

Department of Social Systems and Management

Discussion Paper Series

No.1277

マイクロブログ Twitter における日本人利用者の特徴
(Understanding Japanese Users on Microblog Twitter)

by

石井 健一
(Kenichi ISHII)

August 2011

UNIVERSITY OF TSUKUBA
Tsukuba, Ibaraki 305-8573
JAPAN

要約

マイクロブログとして世界で最も使われている Twitter の API を利用して発言を 19,409 件収集し、発言頻度やフォロー関係などを日本語利用者と英語利用者の間で比較した。その結果、次のような知見が得られた。

- (1) 日本語利用者は、フォロワー、フレンドの数は英語利用者よりも少ない。
- (2) 日本語利用者は発言頻度が多く、「お気に入り(like)」を使う頻度も多い。
- (3) 日本語利用者は、フォロー関係の相互性が高く、さらにフォロー数の多い人ほど相互性が高い傾向がある。

こうした特徴は、日本人利用者の文化的特性を何らかの形で反映していると考えられ、利用者の動機など心理学的な視点からの比較研究が今後の研究課題となる。

研究目的

Twitter は 2011 年現在、日本で最も使われている SNS といえる (石井 2011)。したがって、なぜ日本人が Twitter を多く使うのかについて研究することは、日本人のインターネット行動を理解する上できわめて重要であると考えられる。

Twitter は、API (Application Program Interface) を公開しているため、発言データの自動収集が比較的容易¹であり、いくつかの分析事例も既に発表されている (Naaman, Boase & Lai 2010, Kwak, Lee, Park, & Moon 2010, Java, Song, Finin & Tseng 2007 など)。しかし、これらの報告では Twitter 上のすべての発言(あるいは英語のみ)を対象としており、日本人 (日本語利用者) など言語別の利用パターンの特徴を分析したものはみられない。本研究は、そうした問いに答えるために、Twitter 上における日本人利用者の発言パターンの特徴を他の国の利用者と比較することを目的として行うものである。

方法

Twitter から提供されている API²の一つである public_timeline を使い 2011 年 7 月 29 日から 31 日にかけて Twitter 上のツイート(発言)を 19,409 個収集した。同一 ID の発言を取り除いたところ、ユニークな ID 数は 17,985 個であった。なお、収集したデータの中で ID の重複回数の最大値は 31 だった。

API で収集した発言データには、発言の言語を示すデータがついている。本研究で収集

¹ ただし、時間あたりの利用回数などに制限があり、無制限にすべての発言データが収集できるわけではない。

² <https://dev.twitter.com/>

したデータでの言語の分布をみると英語(en)が 67.4%、日本語(ja)が 18.1%、スペイン語(es)が 7.7%となっていた。ただし、中国語やインドネシア語など Twitter では指定できない言語があり、これらの言語で書かれた発言の多くには、英語になっているとみられる。

さらに、この 17,985 個の ID の中から Twitter で発言した言語(lang)が英語と日本語の利用者でフォロワー数とフレンド数がともに 5000 人未満³という条件を加えて、各 200 件の ID を無作為に抽出した。この 400 件の ID のデータがもつフォロー関係を friends/ids という API によって求めたところ、合計 164,803 件のフォロー関係のデータ(フォロワーまたはフレンドになっている ID のペア)が得られた。このデータを用いて、フォロー関係の相互性に関する比較を行った。本報告で使う主要な変数が表 1 にまとめられている。

表 1 測定した主な変数

変数名	内容
言語	en(英語), ja(日本語), es(スペイン語)など
フォロワー数	フォローしている人数
フレンド数	フォローされている人数
発言数	測定時点までに発言した総回数
お気に入りの数	お気に入り(like)の回数合計
開設日時	Twitter に登録した日時

結果

(1) フォロワー、フレンドの数の比較

フォロワー数(フォローしている人数)とフレンド数(フォローされている人数)の平均値は、表 2 のようになった。標準偏差の値が非常に大きくなっていることからわかるように、右側に歪んだ裾が長い分布であることがわかる(平均値が中央値の 3 倍になっていることから、分布が右に歪んでおり、フォロワー数・フレンド数が極端に多い人が少数いることがわかる)。

日本語利用者と英語利用者での差をみるため、こういう分布の場合に適切とされるノンパラメトリック検定の一つ Mann-Whitney の U 検定を行ったところ、フォロワー数について日本語の平均ランクは 6538.44、英語の平均ランクは 8001.23 となり、両者の間には 0.1% 水準で統計的に有意な差がみとめられた。フレンド数についても、日本語の平均ランクは 6498.88、英語の平均ランクは 8011.86 となり、同様に 0.1% 水準で統計的に有意であった。したがって、日本語利用者は、フォロワー数、フレンド数ともに英語利用者よりも少ないといえる。

³ フォロワーまたはフレンドの数が 5000 人以上であった ID は 224 個(1.2%)であった。

表 2 フォロワー数とフレンド数

	日本語利用者		英語利用者	
	平均値	中央値	平均値	中央値
フォロワー数	398.8 (SD=3048)	110	406.8 (SD=2558)	169
フレンド数	431.5 (SD=3237)	101	585.0 586.0(SD=3912)	172

