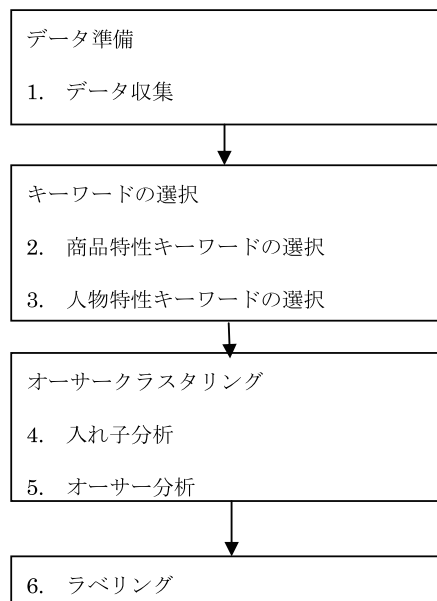
### （1）KIP の概略

KIP の概略は図表 1 のように整理できる。以下で、一つひとつの KIP の手順を説明する。ステップ 1 でデータの収集を行う。データは次のように準備する。まず、分析したい対象を表す単語 1 語（本研究の場合「モナコ」）を定め、goo ブログ検索してその単語を一度でも使った goo ブログオーサーを収集する。次に、このブログオーサーが書いた全ブログ記事を収集する。収集したブログ記事は全て単語にして以下の分析で用いる。これらがステップ 1 である。

KIP はステップ 2 と 3 でブログオーサーをグループにまとめるときに基準となる 2 つの単語群を選択する。グループへのまとめ方の大まかなイメージは次のように説明できる。ブログオーサーがある単語を似通った頻度で使用していれば、それらブログオーサーは同じグループとしてまとめられる。こうしてまとめられたブログオーサーが市場セグメンテーションである。

ステップ 2 では、1 つ目の市場セグメンテーションの基準としてモナコと類似度の高い単語をステップ 1 で収集した単語の中から選択する。これを商品特性キーワードと呼ぶ。ステップ 3 では、2 つ目の市場セグメンテーションの基準としてオーサーを特徴付ける単語をステップ 1 で収集した単語の中から選択する。これを人物特性キーワードと呼ぶ。

図表 1. KIP の概略



ステップ4では、モノコと類似した単語とオーサーを特徴付ける単語の2つの基準で分類するために、商品特性キーワードと人物特性キーワードでオーサーをクラスタリングする。これを入れ子分析と呼ぶ。

なお、ステップ5でオーサー分析を行うならば、ステップ4を実施する必要はない。入れ子分析により、市場セグメンテーションできる。だがマーケティングとしては、ロイヤリティの高いオーサーと低いオーサーに分けて分析するとより有益である。そこでステップ5では、ロイヤリティの高いロイヤルオーサーと低いロングテイルオーサーに分け、オーサーをクラスタリングする。

まず、商品特性キーワードでオーサーをクラスタリングする。次に、クラスタリングされたオーサーからロイヤルオーサーとロングテイルオーサーを選択する。最後に、選択されたオーサーを人物特性キーワードによりクラスタリングする。これをオーサー分析という。

ステップ6では、ロイヤルオーサーとロングテイルオーサーのそれぞれのオーサー分析の結果を受けて、それらを捉える主成分軸を求める。この軸に固有ベクトルを手がかりにラベルを貼る。これをラベリングという。これらがブログ分析の手順となる。次に最後のステップのラベリングについて、より詳細に説明する。

## (2) ラベリングの方法

これまでのラベリングの方法は次のようであった。まずロイヤルオーサーとロングテイルオーサーのそれぞれに対して  $tf \cdot idf$  値により重要度の高い単語に絞る。次いで、それら重要度の高い単語を手がかりにラベリングを行う。このような方法であった。だがあまりに多くの単語からラベルを作成するのは困難な作業であった。そこで本研究では次のような方法を採用する。

ロイヤルオーサーとロングテイルオーサーのそれぞれに対して、ロイヤルオーサー・単語行列とロングテイルオーサー・単語行列を作成する。行列の要素は、全て単語頻度である。この行列に対してカイ二乗統計量とその  $p$  値を計算した。例えば、ロイヤルオーサーが4つのオーサークラスターにクラスタリングされたとすれば、カイ二乗統計量と  $p$  値を求めるデータは次のようになる。

