

テキストマイニングによる香りに関する連想の分析方法

—その手続き・意義について¹

椿田貴史・早川 貴²

1. 背景

「好きな香り」を主題とした連想のアンケート調査結果について、手持ちのソフトを用いて簡易的なテキストマイニングを行った。テキストマイニングについて解説した著書のうち、心理学での調査研究を意識したものとしては、藤井・李・小杉（2005）による『福祉・心理・看護のテキストマイニング入門』（中央法規出版）がある。これは、無料の形態素解析ソフト³と Excel、そして有料のマイニングソフトを用いる手法を中心としたものである。しかし、今回、Keygraph⁴をエンジンとした Polaris というソフトを用いることで、より簡易にテキストマイニングができたので、本稿の「方法」でやや詳しく報告したい。さらに、Polaris から得られた概念マップを PC 上で操作することによって、それぞれの概念（形態素）の関係をよりの確に把握でき、解釈や発見をする上での利便性が増すので、「考察」でその利用方法を報告しておきたい。結論を先取りするならば、このような手軽に用いることができるテキストマイニングのソフトによって、心理学の古典的な研究課題であった「連想」や観念連合について、新たな視点で考察する可能性がもたらされることになる。その前に、まず、(1) テキストマイニングについて概略し、(2) 香りの連想という主題を選んだ理由を述べたい。

(1) テキストマイニングについて

テキストマイニングは、データマイニングの一種である。主にインターネット上の自由記述など大量のテキストデータの分析を目的とした発展途上の技術であると見なしてよい。この分析によって、ある主題についてのキーワードを抽出する、そして、逆に、利用者のキーワードから希望するようなテキスト群を指定するといったことが可能となる。マイニングという言葉から推測されるように、この種の技法は、大量のデータの中から必要かつ有益な情報を見つけだす、という用途に用いられることが多い（岡嶋, 2006）。

この技術は人工知能や web 上の技術以外に、マーケティングの分野でも頻繁に用いられている。たとえば、消費者が特定の商品群に対して抱いている価値観や選択基準を明確にするための技法であるラダリング法（たとえば丸岡, 2002）と組み合わせて、テキストマイニングを実

¹ 本研究は日本心理学会第 70 回大会のポスター発表（「香り」を主題とした連想の分析—関心の高さと連想量—）を基にしている（若干の新しいデータを加えて再度分析をしたものである）。なお、本研究を実施するにあたり、本学の廣島勉氏にテキストデータの分かち書きと語彙カウントおよびデータの整理をお願いした。データ分析については、本学の平井正剛氏にお世話になった。感謝の意をここに記します。

² 愛知県日進市米野木町三ヶ峯 4-4 名古屋商科大学総合経営学部

³ <http://chasen.naist.jp/hiki/ChaSen/>。茶筌 (ChaSen) version 2.1 for Windows。

⁴ <http://www2.kke.co.jp/keygraph/index.html>。また、Keygraph については、大澤幸生・ネルス E. ベンソン・谷内田正彦 (1999) Keygraph: 語の共起グラフの分割・統合によるキーワード抽出, 電子情報通信学会論文誌 D-I Vol. J82-D-I, No.2, pp.391-400 を参照。

施する方法などである（柴田・上田, 2003; 上田・柴田, 2003; 上田・畑井, 2005; 上田, 2005）。また、ブランド戦略などの分析においても、消費者のブランドに対するイメージを抽出するなどの作業で威力を発揮する（豊田, 2005）。

心理学の研究では、少数であっても分析可能なデータの入手にコストと時間がかかるため、データの型や量に関して制約が大きい。言葉がデータとなる場合には、それらはカテゴリ分けをして操作可能な変数へ切りつめたり、内容的に分類するなどの処置がなされる。多くのノン・パラメトリックな手法も開発されているので、工夫しだいで、従来の方法でも様々な分析は可能である。しかし、基本的に心理学の世界で構築された調査手法は、一定の型にはまったデータから仮説について検討したり、何らかの傾向を見いだす、というものである。そして、アンケートの自由記述項目などのテキスト・データは、そうした手法が適用しにくい部分である。いわば、データ分析の脇役であったように思われる。

