

日本語 Twitter ユーザーの 中国人についての言説の計量的分析 — コリアンについての言説との比較⁽¹⁾⁽²⁾ —

高 史 明⁽³⁾

本研究は、Twitter における中国人についての日本語のツイートを集め、高 (2014, 人文研究) が分析したコリアンについてのツイートと比較しつつ計量的な分析を行った。中国人についてのツイートの多くはネガティブなものであり、ポジティブなツイートはコリアンについての言説の場合よりも少なかった。中国人についてのツイートも、少数のユーザーが多くの投稿をしており、コリアンについてのツイートほどではないが、一般的なツイートに比べて“拡散”されやすかった。また計量テキスト分析により、特定の保守的なニュースサイトや活動家、“2ちゃんねる”や“2ちゃんねるまとめブログ”が強い影響力を持つことが示された。これらに加えて、レイシズム、歴史問題や外交問題、政治や経済への言及、マスメディアへの不信、“真実”が隠されているという信念、情報を“拡散”もしくは共有しようという意図、などに注目して計量テキスト分析を行なった。

キーワード：民族偏見、ソーシャル・メディア、ネット右翼、ツイッター、計量テキスト分析

序論

近年の日本では、中国人やコリアン⁽⁴⁾に対するレイシズム⁽⁵⁾が盛んに表出され、深刻な社会問題になっている (自由国民社, 2013; 守, 2014; 高, 2014; 安田, 2012)。例えば、出版業界では 2013 年頃から“嫌中憎韓”が流行になり、中国や韓国への侮蔑がテーマの書籍が多数出版され、書店には専用のコーナーが設けられることも多くなっている (守, 2014)。また、週刊誌のタイトルを両国への侮蔑的言辭が飾ることも多い (守, 2014)。こうした差別的言説はインターネット上ではより早い時期から広範に見られ、00 年代には幾つもの排外主義団体がインターネット上で糾合して結成され、中国人やコリアンを攻撃する街宣・デモを行ってきた (Fackler, 2010; 安田, 2012)。2013 年には、排外主義的言説が一層過激化したこと及びそれに対抗する“カウンター”の動きが活発化したことを受けて“ヘイトスピーチ”の語が一般人の口にも上るようになり、“ユーキャン新語・流行語大賞”のトップ 10 にもランクインした (自由国民社, 2013)。また、警察庁の発行する“治安の回顧と展望”の 2013 年版には“右派系市民グループをめぐる動向”の項が初めて設けられた (警察庁警備局, 2013)。このように中国人・コリアンに対するレイシズム言説は、それを支持する側にとっても批判する側にとっても、あるいはどちらでもない人々にとっても、大きな関心事になっている。

こうした状況を理解するために、高 (2014) および岸・高 (印刷中) は、コリアンについてのツイート (Twitter での投稿) を収集し、計量的な分析を行った。Twitter を特にとりあげて分析したのは、2000 年代のレイシズム言説の伸張にソーシャル・メディアが大きな役割を果たしたこと (荻上, 2011; 安田, 2012)、ソーシャル・メディアの中でも Twitter は使用者が多いこと (許, 2013; 関根,

2013), また他の様々なサイト・サービスとの連携が盛んになされていること, データの大部分が公開データで研究者の手により収集しやすいこと, などが理由である(高, 2014)。これらの研究により, コリアンについてのツイートの大部分はネガティブなものであること, それらの少なくない部分を極端にツイート数とフォロワー数の多い少数のユーザーが行っていることの他に, 計量テキスト分析を用いてコリアンについての言説の様々な特徴が明らかにされた。その特徴として(1) コリアンについてのツイートは“拡散⁽⁶⁾”を呼びかけるものや“話題の共有”を意図したと思われるもの⁽⁷⁾が多いこと, (2) マスメディアへの不信を表出したものが多いこと, (3) 匿名掲示板の“2ちゃんねる”およびその内容を編集してブログ形式で転載する“2ちゃんねるまとめブログ”に情報源を依存する傾向があること, (4) コリアンにとって不都合な“真実”が隠されていると述べるものが多いこと, (5) 何らかの“反日的な”あるいは“売国的な”対象に言及するものが多いことなどが挙げられる(高, 2014)。また, レイシズムに関連するツイートが多いことも示されている。

なお, この研究で分析の枠組みとして用いられたレイシズム概念は, コリアンは劣っているという信念に基づく露骨な偏見である“古典的レイシズム”と, (1) 日本国内においてコリアンに対する差別は既に存在せず, (2) 現存する格差はコリアンの努力不足によるものであるが, (3) それにもかかわらずコリアンは差別に抗議し過剰な要求をし, (4) 不当な特権を得ている, という信念に基づく隠微な偏見である“現代的レイシズム”である。これら2種類のレイシズムは, もととはMcConahay (1986) やSears (1988) らがアメリカでの黒人に対するレイシズムを理解するための概念として提唱したものを, 高らが在日コリアンに対するレイシズムの文脈に適用したものである(高, 2013; 高・雨宮, 2013)。これらのレイシズムに関連するツイートは, コリアンに関するツイートのそれぞれ10.75%, 12.20%を占め(高, 2014), また単にレイシズムに関連するというよりは, コリアンにネガティブな, つまりレイシズムを支持するような投稿が圧倒的多数であった(岸・高, 印刷中)。

このようにコリアンについてのTwitter上の言説の特徴が明らかにされてきた一方で, レイシズム言説のもう一方の対象である中国人についての言説の特徴は, 未だ明らかにされていない。そこで本研究は, Twitter上で中国人に言及する日本語ツイートを収集し, その特徴を分析することを目的とする。比較対象としては, 高(2014)の収集した, コリアン関連ツイートおよびワイルドカード検索により収集されたツイートをを用いる。また, この比較を行うことにより, 高(2014)で示されたコリアン関連ツイートの特徴について, さらに深く理解することも目指す。すなわち, 高(2014)はコリアン関連ツイートとワイルドカード検索の結果の比較しか行っていないため, コリアン関連ツイートの特徴とされたものが本当にコリアンという集団に特異的なものなのか, それとも人間の集団について言及するときには一般的に見られるものであるのかが明らかではない。本研究では, 中国人についての言説との比較を行うことにより, この点についてもさらなる知見を得ることを試みる。

