

■ 論文 ■

読者投稿の記述的計量テキスト分析

－「声」と「気流」－

中 野 康 人
(先端社会研究所)

■ 要 旨 ■ 本稿の目的は、日刊の全国紙である朝日新聞と読売新聞の読者投稿欄について、その投稿者の属性と投稿内容を記述的に分析することにある。中野（2009）で提示された作業にもとづいて、朝日新聞「声」は1989年から2007年まで、読売新聞「気流」は1990年から2007年まで、投稿者の属性と記事内容をデータベース化した。このデータベースを利用して、どのような人が記事を投稿し、どのような内容が語られているのかを、記述的に分析する。

■ キーワード ■ 新聞記事、内容分析、計量テキスト分析

1 問題の所在

日本の日刊新聞には、読者投稿欄というものがある。読者投稿欄の記事は、毎日のように蓄積される人々の意見表明のデータである。もちろん、投稿者の偏り、新聞社や編集者による選別・編集などがあるので、世論全体のそのままの縮図ではない。しかしそこにあらわれる意見が社会のどの部分から滲み出したものなのか、という議論を留保したとしても、長年にわたって蓄積された記事データは社会のある側面を切り取る貴重な資料といえる。

本稿の目的は、日刊の全国紙である朝日新聞と読売新聞の読者投稿欄について、その投稿者の属性と投稿内容を記述的に分析することにある。中野（2009）は、学術用新聞記事データから、読者投稿欄の情報のみを取り出し、投稿者の属性と投稿内容をデータベース化する方針とその可能性を示唆した。ここでいうデータベースは、たとえば竹下（1994）で述べられているような、単に記事を集めただけのものではなく、どのような人がその内容を書いたのか、ということが明確にわかるかたちのデータベースである。本稿では、作成したデータベースをもとにした分析の紹介として、投稿者情報の概要と、投稿記事内容を記述的テキスト分析した結果を報告する。

2 使用するデータ

今回使用するデータは、朝日新聞社発行の「朝日新聞記事データ集 学術・研究用」と、読売新

表1 投稿欄の概要

新 聞	総件数	一日あたり	データ化期間
朝日新聞「声」	52,346	約8件	1988年～2007年
読売新聞「気流」	41,178	約10件	1990年～2007年

表2 投稿者と投稿件数

新聞紙名	投稿者	総数一人あたり投稿件数			
		平均	中央値	最大値	標準偏差
朝日新聞「声」	32,798	1.696	1	72	1.92
読売新聞「気流」	20,826	1.951	1	92	3.25

聞社発行の「読売新聞記事データ集」である¹⁾。これらのデータは、内容分析にかけられることを著作権者が認めた形で販売されているもので、それぞれ、朝日新聞本社版と読売新聞本社版および地方版の記事テキストを基本的にすべて収録したものである。それらの記事データの中から、両紙の読者投稿欄の記事のみを取り出した。

朝日新聞の読者投稿欄は「声」欄を、読売新聞の読者投稿欄は「気流」欄を対象にする。どちらも、記事内容以外に、投稿者の氏名、住所、年齢、職業が基本的に掲載されている。上記のデータからこれらの要素が取り出せるのは、朝日新聞については1998年から、読売新聞については1990年からである。現時点でデータベースとして抽出できている最新の記事は、2007年末のものである。これらの約20年の記事について、中野（2009）の方法にしたがってデータベース化した読者投稿欄の記事を分析していく²⁾。

3 分析

3.1 投稿欄の概要

投稿記事全体の件数は表1の通りである。データの全期間を通して、「声」の記事は5万件強、「気流」の記事は4万件強存在する。「声」「気流」とともに、ほぼ毎日10件弱が掲載されている。これは、この20年程一定しており、月によっても大きな変化はない。

3.2 投稿者の概要

3.2.1 一人あたりの投稿件数

次に、どのような属性の人が投稿しているのか、投稿者の概要をみてみよう。投稿者の氏名で異なるものの頻度をとると、表2のようになる。一人あたりの平均投稿件数は2件弱で、中央値は1件である。同姓同名の区別を行っていないが、「声」では8割弱の投稿者が一度きりの投稿であるのに対して、「気流」で一度きりの掲載経験者の割合は7割強となっている。全期間を通じて、10回以上投稿者として登場する氏名の数をかぞえと、「声」は342名（全投稿者の1%）、「気流」

1) どちらも販売は日外アソシエーツである。

2) 中野（2009）では、朝日新聞の記事データから「声」を抽出する作業の説明を行った。読売新聞「気流」については触れていないが、基本的な作業方針は同一である。ただし、朝日新聞に比べて読売新聞の記事は表記の形式の揺れが激しく、「気流」のデータベース化には大幅な手間を必要とした。

表3 A氏の年齢と年毎の投稿数

	1988	1990	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
31	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	1	6	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	1	7	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