図表2 オーサークラスター・単語行列

	オーサー クラスター0	オーサー クラスター1	オーサー クラスター2	オーサー クラスター3
単語1	単語1の頻度	単語1の頻度	単語1の頻度	単語1の頻度
単語2	単語2の頻度	単語2の頻度	単語2の頻度	単語2の頻度
⋮	⋮	⋮	⋮	⋮
単語N	単語Nの頻度	単語Nの頻度	単語Nの頻度	単語Nの頻度
オーサークラスター毎の単語頻度合計	単語頻度合計0	単語頻度合計1	単語頻度合計2	単語頻度合計3

この図表2から単語1のカイ二乗統計量とp値を計算するときには、次の図表3を作成してカイ二乗統計量とp値を計算する。プログラムを作成して、この計算を単語数分繰り返して計算する。

図表3 単語数削減のためのカイ二乗統計量とp値の計算用の行列

	オーサー クラスター0	オーサー クラスター1	オーサー クラスター2	オーサー クラスター3
単語1	単語1の頻度	単語1の頻度	単語1の頻度	単語1の頻度
オーサークラス ター毎の <u>単語1</u> <u>以外</u> の単語頻度 合計	単語頻度合計0 －単語1の頻度	単語頻度合計1 －単語1の頻度	単語頻度合計2 －単語1の頻度	単語頻度合計3 －単語1の頻度

計算したカイ二乗統計量とp値は各単語に1つずつなので、その値をもとに単語を並び替える。並び替えの基準は、第1基準としてp値の昇順、第2基準としてカイ二乗統計量の降順とした。2つの基準で並び替えた理由は複数の単語のp値がゼロとなることへの対応策である。

並び替えた単語は、ロイヤルオーサーとロングテイルオーサーを特徴付ける程度により並んでいる。この上から0.1%の単語を以後の分析対象とし、それ以外を削除する。これによりオーサーを特徴付ける程度の高い単語へと単語数を減少させた。なお、0.1%という基準は単語数から恣意的に決めており、単語数により分析の対象とする比率を変更する。

それでもまだ単語数が多くラベルを貼るのは難しい。そこで、この0.1%を対象にして主成分分析を行い、4つのクラスターを捉える主成分軸を求め、その主成分軸にラベルを貼る。この主成分分析のデータは、図表3からカイ二乗統計量とp値を計算する時点で同じプログラムにより計算した各オーサークラスターのカイ二乗統計量である。

各オーサークラスターのカイ二乗統計量は、次の図表4のような行列をデータとして計算する。この図表4は、単語1のオーサークラスター0のカイ二乗統計量を計算する状況を例示している。

図表4 主成分分析のためのカイ二乗統計量の計算用の行列

	オーサー クラスター0	<u>オーサークラスター0以外</u> のオーサークラスター
単語1	単語1の頻度	単語1の頻度合計－単語1の頻度
オーサークラスター毎の <u>単語1以外</u> の単語頻度合計	単語頻度合計0 －単語1の頻度	単語頻度総合計－単語頻度合計0－(単語1の頻度合計－単語1の頻度)

このようなデータによるカイ二乗統計量の計算を、単語とオーサークラスターの組み合わせの数だけ繰り返す。この計算は図表3の計算で用いた同じプログラム中で計算しておく。この計算結果を主成分分析にかける。

主成分分析の第1段階は主成分軸の抽出である。ここでの主成分分析は、次のように置き換えて考えれば通常の主成分分析と同様に理解できる。4つのオーサークラスターを4人の被験者と見立

てる。そして、4つのオーサークラスターのオーサーが使用した単語群を4人の被験者が受験した試験科目に見立てる。

すると、4人の被験者が複数の試験科目を受験し、その結果により4人を捉える文系軸と理系軸のような主成分軸を抽出できる。こうした分析と同様に、4つのオーサークラスターで構成される市場を捉える軸を抽出したことになる。

より詳細な主成分分析の説明は例えば三土(1997)を参照いただきたい。市場を捉える主成分軸の数は、累積寄与率を計算することで決定できる。およそ8割程度の累積寄与率に至るまでを採用する。

主成分軸が定まると、第2段階はそれぞれの主成分軸にラベルを貼るときの根拠となる単語の選択である。この単語選択は固有ベクトルを用いて行う。主成分分析により、各単語に対して固有ベクトルの値が計算されている。

この固有ベクトルの値は、その単語にかかる係数の大きさを示している。したがって、固有ベクトルの絶対値の大きい単語に注目することで、主成分軸にラベルを貼るときの根拠として適切な単語を選択できる。

