

Department of Social Systems and Management

Discussion Paper Series

No.1312

Twitter のツイートからみた日本と東アジア諸国の

イメージ

(Measuring Images of Japan and East Asian Countries on Twitter)

by

石井 健一、海後 宗男
(Kenichi ISHII and Muneeo KAIGO)

October 2013

UNIVERSITY OF TSUKUBA

Tsukuba, Ibaraki 305-8573

JAPAN

Twitter のツイートからみた日本と東アジア諸国のイメージ¹

Measuring Images of Japan and East Asian Countries on Twitter

石井健一²・海後宗男³

要約 本報告は、Twitter 上の英語のツイートを用いて海外における東アジア諸国（日本、韓国、北朝鮮、台湾、中国）のイメージの測定結果を報告するものである。API を通じて収集した 218,815 件の Twitter の英語メッセージに対して内容分析とテキスト分析をおこなった。(1)収集したデータからランダムに抽出したツイートを評定者二人に読んでもらい内容を評価した。その結果は、台湾への好意度が最も高く、次いで日本が肯定的に評価されていた。中国や韓国については否定的な評価が大半であった。(2)ツイートのテキスト分析からも同様の結果がみられ、「好き」や「人が好き」という表現の頻度でみると、台湾が最も好意的であり、次いで日本、韓国、中国という順で中国の評価が最も低かった。

1. 研究の背景と目的

最近、「国家ブランディング」の政策への応用が各国で関心を集めている。国家ブランディングとは国を一つのブランドとみなしそのイメージを高めることで、製品の評価や訪問旅行者の増加、外交や国際世論などでの有利な効果を目指す政策である。しかし、日本の国家ブランディングを考えた場合、海外における日本のイメージについての資料は意外にも少ない。たとえば、外務省は毎年アメリカにおいて U.S. POLL ON OPINIONS TOWARD JAPAN 調査を行っていて、日本についての様々な側面についての態度を調査している (<http://www.mofa.go.jp/region/n-america/us/survey/>)。この調査は、きわめて重要なものであるが、他方では海外では日本について具体的にどのようなトピックが語られているのかを把握することが、対日態度の把握のためには有効であると思われる。

こうした観点から本研究では、Twitter の英語メッセージを API(Application Program Interface)を用いて収集し、日本、中国、韓国、台湾のイメージを測定することを目的とすることにした。なお、同じような問題意識に基づいて、日本語のツイートについては、既に内容分析とテキスト分析が行われている（石井 2012）。結果を表 1 に示す。中国に対する評価がきわめて否定的であることがわかる（この分析は日本語ツイートを対象なので日本は分析に含まれていない）。尖閣諸島問題など最近の日中間の政治的な状況が反中的なメッセージの増加として表れているのであると思われる。世界的にも「中国脅威論」がしばしば指

¹本研究は平成 24 年度科学研究費補助金基盤 (B) 『日中の相互国家イメージと「国家ブランディング」の可能性—中国と日本での実証研究』(代表石井健一) によるものである。

² 筑波大学システム情報系

³ 筑波大学人文社会系

摘される状況において、日本や中国が海外の人々からどのようなイメージをもって語られているのかを知ることは重要であろう。

本研究は、このような問題意識に基づき、Twitter の英語ツイートにおける東アジア諸国（日本、中国、韓国、北朝鮮、台湾）のイメージを内容分析とテキスト分析の二つの方法で測定することを目的とする。

表 1 日本語のツイートの内容分析の結果

	ポジティブなメッセージの比率 (好意度) %	国についての評価のあるメッセージの比率 %	(3) 国民性、その国の道徳、 対日感情の側面を内容とする ツイートの%
中国	3.6	12.2	21
台湾	98.1	24.2	23
香港	82.8	12.6	48
韓国	39.1	33.8	66
シンガポール	77.3	9.9	21
マレーシア	81.3	7.0	37
インド	47.8	12.0	13
アメリカ	21.9	34.4	23
イタリア	74.1	13.0	15
ドイツ	75.0	3.1	23
フランス	68.0	12.1	14
ロシア	26.3	10.4	54

