

位置情報付き SNS ログデータにみる使用言語の多様性

—世界都市東京と観光都市京都の比較—

桐村 喬

Diversity of Languages Used in Geotagged Log Data of Social Networking Services in Tokyo and Kyoto

Takashi KIRIMURA

Abstract: This study aims to reveal the diversity of languages used in geotagged log data of social networking services (SNSs) in Tokyo and Kyoto. Twitter's geotagged data, a type of SNS log data, provides a source of knowledge on local geographic features. The data come from geotagged tweets posted in 2014 in Tokyo 23 wards and Kyoto city. A relative variety of languages are used in Tokyo, where many international tourists come and many foreign people live. On the other hand, a variety of languages except for the Japanese language is used in some tourist attractions of Kyoto. Foreign people live in Kyoto are not so influential in the diversity of languages used on Twitter as Tokyo.

Keywords: Twitter, 使用言語 (language used on Twitter), 外国人人口 (foreign population), 外国人観光客 (international tourists)

1. はじめに

本研究は、代表的な SNS (ソーシャル・ネットワークワーキング・サービス) の一つである Twitter のログデータを利用し、都市内部の小地域単位で言語別の使用ユーザー構成比を求め、使用言語の地理的な多様性や特徴を明らかにするものである。

位置情報付きの SNS ログデータは、観光や混雑の予測のような人々の行動に関する研究 (桐村, 2013a) や、投稿される内容に基づく、地域特性や方言・文化の地理的特徴の把握に関する研究 (峪口ほか, 2015) などに利用されている。全世界を対象とするサービスである Twitter では、使

用言語の情報も投稿されるデータに含まれ、その地理的特徴を明らかにすることができる。

都市内部のようなミクروسケールにおける使用言語の地理的な特徴については、日本の場合、国勢調査などで使用言語に関する調査が行われていないことから、看板等に記載された文字についての詳細な調査結果などから把握されてきた (Yamashita, 2013)。しかし、看板等を資料とする場合、広域的に調査することは現実的ではなく、ミクروسケールで使用言語の地理的特徴を把握できる新たな資料が必要と考えられる。

そこで、本研究では、ミクロスケールにおける使用言語の地理的な特徴を捉えるための資料として、位置情報付き SNS ログデータに注目し、その地理的特徴を把握する。日本語以外の使用言語の地理的特徴は、居住する外国人人口や来訪し

桐村 喬 〒516-8555 三重県伊勢市神田久志本町 1704

皇學館大学文学部コミュニケーション学科

Phone: 0596-22-6404

E-mail: t-kirimura@kogakkan-u.ac.jp

た外国人観光客の分布と一定の対応関係をもつものと予想され、比較可能な統計データが存在する、国勢調査結果に基づく外国人人口の分布に注目し、外国人居住者の分布との比較を通じて、使用言語の地理的特徴を把握する。

分析の対象地域は、日本の大都市のなかでも外国人人口の割合が高い、世界都市としての東京 23 区と、多くの外国人観光客が来訪する観光都市としての京都市である。分析に用いる空間単位は、東京 23 区に関しては、2005 年の国籍別外国人人口が把握できる町丁目単位とする。また、京都市に関しては、町丁目では空間単位が過小であることから、国勢統計区に相当する、複数の町丁目からなる元学区単位とする。

分析に利用する位置情報付き SNS ログデータは、Twitter のログデータであり、Streaming API を利用して取得した、2014 年 1 月 1 日から 2014 年 12 月 31 日までに東京 23 区および京都市においてモバイル端末等から投稿されたものである。この 1 年間に東京 23 区内では約 1,370 万件、京都市内では約 178 万件的投稿があった。これらのデータを利用して、各空間単位における言語別の使用ユーザー数を求め、各空間単位での全ユーザーに占める各言語の使用ユーザーの割合を求める。