(2)発言数、「お気に入り」数の比較

次に、発言数の比較を行う。ただし、日本語利用者と英語利用者では、Twitter を始めた時期に少し違いがある。英語利用者が始めてから平均 420.7 日、日本語利用者が平均 388.7 日であり、英語利用者の方が利用期間が長い ($F=33.2$, $DF=1$, $p<.001$)。したがって、利用期間の長さを補正するため、(総発言数)÷(2011年7月31日での利用日数)で、一日あたりの発言数を計算し、これを比較することにした。この結果が表 3 である。やはり標準偏差の値が大きく、分布が大きく右に歪んでいるとみられるので、Mann-Whitney の U 検定を行ったところ、一日あたりの発言数について日本語の平均ランクは 8150.44、英語の平均ランクは 7568.22 となり、0.1%水準で統計的に有意であった。また、「お気に入り」(like)の個数についても同様に検定したところ、日本語の平均ランクは 8302.51、英語の平均ランクは 7527.37 となり、これも 0.1%水準で統計的に有意であった。これらの結果から、日本語利用者は、英語利用者に比べて、一日当たりの発言数が多く、また「お気に入り」を多く使う傾向があるといえる。

表 3 一日あたり発言数とお気に入りの数の平均値

	日本語利用者		英語利用者	
	平均値	中央値	平均値	中央値
一日あたりの発言数	36.3 (SD=69.5)	15.79	28.9 (SD=55.8)	13.12
お気に入りの数	136.5 (SD=336.9)	6	42.1 (SD=137.9)	3

(3)フォローの相互性の比較

Twitter の一つの特徴は、フォローが一方向的でも可能なことである。この点は、Facebook や mixi など主要な SNS で友人関係が必ず双方向的であることと異なっている。Kwak, Lee, Park, and Moon (2010)によると、Twitter でフォローしているユーザー間のペアのうち 22.1%だけが相互にフォローしていたと報告されている。

しかし、今回の分析結果は、ペアの相互性についてこの報告より高い値を示した。表 3

は、いずれかがフォローしている ID 間のペアにおいて、相互にフォローしている比率を示したものである。この比率は、日本語利用者では 63.2%、英語利用者では 37.9%であった。日本語利用者と英語利用者の間にはきわめて大きな差があり、統計的にももちろん有意である($\chi^2=10280$, $DF=1$, $p<.0001$)。日本語利用者の方が英語利用者よりも相互にフォローする比率が高いといえる。

また、フォロー数が少ない人ほどフォロー関係の相互性が高いのではないかという予想をたてたが、この予想はデータからは支持されなかった。分析では、表 4 のように、フォロー数合計=(フォローしている人数+フォローされている人数)によって、利用者を三つのグループに分けて相互にフォローしている比率を求めて比較した。まず、英語の場合、フォロー数と相互性との間に明確な関係が認められなかった。表 3 でもわかるようにフォロー数合計が少ない ID の方が相互性が高いという明確な関係はない。これに対して日本語利用者の場合は、フォロー数合計が多い ID ほど相互にフォローしている比率が高く、予想とは逆の関係がみられた(ピアソンの相関係数を計算すると、 $r=0.222$, $p<.001$)。これは、日本人の場合、多くの人とフォロー関係を結ぶ人ほど、関係の相互性に配慮している傾向があることを示している⁴。

表 4 相互にフォローしている比率

	日本語利用者 (N=68,312)	英語利用者 (N=96,491)
全体	63.2%	37.9%
フォロー数合計		
400 以下	54.7%	38.5%
401-1000	56.6%	36.5%
1001 以上	67.6%	38.3%

N は分析対象のペア数

(4) 変数間の相互の関係

最後に今回の分析で使った変数間の相関関係を表 6 にまとめておく。データ件数が多いため、ほとんどの組み合わせが統計的に有意である。

注目すべき結果は、

- ・フォロワー数とフレンド数の相関係数が英語利用者よりも日本語利用者の方が高い。これは、日本語利用者がフォロー関係で相互性を考慮するため、フォロワー数とフレンド数がバランスのとれている利用者が多いことを反映していると考えられる。

⁴ ただし、これらの関係性を検討したデータは日本語・英語について、ペア数は数万と非常に多いが、もともとは各 200 人の ID のフォロー関係のデータを合成したものである。したがって、フォロー数とフォロー関係の相互性については、統計的に推定される以上に誤差が大きい可能性がある。

・フォロワー数・フレンド数が多いIDの発言数が多いとは限らない。相関係数は正ではあるが、0.1以下であり、両者の関係はきわめて弱い。

・開設してからの日数が長い人は、発言頻度が少なくなる傾向がある。この関係は、英語利用者の方が強い。

・日本語利用者の場合、一日あたり発言数と「お気に入り」個数との相関が0.178と高くなっている。頻繁に使う利用者は、「お気に入り」も多く使う傾向があることがわかる。