表4 A氏の投稿タイトル

1	手助けの基本を学校で教えて	37	長老議員さん、引退考えたら
2	「忙しい」の言葉をもう禁句にしよう	38	挑発的な姿勢、やはり変です石原発言
3	聴障者の不便さ知って、つらさ思う日	39	こんな首相が選ばれる政治
4	字幕付きの放送をもう少し増やして	40	また場あたり、扇氏の建設相
5	会社やめ育児、父親も楽しい	41	名だけの秘書、防ぐ仕組みを
6	聴力障害者の話はていねいに聞いて	42	信用の回復に学界は全体で
7	男性の育児も当然の社会に	43	三宅村は東京都にへつらわずに
8	育児で退職の男の再就職は	44	「三割削減」は簡略化しすぎ
9	聴力障害者のためTVにテロップを提案します	45	JOCが象徴、若手阻む体質
10	男の子育てに社会復帰の壁	46	「忘れた」発言、気楽過ぎます
11	字幕機能内蔵テレビ量産は	47	PTA活動に男性の参加も
12	閣僚ポストに、つられるとは	48	議定書批准し信頼回復せよ
13	謝罪ですまぬ政治家の発言	49	省庁お目付け、外から起用を
14	夫から妻への「育児休暇」を	50	神経逆なでの安全アピール
15	スーツ2着で免職の公務員	51	巨泉氏の辞職、最善の選択か
16	発信側の番号がわかる電話に	52	医療ミス防止、新制度に疑問
17	責任と謝罪に潔さ見えない	53	人を傷つける報道は勇み足メディア
18	納得できない派閥均衡人事	54	参拝した理由、十分に説明を
19	坂本弁護士事件でTBSの内部に批判はなかったのか	55	手話の通訳で娘の門出祝う
20	何とも気楽な政党の社民党	56	疑惑の説明を怠った社民党
21	比例区当選者の離党は許されぬ	57	藤井問題でも傍観的な首相
22	釈然としない自賠責の実態	58	品性の低さが目立つばかりこの1年：中
23	閣僚の年齢に制限必要では	59	三原則にまで要望出すとは
24	大きな励まし、ジャンプのV	60	首相と3閣僚、お粗末すぎる
25	うんざりする天下り正当化	61	格差を是正し、参院の存在を
26	比例区議員の離党に規定を	62	なぜ説明なく国は見直しか
27	国民が参加し首相選べたら	63	言葉足らない答弁いつまで
28	行革や合理化、国がまず示せ	64	年齢の「壁」に育児の空白も
29	急がぬ勇気も報道には必要新聞に：上	65	選挙は情報だショーは不要
30	蔵相の辞任は当然だけれど	66	容疑者を応援、首相の責任は
31	小児科の支援、国の責任では	67	政治家たちはテレビ出過ぎ
32	ゆったり授業、少人数で実感	68	のぞき見的報道はやめて
33	相乗り知事は選挙後が大切	69	派閥の報道は必要ないです
34	塾補助よりも教員の増員を	70	男の育児にも偏見を感じる
35	次官任命した首相らに責任	71	緑資源の談合、監督責任問え
36	信頼の回復を新年の課題にこの一年：下	72	国民への説明、政治家の責任

は 467 名（全投稿者の 2 %）である。「気流」の方が若干常連投稿者が多いことになる。

3.2.2 常連投稿者

ここで、複数回登場する投稿者名が同一人物なのかを確認してみよう。例えば、「声」で最多の投稿数をほこるのは、A 氏³⁾の 72 件である。同姓同名の可能性もあるが、1988 年から 2007 年まで、年齢が段階的に上昇し、住所もほぼ同一の場所であることを鑑みると、同一人物が 20 年間投稿を続けている事が類推される（表 3）。同氏の職業の変化、記事内容の変化を追いかければ、一人の投稿者の 20 年の意見表明史をみることができる（表 4）。こうした分析も、読者投稿欄をデータベース化したからこそできる作業である。

表 5 投稿者の年齢

新聞紙名	平均	最小値	中央値	最大値
朝日新聞「声」	49.45	6	50	99
読売新聞「気流」	49.12	1	49	102

3.2.3 投稿者の年齢

次に、投稿者の年齢分布をみてもみる。「声」「気流」とともに、投稿者の平均的年齢はおおよそ 50 才である。最小年齢は、どちらも 6 才になる⁴⁾。年齢分布のグラフ（図 1）を見ると、「声」は 60 代が突出している分布、「気流」は 30 代から 40 代の山と 60 代の山の双極分布となっていることがわかる。また、特に「声」に関しては、20 代に比較して学齢期の投稿者が多い事が目立つ。投稿者の年齢を時系列的にグラフ化すると、平均年齢が漸次上昇していることがわかる。「声」では、1988 年の 47.13 才から 2007 年の 50.89 才へ、「気流」では、1990 年の 46.28 才から 2007 年の 52.13 才へと高齢化が進んでいる。

3.2.4 投稿者の職業

投稿者の職業も多種多様である。記事中の投稿者の職業は、投稿者の申請によるものと思われ、公的な職業分類等に則ったものではない。「声」には 3,000 を超える職業名が、「気流」では 1,500 弱の職業名が出てくる。どちらについても、一番多いのは「主婦」である。ついで「無職」「会社員」と続く。こうしたかなり一般的な職業名が投稿の半数を占める一方で、中には非常に細かな従業上の地位まで明確にしたものもある。とくに、教育関係の職業は所属機関の詳細や従業上の地位まで明記したものがめだつ。たとえば、「大学教員」という一般的なものもあれば、「大学教授」「大学講師」「大学助教授」「大学非常勤講師」などの地位属性をいれたものまで出てくる。

投稿者の職業について、「声」と「気流」の違いをあげるとすれば、「気流」の方が主婦率が 10% ほど高いということ、「声」の教員率⁵⁾が「気流」の倍であるということ、などがあげられる。

