

# Twitter (X) 上の日本語を対象にした言語学的 研究に関する覚え書き

田川 拓海\*

キーワード：Twitter (X)、マイクロブログ、資料性、CMC、会話の発生、引用

## 要 旨

本論文では、Twitter を対象にした日本語研究について整理・検討を行い、1) Twitter が日本語研究の対象として貴重なメディア・媒体であること、2) それにもかかわらず研究論文がまだそれほど多く出しておらず特に査読付き論文誌掲載論文は少ないこと、を示し、3) 会話の発生と引用というトピックを取り上げて、新語や俗語だけでなくコミュニケーションの観点からも Twitter は興味深い対象であることを見る。

## 1. はじめに

### 1.1 本論文の性格と位置付け

本論文は、正式名称としては「X」となった Twitter 上に現れる日本語を対象にした研究、特に言語学的手法を用いたものを中心に、現在の状況や論点の整理を行うものである。

「X」と正式名称まで変わることになった Twitter のここ最近の急激な変化に迫られて急遽執筆を決めたものであるため、研究論文としてもレビュー論文としても中途半端な内容になっており、取り扱うことができた情報やトピックはごく一部に留まる。内

---

\* tagawa.takumi.kp@u.tsukuba.ac.jp, Twitter アカウントは@dlit

容に含まれる問題の多さはひとえに筆者の力量不足に起因するものであるが、このような複数の性格を持つ文章を公開できるのは紀要論文の柔軟さの利点である。また、機関リポジトリを通じて多くの人がアクセスしやすい形での公開が可能であるのも投稿先に本誌を選択した理由の1つである。

3節で取り上げる日本語研究の対象として Twitter がそれほど取り上げられてこなかったのではないかという調査と考察は、特定の研究分野や研究者集団に対する問題点の指摘というよりは第一に筆者本人の反省として書いたものである。Twitter が和文の査読論文であまり取り上げられず、書籍や紀要論文の存在がそれを補っているという指摘をこのように紀要論文で行うのは奇妙に見えるかもしれないが、現状を鑑みるとまずはできるだけ早く文章化することが優先であると判断した。ここ最近の Twitter (X)の動向を見ているとこの論文が公開されるときにどのような状況になっているかということの予測も難しく、筆者のアカウントを保持することができているかどうかも心許ない<sup>1</sup>。

## 1.2 Twitter に関する用語と最近の動向

本論文では Twitter の歴史や仕組み、関連用語について個別に導入する紙幅の余裕がない。北村・佐々木・河井 (2016)の「第1章 ツイッターとソーシャルメディア」中の「第一節 ツイッターの基本的仕組み」にある簡潔かつ明快な解説を参照されたい。本書は Twitter を対象として行われた研究や調査のレビューとしても優れているが、扱われている研究や調査、データは2010年代中盤までのものが多い。Twitter のソーシャルメディアとしての位置付けは佐々木 (2018)の記述も参考になるがやはり中心的に扱われているのは2010年代中盤までの状況である。

北村・佐々木・河井 (2016: 16-17)で述べられているように 2006 年の最初の投稿が Twitter の開始であると考え、長い歴史を持つ SNS であることが分かる。日本語によるサービス提供が開始された 2008 年からも 15 年が、日本語圏においてその存在感を増すきっかけの1つとなった東日本大震災 (2011 年) からも 10 年以上が経過している。

---

<sup>1</sup> 実際、本論文の執筆を開始した 2023 年 10 月から最終版の修正を行っている同 12 月までの間に限っても、有料化に関する取り組みが進んだり Twitter 外へのリンクの取り扱いに関する仕組みの変更があったりと毎日使用しているようなユーザーであっても現在の Twitter (X)がどのような状況にあるのか把握するのが困難な状態が続いており、本論文でも何をどこまで取り上げるのかの判断が難しかった。

