

ツイッターを利用した流行語の変遷に関する研究

横越 梓・吉田 江依子

1. はじめに

従来の流行語の変遷に関する研究は、主観的記述によるものが多く、反証が難しいという問題がある。本稿では、流行語の変遷を客観的に分析する方策の一つとして、ツイッターから得られる言語データを利用することを提唱する。ツイッターは言語学分野では「新しい」言語資源であるが、どのように利用可能であるのか、言語資源としての特性を考察し、運用面での問題も検証しながら、反証可能な流行語研究の可能性を探る。

2. 先行研究の問題点

流行語については、これまで多くの研究がなされてきた。その論点は、発生の時期、流行の理由、流行源、生産性・創造性、普及性などであるが、従来の分析は、反証可能性が欠如している。例えば桑本 (2000: 13)は、流行語の変遷や消滅の理由について、「使い古された語彙は面白さが無い。そうして若者言葉の語彙は時に劣化していく。新しい語彙にとって代わられるか自然消滅する」と述べている。変遷の理由を、「使い古されることによる新規性の消失」としているが、「使い古される」といった説明の仕方は主観的な記述に過ぎず、どれだけの期間、どの程度使われたら使い古されるのか、何を以てその判断をしているのかについての客観的な基準は示していない。したがって、この主張が正しいものであるかどうかを判断することが出来ない。他の流行語の先行研究においても同様である(cf. 井上 (1998))。反証可能であることは科学研究にとって必須であり、このように反証可能性が欠如している従来の流行語研究は、問題があると考えられる。

3. 言語資料としての特性

反証可能性の欠如という従来の流行語研究にみられる問題点に対して、ツイッターを利用した研究が一つの解決策を与えるのではないかと考える。ここでは、ツイッターの言語資料としての特性について考察する。

3.1 従来の言語資料

従来の流行語研究において用いられてきた言語資料には大きく分けて、聞き取り調査によるものとコーパスの2種類があった。前者の問題点は、データ量の少なさと信頼性の低さである。聞き取り調査では、対象となる語彙を「使うか、使わないか」あるいは「聞いたことがあるか、ないか」という質問形式で調査する。そのため実際には使っているのにそのことに無自覚で、「使いますか」と問われても「使わない」と答える人もおり、データの信頼性が危ぶまれる。一方、コーパスは、データ量も多く、実際に使われたものを収集しているので信頼性の問題もないが、スピーチレベルに関して流行語研究には適当ではないと考える。スピーチレベルとは、人の状況に応じた言葉の使い分けをいい、凍結・正式・諮問・略式・親密スタイルの5段階に分類されている。コーパスに収録される資料は、その性質上、凍結スタイルや正式スタイルがほとんどである。それに対して、流行語は、略式・親密スタイルに分類される類の言葉であり、データの対象が合致しない。コーパス自体が大規模データであっても対象となる流行語のデータを大量に収集することは難しいと考えられる。

3.2 ツイッター

ツイッターは、従来の言語資料に比べて、流行語研究に利用するのにより適当であると考えられる。第一に、そのデータ量の多さである。ツイッターの利用者は世界全体で3億3500万人(2018年)、日本国内では月間アクティブユーザが4500万人、アクティブ率が70.2%となっている(2017年)。彼らのツイートが日々データとして蓄積されることを考えると、コーパスにおけるデータ量をはるかに超える膨大なものであり、その数値の分析は客観性をもった研究に貢献するであろう。第二に、実際に使用されているものなので、コーパスと同様、信頼性の問題もない。第三に、利用者の大部分が若者であるという点である。App Ape Labの調査によると、20代までの利用者が半数近くを占めており、50代以上の利用者はわずか8%である。流行語の発信源および利用者が主に若者であることを考えると、流行語の研究に適したデータを提供することが予測される。第四に、スピーチレベルが合致するという点である。ツイッターで発信される言葉は、熟考して投稿されるというよりは思いつままにつぶやいたものであり、略式スタイル、親密スタイルに分類されるであろう。先に述べたように流行語研究の観点からいえば、まさにこのスタイルが研究対象となる。以上の点から、ツイッターは、言語資料として流行語研究に適切であると言える。