2. 東京 23 区における使用言語の地理的特徴

2.1 言語別使用ユーザー数

表-1 は、東京 23 区における使用ユーザー数の多い上位 10 言語を示したものである。最も多いのは日本語であり、全ユーザーの 79.9% を占めている。日本語以外の言語のうちで最も多いのは英語であり、タガログ語、インドネシア語、ベトナム語、朝鮮語が続いている。2010 年の国勢調査結果によれば、総人口に対する外国人の割合は 3.0% であり、日本語以外の言語の割合は外国人人口の割合と比べて著しく高いことがわかる。

2.2 使用言語の地理的特徴

日本語を除く、東京 23 区において使用ユーザー

の多い上位 5 言語について、その地理的特徴を把握する。図-1 は東京 23 区における英語の使用ユーザーの割合を示しており、港・渋谷・目黒区と台東区に割合が高い地区がまとまっている。1km を近傍の条件に設定して、Anselin の Local Moran's I 統計量 (Anselin, 1995) を求めると、3 区のまとまりは、高い割合の地区が 5% 水準で有意に隣接・集積している状態 (High-High : HH) である。図は割愛するが、上位 5 言語のうちタガログ語、インドネシア語についても、港・目黒区、台東区に割合が高い地区を確認することができる。朝鮮語に関しては、台東区での広がり是一部

表-1 使用ユーザー数上位 10 言語 (東京 23 区)

| 順位 | 言語 | ユーザー数 | 全ユーザー数に占める割合 |
|----|---------|---------|--------------|
| 1 | 日本語 | 328,584 | 79.9% |
| 2 | 英語 | 29,841 | 7.3% |
| 3 | タガログ語 | 5,893 | 1.4% |
| 4 | インドネシア語 | 5,164 | 1.3% |
| 5 | ベトナム語 | 3,358 | 0.8% |
| 6 | 朝鮮語 | 2,880 | 0.7% |
| 7 | スペイン語 | 2,842 | 0.7% |
| 8 | 中国語 | 2,537 | 0.6% |
| 9 | タイ語 | 2,297 | 0.6% |
| 10 | ドイツ語 | 2,148 | 0.5% |
| | 全ユーザー数 | 410,999 | 100.0% |

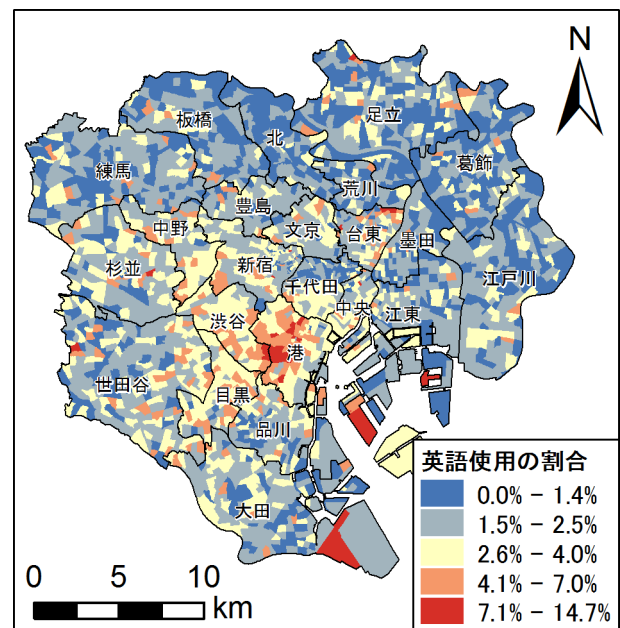


図-1 英語使用ユーザーの割合 (東京 23 区)

重なるものの若干の相違がみられ、新宿・中野区に Local Moran's I 統計量が HH であるまとまりが存在している。一方、残るベトナム語については、東京 23 区の縁辺部を中心に HH の地区が分布するものの、大きなまとまりを確認できない。

