原著論文

テキストマイニングを用いた米に関する消費者 アンケートの解析

磯島昭代*

東北農業研究センター 〒020-0198 岩手県盛岡市下厨川字赤平4

要旨

米に関する消費者意識を解明する研究として、これまで主に選択肢式のアンケート調査やグループインタビューなど少人数を対象とした面接調査が行われてきた。アンケート調査で得られる自由記述回答文は、大量の定性的データであり、新たな知見を得る重要な情報源と考えることができる。これまでこうしたデータを分析する有効な手法がなかったが、テキストマイニング手法の開発により、大量の文章データを計量的に分析することが可能となった。そこで、アンケート調査で得た自由記述回答文にテキストマイニングを適用し、米に関する消費者意識の解明を試みた。その結果、消費者の米に対する関心は「米購入」、「日本の米と農業」、「安全性」に大きく分けられることがわかった。さらに、「米購入」に関しては専業主婦層が、「日本の米と農業」に関しては60歳以上の男性が、「安全性」に関しては農薬の使用に敏感な人がよく記述する傾向にあることが明らかとなった。

キーワード

消費者, 米, アンケート調査, 自由記述回答, テキストマイニング

はじめに

米の消費量の長期的な減少が進むなかで、「売れる米づくり」を推進するために消費者ニーズの的確な把握が求められている。米に関する意識や購入時の選択基準などを明らかにした研究としては、消費者を対象にアンケート調査を行った磯島(1998)や平尾・河野(2000)などがある。こうした調査の多くは、提示した選択肢に対して回答を求める形式で行われており、得られる回答は定型であるため統計的な分析が可能となる。

一方、このような定量的なアプローチに対し、消費者の価値観の多様化を背景として、消費者個人を深く、総合的にとらえられる定性的アプローチの必要性が指摘されている(牧田1994)。米購入に関する消費者意識の把握のために梅本ら(2002)が行ったグループインタビューは、定性的

アプローチの代表的な調査手法といえる。しかし、このような調査は対象が少人数に限られるため、得られた回答を統計処理することができず、論理的で妥当な結論づけが困難であるとされる(梅澤1993)。

ところで、アンケート調査票には先に述べた選択肢式の質問項目に加え、回答者が自由に意見を記入する自由記述回答欄を設けていることがある。このような形式で収集した自由記述回答文には調査者が予想しなかった情報が含まれている可能性がある。また、グループインタビューなどの面接調査と異なり、多数の調査対象者から回答を得ることが可能である。

しかし、自由記述回答欄から得られる文章データは非定型的な定性的データであるため、これまで計量的な分析手段がなく、1つ1つの文章に目を通して分類・整理する以外に情報抽出の方法はほとんどなかった。

*Corresponding Author

E-mail:isojima@affrc.go.jp

近年,大量のテキストデータを分析して内容を瞬時に把握したいというニーズが高まり,そのための自然言語処理技術としてテキストマイニング手法が注目されるようになった(市村ら 2001).本稿では,このテキストマイニング手法を用いて,アンケート調査で得た自由記述回答文を分析し,米に関する消費者意識の解明を試みる.

本稿で用いるテキストマイニング手法の概要

テキストマイニングは,大手企業においては顧客管理やマーケティング戦略などで既に導入が進められているが(石井 2002, 山崎・松田 2002),市販されているテキストマイニング用ソフトウェアの多くは数百万円レベル(林 2002)と非常に

高額であり,誰もが気軽に利用できるような状況にはない.また,社会学など一部の学問領域では,研究の目的に応じたテキストマイニング用ツールを無償で公開する動きもあるが(樋口 2004),一般にはあまり認知されていないというのが現状である.