3) データには個人名が含まれるが、個人を特定する事が目的ではないので、ここでは仮名とする。

4) 「気流」の最小年齢はデータ上は 1 才になっているが、職業が大学生になっていることから判断すると 1 才児の投稿ではないものと思われる。また、4 才の投稿者も一件あるが、こちらも職業が主婦になっているため、誤植と予想される。

5) 以下の職業をまとめて「教員」とした。「教員」、「高校教員」、「大学教員」、「大学教授」、「大学講師」、「小学校教員」、「元教員」、「高校講師」、「中学校教員」、「日本語教師」、「高校教諭」、「大学名誉教授」、「中学教員」、「高校教師」、「大学助教授」、「大学非常勤講師」、「教師」、「短大教員」、「非常勤講師」、「小学校教諭」。こ

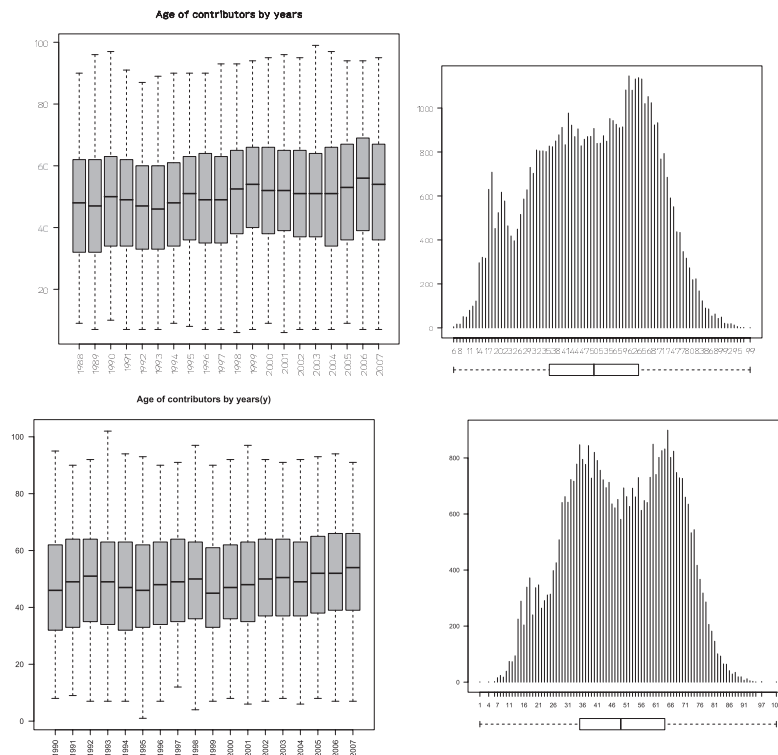


図 1 投稿者の年齢分布

表 6 投稿者の職業

新聞紙名	職業数	主婦率	無職率	会社員率	教員率	小中高生率
朝日新聞「声」	3,134	20.7%	17.2%	9.9%	6.9%	4.6%
読売新聞「気流」	1,482	29.5%	17.7%	12.5%	2.8%	3.8%

表 7 投稿者の「元」職とその頻度（10 回以上）

朝日新聞「声」	「元教員」(150), 「元公務員」(53), 「元教師」(37), 「元会社員」(34), 「元大学教授」(25), 「元高校教員」(23), 「元中学校長」(21), 「元高校教師」(18), 「元小学校長」(18), 「元銀行員」(14), 「元看護婦」(13), 「元商社員」(11), 「元高校教諭」(10), 「元新聞記者」(10)
読売新聞「気流」	「元教員」(55), 「元公務員」(22), 「元中学校長」(22), 「元小学校長」(14), 「元自衛官」(13)

投稿者がどのような職業名を名乗るのか、ということ自体が社会的に規定されたものであることを考慮すれば、投稿者の職業名の変化を時系列的に分析することは興味深い作業になる。例えば、「アルバイト」という職業名は、「声」「気流」ともに、全期間を通じて毎年出現し、頻度は緩やかに上昇している。また、「フリーター」という職業名も両者に登場するが、「声」に始めてフリーターの投稿が掲載されるのは 1990 年、「気流」の場合は 1994 年で、その後継続的に登場している。

投稿者の職業の中には「元〇〇」というように、以前の職業を表記したものがすくなくからず出てくる(表 7)。現在の職業が「無職」なのかそれとも何か別の職業に転職しているのかは定かではない。

れらは、比較的頻度が高い教育的職業である。

表8 記事内容の形態素解析結果

	朝日新聞「声」	読売新聞「気流」
抽出単語数（重複を含む）	14,742,524 語	9,639,830 語
抽出単語数（異なる単語）	69,405 語	52,984 語
最大出現回数	832,592	551,990
平均出現回数	212.413	181.939
出現回数標準偏差	7023.093	5286.124
一記事あたりの単語数	281.642	234.102
名詞		
抽出名詞の総出現回数	4,932,896	3,143,625
抽出名詞単語数	58,877	44,154
抽出名詞の最大出現回数	112,566	74,398
抽出名詞の平均出現回数	83.783	71.197
抽出名詞の出現回数の標準偏差	996.521	744.203

また、記事内容がその「元」職と関係ある場合もあれば、直接関係ない場合もある。医療や看護婦の勤務状況についての投稿者が「元看護婦」と表明しているのは、内容と職業がリンクしている一例である。一方で、「元女優」が戦前の生活を回顧したり、「元衆議院議員」が電車内での座席マナーを嘆いたりという記事は、内容と職業のリンクは強くない。いずれにせよ、「元」職が投稿者の表明されたアイデンティティであることは容易に想像できる。職業とアイデンティティを分析する素材としても、このデータは有用である。「声」では238の、「気流」では101の「元」職が表明されている。