なお, Twitter上で投稿される日本語のツイートは一日におよそ5000万件以上にもなり(NECビッグロブ, 2012), 中国人に言及するツイートも膨大な数にのぼる。そのため, 取得可能なデータの出来る限り多くの部分を用いて分析を行うためには, 質的な分析を行う前にまず情報を量的なものに変換して処理する必要がある。そこで本研究では, ツイート内容の分析の手法としては計量テキスト分析を用いる。この手法は, テキストデータに対してある種の数値化操作を行うことで計量的に把握することを可能にするものである(樋口, 2014)。

方法

2012年9月18日から2012年11月5日⁽⁸⁾にかけて, “中国人”の語を含むツイートを収集した。つまり, 現国籍国や居住国などを限定することなく, “中国人”の語を用いて言及される対象についての

ツイートを収集した。データの収集には、2013年3月まで使用可能であったTwitterのRSS機能を用いた。RSSとは、Webサイトの更新情報を取得するための機能であり、Twitterの場合には、リクエストを送信すると条件に該当するツイートをその時点から遡って15件まで取得できた。データの取得にはThunderbird (Mozilla, Inc.)を用い、15分ごとに自動でリクエストを送信し、ツイートを収集した。つまり、この方法では、前回のリクエストを送信してから今回のリクエストを送信するまでに“中国人”の語を含むツイートが16件以上なされている場合には最新の15件を、15件以下の場合には全てのツイートを、取得することができた。したがってこの方法で取得されたツイートは、全数でもないし無作為抽出でもない。総ツイート数の多い日中から夜間にかけて投稿された特定のツイートが抽出される確率は相対的に低く、逆に深夜から早朝にかけて投稿された特定のツイートが抽出される確率は相対的に高くなる。しかしながら、このような誤差が内容の偏りをもたらすとは考えにくいいため、この方法を用いた。ツイートの一部を抽出して用いたのは、ツイートの全てを用いると、ハードウェア的およびソフトウェア的に分析の困難なデータ・サイズに短期間で達してしまうからである。Twitterでは、その性質上、単一のツイートが爆発的にリツイートされたり、“炎上”と呼ばれる単一のトピック（誰かのツイートやブログ、ニュース等）についてコメントが集中する現象が起こることがあるため、短期間に取得したデータでは、一時的な流行のみを反映してしまう可能性が高い。そこで本研究では、およそ一ヶ月半に渡り一定間隔で少数ずつのツイートを収集し、分析することとした。なお、機材のメンテナンス等のため、上述の期間中にもツイートの取得を中断していた時間帯がある。

なお、RSSでデータを取得した場合には、リツイートされたものであることを示す“RT @ (リツイート元の発言者のID)”を冒頭に加えたときの文字数が140文字を超過した場合、文末の超過した部分が除外されてしまうという制約があった。また、RSSでツイートを収集した場合には、発言者のデモグラフィック変数や心理変数を取得できないという制限もあった。これは、これらの変数を取得するためにはユーザーの許諾を伴う調査を行う必要があるが、そのような手法を取った場合にはできるだけ多くのユーザーの多くのツイートを収集し言説の総体を分析するという目的の達成が不可能になるからである⁹⁾。このような制約はあるのだが、RSSでデータを収集するメリットに鑑み、このような方法を用いた。

比較対象としては、高(2014)で分析されたツイートをを用いた。“コリアン関連ツイート”としては、2012年11月5日から2013年2月16日にかけて“在日 or 韓国人 or 朝鮮人 or チョン¹⁰⁾”のTwitterでの検索結果をRSSで取得したものをを用いた。この場合、いずれかの単語を含んでいる最新15件までのツイートが取得された。“ワイルドカードツイート”としては、コリアン関連ツイートと同様の期間、日本語ツイートのワイルドカード検索結果を取得した。抽出に用いられた検索クエリは、“lang=ja&q=*”である。この場合、ユーザーの言語設定が日本語に設定されている全てのツイートから15件ずつが収集された。

結果

中国人関連ツイートとしては、計67,884件のツイートが取得された。このうち、リツイート以外のツイートは42,745件であった。全ツイート中のリツイートが占める比率は37.0%であった。これは、コリアンについてのツイートにおける比率44.7%より有意に小さく ($\chi^2(1) = 1000.9, p < .001$)、ワイルドカードにおける比率10.4%よりは有意に大きかった ($\chi^2(1) = 21131.9, p < .001$)。したがって、中国人についてのツイートは、一般のツイートに比べると“拡散”されやすいが、コリアンについてのツイートほど“拡散”されやすいわけではない。

発言のシェア

捕捉されたアカウント数は 42,013ID であった⁽¹¹⁾。このうち 81.9% のアカウントは捕捉されたツイート数は 1 件のみであったが、50 件以上のツイートが捕捉されたアカウントも 25 件あった。これらのアカウントの過半数は、中国人に対して明確にネガティブなツイートを繰り返し投稿していた。コリアンについてのツイートを分析した高 (2014) では、ごく一部の差別的なアカウントが大量の投稿を行っていたことが示されている。そこで中国人についてのツイートについても、大量の投稿を行っていた一部のアカウントが投稿全体に占めるシェアを検討する。

本研究において取得されたデータで、捕捉されたツイート数が上位 25 位までのアカウント (捕捉されたアカウントの 0.06%) が全ツイートに占めるシェアは、4.33% であった。この数値は、コリアン関連ツイートでは 10.12%、ワイルドカードツイートでは 0.16% であった (高, 2014)。

また、上位 50 位まで (捕捉されたアカウントの 0.12%) で比較すると、中国人関連ツイートでは 5.81% であった。この数値は、コリアン関連ツイートでは 12.87%、ワイルドカードツイートでは 0.27% であった (高, 2014)。

これらの結果から、少数のユーザーが大量の投稿をする傾向は、コリアンに対するものと同様、中国人関連ツイートでも見出されたといえる。ただしその程度は、コリアンについてのツイートに比べると弱かった⁽¹²⁾。

頻出語の分析

次に、計量テキスト分析を行った。分析には KHCoder 2.beta.30f (樋口, 2001/2013) および形態素解析エンジンとして Mecab 0.996 (工藤, 2006/2013) を用いた。まず KHCoder により分析対象ファイルを自動修正し、次に形態素解析を行った。形態素解析にあたり、最初に高 (2014) で頻繁に登場した複合語を強制抽出語に登録した上で解析を行い、その後本データで頻繁に登場した複合語を新たに強制抽出語に加え、再分析を行った。