こうしたなか,林(2002)は幅広い利用者を想定した,簡単にテキストマイニングを行える手法を紹介している.林の示す手法は手作業が多く複雑な面もあるが,一般的に利用されている表計算ソフト「Excel」でほとんどの作業を行うため,初心者を取り組みやすい手法であるといえよう.その一方で,林の手法は各文章と各キーワードの関連づけがなされておらず,キーワードから元の文章を参照することが困難となっている.また,集計方法はキーワードの出現数を合計するにとどま

図1 簡易版テキストマイニングのプロセス

り、回答者の属性などを考慮に入れた高度な統計処理を行うには限界がある。

そこで本稿では、筆者が考案した簡易なテキストマイニング手法を用いて、自由記述回答文の解析を行う。

本稿で使用しているテキストマイニング（以下、簡易版テキストマイニング）のプロセスは、大きく、1)文章データの入手、2)データファイルの作成、3)データの分析、の3段階に分けることができる（図1）。以下ではデータファイルの作成方法について簡単に説明するが、作業手順の詳細およびマクロのコードについては、磯島（2004）を参照すること。

まず、アンケート調査によって得た自由記述回答文をExcelのファイルに入力する。各文章をフリーソフトウェアの形態素解析ツール「茶筌（ちゃせん）」（奈良先端科学技術大学院大学自然言語処理学講座<<http://chasen.naist.jp/hiki/ChaSen/>>2005年9月12日参照）を用いて形態素に分解する。分解された各形態素には、どの文章に含まれていたかを示す文章番号を付与する。品詞情報や出現頻度をもとにキーワード候補となる語を絞り込み、抽出したキーワード候補と各文章番号とを関連づける「二値データ化」を行う。これを元の文章データと結合することによって、図2のようなデータファイルを完成させる。

このように、各文章におけるキーワード出現の有無を二値データで示すことによって、統計的な

分析が可能となるのである。本稿ではまずクロス集計により出現語と回答者属性の関係を明らかにし、その中からキーワードを抽出して等質性分析を適用する。出現パターンが似ている語をまとめて記述内容を分類するとともに、決定木分析を適用して、各内容を記述する回答者属性の解明を試みる。

米の消費に関する自由記述回答文への適用

キーワード候補の抽出

以下では、簡易版テキストマイニング手法を用いて、米に関する消費者の自由記述回答文の解析を行う。データは、「米の品質と価格に関する消費者意識調査」（2000年10月、東京都の住民3000名に郵送、884部回収）の末尾に設けた「米の品質や価格などについて、日頃からお考えのことがありましたらご自由にご記入下さい」という設問に回答した452名の自由記述回答文である。

これらの自由記述回答文を形態素解析した結果、延べ2万個以上の形態素を得た。これらのうち、単独でも意味をなすような語を比較的多く含んでいる品詞として、形容詞（自立）、動詞（自立）、名詞（サ変接続、一般、形容動詞語幹、固有名詞、副詞可能）を選択し、これらの品詞情報を持つ形態素を抽出した。

次に、抽出した形態素を同じ語でまとめて形態素番号を割り振り、各形態素がそれぞれ出現する

	A	B	C	D	E	F	G	H	I	J	K
1	No	code1	code2	code3	自由記述回答文	米	価格	味	品質	購入	農業
2	1	11	3	2	同じ県産の同一品種でも、DSなどで売っているお米は、	1	0	1	0	0	0
3	2	1	4	1	高齢者の二人家族ですのであまり沢山のお米は食べま	1	0	0	0	0	0
4	3	2	4	1	表示と中味が違ってないかと思うことがある（味が悪か	0	0	1	0	0	0
5	4	1	4	1	良い米を安く	1	0	0	0	0	0
6	5	1	3	4	毎日食べる食品ですので、価格が多少高くても品質（無	0	1	0	1	0	1
7	6	10	6	4	お米屋で買う米がミックスされたもののように思う。おな	1	0	0	0	0	0
8	7	1	4	2	新米の巡回でお米が大変おいしくなりましたが、な	1	0	0	0	0	0
9	8	2	3	4	子どもたちの将来が健康で明るく過ごせるように願う第一	1	1	0	0	0	0
10	9	3	3	2	出来るだけ農薬を使わず手頃な値段のお米を食べたい	1	0	0	0	0	1
11	10	2	3	2	品質が良くなり価格が安くなっているように思うが国（日	1	1	0	1	0	0
12	11	2	3	2	スーパー等の販売商品の品質、内容が正しいものかど	0	0	0	1	0	0
13	12	2	3	3	米の品質が問題となるのは、当然ながら白米について	1	1	1	1	0	1
14	13	2	4	2	昔の米より今の米の方がおいしくない	1	0	0	0	0	0
15	14	9	8	4	無農薬主と考えて生産農家より契約して送っていただ	0	0	0	0	0	1
16	15	3	3	4	健康食品ブームで最近は無農薬の有機米がでまわって	0	0	0	0	0	1