ここで一点、固有ベクトルの符号について注意を要する。負の符号はネガティブな意味を表さない。逆に正の符号はポジティブな意味を持っていない、ということである。したがって、正の符号の最大の絶対値と負の符号の最大の絶対値を比較して絶対値の大きい方の符号により軸を解釈する。

正の符号を採用するとなれば、解釈は正の絶対値の大きい単語から順に注目していなければならない。他方、負の符号を採用するとなれば、解釈は負の絶対値の大きい単語から順に注目していなければならない。

特に根拠はないものの、本研究では絶対値の大きい方から50単語に注目して主成分軸にラベルを貼る。これはあまりに多数の単語からではラベルを貼るのが困難だからである。このようにして主成分軸にラベリングする。以上のようなラベリングをロイヤルオーサーとロングテイルオーサーのそれぞれで別々に行う。

### 3. 結果

ターゲットキーワードは、モノコである。goo ブログ検索を行ない、ブログ記事を得た。次いでそれらのブログオーサーの全ブログ記事を収集した。収集したオーサーは482名である。90%以上の記事が得られなかったオーサーは482名から除外した。その結果、439名になった。

最終的に得られたのは、オーサー439名、延690431記事、出現単語総数は1685540語であった。収集期間は、2004年3月9日から2012年5月8日である。収集したブログ記事から本文テキストを抜き出し、形態素解析を行い、ブログ記事の単語(以下、単語は一部を除き名詞)のみ抽出した。

ステップ2は、商品特性キーワードの設定である。閾値以上にモノコと類似している単語を商品特性キーワードとして採用する。類似度はコサイン尺度を用いた。類似度の累積相対度数を求め、8割以上までの単語を採用することとした。類似度の閾値を0.00442646594136とし、商品特性キーワードとして20754語を選択した。

ステップ3では、人物特性キーワードを設定する。閾値以上にオーサーを特徴付けている程度の高い単語を人物特性キーワードとして採用する。この特徴付けている程度の閾値を379.947312503とし、人物特性キーワードとして20755語を選択した。

ステップ4で、オーサー分析を行った。まずブログオーサーごとにブログ記事をまとめる。ステップ2で作成した商品特性キーワードでオーサーを分類する。この分類により商品特性キーワードに関して類似したオーサーをまとめる。商品特性キーワードによりオーサーを4つにクラスタリングした。

これら4つについてオーサー毎に商品特性キーワードの重要度の相対比率を求め、クラスター単位で上側25%を平均した。平均値の最大のクラスターをロイヤルオーサー、最小のクラスターをロングテイルオーサーとした。ロイヤルオーサーは160名、ロングテイルオーサーは63名となった。

商品特性キーワードで分類されたオーサーごとに、ステップ3で作成した人物特性キーワードで分類する。この分類で、人物特性キーワードがオーサーを特徴付けている程度により、160名のロイヤルオーサー（425884語）と63名のロングテイルオーサー（400841語）のそれぞれを4つにクラスタリングした。

ステップ6は、ラベリングである。ロイヤルオーサーの4つのクラスターは、それぞれ71007語（37名）、117820語（56名）、329581語（52名）、そして47842語（15名）を含んでいた。ロングテイルオーサーの4つのクラスターは、それぞれ155339語（26名）、140635語（6名）、130326語（22名）、そして196633語（9名）を含んでいた。

つづいて、各単語の重要度をカイ二乗統計量のp値により計算した。p値が小さいほど、その単語とロイヤルオーサーまたはロングテイルオーサーとが独立である確率が小さい。したがって、その単語がよりロイヤルオーサーまたはロングテイルオーサーを特徴付けていると解釈できる。

ここから重要度の高いロイヤルオーサー426語（全体の0.1%）とロングテイルオーサー401語（全体の0.1%）に絞った。最後に、426語と401語に対し主成分分析を行い、ロイヤルオーサーとロングテイルオーサーの主成分軸を抽出した。

図表5 ロイヤルオーサーの寄与率と累積寄与率

	PC1	PC2	PC3	PC4
Standard deviation	15.0157	10.4176	9.592	4.696e-15
Proportion of Variance	0.5293	0.2547	0.216	0.000e+00
Cumulative Proportion	0.5293	0.7840	1.000	1.000e+00