総務省情報通信政策研究所による『令和 4 年度情報通信メディアの利用時間と情報行動に関する調査報告書』(2023 年公開)<sup>2</sup>の「【令和 4 年度】主なソーシャルメディア系サービス/アプリ等の利用率 (全年代・年代別)」によると、Twitter の利用者 (全年代) は 45.3%で LINE (94.0%)、YouTube (87.1%)、Instagram (50.1%) に次ぐ数値である。これは令和 3 年度の 46.2%よりは少し数値を下げているものの、植田 (2022a) で言及されている令和 2 年度の 42.3%よりは増えている。年代別に見ても、50 年代が 31.6%、60 年代が 21.0%とやや低いがおのほかの年代ではいずれも 40%を超え安定している。詳しく論じるには利用時間などのデータも見なければならないが、Twitter が今もソーシャルメディア、SNS として一定の存在感を維持していることは確実と言えよう。

Twitter の正式名称は、2023 年 7 月に「X」へと変更され、仕組みや用語についてもその前後に大きな影響があった。しかし本論文で取り扱うのは「X」になる前の Twitter に関することがほとんどであるため、ここまでは文脈に応じて「X」も付記してきたが以降は必要に応じて書き分ける場合を除いて「Twitter」という名称を用いる。また、「リツイート」などの関連用語も基本的には「X」への改称前のものを用い、必要がある場合は適宜使い分けや説明を行う。

### 1.3 「Twitter 上に現れる日本語を対象にした研究」の範囲

「Twitter 上に現れる日本語」を何らかの形で研究の対象にしたものということであれば、「本稿執筆時現在、J-STAGE で「Twitter」と検索すると、5000 件近くヒットする」(植田 2022a: 17) という指摘からも窺えるように、すでにかかなりの蓄積がある。書籍を見ても、たとえば桐村 (2019)、宮下 (2022) などさまざまなトピックにおいて Twitter に投稿された日本語をデータとして取り上げている研究が見つかる。

これは、まず第一に「Twitter 上に現れる日本語」を研究のデータとして用いる (ことのできる) 分野そのものが多岐に渡るということがその要因として挙げられる。Twitter に関するトピックだけで学会の 1 つのセッションが構成されるほど多くの研究が行われてきた自然言語処理をはじめ、心理学、社会学においても比較的容易に研究を見つけることができる。

---

<sup>2</sup> 関係資料は [https://www.soumu.go.jp/ficp/research/results/media\\_usage-time.html](https://www.soumu.go.jp/ficp/research/results/media_usage-time.html) から確認することができる (2023/11/01 確認)。

個別の分野の事情があることは前提とした上で、その理由の1つには言語が言語学だけでなくさまざまな研究分野、研究トピックにおいて重要なデータとなりうるということがまず挙げられる。それに加えて、テキストの収集と処理（テキストマイニング）の手法とツールがより広く比較的手軽に利用可能になったことも関係していよう。たとえば、学会や論文誌においてそれぞれの研究分野の研究者向けのガイドとして鳥海 (2015)、荒川 (2018)が挙げられる。書籍においても、Russell (2011)のようなソーシャルメディアを対象としたものだけでなく、テキストマイニング全般に対する概説書・入門書である那須川 (2020)や石田 (2020)などでも Twitter に投稿されたテキストの収集方法や処理・分析について取り上げられている。

このように複数の大きな研究分野に関わる Twitter を対象にした研究を概観・俯瞰することは言語を日本語に限っても筆者の力量を大きく超えるため、本論文はそれらのうち言語学的研究のさらに一部に言及するに留まる。しかし、取り上げる研究分野を言語学に限ったとしても実際には言語学と他分野との共同研究があるほか、共同研究でなくてもその研究手法が自然言語処理や社会心理学に近いものも珍しくないため明確な線引きは困難である。また、研究分野による言及範囲の限定を行うのはあくまでも議論を拡散させないための便宜的なものである。線引きをより厳密にすることはかえって現状の把握や研究の進展を妨げることにつながるかもしれない。

以上のことを前提とした上で、以降、表現上の煩雑さを避けるために「言語学的手法を用いた日本語研究」を単に「日本語研究」と呼ぶこととする。

## 2. 言語資料としての Twitter

### 2.1 Twitter の資料性

Twitter へ投稿された日本語のテキストを日本語研究の対象として捉えるということについては、すでに岡田・西川 (2016)、岡田 (2019, 2020)、植田 (2022a)によって、その資料性や研究対象として扱う際の注意点、具体的な手順などについての整理がある。