図2 データファイルのイメージ図

注) 表頭のキーワードが各行に示された文章に含まれている場合には1, 含まれていない場合には0が表示される。

回数（出現頻度）をカウントした。この出現頻度を参考に、一定回数以上出現している語をキーワード候補として絞り込むことにする。

図3は、最低出現頻度（最低何回以上出現しているかを示す）ごとの形態素数を示している。最低出現頻度1、すなわち少なくとも1回は記述されたという形態素の数は1680個であるが、これが最低出現頻度2で726個、最低出現頻度3では470個と急速に減少する。つまり、品詞情報によって抽出した形態素のうち、7割以上がわずか1回もしくは2回しか記述されていない語であるということである。

出現頻度によってキーワードを絞り込む場合、出現頻度の上限と下限を設け、その数値を超える高頻度語と低頻度語を採用しないのが一般的であるとされる。しかし、具体的に出現頻度の上限、下限を決める絶対的な基準はなく、「中程度の頻度」を決めるひとつの目安としてはZipfの法則の利用が提案されている（徳永 1999）。

今回の分析では、出現頻度によるキーワードの絞り込みに際して上限は設けず、下限の出現頻度を10とした。その結果、形容詞13、動詞34、名詞74の計121個の形態素を抽出した（表1）。なお、出現頻度10以上の形態素を抽出したにもかかわら

ず、表の「人数」欄にはいくつかそれを下回る数値がある。これは、形態素抽出の段階では、文章中に表れた形態素の数を全てカウントして「出現頻度」としていたが、表中の数値は当該形態素を含む文章の数を示していることによる。すなわち、1人の回答者が1つの文章中で同じ言葉を何度も使用すると、出現頻度はその都度増加するが、その言葉を記入した人数には変動がないため、両者が乖離するということである。以下では、各形態素を含む文章数をカウントしたものを「記述者数」として、「出現頻度」とは区別することにする。

抽出語の特徴

まず、形容詞についてみる。記述者数が多い語は「おいしい」（記述者数135、以下同様）であるが、この他に「美味しい」（48）も6番目に入っている。簡易版テキストマイニングの場合、平仮名と漢字は別の語として認識するため、このように別々に算出されることになる。「良い」（64）、「よい」（28）、「いい」（9）も同様である。

これらの語を同じ語として扱うには、最初の文章の段階で平仮名と漢字をあらかじめ統一しておく方法と、データファイルを作成してから同じ語

図3 最低出現頻度によって抽出される形態素の数

でまとめる方法とを考えることができる。前者の場合、統一すべき語を決定するためには全ての文章に目を通す必要があり、文章数が多くなるほど作業量は膨大になる。後者の場合は、既にキーワード候補として抽出された語の中で統一すべき語を判断できるので、比較的容易に作業が行えるが、出現頻度の低い語はデータファイル作成の過程で切り捨てられてしまうため、重要語の一部が欠損する可能性も否定できない。

今回の分析では後者の方法を採用したため、重要語とはいえないまでも、いくつかの語がキーワード候補からはずれるという事態が生じている。例えば、「うまい」(9)という語は、出現頻度が10以上あったためキーワード候補として抽出されたが、抽出されなかった語の中には、「旨い」と「美味しい」(出現頻度はいずれも2)が存在している。表記方法が異なるだけで、一方は当該キーワードの「記述あり」、他方は「記述なし」となる不整合が生じることになる。また、これは形容詞ではないが、「子ども」(出現頻度9)と「子供」(出現頻度7)は2通りの記述がなされたことにより、どちらかに統一していれば出現頻度16でキーワード候補となっていたものが、抽出されなかったことになる。