2.3 外国人人口割合と言語別使用ユーザー割合

東京 23 区に関しては、東京都が独自に集計した 2005 年の国勢調査結果に関する国籍別外国人人口を把握できる（桐村, 2013b）。桐村（2013b）で示された港・渋谷区に広がるアメリカ国籍・イギリス国籍の割合の高い地域は、英語使用ユーザーの分布と重なっており、これらの地域では、両国籍を中心とする居住者による英語の使用が多くみられるものと思われる。一方、同様に英語使用ユーザーの割合が高い台東区では、アメリカおよびイギリス国籍の外国人人口の割合は低く、英語の使用ユーザーは居住者よりも、観光客を中心としたものであると推測される。台東区のなかでも英語使用ユーザーの割合が高い地域は、観光地である浅草や、安価な宿泊施設が集積する山谷などであることも、英語使用ユーザーの主体が観光客であることを示唆している。上位 5 言語のうち、インドネシア国籍の外国人人口割合とインドネシア語使用ユーザーの割合とでは大使館等の存在する目黒区で、韓国・朝鮮国籍の外国人人口の割合と朝鮮語使用ユーザーの割合とではニューカマーを中心とする集住地区を含む新宿区で、それぞれ割合が高い地域が重なっている。ただし、韓国・朝鮮国籍のオールドカマーが集住する地区においては、使用言語との対応関係はみられない。残るフィリピン国籍とベトナム国籍の外国人人口の割合と、対応する言語別の使用ユーザーの割合との対応関係はそれほど明瞭ではない。

3. 京都市における使用言語の地理的特徴

3.1 言語別使用ユーザー数

京都市における日本語使用ユーザーの割合は、東京 23 区よりも高い 84.9%である（表-2）。使用

ユーザーの割合が高いその他の言語は、英語、タガログ語、インドネシア語、スペイン語などであり、東京 23 区とは順位が若干異なる。京都市の 2010 年の外国人人口の割合は 2.2%であり、東京 23 区ほどではないにしても、外国人人口の割合と比べて、日本語以外の使用ユーザーの割合は高い。

3.2 使用言語の地理的特徴と外国人人口

京都市の場合、元学区数が少ないことから、Local Moran's I 統計量による分布の検討は行わない。また、国籍別外国人人口を把握できないため、使用ユーザーの多い言語の分布と、総人口に占める外国人人口の割合の比較を行う。

表-2 使用ユーザー数上位 10 言語（京都市）

| 順位 | 言語 | ユーザー数 | 全ユーザー数に占める割合 |
|----|---------|--------|--------------|
| 1 | 日本語 | 58,602 | 84.9% |
| 2 | 英語 | 3,726 | 5.4% |
| 3 | タガログ語 | 682 | 1.0% |
| 4 | インドネシア語 | 607 | 0.9% |
| 5 | スペイン語 | 482 | 0.7% |
| 6 | 中国語 | 478 | 0.7% |
| 7 | ベトナム語 | 407 | 0.6% |
| 8 | タイ語 | 399 | 0.6% |
| 9 | 朝鮮語 | 346 | 0.5% |
| 10 | ドイツ語 | 280 | 0.4% |
| | 全ユーザー数 | 69,031 | 100.0% |