4 記事内容のテキスト分析

次に、具体的な投稿記事の内容を、形態素解析にかけ、テキスト分析を行う。分析には、MeCab⁶⁾を利用した。その解析結果は、表8の通りである。「声」も「気流」も、相異なる単語が万単位で出現する。出現回数の分布は、Zipの法則に従っており、一回しか出現しない単語が約4分の1を占め、半数の単語は五回以下の出現である。品詞別の出現単語数は図2の通りで、名詞系の単語が一番多く出現している。

「声」と「気流」を比較すると、「声」の方が単語の種類が多い。「声」の方が期間が長く件数が多いので、その分単語の種類が増えているのかもしれない。そこで、「声」についても「気流」と同じ1990年からの期間のデータで形態素解析をおこなってみた。すると、抽出単語の種類は66,264語（抽出単語数13,205,397語）となり、依然として「声」の記事の方がより多くの種類の単語を使っていることがわかる。一記事あたりの単語数（重複を含む）は、「声」が「気流」の1.2倍になっているので、記事の長さ按比例して単語の種類が増えているのかもしれない。しかし抽出された単語の種類は、「声」が「気流」1.25倍となっている。言語学で使用されている、語彙の豊かさを示す指標のいくつかを計算してみると表9のようになる⁷⁾。TTR (Type Token Ratio)を除いて、「声」の方が「気流」よりも指標値が若干高くなっている。TTRについては「気流」の方が高

6) MeCabは、京都大学情報学研究科－日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクトを通じて開発されたオープンソース形態素解析エンジンである。(http://mecab.sourceforge.net/)

7) 指標の詳細は、金（2009）を参照。

表 9 記事内容の語彙の豊かさ (1990-2007)

	朝日新聞「声」	読売新聞「気流」
抽出単語数 (重複を含む)	13,205,397	語 9,639,830 語
抽出単語数 (異なる単語)	66,264 語	52,984 語
TTR	0.00502	0.00549
Guiraud's R	18.234	17.065
Herdan's C	0.677	0.676
Dugast's k	3.969	3.916

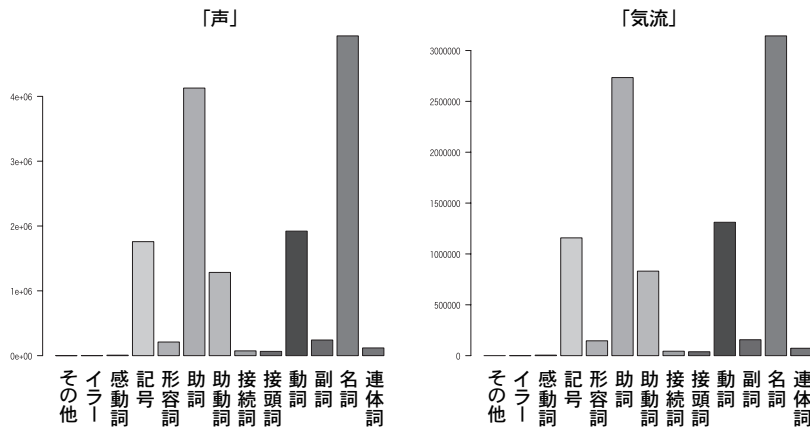


図 2 品詞別の抽出単語数

い値を示すということは、「声」の方が全体として言葉の使い方に比較的冗長性があるということだろう。しかし、単語の種類やその他の指標値は「声」の方が高いということは、若干ではあるが、「声」の投稿者の方が「気流」の投稿者よりも広範な話題・語彙に触れていると解釈できるだろう。

では、「声」と「気流」で、語られている内容に差はあるのだろうか。記事内容の傾向をつかむために、それぞれの頻出単語上位 100 語をあげたのが表 10 である。ここで抽出しているのは、McCab が名詞系の単語として認識したもので、上位を占めるのは「の」「こと」「私」「人」など、非常に一般的な単語であり、「声」「気流」に共通している。

単純な比較ではあるが、この表 10 について、一方には出現するがもう一方には出現しない単語をピックアップすると表 11 のようになる。たとえば、「戦争」は「声」では 49 番目に多い単語で 8,834 回出現しているが、「気流」では 315 番目の 2,436 回である。逆に、「車」は「気流」では 55 番目に多い単語で 5,369 回出現しているが、「声」では 225 番目の 4,863 回である。このように、相対的に「声」でよく話題になるもの、逆に「気流」によくでてくるもの違いが明確になる。「声」では、政治や国家、教育などにかかわる単語が特徴的であるのに対して、「気流」では、身近なメディアの話題が特徴的である。

図 3 は、「声」の記事中に含まれる単語について、単語間のつながりを抽出し、最低 100 回は出現するものだけをネットワークグラフ化したものである。「声」では 2,706 の、「気流」では 1,565 の bigram が抽出されている。抽出単語の bigram にもとづくネットワーク図は、テキストデータに含まれる単語の文脈を概観するうえで有用なツールではあるが、このくらい単語の数が多くなると、ネットワークが密になりすぎて判別が不可能になってしまう。記事内で語られている内容を分析す