もちろん、そうでない場合もある。定性的考察（たとえば臨床の事例研究など）では、被験者や患者が表出する発言や行動が一種のテキストとして解釈されることになる。ここでは、自由な言語的表出こそが、もっとも重要なデータとなる。しかし、そうした言語的表出の分析は、事例性という概念の下でこそ可能であり、同じ解釈手法が他のケースで適用可能かどうかは未知数である。さらに、客観性を重んじる行動科学的文脈では、そうした解釈が解釈者の恣意にゆだねられた勝手な判断と見なされてしまうことも少なくないと思われる。

テキストマイニングは、主観的になりがちなテキスト解釈について、一定の視点からの客観的な指標を与えることができる。具体的には、

- 1) 同じ主題に関する多人数の（つまり横断的な）テキストを特定のパラメータを設定して分析する
- 2) 縦断的に蓄積された特定の被験者によるテキストを、任意の期間に区切って分析する

などである。解釈を行うべき対象であるテキストについて、一定の量と条件がそろえば、客観的な分析を行うことができ、その結果を加味して主観的な解釈を行う、という手続きが可能となる。

（2）「香り」を研究の主題に据えた理由

既に拙論（椿田, 2006）で結論づけたように、香りについての心理学的研究は、“香りの生理学”とでも言うべき領域に縛られてきた。心理学が本来得意としているような“表現行為”に関しては、SD法の結果を因子分析するなどの方法を活用した研究があるが、今日では、やや手詰まりの感が否めない。“香りの心理学”は、次々と発表される嗅覚生理学の成果に圧倒されて、自らのアイデンティティを失いかけているように思われる。本研究はテキストマイニングの技術を活用して、香り心理学に新たな研究のフロンティアを見出したい、という希望を抱きつつ行われる探索的な調査である。

今回は手始めに、香りに関する言語的連想をある程度収集してマイニングを行った。もちろん、データをマイニングソフトにかけるだけで、そこから法則性や連想構造が浮かび上がる、などということは期待できない。ここで期待できるのは、仮説の検証や従来の理論の反駁ではなく、特定の量のサンプルを一次的に「重ね合わせる」こと（概念マップの作成）によって

見えてくる意外な概念（言葉・語彙）同士の結びつきである。このような結びつきが、新たな研究の素材として、あるいは、これまでとは異なる観点につながる何かとしてマイニングされる対象となる、というのがさしあたっての作業仮説である。

2. 方法

（1）質問紙の構成

今回は大学生を対象に、1) 彼らが香りについてどのような嗜好を抱き、2) その香りによってどのような気分になるか、さらに、3) 香りに関する製品の純粋想起を調査する。他の従属変数としては、男女、香り製品に対する投資比率（小遣いに対する）としたが、本稿では1) 2) を分かち書きし、全被験者の記述に関して共起語を調べることを目的としている。語彙数、連想量その他と投資率などに関する分析は割愛させていただく。

共起語とは、異なる二つ以上の形態素が同じ人物によって記述された場合に、それらの語を共起語と定義する。

（2）調査対象者

大学生 442 名に対し質問紙を配布し 409 名（男子 278 人）から回収した。（回収率 92.5%）。

（3）手続き

対象者に質問紙を直接配布して記入をお願いした。質問紙は無記名とし、回答時間に制限は与えなかったが、すべて配布したその場で回収が可能であった。なお、大学は中部圏内の短期大学を含めた 3 大学である。平均年齢 19.8 歳（SD=1.8）。

（4）分析のためのツール

分かち書きのための形態素分析ツールには茶筌 (ChaSen) version 2.1 for Windows を用いた。また、共起関係を調べるためのツールは Keygraph をエンジンとして動く Polaris を用いることにした。

（5）分析の前に行う作業

1) データ入力に関して

データはテキストファイルでも Excel に入力してもどちらでもかまわないが、一定のルールに従って入力する必要がある。今回のように、一つのテーマに関する各被験者の反応（連想）を大量に分析するのであれば、テキストエディタに入力の方が楽である。1 件のデータは改行で区切るなどすればよいが、Polaris ではカンマ、ピリオド、タブ、読点でもデータの区切り（Polaris 上では、1 件のデータを「バスケット」と呼んでいる。また、分かち書きしたそれぞれの形態素を「アイテム」と呼んでいる）でも構わないことになる。しかし、本研究のように、複数の文や語句から成るテキストを分析するのであれば、当然、改行かタブがもっとも都合が良いであろう。