この結果を元に作成した頻出語のリストを、表 1 に示す。表 1 は、名詞 (“国” “人” “日” など 1 文字で、様々な複合語に用いられるために解釈の困難な名詞を除く) および解釈の可能な記号・文字列 (“RT” および “@” を除く⁽¹³⁾) のみの上位 60 件とそれを含むツイート数である。

ツイートの抽出に用いられた語の “中国人” は当然ながらほぼ全てのツイートに用いられている (“中国人” の語を含まないとされたツイートが一定数あるのは、“中国” と “人民” の語に分解される “中国人民” のような語を含むツイートも抽出されているからである)。また、“中国” (19,042 件) “中華” (976 件) の語も多く見られた。中国以外の国・地域では、“日本” (16,674 件)、“日本人” (13,507 件)、“反日” (4,401 件)、“日本語” (1,424 件) と、日本に関する語も多いが、これは高 (2014) の分析したコリアン関連ツイートと同様の結果である。“韓国人” (5,182 件)、“韓国” (3,360 件)、“朝鮮人” (1,892 件) と、韓国・朝鮮への言及も多い。コリアンと中国人が併記されたツイートが多いという点も、高 (2014) と同様の点である。両者はともに東北アジアに位置する日本の隣国であり、また日本人にとってネガティブな集団であるが、このような類似性が、両者を併記するようなツイートの産生に関わっているのであろう。その他に、“アメリカ人” (2,178 件)、“フランス” (2,111 件)、“ドイツ” (1,778 件)、“イタリア” (1,228 件) など、欧米諸国も頻繁に言及されている。これは、中国人 (およびコリアン) を “オチ” にするエスニックジョークの類 (例えば、“製品とは、ドイツ人が発明し、アメリカ人が製品化し、イギリス人が投資し、フランス人がデザインし……中国人が海賊版を作る” といったもの) が盛んにツイート・リツイートされていたことによるものようである。また、中国が民族問題・領土問題を抱える “チベット” (403 件)、“ウイグル” (197 件) や、“台湾” (843 件) も、ランク外だが比較的多く見られた。

表1 頻出語とそれを含む投稿数

順位	単語	投稿数	順位	単語	投稿数	順位	単語	投稿数
1	中国人	67,301	21	ドイツ	1,778	41	今日	1,175
2	http://	28,905	22	報道	1,768	42	学生	1,162
3	中国	19,042	23	事件	1,742	43	CHINA	1,154
4	日本	16,674	24	経済	1,733	44	大学	1,146
5	日本人	13,507	25	観光客	1,688	45	MSN産経ニュース	1,110
6	#	7,212	26	企業	1,672	46	救助	1,075
7	韓国人	5,182	27	政治	1,609	47	国民	1,069
8	反日	4,401	28	世界	1,508	48	逮捕	1,040
9	デモ	4,249	29	日本語	1,424	49	情報	1,001
10	尖閣	3,746	30	関係	1,418	50	(TwitterID A)	994
11	韓国	3,360	31	沖縄	1,415	51	整形	988
12	問題	3,149	32	NEWS	1,402	52	中華	976
13	留学生	2,696	33	外国人	1,385	53	批判	976
14	ニュース	2,465	34	歴史	1,343	54	人間	973
15	アメリカ人	2,178	35	尖閣諸島	1,292	55	2ch	966
16	フランス	2,111	36	女性	1,279	56	民族	962
17	政府	1,901	37	ネット	1,231	57	領土	959
18	朝鮮人	1,892	38	イタリア	1,228	58	ノーベル	954
19	自分	1,803	39	速報	1,225	59	社会	951
20	在日	1,792	40	メディア	1,197	60	戦争	944

以下に高（2014）のコリアン関連ツイートで頻繁に用いられていた語のカテゴリに注目しつつ、本研究で多く見られた語を挙げる。

(1) レイシズム関連語 先に述べたレイシズムに関する語としては、“逮捕”（1,040件）の語が、中国人の“道徳的劣等性”を強調するツイートに頻繁に用いられていたようである。高（2014）では古典的レイシズムに関連するとコーディングされた“事件”（1,742件）の語は、“通州事件”“柳条湖事件”“南京事件”などの歴史的な出来事を指し古典的レイシズムとは関係が弱いものも多かった。“女性”（1,279件）の語も、コリアンに関連するツイートでは“コリアン男性が日本人女性に性的暴行を加えている”という主旨のツイートが少なくなかったのだが（高，2014），中国人に関する場合には，中国人女性の言動に言及したものが多かった。

現代的レイシズムに関連する語は60位以内には見られなかったが，ランク外にはコリアンに関するツイートで頻繁に見られた“生活保護”（758件）が比較的多く見られた。また，同じくランク外だが，“奨学金”（737件）の語も，日本人はもらえない奨学金を中国人留学生はもらえるという主旨のツイートが多く，現代的レイシズムに関係すると考えてさしつかえはなさそうである。

(2) 政治・経済関連語 “政府”（1,901件），“政治”（1,609件）などの政治に関する語も多く見られた。これは，尖閣諸島の領土問題やいわゆる“反日デモ”を巡って日中の政府を批判するものが多かった。また，高（2014）には見られなかった特徴として，“経済”（1,733件），“企業”（1,672件）などの経済活動に関する語も多く見られた。これらを用いたツイートには，いわゆる“反日デモ”により日本企業が被害を受けることや，経済面に悪影響が出る懸念に言及したものが多かった。

(3) 国際問題関連語 外交問題に関連する語としては，“尖閣”（3,746件），“尖閣諸島”（1,292件），“領土”（959件）が多く見られた。また“沖縄”（1,415件）の語も，“沖縄の尖閣諸島”への言及や“中国が沖縄の領有権を主張すること，もしくは主張するに違いないこと”への言及で多く用いられており，外交問題に関連する語と考えてよさそうである。尖閣諸島の中国語の別名である“釣魚島”もしくは“釣魚台”の一部である“釣魚”（457件）や，“魚釣島”（99件）もランク外に見られた。

歴史問題に関連する語としては、“歴史” (1,343 件) が多く見られた。“戦争”の語は、日中間に今後戦争が起こるであろうということに言及したツイートで多く用いられており、歴史問題に関係しているとは言いがたかった。