また、本手法では、文章を分解して出現した語を拾い上げるため、肯定文・否定文の区別がない。「おいしい」「美味しい」という語では、実際に「おいしい米」「美味しかった」などのように肯定的

に使われている文章は174件中146件、「おいしくない」「美味しいと思わなかった」など否定的に使われている文章は21件、残り6件は肯定・否定の両方に使われていた。一方、「まずい」(14)も食味に対する不満を表す語であり、「おいしくない」とほぼ同義であると考えられる。すなわち、「おいしい」と「まずい」は反対の意味をもつ語であるが、前者が否定語とあわせて使用される可能性を考えると、どちらも「味」に関する評価を示す語としてまとめてしまう、という考え方もできる。

「安い」(104)と「高い」(75)についても同様のことがいえる。ただし、「安い」は全て米の価格について述べているのに対し、「高い」は安全性(1件)や湿度(1件)の高さについて言及しているものもあり、必ずしも記述された内容は一定ではない。このようなことは、他の形容詞にも見られる。例えば「多い」(34)、「少ない」(20)は、食べる量や家族の人数から、品質の悪い米、特定の米を販売する店舗の数、銘柄の種類、自身の行動(～することが多い、など)と対象は様々である。同様に、「良い・よい・いい」(95)も、その多くは、米の品質に関することを形容しているが、「よく～する」と頻度を表したり、「～がいい」「～した方がよい」と好みや評価を表すこともある。このように形容詞の場合、形容する対象が多岐にわたるとキーワードとしての利用が難しくなる。

表1 抽出した形態素と記述者数

形容詞 (13語)		動詞 (34語)		名詞 (74語)									
抽出語	人数	抽出語	人数	抽出語	人数	抽出語	人数	抽出語	人数	抽出語	人数	抽出語	人数
おいしい	135	する	262	いる	15	米	302	米屋	24	種類	15	希望	10
安い	104	思う	198	使う	15	価格	92	表示	23	食	15	古米	10
高い	75	ある	139	こだわる	13	味	78	日本	23	水	14	香り	10
良い	64	食べる	123	書く	13	品質	77	新米	22	洗米	14	人	10
ない	59	買う	104	出る	12	コシヒカリ	71	最近	21	生	14	専門	10
美味しい	48	なる	89	聞く	12	購入	67	店	20	岩手	14	農業	10
多い	34	できる	43	変わる	12	新潟	51	毎日	20	一番	14	量	10
よい	28	考える	31	いただく	11	値段	47	ブレンド	19	質問	13	以前	10
少ない	20	わかる	28	求める	11	スーパー	46	安全	19	あきたこまち	13	地方	9
まずい	14	違う	25	答える	10	販売	43	販売	18	大麦	13	粘り	9
悪い	11	出来る	24	入れる	10	産地	39	アンケート	18	現在	13	大切	9
いい	9	作る	19	見る	9	銘柄	39	ササニシキ	18	差	12	宮城	9
うまい	9	炊く	19	言う	9	生産	37	ブランド	18	主食	12	胚芽	8
		入る	18	知る	9	精米	32	家族	17	好き	12	白米	6
		売る	18			農家	29	手	17	ごはん	11		
		比べる	18			品種	28	袋	17	ご飯	11		
		感じる	17			魚沼	26	東京	17	家	11		
		いう	16			消費	25	安心	16	気	11		
		選ぶ	16			今	25	信用	16	日本人	11		
		送る	16			他	24	秋田	16	場合	11		