図表5によれば、第2主成分まででロイヤルオーサーの累積寄与率の約8割を捉えている。したがって、ロイヤルオーサーを捉える主成分軸は第2主成分までとする。

図表6 ロイヤルオーサーの主成分得点

	PC1	PC2	PC3	PC4
[1,]	-11.950867	7.567350	-10.009062	-1.116642e-14
[2,]	15.925966	-8.606149	-6.381342	-7.545093e-14
[3,]	9.675543	10.360798	8.820629	1.088448e-13
[4,]	-13.650641	-9.322000	7.569776	-2.198502e-14

図表6によれば、第1主成分は正の値で15.925966と負の値で-13.650641と-11.950867である。特に正の値が最大である。第2主成分も同様の傾向を示している。

図表7 ロイヤルオーサーの第1主成分の固有ベクトル（絶対値の大きい方から50語）

季節 0.06652818	部屋 0.06569606	蕾 0.06557826	葉 0.06533240	我が家 0.06526475
サラダ 0.06524242	苗 0.06511473	鉢 0.06504760	ピンク 0.06500885	野菜 0.06493324
白 0.06490336	手作り 0.06486896	窓 0.06483573	時間 0.06482791	香り 0.06478335
写真 0.06468720	テーブル 0.06461001	店 0.06455197	トマト 0.06440058	布団 0.06413576
パスタ 0.06405774	木 0.06402304	庭 0.06396192	スープ 0.06389088	ゴルフ 0.06363900
花 0.06345813	味 0.06316209	競馬 0.06310320	ワイン 0.06288759	歌手 0.06282399
ランチ 0.06281815	教室 0.06279059	カフェ 0.06277345	綺麗 0.06269654	馬券 0.06253614
お茶 0.06247650	紅茶 0.06236529	夕食 0.06231654	感じ 0.06222758	水曜日 0.06213619
レストラン 0.06208713	チーズ 0.06200532	チョコ 0.06189683	デザート 0.06188213	皆さん 0.06179099
土 0.06171167	フード 0.06169734	チョコレート 0.06166727	種類 0.06157188	新宿 0.06148441

図表7によれば、ロイヤルオーサーの第1主成分を特徴付ける単語として、「季節、我が家、サラダ、野菜、手作り、トマト、パスタ、スープ、味、ワイン、ランチ、カフェ、お茶、紅茶、夕食、レストラン、チーズ、チョコ、デザート、フード、チョコレート」などがある。



図表8 ロイヤルオーサーの第2主成分の固有ベクトル（絶対値の大きい方から50語）

バラ 0.08267026	楽しみ 0.08094366	主人 0.07792500	夕方 0.07764271	今日 0.07754695
天気 0.07744579	家族 0.07736885	今朝 0.07615509	雨 0.07407796	家 0.07315218
気持ち 0.07299057	朝 0.07091652	毎日 0.07090395	たくさん 0.07082058	品種 0.06973020
明日 0.06895345	そう 0.06774063	大好き 0.06604074	ほん 0.06579327	ドイツ 0.06464629
S 0.06403868	元気 0.06348305	友達 0.06258098	きれい 0.06226917	月 0.06051629
子ども 0.06006079	軽井沢 0.06002147	ロ 0.05968864	イ 0.05883313	猫 0.05876744
Salon 0.05869961	じい 0.05869405	バロック 0.05863340	バレエ 0.05862930	ジブシー 0.05821053
フィル 0.05820342	モナコ 0.05819641	バリトン 0.05811496	追分 0.05809071	レーベル 0.05790207
変 0.05787060	web 0.05780883	テノール 0.05780191	ハ 0.05776584	Goo 0.05773217
歌劇 0.05770881	ライブ 0.05770737	Vn 0.05770667	Pf 0.05770667	SACD 0.05770667

図表8によれば、ロイヤルオーサーの第2主成分を特徴付ける単語として、「主人、夕方、今日、天気、家族、今朝、雨、家、朝、明日、元気、友達、子ども、じい」などがある。

図表9 ロングテイルオーサーの寄与率と累積寄与率

	PC1	PC2	PC3	PC4
Standard deviation	13.523	10.8980	9.9682	4.074e-15
Proportion of Variance	0.456	0.2962	0.2478	0.000e+00
Cumulative Proportion	0.456	0.7522	1.0000	1.000e+00

図表9によれば、第2主成分まででは累積寄与率の約8割を捉えていない。したがって、ロングテイルオーサーを捉える主成分軸は第3主成分までとする。

図表10 ロングテイルオーサーの主成分得点

	PC1	PC2	PC3	PC4
[1,]	-7.305577	13.7143882	6.100070	5.376016e-16
[2,]	-4.666794	-0.3757234	-14.547208	1.071341e-13
[3,]	20.159853	-0.3799995	1.616959	1.020538e-14
[4,]	-8.187481	-12.9586653	6.830179	-1.166487e-13