(4) メディア関連語 マスコミに関連する語としては、“ニュース” (2,465 件)、“メディア” (1,197 件) が多く見られた。

この他に、“2ch” (966 件)、“速報” (1,225 件) の語も多く見られている。前者は、匿名掲示板の2ちゃんねるもしくはその投稿を編集して掲載した“2ちゃんねるまとめブログ”と呼ばれるサイト群に関連するものに多く用いられており、多くは、ハッシュタグ“#2ch”の形で用いられていた。後者は、2ちゃんねるまとめブログの名前で頻繁に用いられている語であった。

(5) 拡散・話題の共有関連語 ハッシュタグに用いられる“#” (7,212 件) の記号も多く見られた。“拡散”を呼びかけるために“【拡散】”“【拡散希望】”などの形で用いられる“拡散” (904 件) の語も、リンク外だが比較的多く見られた。

(6) 中国人関連ツイートに特異的な語 高 (2014) では見られなかった特徴としては、日本にいる中国人への言及も多く、“留学生” (2,696 件)、“観光客” (1,688 件) などが頻繁に用いられていた。また、データ収集時盛んに行われていたいわゆる“反日デモ”に関連して、“デモ” (4,249 件) の語や、リンク外だが“暴動” (820 件) の語なども多く見られた。

また、“MSN 産経ニュース” (1,110 件) も多く見られている。これは、産経新聞のウェブ媒体である MSN 産経ニュースのサイト上から記事へのリンクをツイートしたと思われるものが大部分であった。同メディアは中国に対して否定的な報道を頻繁に行なってきた。さらに、極右の元自衛隊高官の TwitterID である“(TwitterID A)”⁽¹⁴⁾ (994 件) の文字列も、多く見られている。これは、同氏のツイートの幾つかが爆発的にリツイートされていたことに依っていた。同氏もやはり中国に対して否定的な主張を繰り返して行なってきたのだが、これら少数のメディア及び個人の強い影響力が示されたことになる。

言説のテーマ

次に、より詳細に言説の内容を分析するために、幾つかのテーマに関連するツイートで頻繁に用いられる語を含むツイートにコードを割り当てて分析を行う。コードの生成にあたってはまず高 (2014) が用いたコーディング・ルールを下敷きに、頻出語の分析結果を加味して修正を行なった。本研究で最終的に用いられたコードと分類に用いられた単語は、表 2 の通りである。

“中国”“韓国朝鮮”“日本”のそれぞれのコードは、それぞれの国家や国民、文化に言及すると考えられるツイートに用いられた。

“歴史問題”コードは日本の中国もしくは韓国・朝鮮への加害の歴史、“外交問題”コードは日本が中国および韓国・朝鮮との間に抱える問題に言及すると考えられるツイートにそれぞれ用いられた。

“古典的レイシズム”“現代的レイシズム”“政治”“経済”“マスコミ”“2ちゃんねる”の各コードは、それぞれ関連する用語を用いて分類した。ただし“マスコミ”は、特定のメディアに関する語ではなく、メディア一般を指す語を用いて分類した。

特殊なコードとしては、“反日系”コード (何らかの対象を“反日的”あるいは“売国的”と糾弾するツイート)、“真実系”コード (何らかの隠された真実がありそれを明らかにするとするツイート)、“侮辱語”コード (何かを激しく侮辱するツイート)、“拡散”コード (“拡散”を呼びかけるツイート)、“話題の共有”コード (ハッシュタグを用いているツイート) がある。これらは、高 (2014) で分析に用いる必要性が示されたものである。

また、日本または日本でも中国でもない国・地域にいる中国人に言及するとき頻繁に用いられる語

を含むツイートには、新たに“外国人の移動”コードを割り当てた。このコードがコリアンについて用いられる場合には、日本または日本でも韓国・朝鮮でもない国・地域にいるコリアンに言及したものである可能性が高いことを指す。

なお、これらのコードがワイルドカードツイートにおいて用いられる場合には、コードの名称はツイートの内容の説明として必ずしも適切ではないものもあった（例えば、ワイルドカードツイートにおいて“店員”の語が用いられている場合には、“外国人の移動”のコードが割り当てられるが、これらのツイートには外国人への言及を含まないものが多かった）。したがってワイルドカードツイートにおいては、これらのコード名は言説のテーマを表したものというよりは、コーディングに用いられた単語の出現率のベースラインを表したものと考えられたい。

これらのコードを用いて、コードが出現したツイート数およびそれが収集されたツイートに占める比率を、表3に示す。

中国人関連ツイートにおける“中国”コード、コリアン関連ツイートにおける“韓国朝鮮”コードが100%にならないのは、抽出語が複数の語に分割されていたり、複合語の一部とされていたりする場合があるためである。

その他のコードを見ると、中国人関連ツイートにおける出現率は1.33%から39.54%であり、いずれもワイルドカードツイートにおける出現率を大幅に上回っていた⁽¹⁵⁾。したがって、これらのコードが中国人についての言説の特徴を捉えていることが示された。

一方で、コリアン関連ツイートと比較してみると、コードの出現率にはある程度の差があった。中国人関連ツイートで多かったものには、“外交問題”“経済”“デモ及び暴動”の各コードがあった。これには、日本政府による尖閣諸島国有化の方針を受けて中国国内でいわゆる“反日デモ”が行われており、日本経済や日本企業への悪影響が懸念されていたという、時期的な要因も影響していたのかもしれない。ただし、コリアン関連ツイートが収集された時期も竹島（独島）問題などが紛糾していたことを考えると、コリアン関連ツイートとの差異は、比較的恒常的な言説の差異を反映したものである可能性もある。つまり、中国人関連ツイートの方が“外交問題”が多く“歴史問題”が少ないという違いは、両国との間に日本が抱える問題のうちどちらにより焦点が当てられやすいのかを反映しているのかもしれない。他に中国人関連ツイートで多かったものとしては、“外国人の移動”コードがある。中国人の方が、日本にいる外国人としては意識されやすいのかもしれない。これは、中国人の方がニューカマーや一時滞在者が多く、言語などで外国人だと容易に判断できることが多いことを反映しているのではないだろうか。なお、中国人関連ツイートとコリアン関連ツイートでほとんど差がなかったものとしては、“日本”“反日系”“2ちゃんねる”などのコードがある。