動詞はさらにキーワードとしての利用が難しい語が多い。抽出された動詞の中には、「食べる」(123)、「買う」(104)、「炊く」(19)など米に関する記述としての特徴を示す語や、「比べる」(18)、「選ぶ」(16)、「こだわる」(13)など、文章を読み込む際の有効な切り口となる可能性のある語がいくつか存在する。その一方で、「する」(262)、「ある」(139)、「なる」(89)など単独では意味の取りづらい語も多い。また、「思う」(198)、「考える」(31)、「感じる」(17)なども、具体的内容を想起するものではない。従って、動詞の中からキーワードを選択する際には、かなり明確な分析視点を持つことが必要であると考ええる。

最後に名詞についてみる。名詞は、今回もっとも多く語が抽出された品詞である。その中でも記述者数が多いのは「米」(302)であるが、これは設問から考えて当然といえよう。次に多い語は「価格」(92)、「味」(78)、「品質」(77)などである。また、「コシヒカリ」(71)、「ササニシキ」(18)、「あきたこまち」(13)など米の品種を表す語や、「新潟」(51)、「魚沼」(26)、「秋田」(16)など米の産地としてよく知られる地名も多く出現している。このように、名詞の抽出語は、形容詞・動詞に比べて記述内容を想起しやすい語が多く、キーワードとしては最も適していると考ええる。

一方で、注意しなければならない点もある。抽出した名詞の中には、「生(文章中では『生協』、以下同じ)」「洗米(『無洗米』)」「専門(『米穀専門店』)」など、実際の記述内容とは若干異なる語として表記されるものがある。また、今回の分析では特に不都合は生じなかったが、「米」が「名詞-固有名詞-地域-国」に分類されるなど、明らかに記述内容と異なる解析結果が生じることもある。このようなことは、形態素解析ツールに搭載されている辞書が分析用途と合致しない場合があること、辞書に未登録の語彙には対応できないことなど、形態素解析を利用する上での限界といえよう。

形態素解析の辞書に農業関連用語を組み込む試みがなされている(法隆ら2004)ことから、将来的にはこうした研究成果の活用によって解決できる部分もあるが、現段階としては、意味がとりに

くい形態素は元の記述文を確認する、データファイルを作成する過程で未知語にも配慮する、などの作業を行う必要がある。

以下では、比較的文章の内容を想起しやすい語が多いという理由から名詞を対象として分析を進めることにする。キーワードの抽出に際しては、上記に示したように必要に応じて元の文章を確認するなどし、解釈に誤りがないよう配慮しながら行った。

統計的手法の適用によるキーワード出現状況の分析

クロス集計の適用

まず、各抽出語について、記述の有無に回答者の属性による違いがあるかを調べるためにクロス集計を行った。表2は、カイ2乗検定の結果、記述が有意に多い語を属性別に示している。

これを見ると、性別では男性が「農業」「生産」「品質」などの語を多く記述しているのに対し、女性は「農業」「安心」「購入」などの語を多く記述していることがわかる。また、年齢では50歳代までの回答者は「購入」「表示」「スーパー」などの語を多く記述しているが、60歳以上では「消費」「生産」「品質」などの語の記述が多い。このほか、子どもや高齢者の有無、就業状況、世帯員数、米購入時の考えなど、それぞれの属性区分によってよく出現する語が異なっており、回答者の属性によって自由記述回答文の内容には異なる傾向があることを示唆している。

等質性分析の適用

前述のクロス集計では、回答者属性別の記述傾向の違いを個々の抽出語について明らかにした。次に、抽出語相互の関連を調べるために、等質性分析の適用を試みる。等質性分析は多重対応分析や数量化3類と同じ分析手法であるが(石村2001, 柳井ら2002)、分析で使用したSPSS 12.0J Categoriesにおいて「等質性分析」と表記しているため、本稿ではこの用語を用いることにする(なお、SPSS 13.0J Categoriesでは「多重応答分析」と表記している)。