出典 石井(2012)より作成

2. 方法

2.1 ツイートの収集

2012年11月28日から2013年2月17日にかけて以下のキーワードでTwitterのAPI経由でメッセージを検索した。用いたAPIのコマンドは、以下のとおりである。

http://search.twitter.com/search.json?lang=en&q=検索語&result_type=recent&rpp=100
 本分析で用いた検索語は、"Japan", "China", "Korea", "Taiwan", "Japs", "Chink"の6つである。なお、登録された言語が英語(en)のツイートのみをプログラムで選択した。ただし、Twitterのsearch.jsonは検索語を含むtwitter上の全てのメッセージを出力するわけではない。出力数に上限があり（一日あたり100個）、ヒット数がそれを越える場合は各日について時間が古いものから（つまり0:00後のメッセージで時刻が近いものから）順番に上限まで表示される。したがって各検索語について収集できるツイート数は、一日あたり最大100である。

こうした収集されたツイート数は合計で 218,815 件であった(表 2)。内容分析を行うコーダーの作業量を考慮して、収集されたツイートの中から国名 (China, Japan, Korea, Taiwan) については 2.3%、差別語 (Japs, Chink) については 1.8%の確率でランダムにさらにツイートを抽出し、これを内容分析の対象として二人のコーダーに評価してもらった。二人の評定者のうち一人はアメリカ人 (女性) であり、もう一人はアメリカ人ではないがほぼネイティブスピーカーに近い欧米出身の外国人(男性)である。

表 2 収集したツイート数

	収集したツイート数	内容分析のため抽出した個数
China	39,246	913
Chink	36,392	665
Japan	39,336	906
Japs	25,213	461
Korea	39,317	945
Taiwan	39,311	927
合計	218,815	4817

2.2 内容の評価の手続き

まず、こうして抽出されたツイートが意図通りにアジア四カ国に関係するものであるかどうかを判定してもらった (表 3 の項目 A)。その結果、「検索語と全く関係のないもの」が 2%、「意味不明なツイート」1%、「ツイートの本文ではなくアカウント名中の検索語にヒットしているもの」7%があった。また、これら以外に、検索語はツイート本文に含むが内容的に主要な対象が上記の四カ国以外であるもの(つまり、Japs で日本、Chink で中国、Korea で韓国か北朝鮮以外であるもの)が 11%あった(表 5)。

表 3 内容分析で評定した各項目の説明

項目	説明	コーディング方法
A	No relation to the target nation (検索語)	1. No relation (eg. China as porcelain), 2. unable to understand (eg. Not English, incomplete sentence) 3. only account name
B	Main target nation	1 Japan 2 South Korea 3 North Korea 4 China 5 Taiwan 6 Other Nation 7 No reference to any nation 8 multiple nations
C	Topic (一つのみ)	下の表 2 参照
D	Any emticons or ASCII art	1 Yes
E	Economy/technology	1. Positive 2. Negative
F	National character/morals	1. Positive 2. Negative
G	Overall evaluation of the target nation	1. Positive 2. Negative

表 4 トピック(c)の項目

番号	トピック	例	番号	トピック	例
1	economy	company	15	TV drama	
2	products	cameras	16	popular music	
3	technology		17	game	
4	politics	election	18	anime/manga	
5	military		20	travel	
6	society/customs		21	food	
7	looks		22	disaster	
8	language		23	crime	corruption
10	books		24	environmental issues	whaling
11	academic related		25	greetings/comments of weather	"good morning"
12	traditional culture		88	other topics	
13	sports		99	unable to define	
14	movies				

各ツイートについて表 3 の 7 項目を評定してもらった。項目 B では、そのツイートの主

要対象国、項目 C ではトピックを表 4 の表から最も近いものを一つ選んでもらった。さらに、絵文字またはアスキーアートが含まれるかどうか (項目 D)、経済・技術的側面での評価が含まれるかどうか (項目 E)、国民性・道徳の側面での評価が含まれるかどうか (項目 F)、全体的な評価 (項目 G) について評価してもらった。

表 5 は、項目 A の結果を示したもので、さらに具体的にどこの国が主要な対象となっているのかを検索語と主要対象国の関係を示したものが表 6 である。以下の内容分析に関しては、表 6 でセルを塗りつぶした組み合わせ、つまり検索語と主要対象国が完全に一致すると評定されたツイートに限定して分析結果を報告することにする。