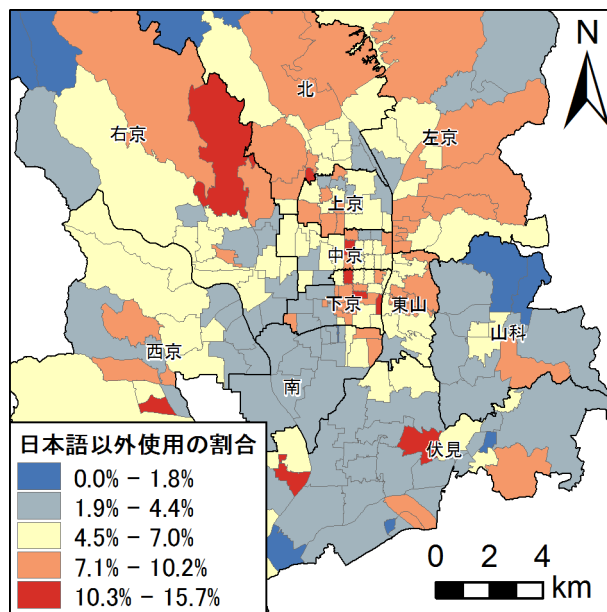


図-2 日本語以外の言語の使用ユーザーの割合（京都市主要部）

京都市における日本語以外の言語の使用ユーザーの割合は、中心市街地であり、観光地でもある上京・中京・下京・東山区と、京都駅周辺や、郊外の観光地周辺で高い傾向にある(図-2)。2010年の外国人人口の割合と比較すると、京都駅周辺と左京区南部において割合の高い地域が一部重なるものの、外国人人口の割合が高いのは市の南西部であり、日本語以外の使用ユーザーの割合の分布とは対照的なものとなっている。従って、京都市に関しては、日本語以外の言語の使用ユーザーの多くが居住者ではなく、主として観光客であると考えられる。

日本語以外で使用ユーザーの多い上位5言語を言語別にみると、英語の使用ユーザーの割合は、日本語以外の使用ユーザーの割合の分布に近い。一方、英語に次いで多いタガログ語以下の言語は使用ユーザー数が少なく、言語別の使用ユーザーの割合を正確に読み取ることは難しい。例えば、タガログ語の使用ユーザーが5以下の元学区は185地区であり、全225地区の大多数を占めている。そのため、分母となる全ユーザー数の多寡が各言語の使用ユーザー割合に大きな影響を与えやすく、全ユーザー数の少ない周辺部ほど外れ値も多くなる。

4. おわりに

紙幅の都合上、多くの図を割愛せざるを得なかったが、これまでの分析の結果として、国籍別外国人人口の割合が把握できる東京23区においては、居住する外国人人口と言語別使用ユーザー割合に関して、一定の地理的な対応関係が存在することを確認できた。外国人人口の割合との対応関係がみられない地域もあるものの、主要な観光地や宿泊施設が集積する地域であることから、主として観光客による当該言語の使用と考えられた。京都市に関しては、外国人人口の割合との対応関係はむしろ逆転しており、多くは観光客が様々な言語を使用しているものと考えられた。ただし、

東京23区と比べて全ユーザー数が非常に少なく、英語を除けば分布パターンを十分に吟味できるほどの使用ユーザー数を抱える言語は少ない。

なお、本研究では、Twitterのログデータを利用したため、Twitterに代わる別のSNSが多く利用されている中国などからの観光客による各言語の使用状況は把握できなかった。また、本研究は使用言語の地理的特徴に関して基礎的な整理を行ったのみであり、居住者と観光客の区別をより厳密に行うことができれば、外国人観光客の分布の推定など、社会的・経済的意義をもつ研究を展開できると考えられる。

謝辞

本研究はJSPS科研費15K12886の助成を受けたものである。

参考文献

- 桐村 喬 (2013a) 位置情報付きツイッター投稿データにみるユーザー行動の基本的特徴—観光行動分析への利用可能性—。「地理情報システム学会研究発表大会講演論文集」, **22**, CD-ROM.
- 桐村 喬 (2013b) 居住地域構造との関係からみた東京23区における国籍別外国人集住地区の社会経済的特徴。「人文地理」, **65**, 29-46.
- 峪口有香子・桐村 喬・岸江信介 (2015) ツイッター投稿データにもとづく「気づかない方言」の分布解明, 日本語学会2015年秋季大会発表原稿集, 193-198.
- Anselin, L. (1995) *Local Indicators of Spatial Association - LISA. Geographical Analysis*, **27**, 93-115.
- Yamashita, A. (2013) The Linguistic Landscapes of Multicultural Living Areas in Japan, *Meikai Nihongo (Meikai Japanese Language Journal)*, **18** (suppl.), 207-221.