

日本のソーシャルメディア利用者の対「外国人」感情  
—機械学習モデルによる日本語テキストの感情分析—

Sentiment toward “Foreigners” among Twitter Users in Japan: A Japanese-text Sentiment Analysis using the  
Machine Learning Model

大茂矢由佳 Yuka Omoya  
筑波大学 人文社会科学研究科  
University of Tsukuba, Graduate School of Humanities and Social Sciences

キーワード: ソーシャルメディア、外国人、感情分析、テキストマイニング

## 1. 研究目的とその背景

本研究の目的は、外国人に関連した日本語ツイートの感情分析から、ツイッターにおける日本人の対「外国人」感情を解明することである。2010年代以降、欧米の移民研究においてソーシャルメディア上の言説に着目した研究が急速に増加している(Backfried & Shalunts, 2016; Ekman M., 2019; Evolvi, 2018)。日本においては、在日コリアンに対するネット空間での差別を分析した高(2015)が先駆的であるが、欧米と比較すると十分な研究蓄積が存在しないのが実情である。

近年のメディア環境に鑑みると、日本人の「外国人」に対する感情や意識を考える際に、ソーシャルメディアの影響力は無視できない。そこで本研究では、日本でもっとも利用者の多いツイッターに焦点をあて、「外国人」に関する言説を総体的に分析することを試みる。移民という言葉の使用が避けられる日本においては、外国人材や外国人労働者、技能実習生など複数の代替語が存在する。こうした事情から、本研究では「外国人」を意味する語をキーワードにツイートを収集し、各語が使用される際に発信者がいかなる感情を抱いているかを解析する。また、これらの語とともに出現する言葉を分析し、こうした感情が生み出される要因や背景について考察する。オンライン空間で広がる対「外国人」感情を解明するという本研究の着眼点とアプローチは、日本の移民研究分野でははじめてであり学術的貢献が大きい。

## 2. 分析

### 2-1. データ

本研究では、統計ソフト R の *rtweet* パッケージ (Kearney, 2019) を用いて「移民」、「外国人」、「外国人材」、「外国人労働者」、「技能実習生」、「難民」の6つのハッシュタグ<sup>1</sup>を含むツイートを収集した。データの収集期間は2020年2月7日から24日の18日間である。収集されたツイートは移民が578、外国人が1,689、外国人材が96、外国人労働者が367、技能実習生が260、難民が500であった<sup>2</sup>。

### 2-2. 分析方法

まず、*sentiment\_ja*<sup>3</sup>による感情分析を行なった。これは、喜び、悲しみ、怒り、恐怖、驚き、嫌悪という人間の6つの普遍的感情(Ekman P., 1970)に基づいて生成されたオープンソースの機械学習モデルである。本研究ではプログラミング言語 Python にこのモデルをインポートし、収集した全てのツイートに対して感情スコアを算出した<sup>4</sup>。

次に、日本語テキストマイニングのための R パッケージ *RMeCab* (石田, 2017) を用いて、ツイートのなかに頻繁に出現する名詞のペアを抽出し、共起ネットワーク分析を行なった。

### 3. 分析結果の考察と課題

感情分析の結果、「喜び」を多く含むのは、外国人 (29.5%)、技能実習生 (27.4)、外国人材 (25.8) であった。「悲しみ」は難民 (17.4) と移民 (8.1) 以外のツイートではほとんど観測されなかった。「怒り」は移民 (58.2) が突出して多く、以降は外国人労働者 (39.5)、技能実習生 (27.5)、難民 (24.9) が続いた。「恐怖」は外国人材 (55.8)、外国人労働者 (34.1)、外国人 (29.8) で多く観測された。「驚き」は外国人で 12.2% が観測されたものの、その他のツイートでは 10%未満であった。「嫌悪」は難民 (27.2) がもっとも多く、以降は移民 (8.5)、技能実習生 (5.8) と低い割合に留まっていた。

共起ネットワーク分析の結果、移民は「反感」や「問題」など、「怒り」に結びつく語との共起性が確認された。加えて、「政策」などの政治的な語の出現が多いことも、移民のツイートに「怒り」が多い理由といえる。難民のツイートでは、「デンマーク」や「スウェーデン」など具体的な難民受け入れ国名が出現している。したがって、他国の状況を踏まえた上で、難民受け入れが「大変」であるとの主張や、受け入れ「反対」の表明がなされていると推察される。感情分析で「怒り」よりも「嫌悪」が勝った理由として、日本のツイッター利用者が難民をめぐる問題を他人事として認識していることが指摘できる。外国人材のツイートでは「在留資格」や「PR ビザ」などが多く、労働市場における外国人材のプレゼンスの高まりが「恐怖」へと結びついていると推察される。一方、外国人労働者のツイートには「投票」や「入管」など政治的な語が多く、これが感情分析で「怒り」が多かった要因と考えられる。

本研究の限界として、データの収集期間が限られていた。また、人間の感情分類は Ekman の他にも複数の研究者が試みている。より多くの感情を検討に加えることで、よりの確で詳細な分析・考察が可能になる。

### 参考文献

- Backfried, G., & Shalunts, G. (2016). Sentiment analysis of media in German on the refugee crisis in Europe. *Lecture Notes in Business Information Processing*, 265, 234–241.
- Ekman, M. (2019). Anti-immigration and racist discourse in social media. *European Journal of Communication*, 34(6), 606–618.
- Ekman, P. (1970). Universal Facial Expressions of Emotion. *Mental Health Research Digest*, 8(4), 151–158.
- Evolvi, G. (2018). Hate in a tweet: Exploring internet-based islamophobic discourses. *Religions*, 9(10).
- Kearney, W. M. (2019). rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software*, 4(42), 1829.
- 石田基広 (2017) 『Rによるテキストマイニング入門 (第2版)』森北出版.
- 高史明 (2015) 『レイシズムを解剖する—在日コリアンへの偏見とインターネット』勁草書房.

<sup>1</sup> ソーシャルメディアのタグ機能のこと。特定のキーワードの前に「#」をつけることで投稿がタグ化される。

<sup>2</sup> なお、難民に関しては同一内容のツイートが大量に抽出されたため、感情分析の結果への影響を考慮し、その他のツイートとは別途分析を行なった。

<sup>3</sup> モデルは以下のリンクから入手可能。 [https://github.com/sugiyamath/sentiment\\_ja](https://github.com/sugiyamath/sentiment_ja) (2020年2月29日最終アクセス)

<sup>4</sup> たとえば、2月17日 13:21:49 のツイート「行政文書を優しい日本語で書くのは、日本人にとってもありがたい。法務省「やさしい日本語有識者会議」に期待すること | にほんごぶらっと <https://t.co/LgNhsJU6xm> #在留資格 #技能実習生 #特定技能 #外国人材」の感情スコアは、喜び(13)悲しみ(0)怒り(1)恐怖(3)驚き(0)嫌悪(0)

---

のようになる。