表 5 各検索語の抽出結果

検索語		分析対象	検索語と 無関係	意味不明	アカウン ト名のみ	他の国が 主要対象	合計
China		706	38	6	110	53	913
	%	77%	4%	1%	12%	6%	100%
Chink		284	1	4	78	298	665
	%	43%	0%	1%	12%	45%	100%
Japan		829	1	9	23	44	906
	%	92%	0%	1%	3%	5%	100%
Japs		270	21	4	12	154	461
	%	59%	5%	1%	3%	33%	100%
Korea		923	1	8	13	0	945
	%	98%	0%	1%	1%	0%	100%
Taiwan		811	11	5	100	0	927
	%	88%	1%	1%	11%	0%	100%
合計		3823	73	36	336	549	4817
	%	79%	2%	1%	7%	11%	100%

表 6 検索語とツイートの主要対象国の関係

検索語		日本	韓国	北朝 鮮	中国	台湾	その 他の 国	国へ の言 及な し	複数 の国	合計
China	度数	0	2	0	661	0	3	8	41	715
	%	0%	0%	0%	92%	0%	0%	1%	6%	100%
Chink	度数	3	0	1	167	0	0	292	2	465
	%	1%	0%	0%	36%	0%	0%	63%	0%	100%
Japan	度数	823	0	0	0	0	0	2	42	867
	%	95%	0%	0%	0%	0%	0%	0%	5%	100%
Japs	度数	175	0	0	0	0	0	151	3	329
	%	53%	0%	0%	0%	0%	0%	46%	1%	100%
Korea	度数	1	554	283	1	3	1	6	66	915
	%	0%	61%	31%	0%	0%	0%	1%	7%	100%
Taiwan	度数	3	4	1	3	694	0	2	81	788
	%	0%	1%	0%	0%	88%	0%	0%	10%	100%
合計	度数	1005	560	285	832	697	4	461	235	4079
	%	25%	14%	7%	20%	17%	0%	11%	6%	100%

2.3 評定者の信頼性

二人の評定者の一致度をみるため Cohen の kappa を計算した(表 7)。トピック (変数名 C) のみカテゴリー数が 26 と多いため値が 0.583 と低いが、これ以外は最低でも 0.686 であり評定者の信頼性は問題ないものと思われる。一致率でみた場合、トピックについては 61%であったがそれ以外の項目は全て 85%以上である。二人の評定者の信頼性は高いと考えられるので、二人のデータを以下のような手続きで統合した。(1)一人しか評定していないデータについては、そのまま使う。(2)二人が評定しているデータについては、偏りがないようにランダムに評定者を選んだ。

表 7 評定の信頼性

変数名		N	Cohen's Kappa	単純な一致度
A	検索語と主要国の関係の有無	211	0.956	99.1%
B	主要対象国	1901	0.836	86.6%
C	トピック (26 カテゴリー)	1737	0.583	61.2%
D	絵文字か ASCII アートの有無	60	1	100%
E	経済・技術の評価	41	0.688	85.4%
F	国民性・道徳の評価	188	0.976	98.9%
G	全体的な対象国の評価	424	0.904	96.0%

注 評定対象外のデータは除外して計算している(たとえば、E で経済・技術の評価を含まないもの)

3. 分析結果

以下、まず評定者による内容分析の結果を見て、次にツイートをテキスト分析した結果をみる。

3.1 ツイート内容の分析

・トピック

当該国についてどのようなトピックが語られているのかを、評定者に一つ選んでもらった。各国について特に多いトピックをみていくと、中国(China)の場合、「経済」「製品」「犯罪」「環境問題」関係のトピックが多い(表 8)。日本(Japan)の場合は、「災害」が多いがそれ以外ははっきりした傾向は見られない。韓国(Korea)は、「政治」「軍事」のトピックが多い。台湾は「ポピュラー音楽」が多い。

・絵文字・アスキーアート

各ツイートの中で絵文字またはアスキーアートが使われているかどうかをみた(表 9)。台湾のツイートの場合に最もよく使われている(8%)。これに対して中国の場合は、ほとんど使われていない(1%)。