筆者は Twitter (を含むソーシャルメディア) が日本語研究の対象として貴重な資料である可能性があるにもかかわらず実際の取り組みはそれほど進んでおらず、また研究を行うための前提や方法論の整備も進んでいないという岡田 (2019, 2020)の指摘に賛同する。

SNS には日本語研究の新たな展開の可能性が秘められているとも思われるのである。(岡田 2019: 57)

ここで紹介した研究はすべて Twitter を日本語研究の用例収集の対象として利用したものであり、Twitter 上でのコミュニケーションの実態を日本語・言語研究の立場から分析、考察の実態を考察したものを試みた研究は多くない。(岡田 2019: 61)

ただし、岡田 (2019, 2020) は Twitter などのソーシャルメディア、SNS を日本語研究の対象とした場合にそれぞれのメディア、資料としての特性により注意を払うべきであるということについても明確かつ具体的に論じており、その視点も重要であることはここでも改めて強調しておく。

また、すでに公開されている Twitter を対象にした日本語研究については後でも少し触れるが、ここで挙げた岡田・西川 (2016)、岡田 (2019, 2020)、植田 (2022a) が挙げている文献も参照されたい。

## 2.2 Twitter の変化と言語への影響

Twitter の基本的な仕組みである「フォロー (関係)」「リプライ」「リツイート」などと Twitter を日本語研究の対象にする時のポイントについては、基本的なデータの収集からコミュニケーションの関係まで岡田・西川 (2016) が具体的に丁寧な整理になっており、ここで改めて検討し直すことはしない。

ここで取り上げるのは、Twitter の仕組みの変化が言語使用やコミュニケーションに及ぼす影響である。岡田・西川 (2016) や岡田 (2019) で整理されているように Twitter の仕組みが Twitter 上の日本語の現れ方に関係するのであれば、仕組みが変化した際にその関係にも影響する可能性がある。

2016 年頃までの基本的な仕組みと用語、主要な仕組みに関する簡単な経緯についての整理は、北村・佐々木・河井 (2016: 17-23) を参照されたい。そこで触れられている変化には、たとえば「リツイート」が公式の機能として採用され「非公式リツイート」との使い分けができたことがある。

ほかの言語使用に影響を与えそうな変化で主なものとしては、ハッシュタグに仮名などの日本語表記用の文字が使えるようになったこと (2011 年)、「引用ツイート」の

採用 (2015 年)、「お気に入り」の名称が「いいね」に変更 (2015 年)、「@ユーザー名」と「.@ユーザー名」の使い分けの廃止 (2016 年)、リプライ関係にあるツイートのツリー表示の採用 (2020 年) などが挙げられる。また、日本語は 2017 年に行われた 1 投稿当たりの制限字数が 140 字から 280 字へ増加するという大きな変更の対象ではなかったが、「@ユーザー名」や画像などの添付が文字数としてカウントされなくなったという変更 (2016 年) は実質的な字数制限の緩和と見ても良いだろう。なお下でも触れているように、現在有料プランのユーザーについては字数制限が 1 投稿当たり数千字と大幅に緩和されている。

ただ、これらの変化が Twitter 上の日本語に偶然ではないほどの影響をもたらしたかどうかについては個別に確かめる必要がある<sup>3</sup>。可能性だけ考えても、Neubig and Duh (2013) で示されたように 1 投稿当たりの文字数がそもそも比較的少ない傾向にある日本語においては字数制限の緩和が言語使用にもたらす影響はあまり大きくないかもしれないし、「@ユーザー名」と「.@ユーザー名」の使い分けが元々それほど広くは使用されていなかったのであれば、廃止されてもその影響は大きくなかったかもしれない。引用ツイートの採用の影響については個別のトピックとして引用を扱うところで具体的に述べる。