一方、中国人関連ツイートでコリアン関連ツイートでよりも少なかったコードは、“古典的レイシズム”“現代的レイシズム”“マスコミ”“真実系”“侮辱語”“拡散呼びかけ”“話題の共有”などである。2つのレイシズム概念は、コリアンに対する言説を理解する場合のほうが、より有効なようである。また、コリアンについてのツイートでは現代的レイシズムの方が古典的レイシズムよりも顕著だった（高，2014）のだが、中国人では逆であり、古典的レイシズムの方が現代的レイシズムよりも頻繁に見られた。つまり、コリアンに対する偏見を特徴づけるのは“もはや差別は存在せず不当に優遇されている”という認識であるのに対し、中国人に対しては“民族的に劣っている”という認識の方が特徴的であると言える。現代的レイシズムはその内容上、日本国内にいる民族的マイノリティに向けられるものである。日本における民族的マイノリティとしては、既に中国人の数がコリアンの数を上回っている（法務省入国管理局，2013）のだが、“特権言説”は、人口が多い中国人に比べて、人口は相対的に少ないが長く複雑な歴史的経緯のあるコリアンについてなされやすいようである。また、“マスコミが真実を隠している”というようなツイートや、他のユーザーに何らかの情報を知らしめようというツイート

表2 ツイート分類コード⁽¹⁶⁾

コード	分類に用いられた単語
中国	中国, 中国人, 中国語, <u>中華</u> , <u>チャイナ</u> , シナ, 支那, シナ人, 支那人
韓国朝鮮	韓国人, 朝鮮人, 在日, チョン, 韓国, 朝鮮, 韓, 朝鮮語, ハングル
日本	<u>日本</u> , <u>日本人</u> , <u>我が国</u>
古典的レイシズム	犯罪, <u>事件</u> , 逮捕, 悪事, 凶悪, 強姦, レイプ, 犯す, 襲う, 強盗, 整形, 劣る, 劣等, ヤクザ, <u>うるさい</u> , 火病, (<u>事件を含み通州, 柳条湖, 南京を含まないもの</u>)
現代的レイシズム	奨学金, 特権, 生活保護, 受給, 人権, 通り名, 通名, 参政権, 年金
歴史問題	歴史, 大日本帝国, 植民地, 統治, 併合, 支配, 侵略, 加害, 南京, 満州, 柳条湖, 強制連行, 連行, 慰安婦, 従軍, 太平洋戦争, <u>戦争</u> , 戦前, 戦後, 敗戦, 関東大震災, 謝罪, 賠償, 補償, <u>靖国</u> , <u>靖国神社</u>
外交問題	尖閣, 尖閣諸島, 釣魚, 魚釣島, 沖縄, 竹島, 独島, 領土, 東海, 日本海, ミサイル, 拉致, 条約
マスコミ	マスコミ, マスゴミ, <u>メディア</u> , ニュース, テレビ, テレビ局, 新聞, 放送, 報道, 記事, 記者
政治	政治, 政党, 政府, 政権, 議員, 首相, 総理, 書記, 大統領, 民主党, 民主, 自民党, 自民, 維新, 安倍, <u>河野</u> , <u>石原</u> , <u>野田</u> , <u>野中</u> , <u>田中</u> , 選挙, 法律, 法案, 政策, 憲法
経済	<u>経済</u> , <u>経営</u> , <u>会社</u> , <u>企業</u>
反日系	反日, 売国, 売国奴
真実系	実態, 真実, 本性, 正体, 発覚, 疑惑, 暴露, 嘘, ウソ, 嘘八百, 捏造, 自演, 歪曲, 隠蔽, 矛盾, 陰謀
デモ及び暴動	デモ, <u>暴動</u> , 暴徒, <u>愛国</u> , <u>破壊</u> , 製品
外国人の移動	<u>留学生</u> , <u>留学</u> , <u>観光</u> , <u>観光客</u> , <u>移民</u> , <u>店員</u> , <u>バイト</u>
侮辱語	馬鹿, バカ, アホ, ボケ, カス, クズ, ゴミ, ゴキブリ, 寄生虫, ウジ, ウジ虫, ハエ, 害虫, ブタ, 鬼畜, 糞, ガン, 嘘つき, 童貞, 底辺, 乞食, 変態, デブ, 気違い, 基地外, マジキチ
2ちゃんねる	2ch, 速報, ニュー速, v速, 嫌儲, 鬼女, および14個の代表的な2ちゃんねるまとめブログ名, 2ちゃんねるまとめブログの管理人のTwitterID5つ
拡散呼びかけ	拡散, RT希望, RTを, RT望, RT願
話題の共有	#, #

下線は高(2014)では用いられておらず本研究で用いられたコードもしくは分類語, 斜体は高(2014)では用いられたが本研究では用いられていない分類語である。

表3 コード別出現ツイート数と出現率

	中国人関連ツイート		コリアン関連ツイート		ワイルドカードツイート	
	ツイート数	出現率(%)	ツイート数	出現率(%)	ツイート数	出現率(%)
中国	67,540	99.49	9,957	9.09	180	0.16
韓国朝鮮	10,940	16.12	108,794	99.27	225	0.20
日本	26,841	39.54	40,095	36.59	760	0.66
古典的レイシズム	4,034	5.94	10,408	9.50	371	0.32
現代的レイシズム	2,159	3.18	13,365	12.20	60	0.05
歴史問題	4,819	7.10	12,715	11.60	258	0.22
外交問題	7,349	10.83	5,703	5.20	181	0.16
マスコミ	4,553	6.71	14,093	12.86	1,101	0.96
政治	6,899	10.16	23,178	21.15	893	0.78
経済	4,212	6.20	4,878	4.45	396	0.34
反日系	5,337	7.86	8,280	7.56	25	0.02
真実系	2,541	3.74	8,964	8.18	361	0.31
デモ及び暴動	6,404	9.43	5,565	5.08	115	0.10
外国人の移動	6,497	9.57	2,373	2.17	748	0.65
侮辱語	2,202	3.24	5,523	5.04	1,197	1.04
2ちゃんねる	3,101	4.57	5,721	5.22	317	0.28
拡散呼びかけ	904	1.33	6,348	5.79	461	0.40
話題の共有	7,343	10.82	18,991	17.33	7257	6.31
総数	67,884		109,589		114,932	

コリアン関連ツイート, ワイルドカードツイートのデータは, 高(2014)のものを用いている。ただし, 中国人についての言説を分析するのに適するように一部のコードのコーディング・ルールを修正している(表2参照)ため, 本表中の一部の数値は, 高(2014)と一致しない。