図表10によれば、第1主成分から第3主成分まで主成分得点は次のような特徴を持っている。第1主成分は正の値で他の数値と比較して絶対値が大きい値を示している。特にオーサーグループ3の値が20.159853である。

第2主成分は、オーサーグループ1の主成分得点が13.7143882で、オーサーグループ4の主成分得点が-12.9586653である。第3主成分は、オーサーグループ2の主成分得点が-14.547208である。

図表11 ロングテイルオーサーの第1主成分の固有ベクトル（絶対値の大きい方から50語）

映画 0.07391528	名無し 0.07373297	昨日 0.07369056	人間 0.07368472	自己 0.07366165
先生 0.07362137	自分 0.07361000	ウソ 0.07360932	ひとつ 0.07360550	そのもの 0.07359694
It 0.07356587	あと 0.07355089	ハ 0.07354261	gt 0.07354026	心 0.07353661
常識 0.07353029	人格 0.07352135	先 0.07351388	今朝 0.07351312	子 0.07350907
大切 0.07350705	人生 0.07350163	amp 0.07349775	教師 0.07349532	欲望 0.07348530
心理 0.07348393	いろいろ 0.07346439	前後 0.07346392	クラブ 0.07346130	youtube 0.07345883
静か 0.07344947	人 0.07344596	友人 0.07343536	症状 0.07342721	孤独 0.07342603
昨夜 0.07342116	自身 0.07342097	夢 0.07340863	あい 0.07340807	ゝ 0.07340611
親子 0.07340204	言葉 0.07340111	損 0.07339997	年齢 0.07339853	マガジン 0.07339811
母親 0.07339641	母 0.07339430	善人 0.07339310	実家 0.07339119	親 0.07338891

図表11によれば、ロングテイルオーサーの第1主成分を特徴付ける単語として、「人間、自己、先生、自分、ウソ、心、常識、人格、子、人生、教師、欲望、心理、静か、人、友人、症状、孤独、自身、夢、あい、親子、言葉、年齢、母親、母、善人、親」などがある。

図表12 ロングテイルオーサーの第2主成分の固有ベクトル（絶対値の大きい方から50語）

日本人 0.08498037	大統領 0.08339857	米国 0.08210147	日本 0.08187932	企業 0.08030373
オバマ 0.07895450	政権 0.07890513	小泉 0.07888712	戦略 0.07884747	被告 0.07870745
大臣 0.07858429	ロシア 0.07851758	議員 0.07848656	首相 0.07835657	中国 0.07832341
内閣 0.07819881	郵政 0.07814909	政治 0.07814487	自民党 0.07803097	予算 0.07794323
国民 0.07785240	確か 0.07778282	利権 0.07762481	マスコミ 0.07739520	総理 0.07724740
アメリカ 0.07711269	資金 0.07694727	鳩山 0.07686590	経済 0.07675965	財政 0.07674514
官僚 0.07672838	インド 0.07658306	政策 0.07646508	子供 0.07635413	幹事 0.07625754
国家 0.07617676	民主党 0.07612976	野田 0.07605644	銀行 0.07601339	前原 0.07592694
税金 0.07585449	財務省 0.07583094	ひと 0.07577525	税 0.07571302	金融 0.07566450
板垣 0.07563838	アメリカ合衆国 0.07552961	ユーロ 0.07552867	一郎 0.07549215	サン 0.07545849

図表12によれば、ロングテイルオーサーの第2主成分を特徴付ける単語として、「日本人、大統領、米国、日本、企業、オバマ、政権、小泉、戦略、大臣、ロシア、議員、首相、中国、内閣、郵政、政治、自民党、予算、国民、利権、総理、アメリカ、資金、鳩山、経済、財政、官僚、インド、政策、幹事、国家、民主党、野田、銀行、前原、税金、財務省、税、金融、板垣、アメリカ合衆国、ユーロ」などがある。