表 10 頻出単語上位 100 語

朝日新聞「声」										
の	こと	私	人	よう	たち	者	日	1	もの	日本
112566	110649	73671	59253	51042	35847	33146	31819	30829	28150	28146
年	的	一	0	ため	中	それ	2	時	十	今
28137	25406	24780	22750	21958	21619	21446	21147	20773	20327	18835
何	自分	さ	問題	二	これ	子供	国民	3	前	政治
18353	16471	16075	15795	15650	15284	15014	14198	14039	13725	13180
さん	方	国	社会	家	心	生活	5	目	三	学校
12697	12456	12006	11638	11024	10655	10569	10293	10213	10082	9929
必要	歳	化	4	戦争	言葉	円	時間	後	教育	五
9865	9517	9286	8905	8834	8808	8767	8668	8402	7877	7870
数	母	声	仕事	そう	気	多く	手	選挙	世界	時代
7523	7487	7463	7401	7399	7098	7090	7049	6939	6852	6843
会	先生	性	ところ	人間	政府	6	力	気持ち	家族	氏
6629	6565	6432	6362	6325	6315	6254	6155	6124	6102	5993
日本人	思い	娘	上	企業	米	はず	万	子	四	大学
5938	5900	5888	5881	5875	5872	5799	5792	5743	5709	5663
女性	子ども	度	以上	責任	9	会社	話	夫	事件	父
5659	5572	5567	5531	5526	5500	5463	5376	5351	5334	5329
関係										
5260										
読売新聞「気流」										
こと	の	私	人	よう	一	たち	者	十	子供	日
74398	70076	42608	42294	38523	22586	22099	20695	18999	18144	18060
もの	二	年	時	ため	日本	中	的	今	何	さ
16974	16574	16141	15338	15283	14349	14076	13378	11547	11048	11015
それ	三	自分	問題	前	これ	方	1	五	さん	心
10879	10604	10466	9983	9227	8946	8383	8245	7735	7653	7595
国民	必要	0	歳	目	社会	家	2	そう	四	学校
7557	7362	7057	7014	6930	6827	6647	6553	6416	6351	6309
政治	時間	生活	言葉	国	気	化	娘	円	後	車
6297	6237	6004	5895	5749	5700	5657	5640	5388	5373	5369
気持ち	教育	手	家族	数	母	電話	親	大切	事件	仕事
5155	4991	4967	4964	4931	4923	4839	4817	4775	4737	4714
息子	多く	声	ところ	女性	最近	六	性	百	3	先生
4686	4537	4478	4427	4387	4372	4350	4327	4233	4168	4166
時代	先日	ころ	会	度	夫	子	テレビ	今年	以上	話
4096	4001	3982	3950	3910	3878	3873	3862	3853	3827	3820
思い	力	はず	姿	万	利用	今回	事故	八	七	関係
3770	3762	3709	3679	3625	3594	3593	3590	3571	3520	3517
上										
3514										

表 11 頻出単語：「声」と「気流」の違い

「声」の特徴的単語	戦争, 選挙, 世界, 人間, 政府, 氏, 日本人, 企業, 米, 大学, 子ども, 責任, 会社, 父
「気流」の特徴的単語	車, 電話, 親, 大切, 息子, 先日, ころ, テレビ, 今年, 姿, 利用, 今回, 事故

るのであれば、単語（群）を特定し、それについて関連する語の距離を測ったり、属性との関係を調べるのが現実的であろう。

例えば、「日本人」という言葉が投稿記事内のどのような文脈で使用されているのかを調べてみ

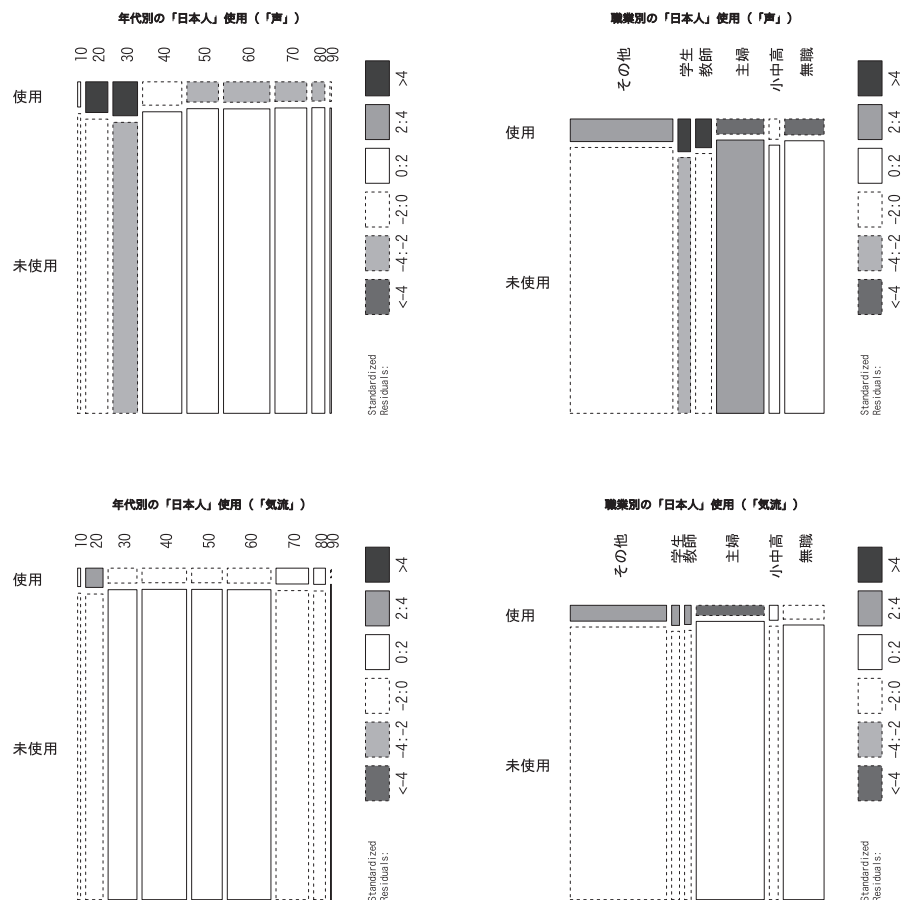


図5 投稿者の属性と「日本人」使用の関係

件である。「日本人」と「外国」が同時に観察されるのは、281件ある⁹⁾。これをもとに T 値を算出すると、 $T=14.87$ となり、「日本人」「外国」は強い共起関係にあるといえる。