2) 形態素分析と分かち書き

テキストマイニングは、準備したテキストに含まれる全ての文章を形態素レベルで分かち書きし、全ての形態素について出現頻度を確認することから始まる。この作業については、本巻に収められている廣島氏の論文を参照されたい。

本研究で用いたマイニングソフト **Polaris** は、テキストデータの取り込み、形態素分析、統計、共起語その他の関数を利用した概念マップの作成が行える統合環境とでも言えるものである。よって、データ入力さえきちんに行えば、作業中のミスはほとんど起こらないというメリットがある。具体的には、**Polaris** 上から分析対象となるテキストデータのファイルを指定して実行しさえすれば、茶筌を経由した形態素解析の結果が **Polaris** 上に出力される。

テキストマイニングでは、助詞や句読点などは分析対象にしないことが多い。**Polaris** では、茶筌を経由して出力される形態素を「動詞」「名詞」「形容詞」に限定すれば良い。取り立てて削除したい動詞や名詞、形容詞、その他がある場合には、「削除フィルタ」で削除するリストを指定すれば良い。今回は、削除辞書や類義語を統一する設定はせず、分析対象の形態素を「動詞」「名詞」「形容詞」に限った。「香り」「におい」などの語を統一しても良かったが、手書きによるアンケートなので、念のためそのままにしておいた。

3. 結果

マイニング結果として選ばれる語は、単純に考えると、サンプル全体を通じて、1) 共起語としての出現回数が多い、2) 他の語との結びつきが多い、という二つの条件で拾い上げられると考えられる。**Polaris** ではこれは黒いノードで示される。さらに、黒いノードほど出現頻度が多いわけではないが、黒に結びつけている潜在的な語も考えられる。後者は赤いノードで示される。できるだけ、多くの情報量を得たい場合には、設定をたとえば以下のようにして、ノードやリンク数を増やせばよい。分析により、男女合わせたサンプル (図1)、男性 (図2)、女性 (図3) の概念マップを得た。

Polaris の設定 (例)

黒ノード 30

黒リンク 40

赤ノード 60

黒リンク計算方法 共起頻度

赤リンク計算方法 共起頻度

るかもしれない。マイニングによる処理によって、こうした共通する語彙はもちろん、それらに近い語彙を析出する、ということが可能になる。

もちろん、何らかの手作業によって、たとえばKJ法のような方法によっても人の連想構造を明確化することはできるかもしれないし、今後はコンピュータに頼らなくても、何らかの方法論を洗練させてゆくことで読み取りができるかもしれない。しかし、そうした人海戦術で可能なのは、少量の質的データに限られる。

今回はたった400名ほどのテキストデータであったが、これだけでも手作業でキーワードの析出を行うことは困難である。このように、人の手では不可能な量の被験者について、「連想の重ね合わせ」を機械的にすることの意義は、大きい。考えてみると、集会的な連想を扱う方法論的枠組みは従来、民族心理学やユング流の集会的無意識などの理論を前提とした物語解釈、さらにはウラジミール・ブロップなどに代表される物語の形態分析などが提供していた。それらは複数の物語に共通のマクロな構造を抽出する解釈技法であった。しかし、これらの理論の適用は、理論を適用する個人の知識や能力に過度に依存するおそれがあり、不毛な方法論的論争に終わることもしばしばである。ある程度まとまったテキストの通時的変化を調べるような場合には、個人の“読み取り”能力とは関係なく結果を提供してくれるテキストマイニングが、分析に新たなヒントを与えてくれると思われる。ただし、パラメータの設定や最後の読み取りには、最終的に使用者の力量が試されることは当然である。定性的な分析と定量的な分析の双方をバランスよく保つためのメタレベルでの方法論を今後探求してゆく必要がある。

一方、マーケティングの立場からこうしたマイニング結果を考察するとすれば、研究の面では発見的な過程における新たな分析ツールの一つとしての利用法が、実践の面ではブランドポジショニングやコミュニケーション開発など顧客の心理を探る必要のあるプロセスの他、製品開発分野に関わる市場調査のための分析ツールとしての利用法が考えられる。本稿で用いた技法の大きな特色の一つは「赤ノード」の存在だが、小さな出現頻度ながら思考の連鎖の中では重要な役割を果たしているかもしれないキーワードを抽出し、概念のネットワークの中に位置づけることができる「赤ノード」は、研究の面でも実践の面でも重要な役割を果たすだろう。