図表13 ロングテイルオーサーの第3主成分の固有ベクトル（絶対値の大きい方から50語）

竹島 -0.10003716	韓 -0.09920429	店 -0.09895784	キムチ -0.09880693	選手 -0.09826815
韓国 -0.09815356	書記 -0.09813647	msn -0.09795044	sankei -0.09791940	総 -0.09789110
客 -0.09782382	n -0.09761161	今年 -0.09759528	平壤 -0.09739820	キム -0.09722634
バスターミナル -0.09719204	李 -0.09717168	朝鮮 -0.09717081	置屋 -0.09712935	?? -0.09711344
仁川 -0.09710960	海雲台 -0.09708231	昨年 -0.09707849	北 -0.09707770	木浦 -0.09707248
麗水 -0.09705853	? -0.09704117	朴 -0.09702648	cgi -0.09700104	恩 -0.09698772
ハンブル -0.09697254	釜山 -0.09696634	鄭 -0.09693920	ロッテ -0.09691367	洞 -0.09691180
サムスン -0.09688425	midot -0.09687895	ソウル -0.09687692	光州 -0.09686835	ch -0.09681187
yutori -0.09679536	kr -0.09679157	大邱 -0.09679049	test -0.09676415	文化 -0.09675345
済州 -0.09671340	ソース -0.09668147	nbsp -0.09668059	net -0.09665600	京畿 -0.09664498

図表13によれば、ロングテイルオーサーの第3主成分を特徴付ける単語として、「竹島、韓、キムチ、韓国、書記、平壤、キム、李、朝鮮、仁川、海雲台、北、木浦、麗水、朴、ハンブル、釜山、鄭、ロッテ、サムスン、ソウル、光州、大邱、済州、京畿」などがある。

以上のように、ロイヤルオーサーとロングテイルオーサーを特徴付ける単語が明らかとなった。続く考察で、これら単語をもとにして各主成分にラベルを貼り消費者ニーズを解釈する。

#### 4. 考察

先の結果で明らかとなった固有ベクトルの絶対値の大きな単語をもとにして、各主成分軸にラベルを貼る。まず、ロイヤルオーサーを捉える2軸に対してラベルを貼る。

ロイヤルオーサーの第1主成分を特徴付ける単語は、「季節、我が家、サラダ、野菜、手作り、トマト、パスタ、スープ、味、ワイン、ランチ、カフェ、お茶、紅茶、夕食、レストラン、チーズ、チョコ、デザート、フード、チョコレート」などであった。これらの単語から、ロイヤルオーサーの第1主成分は「食事軸」と名付ける。

ロイヤルオーサーの第2主成分を特徴付ける単語は、「主人、夕方、今日、天気、家族、今朝、雨、家、朝、明日、元気、友達、子ども、じい」などであった。これらの単語から、ロイヤルオーサーの第2主成分は「家族・天候軸」と名付ける。

次に、ロングテイルオーサーの第1主成分から第3主成分までにラベルを貼る。ロングテイルオーサーの第1主成分を特徴付ける単語は、「人間、自己、先生、自分、ウソ、心、常識、人格、子、人生、教師、欲望、心理、静か、人、友人、症状、孤独、自身、夢、あい、親子、言葉、年齢、母親、母、善人、親」などである。これらの単語から、ロングテイルオーサーの第1主成分は「心理軸」と名付ける。

ロングテイルオーサーの第2主成分を特徴付ける単語は、「日本人、大統領、米国、日本、企業、オバマ、政権、小泉、戦略、大臣、ロシア、議員、首相、中国、内閣、郵政、政治、自民党、予算、国民、利権、総理、アメリカ、資金、鳩山、経済、財政、官僚、インド、政策、幹事、国家、民主党、野田、銀行、前原、税金、財務省、税、金融、板垣、アメリカ合衆国、ユーロ」などである。これらの単語から、ロングテイルオーサーの第2主成分は「政治・経済軸」と名付ける。

最後に、ロングテイルオーサーの第3主成分を特徴付ける単語は「竹島、韓、キムチ、韓国、書記、平壤、キム、李、朝鮮、仁川、海雲台、北、木浦、麗水、朴、ハンゲル、釜山、鄭、ロッテ、サムスン、ソウル、光州、大邱、済州、京畿」などである。これらの単語から、ロングテイルオーサーの第3主成分は「韓国軸」と名付ける。

以上で、全ての主成分へラベルを貼り付けた。ここから、ロイヤルオーサーとロングテイルオーサーのニーズを解釈する。先ずロイヤルオーサーのニーズから解釈する。

モナコへ高いロイヤリティをもつ日本の消費者は食事軸、家族・天候軸によって捉えられた。ここから日本の消費者は、おいしい食事、家族で過ごす時間、穏やかな天候にモナコの魅力を感じていると考えられる。