どちらの紙面でも、「外国」や「海外」、その他の国名は関係の強い語として上位にあがっている。「声」と「気流」の顕著な違いは次のとおりである。まず「気流」で「拉致」が非常に強い共起語になっている。「声」でも「拉致」は共起語であるが、 T 値は 8.634 と 4.155 と倍以上の差がある。また、「気流」には、「選手」や「力士」などスポーツ関係の単語が強くでている。「イチロー」に至っては、「声」のリストには出現していない。「声」では、「アジア」や「韓国」、「中国人」といった周辺諸国に関する単語が比較的強く出現している。また、「声」のリストにあって「気流」に無いものとしては、「アイデンティティ」があげられる。

頻度分析では「日本人」という単語は、「気流」（3,062 件の 233 番目）に比べて「声」（5,938 件の 78 番目）に比較的多く出現するという特徴があった。このように共起関係の分析を行うと、そうした頻度の特徴だけでなく、語られている文脈の違いも明確になる。

さらに、投稿者の属性と「日本人」という言葉との関係をみてみよう（図 5）。年齢別の分析では、

9) ここでは、「日本人」の前後 10 語に出現すれば、「同時」とみなす。

いくつかの年代で有意に「日本人」を使用する傾向がみてとれる。「声」では20代、30代の使用者が有意に多い。逆に、50代以上では使用者が減る傾向にある。「気流」では、20代のみ有意に使用者が多い。職業別の分析でも、「日本人」を多く使用する職業とそうでない職業があることがわかる。「声」では、「学生」「教師」「その他」に多く、「主婦」や「無職」では使用者が減る。「気流」では、「学生」や「教師」そして「その他」が使用する傾向にある。

共起分析の結果、属性との関連分析の結果をまとめてみると、投稿欄における「日本人」という言葉の語られ方がみえてくる。比較的若い層が、自分自身の経験やメディアで活躍する人間を通して海外との対比をする、というのがまず頭に浮かぶ解釈である。もちろんこれはひとつの解釈であって、「日本人」という内容のすべてではない。全体像を把握するには、さらなる分析が必要であるが、それは本稿の範囲を超えるので、ここまでにとどめておく。

表12 「日本人」の共起語

「声」単語	T 値	MI 値	「気流」単語	T 値	MI 値
外国	14.867	3.144	外国	11.25	3.47
多く	12.404	2.163	拉致	8.63	4.11
アメリカ	9.399	2.262	海外	7.84	3.40
海外	8.981	2.799	選手	7.18	2.21
文化	8.425	2.320	日本語	6.83	3.09
英語	8.348	2.562	力士	6.68	4.17
韓国	8.071	2.504	北朝鮮	6.34	2.95
朝鮮	8.014	3.190	誇り	6.29	4.02
意識	7.809	2.244	英語	6.26	2.68
留学生	7.493	3.564	客	6.18	2.81
日本語	7.294	2.362	観光	5.92	3.37
観光	7.032	3.132	文化	5.72	2.51
恥ずかしい	6.994	3.072	旅行	5.64	2.37
観	6.653	2.825	米国	5.54	2.21
欧米	6.627	3.128	アメリカ	5.51	2.34
中国人	6.614	3.760	活躍	5.32	2.87
客	6.603	2.693	らち	5.30	5.91
アジア	6.451	2.368	意識	5.26	2.00
誇り	6.261	2.837	韓国	5.24	2.74
我々	5.934	2.096	恥ずかしい	5.04	2.89
現地	5.542	2.524	現地	5.02	3.14
われわれ	5.534	2.594	欧米	5.00	3.53
民族	5.496	2.544	マナー	4.99	2.40
残留	5.268	3.863	留学生	4.97	4.06
在日	5.189	2.660	中国	4.96	2.10
帰国	5.012	2.141	イラク	4.82	2.90
同士	4.959	2.527	孤児	4.69	4.52
勤勉	4.954	5.138	残留	4.15	4.35
墓地	4.761	3.913	訪れる	4.10	2.15
活躍	4.641	2.477	中国人	4.07	3.91
感情	4.590	2.089	我々	3.96	2.07
感覚	4.578	2.016	タイ	3.93	3.33
習慣	4.466	2.437	考え方	3.89	2.41
マナー	4.448	2.413	情けない	3.80	2.74
船員	4.392	4.585	礼儀	3.77	3.53
白人	4.235	4.238	語	3.77	2.23
孤児	4.212	3.037	美德	3.75	5.02