科学研究としてのマーケティングのプロセスは、他の諸科学同様、所謂「発見の文脈」と「正当化の文脈」または「非正当化の文脈」の二つのプロセスに大別して考えることが妥当である。このうち「発見の文脈」は方法論的な自由度が大きく、方法としての優劣の評価は後に続く「正当化の文脈」または「非正当化の文脈」に耐えうる仮説の構築に如何に効果的に貢献できるかにかかっていると行って良い。

従来用いられてきた因子分析等の発見的分析手法が、いわば「黒ノード」同士とってよい主要因子間の関係の強弱に主として焦点を当てるのに対し、本稿で用いた手法は主要因子間の関係の意味に主に焦点を当て、主要因子やその関係を構成する多様な泡沫因子を可視化することで、見落とされてきた潜在変数の発見を助けるかもしれない。すでに確立された発見的分析手法を併用することで、新たな潜在変数の発見をより効果的に導くことが期待される。

実践においては、本稿が採用した手法は、グループインタビュー等の質的情報収集を目的とした市場調査の際に用いる分析手法としての利用が期待され、従来、分析者の経験と属人的能力に大きく依存して来たブランド表現や製品開発等に関わる調査と分析の分野で、ある程度、分析ロードを軽減するとともに、成果の平均化を図ることを可能とするだろう。こうした分野では通常、集会的な連想の方向性は相当程度予測の範囲内にあり、むしろ周辺の個別多様な断

片的連想と集合的連想の関係から、市場の変化の兆しや開発のヒントを得ることが重視される。「赤ノード」はまさに、そうした個別多様な断片的連想のうち集合的連想に関連の深いものを抽出しており、経験と属人的能力に恵まれた分析者の分析ロードを軽減するとともに、能力に恵まれない分析者にも有意な断片の抽出を可能とすることで、分析成果の平均化に貢献すると考えられる。

参考文献

- 藤井美和・李政元・小杉考司 (2005) 『福祉・心理・看護のテキストマイニング入門』中央法規出版.
- 廣島勉 (2006) 「プログラム言語を使った日本語テキスト処理技術について」NUCB JOURNAL OF ECONOMICS AND INFORMATION SCIENCE (本書)
- 丸岡吉人 (2002) 「手段目的連鎖モデルで消費者を理解する」『一橋ビジネスレビュー』Vol. 50, No.3, pp.48-56. 東洋経済新報社.
- 岡嶋裕史 (2006) 『数式を使わないデータマイニング入門 隠れた法則を発見する』光文社新書
- 大澤幸生・ネルス E. ベンソン・谷内田正彦 (1999) 「Keygraph: 語の共起グラフの分割・統合によるキーワード抽出」『電子情報通信学会論文誌』D-I Vol. J82-D-I, No.2, pp.391-400.
- 柴田典子・上田隆穂 (2003) 「WEB テキストマイニング型ラダリング法による広告制作・新製品開発～消費者セグメント別製品利用オケーションと価値体系～」『学習院大学 経済論集』第 40 巻 第 2 号 pp.113-130.
- 豊田裕貴 (2005) 「自由回答調査によるブランドイメージ分析」『テキストマイニングによるマーケティング調査』所収 ,pp.129-146, 上田隆穂・黒岩祥太・戸谷圭子・豊田裕貴, 講談社サイエンティフィック.
- 椿田貴史 (2006) 「嗅覚心理学の展開と課題」NUCB JOURNAL OF ECONOMICS AND INFORMATION SCIENCE Vol51, No.1, pp.33-42.
- 上田隆穂・柴田典子 (2003) 「製品利用におけるオケーションと価値体系：ラダリング法とテキストマイニングの活用～ビール・発泡酒を事例として～」『マーケティングジャーナル』87, Vol.22, No.3, pp.18-32.
- 上田隆穂・畑井佐織 (2005) 「類似価値体系セグメント発見の WEB ラダリング調査～テキストマイニングの活用～」『マーケティングジャーナル』96, Vol.24, No.4, pp.4-17.
- 上田隆穂 (2005) 「消費者価値体系を探る Web ラダリング法の利用」『テキストマイニングによるマーケティング調査』所収 ,pp.104-128, 上田隆穂・黒岩祥太・戸谷圭子・豊田裕貴, 講談社サイエンティフィック.