次にモナコへの低いロイヤリティを持つ日本の消費者は、心理軸、政治・経済軸、そして韓国軸によって捉えられた。これらの軸から日本の消費者は、心理的・精神的な癒しを求めており、自分の身の回りの出来事だけでなく政治・経済の出来事に関心を持ち、モナコだけでなく韓国にも関心を持っていると考えられる。

続く結論で今一度問題意識に戻り、ここまでで行った解釈から実践的な洞察を導く。

## 5. 結論

本研究により期待される貢献は大きく2つであった。1つ目の問題意識は、本研究によりモナコ公国に対する日本人の消費者ニーズを明らかにすることで、モナコ公国はより一層の観光客を集める上で重要な消費者ニーズを把握できる、であった。

この問題意識について、ロングテイルオーサーのニーズから洞察を導ける。ロングテイルオーサーの軸から、日本の消費者は心理的・精神的な癒しを求めており、自分の身の回りの出来事だけでなく政治・経済の出来事に関心を持ち、モナコだけでなく韓国にも関心を持っていると考えられる。

したがって、モナコ公国の観光担当者は、心理的あるいは精神的な癒しのスポットを紹介することでより多くの集客を望める。政治・経済への関心は、Kato (2012) のハウステンボスへの消費者ニーズの調査においてもロングテイルオーサーが政治経済軸（例えば、党、昨年、銀行、金融、国会、大震災、米国、衆院、3月、委員、事件、政権、首相、国民、政策、法案など）により捉えら

れていた。

このような重複から見ても、ヨーロッパ的なリゾート地に関心を持つ日本の消費者の知的水準は高いと想定できる。したがって、それを満たすようなモナコでの余暇の一方で、Executive MBA などの短期で政治・経済を学べるプログラムなどの提供によりこれらの日本の消費者を惹きつけばより多くの集客につながると考えられる。

最後に、日本の消費者はモナコだけでなく韓国への関心をもっていた。日本の消費者はモナコ同様にカジノを楽しめ、モナコよりも地理的に近い韓国へ関心があると考えられる。したがってモナコ公国の観光担当者は、この地理的な近さに対抗できる魅力を提供する必要がある。引き続いて韓国をモナコ公国の競争相手の一つと想定して、韓国への日本人の消費者ニーズを調査してみる必要がある。

2つ目の問題意識は、テーマパークの経営者がハウステンボスと同様にヨーロッパ的リゾート地を競争相手としているならば、消費者が競争相手にどのようなニーズをもつのかを把握できる、であった。これは同様の競争相手を持つテーマパークの経営者にとって経営戦略立案の一助となる。

この問題意識について、ロイヤルオーサーのニーズから洞察を導ける。日本の消費者は、おいしい食事、家族で過ごす時間、穏やかな天候にモナコの魅力を感じていると考えられた。したがって、おいしい食事という観点からは、モナコで著名なレストランを自身のテーマパークに出店させることによりモナコを訪れなくとも国内のリゾート地で堪能できるようにすることで競争力を獲得できる。

家族で過ごす時間という観点からは、若者や高齢者といった年齢層に限定するのではなく家族連れで訪れたときにそれぞれが楽しめる施設が必要となる。例えばモナコ同様にカジノやレース観戦のような大人が楽しむ娯楽を提供しているとする。

大人がそのような娯楽へ出かけている間に、子供たちを楽しませる企画が必要とされる。こうした家族全員をターゲットとしたサービスを準備することで、モナコと競合するテーマパークは日本の消費者をより惹きつけられると考えられる。

天候という観点からすれば、悪天候時の企画があることで、操作できない天候により旅行者の期待を損ねないよう予防策を講じる必要がある。これには屋内型の施設の充実などが考えられる。

本研究は当初の問題意識に対して、ブログテキストデータを KIP により分析することで以上のような貢献を果たした。だが未だ解決されるべき課題も残されている。

第1に、時系列分析への拡張の必要性である。消費者のニーズは時間と共に変化する。その変化点前後での市場の変化の解明は、市場創造プロセスの理解にとって重要である。

そのためには時間をずらしながら、繰り返し KIP を時系列テキストデータへ用いる必要がある。これを実現するには KIP を完全自動化しなければならない。KIP は未だ完全な自動化ではなく所々で実際の数値を見ながら基準となる値を決定しつつ分析を進めなければならない。

したがって、データを時間で分け同じ分析を繰り返すだけで時系列の分析ができるというわけではない。まず KIP のデータ収集を除く全ての分析手順を完全自動化し、その上で時系列の分析へと進む必要がある。