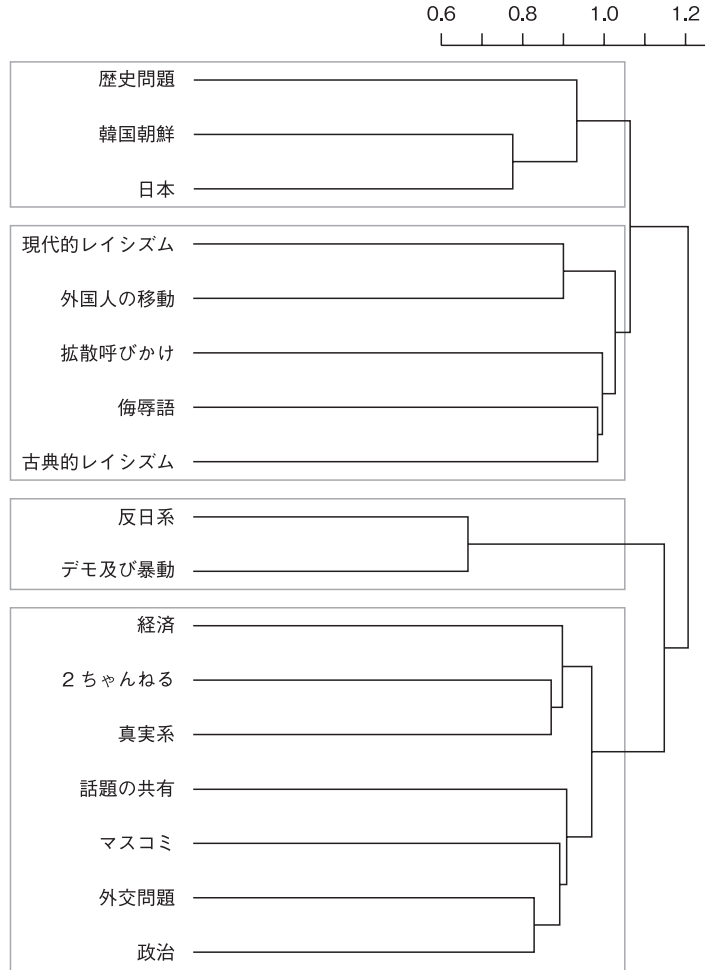


図1 コードを用いた階層的クラスター分析

も、中国人に対してはコリアンに対する場合よりも相対的に少ないことが示唆された。

このように、中国人についての言説とコリアンについての言説には、共通点と相違点がそれぞれ見出された。両者の特徴をより明確に理解するためには、今後、ポジティブな外集団についての言説や、内集団である日本人についての言説と比較する必要があるだろう。

テーマ間の関連性

次に、これらのコードを用いた言説の特徴をより詳細に検討するために、コードを用いた階層的クラスター分析を行なった結果を、図1に示す。なお、出現頻度が極端に高い“中国”のコードは除外してある。

“歴史問題”は、やはり“韓国朝鮮”と同一のクラスターを形成している。中国人とコリアンの両方に言及するときには歴史問題が焦点となりやすいというのは、高(2014)と一貫する結果である。

“古典的レイシズム”は“侮辱語”と結びつきが強く、感情を強く刺激するような言葉を用いて中国人の劣等性が語られていたことが示唆されている。またこれらは“拡散呼びかけ”ともクラスターを形

成しており、積極的に広めようという意図を伴って投稿されていたことが示唆されている。“現代的レイシズム”は“外国人の移動”と最も結びつきが強かった。現代的レイシズムは日本国内における人種・民族的マイノリティに対し向けられるものであるため、日本にいる中国人に明示的に言及するような語と一緒に用いられていたのであろう。これらのコード全体で一つのクラスターを形成していたのだが、“古典的レイシズム”と“現代的レイシズム”がある程度共起するという点は、高（2014）と同一である。なお、コリアンについての言説でなぜ2つのレイシズムが共起するのかは、岸・高（印刷中）が詳細に分析している。

また、“2ちゃんねる”と“真実系”との間にも比較的強い結びつきが見出された。これは、2ちゃんねるもしくは2ちゃんねるまとめブログを情報源として中国人にとって不都合な真実を明らかにするというツイートが多く投稿されていたことを示唆している。“2ちゃんねる”と“真実系”との結びつきはコリアンについての言説でも見出されている（高，2014）のだが、非伝統的なメディアの本来疑わしい情報に依存する傾向は、中国人についての言説でも共通していた。

“政治”コードが最も結びつきが強かったのは、“外交問題”のコードであった。高（2014）のコリアン関連ツイートでは、“政治”コードが共起していたのは“反日系”“現代的レイシズム”“話題の共有”などであったのだが、この点で異なっている。コリアンについてのツイートにおいては政治家や政府が日本の利益に反しコリアン（特に在日コリアン）の利益に沿うように動いているという言説が多く見られた（高，2014）のだが、中国人を巡って政治家や政府が話題になるときは、外交上の、領土問題などを巡ってであることが示唆されている。また“政治”“外交問題”のコードは“マスコミ”“話題の共有”などとも結びつきが強く、これらの話題を巡って議論したり、他人に情報を伝えたりしようとする意図を伴って投稿がなされていたことが示唆される。

発言の感情価

最後に、ツイートが中国人に対してどのような感情を表出するものであったかを分析した。収集されたツイートから無作為に150件を抽出し、ツイートが中国人に対して（1）ネガティブである、（2）ニュートラルであるもしくは判断できない、（3）ポジティブである、のどれに該当するかを、訓練された2人の評定者が独立に評定した。その結果、カッパ係数は.55であり“だいたい一致する”結果であったので、二人の評定値を平均してそのツイートの感情価とした。

その結果、最も多かったのはネガティブ（感情価<2）なツイートで、67.3%（95%CI=59.8%~74.8%）であった。次に多かったのはニュートラル（感情価=2）なツイートで、26%（95%CI=19.0%~33.0%）であった。最も少なかったのはポジティブ（感情価>2）なツイートで、6.7%（95%CI=2.7%~10.7%）であった。高（2014）が示したコリアンについてのツイートにおける比率（ネガティブ70.0%、ニュートラル12.7%、ポジティブ17.3%）と比較すると、2×3のカイ二乗検定の結果は有意であった（ $\chi^2(2)=14.1, p<.001$ ）。残差分析の結果、中国人についての言説の方がニュートラルなツイートが多く、ポジティブなツイートは少なかった。ネガティブなツイートの比率には差がなかった。したがって、中国人についてのツイートもコリアンについてのツイートも、対象の集団に対してネガティブなものが大多数を占めるのだが、コリアンについてのツイートではポジティブな立場からのものも一定量見られるのに対して、中国人に対するポジティブなツイートは非常に少なかったと言える。これは、コリアンについてのツイートは在日コリアン自身や関係者によるものも少なくないことや、在日コリアンへの差別に対する問題意識の方がより広く受け入れられており、反差別的な言説がより盛んになされていることなどを反映しているのかもしれない。