「声」単語	<i>T</i> 値	<i>MI</i> 値	「気流」単語	<i>T</i> 値	<i>MI</i> 値
拉致	4.155	2.903	在住	3.73	4.80
抑留	4.089	3.214	疑惑	3.70	2.72
人質	4.013	3.654	帰国	3.68	2.34
在住	3.932	3.773	低い	3.66	2.19
滞在	3.830	2.446	観	3.60	2.52
働き	3.770	2.887	人質	3.57	4.41
貧しい	3.683	2.107	イチロー	3.40	4.10
話せる	3.675	2.902	ロシア	3.38	2.47
留学	3.646	2.060	話せる	3.35	3.24
平均	3.640	2.158	台湾	3.34	3.22
美德	3.635	4.022	抑留	3.31	3.62
情けない	3.633	2.149	食生活	3.29	3.53
アイデンティティー	3.609	4.814	解放	3.20	3.17
下手	3.567	3.206	感覚	3.17	2.12
捕虜	3.496	2.989	とかく	3.15	3.48
ロシア	3.493	2.072	感情	3.11	2.34
大戦	3.484	2.061	宇宙	3.10	3.25
グループ	3.437	2.000	恥	3.10	3.92
黒人	3.398	3.446	民族	3.06	3.08
人種	3.372	2.671	アジア	3.03	2.20
国籍	3.346	2.106	射殺	3.03	4.63
特有	3.323	3.673	カンボジア	2.97	3.26
来日	3.300	2.324	大リーグ	2.93	3.10
礼儀	3.257	2.948	相撲	2.92	2.19
シベリア	3.253	2.644	文民	2.91	5.12
心情	3.247	2.629	母国	2.90	3.57
タイ	3.229	2.376	初	2.86	2.08
気質	3.154	4.354	宗教	2.84	2.05
無神経	3.154	3.480	飛行	2.78	3.75
補習	3.106	3.977	殺害	2.75	2.55
シンガポール	3.091	3.215	勤勉	2.70	4.42
オーストラリア	3.085	2.791	メジャー	2.69	4.30
ユーモア	3.061	3.103	金持ち	2.67	4.11
尊敬	3.050	2.235	訪日	2.63	3.81
力士	3.049	2.433	横綱	2.62	2.25
地球人	3.035	4.639	サポーター	2.59	3.58
日系	2.954	3.192	映る	2.58	2.44
憎しみ	2.946	3.163	ツアー	2.57	2.79
フィリピン	2.944	2.447	人種	2.57	3.47
嫌い	2.938	2.051	われわれ	2.55	3.33
主食	2.919	3.698	モラル	2.55	2.37
外人	2.906	3.625	主食	2.55	4.79
あいまい	2.886	2.128	各国	2.53	2.08
コンプレックス	2.858	4.403	古来	2.51	4.33
母国	2.857	2.851	姓	2.51	3.15
金持ち	2.850	2.828	留学	2.51	2.28
大リーグ	2.835	3.272	来日	2.49	2.55
宇宙	2.831	2.218	拘束	2.48	3.02
ツアー	2.787	2.356	平均	2.47	2.20
野茂	2.775	3.739	安否	2.46	3.86
謝る	2.753	2.284	冷たい	2.42	2.09
憎む	2.736	2.891	英会話	2.39	3.37

「声」単語	<i>T</i> 値	<i>MI</i> 値	「気流」単語	<i>T</i> 値	<i>MI</i> 値
射殺	2.732	3.487	友好	2.38	2.67
移民	2.716	3.403	日の丸	2.37	2.63
肉親	2.707	2.443	乏しい	2.34	2.52
トルコ	2.700	3.324	体形	2.34	4.53
駐在	2.690	3.273	移民	2.33	4.37
愛国心	2.685	2.392	大戦	2.31	2.46
寿命	2.683	2.721	離れ	2.31	2.12
帰化	2.672	4.180	一行	2.30	4.01
風土	2.672	2.690	人力	2.30	4.01
殺害	2.629	2.569	在日	2.28	2.85
とかく	2.606	2.508	野茂	2.28	3.82
モンゴル	2.599	2.902	カナダ	2.27	2.82
籍	2.596	2.892	曙	2.27	3.78
在米	2.594	5.681	国技	2.25	3.60
高慢	2.590	3.569	敗戦	2.24	2.69
ホームステイ	2.583	2.845	トルコ	2.23	3.50
モラル	2.574	2.158	飽食	2.23	3.50
在外	2.550	3.346	州	2.22	2.64
植民	2.547	2.107	落書き	2.22	3.43
東洋	2.545	3.317	古く	2.21	3.33
おおかた	2.521	4.411	在米	2.21	6.46
ハーフ	2.512	4.309	箸	2.21	3.38
ガイド	2.507	2.272	国旗	2.20	2.58
軽べつ	2.503	4.214	公德	2.19	5.54
ユダヤ	2.491	3.066	大リーガー	2.19	5.46
進出	2.491	2.007	捜す	2.17	3.11
無知	2.484	2.538	ガイジン	2.16	4.93
世界一	2.468	2.494	劣る	2.16	3.10
訪日	2.468	2.972	死因	2.15	4.77
狙う	2.467	2.185	工作	2.14	2.99
好む	2.454	2.159	朝鮮	2.14	2.38
すぎ	2.443	2.430	シベリア	2.13	2.96
カナダ	2.436	2.123	奮起	2.12	4.30
苦手	2.434	2.118	知り合う	2.12	2.90
閉鎖	2.434	2.118	肉親	2.12	2.91
古来	2.433	3.639	会談	2.11	2.31
独特	2.409	2.343	寿命	2.10	2.82
恥	2.406	2.336	器用	2.09	3.95
関取	2.397	5.546	体格	2.08	3.87
食生活	2.397	2.713	日系	2.07	3.79
買春	2.381	5.154	忘れかける	2.06	3.70
恨む	2.376	3.293	英国	2.06	2.18
見習う	2.366	2.612	見なす	2.05	3.60
美意識	2.364	4.845	銃	2.05	2.15
好み	2.363	2.603	無知	2.04	3.54
居住	2.363	2.236	賢い	2.04	3.52
当地	2.358	2.224	反日	2.03	3.44
しゃべれる	2.322	4.261	犯す	2.03	2.10
信仰	2.315	2.461	知らず	2.03	2.55
排他	2.312	4.154	独特	2.02	2.51
上海	2.309	2.974	純粹	2.02	2.50
レストラン	2.299	2.097	見方	2.01	2.05