・経済的・技術的な評価

当該国の経済的・技術的な側面での評価が含まれている場合、肯定的か否定的なのかを評価してもらった(表 10)。最も評価が高いのが台湾(80%が肯定的)であり、次いで中国の評価が高い(56%が肯定的)。日本と韓国については、否定的な評価の方が多い。

・国民性・道徳に関する評価

中国(93%が否定的)と韓国(91%が否定的)の評価が低い(表 11)。これに対して、日本の評価が最も高く(58%)、台湾(46%)がこれに次いでいる。Chink と Japs は一種の差別語であるので否定的な評価が多いことは当然でもあるが、前者は全て否定的、後者も 84%が否定的な評価となっている。

・全体的な評価

これらの側面を総合的にみた場合、そのツイートが当該国をどのように評価しているのかを見たのが「全体的な評価」である(表 12)。台湾の評価が最も高く(70%)、次いで日本(54%)、韓国(35%)、中国(25%)となっている。また、中国は肯定・否定を合計した評価を含むツイートの比率が四カ国の中で最も高い。一方、台湾は評価を含むツイートの比率が最も低い。

表 8 検索語とトピックの関係

	China		Chink		Japan		Japs		Korea		Taiwan		合計
経済	76	12%	1	1%	81	10%	3	2%	46	5%	72	9%	279
製品	43	7%	0	0%	16	2%	5	3%	12	1%	41	5%	117
技術	35	5%	0	0%	25	3%	2	1%	57	6%	26	3%	145
政治	47	7%	1	1%	55	7%	0	0%	117	13%	45	6%	265
軍事	6	1%	0	0%	5	1%	10	6%	74	8%	6	1%	101
社会・慣習	33	5%	8	5%	41	5%	17	10%	34	4%	20	3%	153
外見	6	1%	38	23%	1	0%	4	2%	17	2%	2	0%	68
言語	3	1%	5	3%	7	1%	4	2%	3	0%	4	1%	26
本	0	0%	0	0%	0	0%	1	1%	3	0%	2	0%	6
学術関係	3	1%	0	0%	3	0%	0	0%	9	1%	10	1%	25
伝統文化	1	0%	0	0%	1	0%	0	0%	1	0%	2	0%	5
スポーツ	18	3%	2	1%	27	3%	2	1%	23	3%	9	1%	81
映画	4	1%	2	1%	7	1%	5	3%	16	2%	4	1%	38
テレビドラマ	2	0%	0	0%	1	0%	1	1%	12	1%	2	0%	18
ポピュラー音楽	7	1%	0	0%	63	8%	9	5%	83	9%	115	15%	277
ゲーム	1	0%	0	0%	12	2%	3	2%	5	1%	1	0%	22
アニメ・漫画	1	0%	1	1%	9	1%	1	1%	2	0%	0	0%	14
旅行	24	4%	0	0%	65	8%	2	1%	104	11%	84	11%	279
食事	28	4%	18	11%	19	2%	7	4%	18	2%	32	4%	122
災害	2	0%	0	0%	32	4%	0	0%	4	0%	5	1%	43
犯罪	96	15%	0	0%	65	8%	11	6%	5	1%	6	1%	183
環境問題	18	3%	0	0%	45	6%	1	1%	10	1%	8	1%	82
挨拶・季節	7	1%	17	10%	52	6%	11	6%	44	5%	61	8%	192
その他	104	16%	42	25%	85	10%	33	19%	100	11%	83	11%	447
不明	96	15%	32	19%	106	13%	43	25%	110	12%	148	19%	535
N	661	100%	167	100%	823	100%	175	100%	909	100%	788	100%	3523

表 9 絵文字・Ascii art の有無と検索語のクロス表

		なし		あり		合計
			%		%	
検索語	China	847	99%	12	1%	859
	Japan	828	96%	34	4%	862
	Korea	901	95%	44	5%	945
	Taiwan	851	92%	76	8%	927
	Chink	355	97%	12	3%	367
	Japs	299	97%	8	3%	307
合計		4081	96%	186	4%	4267

表 10 経済的・技術的な評価と検索語のクロス表

		肯定的		否定的		合計
			%		%	
検索語	China	34	56%	27	44%	61
	Japan	25	46%	29	54%	54
	Korea	19	37%	32	63%	51
	Taiwan	32	80%	8	20%	40
	Chink	0		0		0
	Japs	0		0		0
合計		110	53%	96	47%	206