表6 変数間の相関係数

変数名	フォロワー数	フレンド数	一日あたり発言数	お気に入りの数	開設してからの日数
日本語利用者 (N=3257)					
フォロワー数	1	0.950***	0.089***	0.027	0.053**
フレンド数		1	0.082***	0.024	0.043*
一日あたり発言数			1	0.178***	-0.096***
お気に入りの数				1	0.266***
開設してからの日数					1
英語利用者 (N=12125)					
フォロワー数	1	0.737***	0.081***	0.059***	0.088***
フレンド数		1	0.077***	0.042***	0.073***
発言数			1	0.090***	-0.178***
お気に入りの数				1	0.153***
開設してからの日数					1

結論

これらの結果は、同じTwitterシステムを使っている、日本人(日本語利用者)と英語利用者の間には、利用パターンに顕著な違いがあることを示している。最も顕著な違いは、フォロー関係の相互性である。日本語利用者の場合、英語利用者に比べて20%以上もフォロー関係の相互性が高く、日本人のTwitter利用者が「フォロー」された場合、相手との人間関係を気にして「フォロー」を返すのかもしれない。こうした日本人の特徴は、「お気に入り」を使う頻度が多いこととも関連していそうである。また、もう一つの日本人の利用パターンの特徴は、フォロワー・フレンド数が英語利用者よりも少ないにも関わらず、発言の頻度が多いことである。これらの特徴は、日本人のコミュニケーションの文化的な特徴を反映している可能性がある(Ishii & Ogasahara, 2007; Ishii & Wu, 2006)。今後は、こうした特徴がどのような心理的な傾向と関連しているのかを、発言内容の具体的な比較や

利用者のアンケート調査などで解明することが必要であろう。

最後に、今回の分析についての問題点と今後の課題をあげておく。まず、今回の分析では、2011年7月末の数日間に発言を収集したが、この時期に発言を収集したことによってデータにどのようなバイアスがかかったかは明らかではない。また、Twitterの言語コードが日本語の利用者が日本語を使っていることはほぼ確かとみられるが、英語については実際は中国語など他の言語の利用者も含まれているようである（Twitterではこれらの言語に対するコードがないため、中国語などの利用者は英語に設定して使っていると考えられる）。したがって、今回の比較は必ずしも純粋に日本語と英語利用者の比較とは言えない面がある。

また、TwitterはAPIを公開しているため、プログラムによって発言を自動的に投稿することが可能であり、BOTといわれるプログラムを使った自動投稿が、特に発言頻度の多いIDには多く含まれていると考えられる。表5は、今回収集したデータの中から、開設日から50日以上経過し、かつ一日あたりの発言回数が多い上位IDを示したものである。各種の商品・サービス情報を一日1000回近くも発言し続けているIDもあり、人間が通常の方法で書き込んでいるとは考えられないので、プログラムによる自動投稿と考えられる。こうした投稿が含まれているということは、利用者の行動を心理的なモデルで説明しようとすることに對してノイズ要因になる可能性があることにも注意すべきであろう。

表5 一日あたりの発言回数が多いIDの上位(開設日から50日以上)

順位	ID	1日あたり 発言回数 平均	開設日	内容
日本語利用者				
1	Pcsoftwarekenky	1018.7	2011/ 1/26	PCソフトウェア情報
2	Nowwwwww	996.0	2010/10/27	芸能人情報
3	oirase5963	967.3	2011/ 5/27	各種の商品情報
4	i_appinfojp_pdf	793.6	2010/10/17	iPhone関係の情報
5	maichan_001	704.1	2011/ 2/23	風俗関係情報
英語利用者				
1	Jobs_Australia	849.1	2011/ 5/26	求人情報
2	Jtorrent	795.7	2009/ 5/12	不明(ほぼ同じ文を繰り返している)
3	TaraBluth	788.0	2011/ 4/23	医学研究情報
4	tebakanlagu	727.7	2011/ 1/ 5	不明(インドネシア語)
5	wilmanstandard	715.1	2010/ 6/23	ポルノ情報

参考文献

- Akshay Java, Xiaodan Song, Tim Finin & Belle Tseng (2007) Why We Twitter: Understanding Microblogging Usage and Communities, Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007.
- Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon (2010) What is Twitter, a Social Network or a News Media? WWW 2010, April 26–30, 2010
- 石井健一(2011)「強いつながり」と「弱いつながり」の SNS—個人情報の開示と対人ネットワークの比較—、情報通信学会第 28 回学会大会発表用資料 (2011 年 7 月 3 日)
- Ishii, Kenichi & Ogasahara, Morihiko (2007) Links between Real and Virtual Networks: A Comparative Study of Online Communities in Japan and Korea. *CyberPsychology & Behavior*, 10(2), pp.252-257.
- Ishii, Kenichi & Wu, Chyi-In Wu (2006) A comparative study of media cultures among Taiwanese and Japanese youth. *Telematics and Informatics*, 23(2), 95-116.
- Mor Naaman, Jeffrey Boase, Chih-Hui Lai (2010) Is it Really About Me? Message Content in Social Awareness Streams, CSCW 2010 Proceedings of the 2010 ACM conference on Computer supported cooperative work, February 6–10.