等質性分析は反応パターンを調べ、反応が似て

表2 回答者の属性別にみた出現語

属性区分		人数	よく出現する語
性別	男性	153	農業**, 東京**, 生産*, 品質*, 価格, 日本
	女性	299	農薬**, 安心*, 購入*, 質問*, 洗米*, ブレンド, スーパー, 安全
年齢	40代まで	127	ブランド**, 生**, 購入*, 表示*, スーパー*, 好き*, 気*, 水*, 洗米*, 品種, 差, コシヒカリ, 毎日
	50代	132	購入*, 表示*, スーパー*, 水*, 好き*, 毎日, 精米
	60歳以上	193	消費*, 生産*, 品質*, 日本, 農業
18歳未満のこども	いる	122	ブランド**, 生*, 洗米, 家族, 差, 手, 東京
	いない	330	—
65歳以上の高齢者	いる	205	購入*, 大切*, 食, 胚芽
	いない	246	スーパー*, 洗米*, 生
就業状況	有職	225	表示, アンケート, 米, 現在, 今
	無職	223	銘柄**, 価格, 白米, 安全
世帯員数	3人以下	265	香り*, 日本*, 質問, 販売
	4人以上	187	家族**, 信用*, 洗米*, 農薬*, 生*, 米
購入時の考え	高くても味	124	粘り*, コシヒカリ*, 新潟*, 量
	値段も味も	237	洗米
	なるべく安く	21	白米**, 胚芽**, 農業
	無(減)農薬	50	安心**, 農薬**, 白米**, 胚芽**, 安全**, 毎日**, 岩手*, 洗米, 量

注1)カイ2乗検定の結果, 他の属性に比べ記述の多かった語を示す。「**」は1%, 「*」は5%水準, 無印は10%水準で有意。

注2)「18歳未満のこどもがいない世帯」は, 他方の属性と比較して特に記述の多い語がなかった。

注3)「購入時の考え」は, 平成9年度第1回食糧モニター定期調査(食糧庁)においてなされた「あなたは, 米を購入する場合, 品質及び価格等についてどのように考えていますか」という設問とその選択肢「値段が高くて味の良いものを選ぶようにしている」「値段も味もまあまあのもものを選ぶようにしている」「できるだけ値段の安いものを選ぶようにしている」「値段が高くて無(低)農薬などのものを選ぶようにしている」に従って設問をした。

いるカテゴリーやサンプルを座標面上の近くに布置することによって, 視覚的に類似度を示す。したがって, 近くに布置されたキーワードは同じ文章に出現することが多く, 互いに関連があるものと理解することができる。

今回は, 抽出語の中から, 「価格・値段」「ブランド・銘柄」のように同じ意味をなす語や, 「スー

パー・専門(米穀専門店)・店・生(生協)・米屋」「コシヒカリ・ササニシキ・あきたこまち」のように同じカテゴリーで括れるもの, 「販売・購入」「生産・消費」のように対をなしている語をまとめ, さらにキーワードになりそうな語をいくつか拾い上げて20項目について分析を行った(表3, 図4)。

表3 分析に用いたキーワード

キーワード	記述者数
価格・値段	137
スーパー・専門・店・生・米屋	89
あきたこまち・ササニシキ・コシヒカリ	84
販売・購入	80
味	78
品質	77
生産・消費	57
ブランド・銘柄	57
産地・地方	47
農薬	43
袋・表示	38
農家・農業	37
日本・日本人	33
安心・安全	33
精米	32
品種	28
ブレンド	19
粘り・香り	17
信用	16
差	12

その結果、以下の3つの内容を抽出することができた。次元1のマイナス方向に布置されたのは、米の購入先や「ブランド・銘柄」、味などの「差」, 「信用」および「袋・表示」「ブレンド」など、米の購入に関わる語が多いので、「米購入」に関する内容といえる。次元2のマイナス方向には「粘り・香り」や「品質」など米の品質そのものに関する語や、「農業・農家」「生産・消費」「日本・日本人」など日本の産業としての農業という視点を想起させる語が多く布置されていることから「日本の米と農業」に関する内容であると考えられる。次元3はプラス方向に「農薬」と「安心・安全」が布置されているので、「安全性」に関する内容といえる。452件の書き込みは、大別するとこの3つの内容について書かれているということがわかった。