表 11 国民性・道徳の評価と検索語のクロス表

		肯定的		否定的		合計
			%		%	
検索語	China	8	7%	109	93%	117
	Japan	72	58%	53	42%	125
	Korea	4	9%	40	91%	44
	Taiwan	12	46%	14	54%	26
	Chink	0	0%	7	100%	7
	Japs	3	16%	16	84%	19
合計		99	29%	239	71%	338

表 12 全体的な評価と検索語のクロス表

		肯定的評価		否定的評価		評価を含むメッセージの%
		%		%		
検索語	China	72	25%	220	75%	34.0 %
	Japan	153	54%	128	46%	32.6 %
	Korea	97	35%	177	65%	29.0 %
	Taiwan	142	70%	61	30%	21.9 %
	Chink	1	2%	67	99%	18.5 %
	Japs	8	9%	78	91%	28.0 %
合計		473	39%	731	61%	28.2%

3.2 ツイートのテキスト分析

テキスト分析ソフトを用いて各国をトピックとするツイート内の頻度を分析した。ツイート分析対象期間内の Japan (合計 39336 ツイート), Korea (合計 39317 ツイート), China (合計 39246 ツイート), Taiwan (合計 39311 ツイート) のそれぞれのツイート内で、ツイッター機能のリツイートやリンク等に関連する文字列 (`http`, `rt`, `co` 等) や数字、接続詞、冠詞、また、英語において意味が曖昧な単語 (`good` 「良い」という場合以外に高い頻度で `good morning` 「おはよう」として用いられるもの) として使用される単語を除外した集計結果が表 13 である。

表 13 各国ツイートのもっと頻度の多かつた単語（1～20位）の集計結果

JAPAN		KOREA	
単語	度数	単語	度数
1 America	1562	north	11289
2 Japanese	1475	kimjongnumberun	1425
3 Year	1451	korean	1413
4 Taiji	1191	japan	1286
5 China	1166	nuclear	1116
6 World	1060	seoul	1095
7 Tokyo	1049	kim (jung un)	1038
8 Amour	997	china	1022
9 People	975	world	1011
10 Canada	902	news	971
11 News	861	america	970
12 Women	779	video	794
13 Radiation	752	test	756
14 Uberfacts	748	super junior	696
15 Music	742	south	688
16 Earthquake	741	Schmidt	684
17 Children	703	love	663
18 Killed	683	rocket	660
19 March	682	launch	619
20 Sweden	672	everyone	594

CHINA		TAIWAN		
単語	度数	単語	度数	
1	Chinese	1484	china	2493
2	Man	1301	amour	1477
3	World	1203	taipei	1417
4	People	1151	korea	1369
5	America	1099	japan	1259
6	Children	1041	news	1162
7	Lol	1029	concert	1112
8	News	792	happy	981
9	Kids	783	hong kong	931
10	Everything	764	thank you	899
11	Knife	759	lol	881
12	Japan	758	today	848
13	English	675	asia	832
14	Girl	648	kids	816
15	Amour	580	singapore	771
16	India	574	lmfao	751
17	Stabbed	571	thai	751
18	Dead	553	died	740
19	elementary school	519	bro	728
20	Crazy	517	virgins	719

Japan

Japan のツイート集計結果で、最も多かった単語が **america** (1位) であったが、これは日本に関係するニュース等のリツイートで米国の記載が多いことと関係していると示唆できる。ツイートの上位に出現する他の国として **china** (5位)、**canada**(10位)や **Sweden**(20位)がある。**Sweden** が 20位に登場している理由として、同国と日本において空からの落下物の噂がツイッター上に出現し、そのことに関係していることが考えられる。また地域を表す単語に関して、**World** が 6位、**tokyo** は 7位であった。震災に関係する単語として **radiation** (放射線) が 13位、**earthquake** (地震) が 16位、**march** (3月) が 19位であったことにより、東日本大震災とその後に関する関心の高さが窺える。