まとめ

本研究では、Twitter に日本語で投稿された中国人についての言説を計量的に分析し、高（2014）の分析したコリアンについての言説と比較した。

ごく一部のアカウントが極端に多くの投稿を行なっているという傾向は、本研究でも見出された。また、ごく一部の保守メディアや右翼活動家の言説、2ちゃんねるや2ちゃんねるまとめブログに由来する情報が盛んに広められていることも示された。

ツイートの大部分は中国人に対してネガティブなものであり、ポジティブなものはコリアンに対する言説における比率に比べて少なかった。また、中国人についてのツイートも、コリアンについてのツイートほどではないが、“拡散”されやすかった。ただし、“拡散”しようという意図の積極的な表出は、コリアンについてのツイートに比べると格段に少なかった。

また、本研究で用いられたコードが、コリアンについての言説とは多少相違点はありつつも、中国人についての言説の特徴も捉えていることが示された。

なお、本研究で扱われた中国人も、高（2014）の扱ったコリアンも、いずれも現在、日本人にとってのネガティブな外集団である。今後、よりポジティブな外集団や、内集団である日本人についての言説と比較する研究も、行われる必要があるだろう。

注

- (1) 本研究の分析にあたり、日本学術振興会より給付された平成 26 年度科学研究費助成事業助成金が用いられた。
- (2) 本研究は、著者が東京大学大学院に提出した学位申請論文“在日コリアンに対するレイシズムの研究：現代的レイシズム理論に着目して”の研究 2 を元に新たな分析を加え、加筆修正したものである。また、2013 年の日本心理学会第 77 回大会において、“Twitter における中国人についての言説の計量テキスト分析”として発表された。
- (3) 研究実施時の所属は東京大学大学院人文社会系研究科。著者への連絡用 E-mail アドレスは FumiakiTaka@gmail.com。
- (4) コリアンの呼称としては“韓国人”“朝鮮人”なども用いられるのだが、それらはしばしば特定の国・地域と結びつけて用いられる。そこで本研究では、総称としては“コリアン”の語を用いることとした。本研究との比較に用いた高（2014）のデータは、国籍国・地域、居住国・地域を限定することなく“韓国人”“朝鮮人”などの語を用いて言及される対象についての言説を収集したものである。
- (5) 本研究では、民族偏見・人種偏見の総称として、“レイシズム (racism)”の語を用いる。“民族 (ethnicity)”と“人種 (race)”を、生物学的基盤の有無という観点から異なる概念として定義する場合もあるが、その区別は、人種間の差異も民族間の差異と同様に生物学的基盤を欠いているという科学的知見から見ても、一般人の理解という素人理論の観点から見ても、さほど有益ではない (Zarate, 2008)。実際、The New Oxford American Dictionary などにおいても、race の語は ethnicity の意味で用いられる場合があると定義されている (“race,” 2005/2008)。
- (6) インターネット用語では、他の多くのユーザーの目にも触れるようにすること。Twitter においては、“リツイート”と呼ばれる、そのツイートが自分のフォロワー（講読者）の画面にも表示されるようにする特殊な投稿形式を用いたり、ツイート中に含まれる URL を転載したりすることによってなされることが多い。
- (7) 具体的には、“ハッシュタグ”と呼ばれる機能を用いるための文字列を含む投稿が多く見られた。この機能は、“#”の後に任意の文字列を加えて投稿すると、同じ“ハッシュタグ”を含むツイートが容易に検索できるようになるものである。
- (8) 時期としては、2012 年 9 月に日本政府が尖閣諸島を国有化する方針を固めたのを受けて、中国国内でいわゆる“反日デモ”が大規模に行われていた時期を含んでいる。そのため本研究で収集された言説の性質は、時

期的な要因に影響されている可能性もある。

- (9) 本研究では個々のユーザーの承諾を得ずにツイートを集めているのだが、Twitterのプライバシーポリシーは公開されているプロフィール情報やツイートをユーザーの承諾を受けず収集し利用することを承認しており (Twitter, Inc., 2012), 本研究はユーザーがこのプライバシーポリシーのもとで公開したプロフィール及びツイートのみを集めているため、倫理的問題は無い。
- (10) コリアンに対する蔑称であり、インターネット上でもしばしば用いられるものである。本研究では、蔑称やスラングなどの社会的に望ましくない語も分析に用いているのだが、論文の資料的価値に鑑み、極力編集せずにそのまま記載した。
- (11) ツイート取得期間中にアカウント名 (全角・半角の両方を使用できる、容易に変更可能な識別子) を変更した場合でも、同一のアカウントとして扱うことができている。ただし、期間中に Twitter ID (半角文字のみが使用可能な、容易に変更できない識別子) を変更したユーザーがいた場合には、別個のアカウントとして計上されている。なお、一人のユーザーが複数のアカウントを使用していた可能性もある。
- (12) 高 (2014) では発言数のシェアの分析に加え、フォロワー数の分析も行なっているのだが、本研究の分析を行ったのはデータ取得からおおよそ2年が経過した時点であり、したがってフォロワー数の構成もデータ取得時とは異なっていると考えられるため、そうした分析は行わなかった。
- (13) “RT”は“リツイート (retweet)”の略で、他のツイートをリツイートしたものであることを示すために、ツイート冒頭もしくはツイート中で用いられる文字列である。RSSで取得したデータでは、“公式リツイート”と呼ばれる形式のツイートの場合、自動でツイート冒頭に加えられた。“@”は、“@ (Twitter ID)”の形で用いられ、フォロー関係にかかわらず相手の画面にそのツイートを表示させる機能がある。これらの文字列および記号は、ツイートの内容を分析するためには有益ではないため、ここでは除外された。
- (14) この人物は著名人だが現在は公人ではないため、本文中及び表1では個人を特定できる Twitter ID は伏せた。以下、表2のブログ名、Twitter ID についても同様である。
- (15) なお、サンプルサイズが大きくカイ二乗検定の結果は容易に有意になるため、以下では統計的有意差ではなく絶対的な数値の大小関係についてのみ示す。
- (16) 本表では、以下のスラング及び蔑称が記載されている。“火病”はネット・スラングで、コリアンに特有の精神疾患とされているものである。“支那”“シナ”は中国の蔑称である。“基地外”は“間違い”を意味するネット・スラング、“マジキチ”は“マジで間違い”を省略したネット・スラングである。“ニュー速”“v速”“嫌儲”“鬼女”は、それぞれ2ちゃんねるの“板”と呼ばれる掲示板群の小分類の呼称である。