決定木分析の適用

決定木分析は、母集団を基準変数の値が有意に異なるように、説明変数を用いて2つ以上のグ

図4 抽出語の等質性分析結果

注1)SPSS 12.0J Categories による分析。

注2)各項目の次元1と次元2のカテゴリ数量化。楕円は各次元の絶対値が大きいものをまとめた。

図5 各記述内容の決定木分析結果

注1)SPSS AnswerTree3.0 のCHAIDによる。

注2)分岐の有意水準0.05, 樹木の深度3, 親ノード・子ノードの最小ケース数100・50と設定。

注3)「米購入」と「日本の米と農業」はマイナスの絶対値が大きい方が, 「安全性」はプラスの絶対値が大きい方が, それぞれ関連する語を記述する傾向の強いことを示す。

ループに逐次分割していく手法である。農産物に関する研究では、米小売業者の意向調査結果に適用した平尾（1998）がある。ここでは、等質性分析によって明らかとなった記述傾向の違いが、回答者のどのような属性と関係しているかを明らかにするために、各次元のオブジェクトスコアを基準変数とし、回答者の性別、年齢、子ども・高齢者の有無、就業状況、世帯員数、購入時の考えを説明変数として決定木分析を行った（図5）。

その結果、「米購入」と関連の強い語を記述するのは、「60歳未満」の「無職」の回答者がもっとも多いことがわかった。年齢層からみておそらく大半が専業主婦層であろうと判断する。次に多いのが「60歳未満」の「有職」、次に「60歳以上」の「女性」、一番少ないのが「60歳以上」の「男性」であることがわかった。

「日本の米と農業」に関する記述は、「米購入」とは反対に、「男性」の「60歳以上」がもっとも多くなっている。次いで「女性」が多く、最も少ないのが「男性」の「60歳未満」となっている。

「安全性」については、回答者の性別や年齢などの属性ではなく、米購入時の考えによって記述傾向が異なっていた。すなわち、米を購入する時の考えとして「値段が高くて無（低）農薬などのものを選ぶようにしている」と答えた回答者が、他の考えの回答者に比べ、この内容と関連の強い語を記述する傾向にあることが明らかとなった。

分析結果の検証

最後に、これらの分析結果が実際の記述内容や回答者の特徴をどの程度示しているかを検証する。3つの内容のうち、「米購入」および「安全性」に関する内容については、比較的理解を得やすい結果であったといえよう。一方、「日本の米と農業」については、同じ男性でありながら60歳以上と60歳未満で大きく記述傾向が異なるという結果になっている。そこで、ここでは男性によって記述された自由記述回答文に目を通して確認することにする。

男性の回答者は60歳以上が81人、60歳未満が72人である。そのうち、日本の米の品質や農業政策などについて言及している回答者の数は、前者が20人（24.7%）、後者が8人（11.1%）であり、

5%水準で有意に前者が多くなっている。これは、筆者が各文章に目を通し、判断に迷うものについては除外した数であるため、分類の基準によっては若干の違いが生じる可能性はあるが、男性の中で年齢によって「日本の米と農業」に関する内容を記述する傾向に差があることは確かである。

以上のことから、今回の分析結果は実際の文章内容をかかなりの確に示していると考えることができる。

おわりに

本稿では、アンケート調査で得た米に関する自由記述回答文にテキストマイニングを適用し、米に関する消費者意識の解明を行った。

分析の結果、452件の自由記述回答文は、「米購入」、「日本の米と農業」、「安全性」の大きく3つに分類できることがわかった。さらに、「米購入」は専業主婦層が、「日本の米と農業」は60歳以上の男性が、「安全性」は米を購入する時に「値段が高くて無（低）農薬などのものを選ぶようにしている」と考えている人がよく記述する傾向にあることが明らかとなった。