「声」単語	<i>T</i> 値	<i>MI</i> 値	「気流」単語	<i>T</i> 値	<i>MI</i> 値
真珠湾	2.273	2.826			
狂う	2.270	2.814			
観客	2.267	2.033			
誇れる	2.267	2.803			
恥づ	2.256	3.663			
屈辱	2.254	2.757			
お上	2.239	2.702			
ブラジル	2.238	2.261			
品位	2.227	3.458			
西洋	2.224	2.649			
溶け込む	2.223	3.437			
振る舞い	2.220	3.417			
鈍感	2.214	3.376			
ロンドン	2.188	2.530			
熱す	2.186	5.476			
シャイ	2.182	5.376			
誠実	2.164	2.091			
在留	2.148	3.022			
死生	2.143	4.582			
バリ	2.139	2.385			
人情	2.133	2.367			
特性	2.132	2.946			
風習	2.132	2.946			
冷める	2.128	2.931			
元号	2.119	2.888			
品性	2.114	4.195			
情	2.082	2.739			
ペルー	2.069	2.688			
弱点	2.067	3.728			
姓名	2.064	3.698			
西欧	2.063	2.663			
ルーツ	2.057	3.639			
パチ	2.053	3.610			
あこがれ	2.042	2.132			
大連	2.042	3.528			
行儀	2.035	3.476			
感銘	2.033	2.110			
訳す	2.031	3.450			
ゲリラ	2.028	3.425			
ボケ	2.027	2.096			
運航	2.021	3.376			
教徒	2.010	3.306			
永住	2.010	2.480			

5 まとめと今後の課題

ここまで、データベース化した読者投稿欄の記述的分析を紹介してきた。投稿者の概要、そして投稿記事に含まれる言葉の概要を概観した。どのような投稿者がどのような内容を投稿しているのか、記事データの計量テキスト分析を行うことで明確に見えてくるものがあることを示唆できた。しかし、今回提示できたのはデータの概略だけであり、膨大なデータの全容を詳細に分析できてい

るわけではない。記事内容が広範にわたることを考えれば、詳細な分析は、田中(1993)や樋口(2004)のように内容に関するテーマを設定して、そのテーマについての投稿者の属性や関連する言葉の分析を行う、という方針が現実的であろう。そういう意味で、このデータをさらに利用していくことが今後の課題である。

また、このデータの外部的な課題にも触れておきたい。このデータを利用することによって、職業や年齢と投稿内容の関連があきらかになるが、それはあくまで各新聞社の投稿記事内だけでのなしである。投稿者さらには投稿内容が、投稿者以外の一般的な人々とどのようなずれがあるのか、ここでは明確にできない。このことは、それぞれの新聞社の読者層にどのような偏りがあるのか、どのようなプロセスで投稿者が選別され、また投稿内容に編集の手が加わっているのか、といった情報なしでは考えることができない。編集プロセスまで視野にいった、新聞の読者投稿に関する総合的な研究も課題として残っている。

参考文献

- [1] 樋口耕一, 2004, 「計算機による新聞記事の計量的分析－『毎日新聞』にみる「サラリーマン」を題材に－」, 『理論と方法』 19 (2) :161-176.
- [2] 石田基広, 2008, 『R によるテキストマイニング入門』, 森北出版.
- [3] 金明哲, 2009, 『テキストデータの統計科学入門』 岩波書店.
- [4] 中野康人, 2009, 「社会調査データとしての新聞記事の可能性－読者投稿欄の計量テキスト分析試論－」, 『関西学院大学先端社会研究所紀要』 1:71-84.
- [5] 竹下俊郎, 1994, 「内容分析のツールとしての新聞記事データベース」, 『新聞研究』 516:60-63.
- [6] 田中伯知, 1993, 「日本の新聞論調にみる太平洋戦争史観: 社説, 連載, 談話・転載記事の内容分析」, 『慶應義塾大学新聞研究所年報』 40:27-42.

Abstract

Quantitative Text Analysis of readers' columns: Asahi's "KOE" and Yomiuri's "KIRYU"

NAKANO Yasuto¹

The purpose of this paper is to report results of descriptive analysis on readers' columns of ASAHI and YOMIURI. Each column contains a contributor's name, address, occupation and age. Including these contributors informations, Nakano (2009) sets up a dataset of readers' columns from 1989 to 2007. From this dataset, characteristics of contributors are explored. Utilizing morphological techniques, contents of columns are quantitatively analyzed. Through the analysis differences between Asahi and Yomiuri are clarified.

Key words: newspaper, quantitative text analysis, content analysis

¹ Kwansei Gakuin University