第2に、データクリーニングの方法を一層洗練させなければならない。我々は Kato and Imanishi (2012)においてナイーブベイズを用いたデータクリーニング法を提案し、加藤・今西 (2012)において改善を行った。だが未だ十分な成果を出せておらず、本研究ではデータクリーニングを行わなかった。今後特に確率計算方法を改善することにより、より適切なデータクリーニングができるようにしなければならない。

これらの課題はあるものの、本研究は当初に掲げた問題意識にたいしては貢献できた。したがって、今後これらの課題一つひとつを解決することで一層魅力的な手順へと改善し、更なる経験研究を積み重ねることで KIP のマーケティング研究の方法としての地位を確立していきたい。

#### 参考文献

- [1] Carson, Dean., (2008), "The 'Blogosphere' as a Market Research Tool for Tourism Destinations: A Case Study of Australia's Northern Territory," *Journal of Vacation Marketing*, April, Vol. 14, No. 2, pp. 111-119.
- [2] 千野珠依(2011), 「国内観光市場の見通しと雇用への影響」, 『みずほりポート』, みずほ総合研究所。
- [3] 加藤淳一, 石川雅弘(2011), 「大量ブログ記事をデータとした市場セグメンテーションの半自動的分析手法」, 『2011年春季研究発表会アブストラクト集 (日本オペレーションズ・リサーチ学会)』, 104-105頁。(http://ci.nii.ac.jp/naid/110009359026/en)
- [4] Kato, Junichi. (2012), "Exploring Keywords to Create Tourism Markets by Using Blog Text Mining", 『第49回 (2012年) 年次大会学術発表論文集 (日本地域学会年次大会)』。(http://jsrsai.envr.tsukuba.ac.jp/Annual\_Meeting/PROG\_49/Resume2/rB04-1%20KatoJunichi.pdf)
- [5] 加藤淳一・今西衛 (2012), 「ディズニーランドに対する消費者ニーズの探索的探求」, 『全国研究発表大会要旨集 (経営情報学会 (アブストラクトでの査読付国内学会))』, 19-22頁。
- [6] Kato, Junichi and Mamoru Imanishi (2012), "An Analysis of Customers' Needs for Smartphone Markets by Using Blog Data and Improvements of Kato & Ishikawa's Procedure from Bayesian Viewpoint", 『2012年秋季研究発表会アブストラクト集 (日本オペレーションズ・リサーチ学会)』, pp. 70-71.
- [7] 三田村保, 岩佐渉, 湯川恵子 (2008), 「ブログを利用した観光情報の調査分析」, 『観光と情報』, 4, 1, 57-65頁。
- [8] 三土修平(1997), 『初歩からの多変量統計』, 日本評論社。
- [9] 日本再生戦略 (http://www.npu.go.jp/policy/pdf/20120731/20120731.pdf 2012年10月9日に確認)。
- [10] 日本新成長戦略 (http://www.kantei.go.jp/jp/sinseichousenryaku/sinseichou01.pdf 2012年10月9日に確認)。
- [11] Pan, Bing., Tanya MacLaurin, John C. Crotts, (2007), "Travel Blogs and the Implications for

Destination Marketing”, *Journal of Travel Research*, August, Vol. 46 No. 1, pp. 35–45.

[12] 上之郷利昭(1992), 『ハウステンボス物語—男たちの挑戦』, プレジデント社。

[13] Volo, Serena. (2010), “Bloggers’ Reported Tourist Experiences: Their Utility as a Tourism Data Source and Their Effect on Prospective Tourists”, *Journal of Vacation Marketing*, Vol. 16, pp. 297–311.

## Explore the Customers' Needs for Foreign Tourism Destinations by Using Blog Text Mining

Junichi Kato

In this research, for clarifying the Japanese customers' needs for Principality of Monaco as foreign tourist city by using Kato and Ishikawa's procedure (KIP), we segment the markets and extract some principal axes.

There are two expected contributions in this research. First contribution is that the person in charge of tourism of Monaco can grasp customers' needs to gather more tourists. Second contribution is that managers of European style theme parks can understand customers' needs about competitors.

We can get insights about these two research questions. Consumers demand psychological healing, are interested in issues of politics or economy, and are interested in not only Monaco but also Korea. If the person in charge of tourism of Monaco takes these needs into considerations, they can gather more customers.

In reference to the second research question, consumers feel the charm of Monaco for a delicious meal, time to spend with families, and calm weather. If managers think of these customers' needs, they can create more charming theme parks for customers.

Keywords: Marketing , Blog, Text Mining, Tourism