本稿で行った分析は、2000年時点における調査データに基づいている。したがって、その後の社会情勢や食を取り巻く環境の変化によって、現在の消費者意識とは乖離する部分もあるものと推測する。今後、こうした消費者意識の変化を捉える上でも、アンケート調査による自由記述回答文の収集とテキストマイニングによる分析が研究上の意義を増すと考える。

引用文献

- 林俊克（2002）Excelで学ぶテキストマイニング入門，オーム社，東京，237pp.
- 樋口耕一（2004）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）：101-115.
- 平尾正之（1998）食糧法下における米小売業の展開方向，農業総合研究，52（3），1-52.
- 平尾正之・河野恵伸（2000）消費者の米購買行動の変化とブランド化，農業政策研究会編，「国境措置と日本農業」，農林統計協会，東京，273-292.

- 法隆大輔・深津時広・大塚彰ら(2004)農業関連文書用形態素解析サーバーの開発とテキストの自動分類による検証, 農業情報研究, 13(2): 127-138.
- 市村由美・長谷川隆明・渡部勇ら(2001)テキストマイニング事例紹介, 人工知能学会誌, 16(2): 192-200.
- 石井哲(2002)テキストマイニング活用法, リックテレコム, 東京, 263pp.
- 石村貞夫(2001)SPSSによるカテゴリカルデータ分析の手順, 東京図書, 東京, 217pp.
- 磯島昭代(1998)米購入時の選択基準と消費者像ー消費者意識調査による分析ー, 農業経営研究, 36(3): 22-32.
- 磯島昭代(2004)テキストマイニングのための1・0データファイルの作成手順, 東北農研総合研究(B)19, 東北農業研究センター総合研究部, 58pp.
- 牧田亮(1994)消費者行動の定性的アプローチ, 鮑戸弘, 消費行動の社会心理学, 福村出版, 東京, 98-119.
- 徳永健伸(1999)情報検索と言語処理, 東京大学出版会, 東京, 234pp.
- 梅本雅・大浦裕二・山本淳子(2002)米購入時における消費者の意思決定過程の実態と特徴, 農業経営研究, 40(3): 26-38.
- 梅澤伸嘉(1993)実践グループインタビュー入門, ダイヤモンド社, 東京, 228pp.
- 山崎秀夫・松田潤(2002)顧客を創造するテキストマイニング, 日本工業新聞社, 東京, 199pp.
- 柳井晴夫・岡太彬訓・繁樹算男ら(2002)多変量解析実例ハンドブック, 朝倉書店, 東京, 889pp.

受付日 2005年9月16日

受理日 2005年12月26日

担当部会 経済・社会情報部会

Analysis of a Consumer Questionnaire Pertaining to Rice by using Text Mining

Akiyo Isojima*

National Agricultural Research Center for Tohoku Region 4 Akahira, Shimokuriyagawa, Morioka, Iwate 020-0198, Japan

Summary

Several investigations have sought to clarify the attitudes of consumers to rice. One such investigation involves a closed-form questionnaire survey, and another is an interview survey, similar to a group interview, that poses questions to people in a group. The responses to the open-ended questions collected from the questionnaire survey constitute a considerable amount of qualitative data, which may provide us with new findings. To date there has not been any effective technique for analyzing such data. However, the development of text mining has now enabled the analysis of large amounts of textual data.

With the aim of clarifying the attitudes of consumers to rice, this study analyzed the responses to the open-ended questions by using text mining. According to the result obtained, the concerns of the consumers

pertaining to rice were classified into three groups: "buying rice," "rice and agriculture in Japan," and "safety." It was observed that many of the responses that fell into the category "buying rice" were those of full-time housewives. A number of the responses that were categorized in "rice and agriculture in Japan" were those of men aged 60 years and above. Finally, many of the responses that were categorized in "safety" were those of people who were sensitive to the use of pesticides.

Keywords:

Consumer, Rice, Questionnaire survey, Open-ended question, Text mining

---

*Corresponding Author

E-mail:isojima@affrc.go.jp