引用文献

- Fackler, M. (2010, August 29). A new wave of dissent in Japan is openly and loudly anti-foreign. *The New York Times*.
- 樋口耕一. (2014). 社会調査のための計量テキスト分析：内容分析の継承と発展を目指して. 東京：ナカニシヤ出版.
- 法務省入国管理局. (2013). 平成24年末現在における在留外国人数について (確定値). Retrieved December 30, 2013, from http://www.moj.go.jp/nyuukokukanri/kouhou/nyuukokukanri04_00034.html
- 自由国民社. (2013). 「現代用語の基礎知識」選ユーキャン新語・流行語大賞全受賞記録. Retrieved January 5, 2014, from <http://singo.jiyu.co.jp/>
- 警察庁警備局. (2013). 治安の回顧と展望 (平成25年版). Retrieved from http://www.npa.go.jp/keibi/biki/kaiko_to_tenbou/H25/honbun.pdf
- 岸政彦・高史明. (印刷中). 日本語 Twitter におけるコリアンに対するレイシズム言説：高 (2014) のさらなる分析. 国際社会文化研究所紀要.
- (Kishi, M., & Taka, F. (in press) Racist Statements against Koreans on Japanese Twitter: Additional Analysis of Taka (2014). *Journal of the Socio-Cultural Research Institute*)
- 許直人. (2013). 【スマホ版】mixi, Twitter, Facebook, LINE, カカオ 2012年12月ソーシャルメディア最新視聴率調査【VRI】 [in the loop]. *in the loop*. Retrieved December 27, 2013 from <http://media.loops.net/naoto/2013/02/12/vri-sm3-2012-12/>
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaert-

- ner (Eds.), *Prejudice, Discrimination, and Racism* (pp. 91-125). Orlando, FL: Academic Press.
- 守真弓. (2014). 売れるから「嫌中憎韓」書店に専用棚／週刊誌, 何度も扱う. 朝日新聞. Retrieved February 28, 2014, from http://digital.asahi.com/articles/DA3S10972937.html?requesturl=articles/DA3S10972937.html&ioref=comkiji_txt_end_s_kjid_DA3S10972937
- NEC ビッグロープ. (2012). ツイッター分析：BIGLOBE が 11 月の Twitter 利用動向を発表：企業の公式アカウントによる“心に直接呼びかける”ツイートが話題に. ついっぶるトレンド. Retrieved January 2, 2014 from <http://tr.twipple.jp/info/bunseki/201211.html>
- 荻上チキ. (2011). 検証東日本大震災の流言・デマ. 東京：光文社.
- race. (2005/2008). In *The New Oxford American Dictionary*. Oxford, UK: Oxford University Press.
- Sears, D. O. (1988). Symbolic racism. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 53-84). Plenum Press: New York & London.
- 関根健介. (2013). mixi, Twitter, Facebook, Google+, LinkedIn 2013 年 3 月最新ニールセン調査. *in the loop*. Retrieved December 27, 2013, from <http://media.loops.net/sekine/2013/05/08/neilsen-netview-201303/>
- 高史明. (2013). 在日コリアンに対する古典的／現代的レイシズム尺度の確認の因子分析と基礎的な検討. 人文研究, 180, 69-86.
- (Taka, F. (2013). Confirmatory Factor Analysis and Basic Investigations of Old-fashioned and Modern Racism Scales against Zainichi Koreans. *Studies in Humanities*, 180, 69-86.)
- 高史明. (2014). 日本語 Twitter ユーザーのコリアンについての言説の計量的分析. 人文研究, 183, 131-153.
- (Taka, F. (2014). A Quantitative Analysis of Japanese Twitter Users' Public Opinions of Koreans. *Studies in Humanities*, 183, 131-153)
- 高史明・雨宮有里. (2013). 在日コリアンに対する古典的／現代的レイシズムについての基礎的検討. 社会心理学研究, 28 (2), 67-76.
- (Taka, F., & Amemiya, Y. (2013). A Basic Investigation of Old-fashioned and Modern racisms against Zainichi Koreans. *Japanese Journal of Social Psychology*, 28, 67-76)
- Twitter, Inc. (2012). Twitter プライバシーポリシー (バージョン 6). Retrieved October 14, 2014, from https://twitter.com/privacy/previous/version_6
- 安田浩一. (2012). ネットと愛国：在特会の「闇」を追いかけて (kindle 版.). 東京：講談社.
- Zárate, M. A. (2008). Racism in the 21th Century. In T. Nelson (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 387-406). New York: Psychological Press.

A Quantitative Analysis of Japanese Twitter Users' Public Opinions of Chinese People: A Comparison with Their Opinions of Korean People

TAKA, Fumiaki

This study gathered Japanese tweets (statements posted on Twitter) regarding Chinese people and conducted quantitative analysis to compare them with tweets regarding Koreans, which was analyzed by Taka (2014, *Studies in Humanities*). The majority of the tweets regarding Chinese people were negative, and there were fewer positive tweets among those regarding Chinese people than those regarding Koreans. There was a tendency for a few users to post large numbers of tweets. In addition, tweets regarding Chinese were diffused more frequently than general tweets, and less than those regarding Koreans. Furthermore, quantitative text analysis revealed a strong influence of a conservative news site, a famous far right activist and an anonymous BBS site "2 channel" and its curator sites ("2 channel-matome-blogs"). In addition, quantitative text analyses focusing on racism, historical or diplomatic issues, references to politics or economics, distrust of mass media, beliefs in the actual "truth" being hidden, and intention to diffuse or withhold information were conducted.

Keywords : ethnic prejudice, social media, Internet right-wingers, Twitter, quantitative text analysis