Korea

Korea のツイートの集計のなかで、north が 1 位、つまり北朝鮮に関する記載が極めて多く、2 位に kimjongnumberun という北朝鮮指導層の架空（パロディー）情報を発信しているアカウントのリツイートが多かった。指導者名 Kim jung un（金正恩）は 7 位であった。国や地域の出現では Japan は 4 位、china が 8 位、america が 11 位で korea というトピックにおいて japan の頻度が比較的高いことがわかった。Nuclear（核）が 5 位で、rocket ロケット 18 位、launch（発射）19 位など、核施設や北朝鮮の軍事動向に関する単語の出現頻度が高かった。北朝鮮に関係する単語の出現が多い中、大韓民国の関連する単語は seoul（ソウル）の 6 位、super junior（韓国アイドルグループ＝スーパージュニア）が 14 位であった。

China

China の集計の中で、chinese(中国人)が 1 位であった。国では、america が 5 位、japan が 12 位、india が 16 位であった。また、japan, korea, taiwan には上位に登場しないが、中国に関係して、「子ども」(children=6 位、kids9 位)を表す単語の頻度が多かった。これは小学校で複数の子どもが刺殺された事件がツイッター上で人気ニュースになったこととも関係していると考えられる。同事件に関係する単語として、次の通りである knife=11 位、stabbed（刺された）=17 位、dead=18 位、elementary school =19 位、crazy(man)=20 位。

Taiwan

Taiwan の集計の中で、国や地域を表す単語の頻度の多さが顕著であり、結果、次の通りであった china（1 位）、korea(4 位)、japan（5 位）、hong kong（9 位）、asia（13 位）、singapore(15 位)、thai(17 位)。

「好き」の出現比較

また、Japan, Korea, China, Taiwan に関して「好き」や「人が好き」という表現の頻度に関する度数では、Taiwan=1477 が最も多く、次いで Japan=997, Korea=663, China=580 であり、中国が最も低かった。

蔑称「Chink」「Japs」

Chink はアジア系に対する蔑称であり、そのため、このトピックでは不敬や猥褻な単語が多く（shit=4 位、fuck=5 位、ass=10 位、bitch=13 位、fucking=15 位）、女性に関する単語も多かった（she=6 位、her=7 位、bitch=13 位、girl=14 位）もっとも多かった lol は「大笑い」という意味であり、2 位の eyes は chink eyes（つり目、裂け目）というスラングのため、多かった。

日本人の蔑称 japs に関しては、1 位（myjaps）や 3 位(bivouac)のアカウント名のツイートやリツイートの頻度は多く、また、chink eye と同様、jap eye は裂け目（猥褻な意味

も含む) ものとして頻度が本集計結果では2位であった。そのため、この集計結果では、日本人の蔑称 **japs** に関するツイートの分析を行うことができなかった。

表 14 トピック **chink** でツイートの最も頻度の多かった単語 (1~15位) の集計結果

	単語	度数
1	lol	2649
2	eyes	1651
3	amour	1124
4	shit	1063
5	fuck	1001
6	she	992
7	her	973
8	food	855
9	people	804
10	ass	789
11	face	722
12	see	720
13	bitch	698
14	girl	680
15	fucking	566

表 15 トピック **japs** でツイートの最も頻度の多かった単語 (1~5位) の集計結果

	単語	度数
1	myjaps	3138
2	eye	3052
3	bivouac	1801
4	love	866
5	cotton	706

結論

本研究は試行的なものであり、いくつかの問題点がある。ひとつはメッセージからは **Korea** が韓国を指すのか北朝鮮を指すのかが、必ずしも明確に分離できないということである。また、**Twitter** は 140 文字以内という制限もあり、意味が明確に把握できないツイートも多数あった。また、今回対象としたのは英語のツイートであるが、発言者の国籍や居住地も把握できない。

本分析の結果は、中国とコリア(韓国・北朝鮮)について否定的な言説が大部分を占めてい

ることを示していた。これに対して、日本と台湾に関しては肯定的な言説が相対的に多かった。今後は、テキスト分析の方法を発展させるとともに質問紙調査によっても同様の結果が得られるのかを検討する必要があるであろう。

参考文献

石井健一 (2012) Twitter上のメッセージによる国のイメージ測定—内容分析とテキストマイニングによる分析 Department of Social Systems and Management Discussion Paper Series, no 128.