一方で、運用の面から、なかなか手が出しにくい言語資料でもある。ツイートの収集方法には、リアルタイ

ム検索、データの購入、Twitter Search API（以後 API）による収集がある。情報工学分野の研究においては、APIによる収集が一般的であるが、これを採用するにはいくつか問題がある。第一に、費用の問題である。APIには、無料版で使用できるものもあるが、利用可能な取得データに制限がある。十分なデータを取得するためには（最大で月額2,500ドルという）高額な契約料が必要となる。第二に、データ解析の際に、プログラミング言語に関する多少の知識を要する。それを専門としない研究者にとってはハードルが高いものとなっている。ツイッターを用いた研究は、情報工学分野においてはすでに数多くなされているが、言語学分野においては「新しい」言語資料であり、コーパスに比べてその整備は進んでいない。そのためツイッターを用いた研究は言語学分野においてはほとんど例がなく、その研究の方法論が確立していないという問題点もある。次節では、この問題に対して、具体的な研究事例を紹介する。この研究は、実際にツイートデータを収集し、客観的な数値を裏付けとした提案を行っており、その手法や分析など、反証可能な流行語研究の一つのモデルになることを期待する。

4. 事例研究

ツイッターを利用した研究例を紹介する。橋本ら(2020)では、ツイッターのツイートデータを用いて流行語を抽出し、使用頻度の推移に基づいて分類し、それぞれにおける語の特徴を調べた。若者によく使われる語句に着目し、2018年のユーキャン流行語大賞の女子中高生部門で1位を獲得している、テンションが上がるという意味の「あげみざわ」、及びそれを基にした「やばみざわ」や「らぶみざわ」などの「〇〇みざわ」を対象に、どのような新語がどのような使用の推移を見せるかを調査した。具体的には、まず、ツイッター社が公開するWebAPIを通じて日本時間で2018年1月1日から2019年12月31日に渡って発信された「みざわ」という文字列を含むツイートを取得し、ツイート本文から「〇〇みざわ」を抽出した。¹そして抽出した語を使用回数の多い順に並べ替え、使用回数が200回を超えていた38語を対象データとし、階層的クラスタリングを用いて、4つのクラスターに分類した。各クラスターには以下のような語が含まれる。

- (1) 1: ちらみざわ、バブみざわ、らぶみざわ… 2: にちみざわ、おめみざわ、よろみざわ…
3: ツラみざわ、やばみざわ、いたみざわ… 4: 嬉しみざわ、よきみざわ、えろみざわ…

各クラスターにおける語の使用頻度の推移は以下のとおりである。クラスター1に含まれる語のピークは最新の月である。クラスター2はすべての語の最新3ヶ月の使用頻度がピーク時の50%未満であり短期型の傾向を示す。クラスター3はピークが比較的早く、ピーク後も比較的安定してピーク時の50%以上使用されている。クラスター4はピークが若干遅めな傾向がみられ、ピーク後はほぼ継続的に60%を超えて使用されている。クラスター3、4に属する語が、ピークを迎えた後も継続的に使用される、あるいは今後も使用される可能性が高い長期型流行語と言えるだろう。また、各クラスターに含まれる語を考察すると次のような傾向がみられる。クラスター1は形容詞・形容動詞語幹でないものに「みざわ」がついたものが多いため、この群では付加対象の品詞に関する制限が弱まりつつあると推測される。クラスター2に属する語はある種の挨拶語に「みざわ」が付いたものが特徴的であり、形容詞・形容動詞語幹でない間投表現に付属するものは不安定になると推測される。クラスター3、4は単独で発話として成立するものやある種の文末表現として機能するものが多い。また、形容詞・形容動詞語幹が多いことから「みざわ」がそれらに付加する「み」に「ざわ」を加えたものと再解釈されたと推測される。

今後もより多くの検証が必要であるが、本節では数量的な調査による大規模データの分析の有効性を示した。

5. まとめ

本稿は、流行語の変遷についての従来の研究が主観的で反証可能性を欠くという問題点を指摘した上で、ツイッターが流行語研究の言語資源として適したものであることを述べ、具体的な研究方法や分析のモデルケースとなる事例研究を示した。

参考文献

- 井上 史雄 (1998) 『日本語ウォッチング』 岩波新書：東京。
桑本 裕二 (2010) 『若者ことば 不思議のヒミツ』 秋田魁新報社：秋田。
橋本受旺, 武藤敦子, 森山甲一, 横越梓, 吉田江依子, 犬塚信博 (2020) 「Twitter データを用いた流行語使用頻度の推移に関する分析」, 令和二年度電気・電子・情報関係学会東海支部連合大会。

¹ リツイートは他人のツイートの引用であり、本人によって純粋に使用されたとは言い難いため除外した。