さらに、特に「X」になってからのごく最近の変化としては<sup>4</sup>、「コミュニティノート」の日本語圏における利用の活性化<sup>5</sup>、有料プランユーザー限定でのいくつかの機能の提供 (既投稿の編集、収益化、長文の投稿)、API の体系と料金の変更などがあり、今後もし実現されたら大きな影響を及ぼしそうなものに全ユーザーに対する有料化が挙げられる。コミュニティノートや収益化はほかのユーザーの投稿に働きかける動機になりうるものであり、全ユーザーの有料化は (アクティブな) ユーザー数そのもの、つまりアカウントを保持し今後も投稿を行うかどうかへの影響が予想される。

---

<sup>3</sup> 関係のありそうな文献は自然言語処理や数理的な手法を用いた社会学的研究などの分野のものと見つかるが、その検討については時間にも紙幅にも余裕がなくここでは行わない。

<sup>4</sup> ここで取り上げている期間は短いにもかかわらずさまざまな変化があり、細かい時系列を追うのが難しくまたその影響もまだよく分からないため、それぞれの変化が発表された日付などの情報は付していない。

<sup>5</sup> コミュニティノートの機能自体は「X」になる前から英語圏では使用されてきたものである (「Twitter の「コミュニティノート」はイーロン・マスクが始めたのではない、ということについて、英語で確認する手順」, <https://hoarding-examples.hatenablog.jp/entry/2023/07/18/Twitter> の「コミュニティノート」はイーロン・マス, 2023/10/01 確認)。

中でも API の体系と料金の変更は Twitter を対象とする研究全般への影響が大きい。2023 年の変更なので具体的にどのような影響が出る (出た) かについてはもう少し時間が経たなければ分からないところもあるが、見通しは明るくないだろう。まず多くの文献で Twitter を研究対象として選択する場合の利点としてあげられている API を利用した大量のテキストデータの取得が難しくなっている。仕組み上不可能になったわけではないが、現在の料金体系では、テキストマイニングや自然言語処理的手法で一般的な研究を行うサイズのデータを得るには一定以上の規模の研究費が必要だろうし、投稿の取得が可能なプランの最低限である Basic プラン (月額\$100) で得られる月ごとに 1 万投稿までという量は、頻度が低い言語表現の振る舞いを調べる場合など研究課題によっては数量的な要求が厳しくない研究手法であっても不足する恐れがある。実際、筆者が学位論文・研究の指導に関わっている学部生・大学院生との相談でも、可能であれば Twitter 以外の資料・データを探したり API の利用を必要としない研究手法を選択するというような話が多くなっており、研究テーマそのものを変更した者もいる<sup>6</sup>。また、研究者も Twitter を対象にした研究を中止したり研究対象を変更したりしているという調査についての報道もある<sup>7</sup>。

なお、API に関する変更は研究においてテキストデータの直接の取得のみに影響があるわけではない。API を利用した非公式のサービス・ツールには大幅な値上げに直面することになったものもあり、Twitter への投稿を公式サイト・公式アプリ以外の手段で確認・検索・収集する方法に影響が出る可能性がある。有名なサービスとしてはアカウントごとに投稿を記録できる「Twilog」において一度投稿の保存が行えなくなり、「togetter」への統合により復活したことが知られている。Twitter 自体に「高度な検索」という検索機能があり、また上記の Twilog のほか Yahoo!リアルタイム検索もまだ利用可能なので今は目に見えて大きな影響はないようだが、今後同じ状況が継続されるかどうかはやや不安の残るところである。

以上概観したように Twitter の仕組みの変化が Twitter 上の日本語や日本語によるコミュニケーションに及ぼす影響にはさまざまな可能性があるが、日本語以外の言語への

---

<sup>6</sup> これには、API に関する変更だけでなく、ここ最近では Twitter そのものの仕組みに急に大きな変更が加えられることがあるので、研究対象として付き合い方が難しいという側面もある。

<sup>7</sup> “Exclusive: Elon Musk’s X restructuring curtails disinformation research, spurs legal fears” (<https://www.reuters.com/technology/elon-musks-x-restructuring-curtails-disinformation-research-spurs-legal-fears-2023-11-06/>, 2023/11/07 確認)

影響が予測できるものも多く、特に研究が先行している英語に関する研究・調査の成果、知見を参考にすることができる<sup>8</sup>。

## 2.3 デジタルデータの継続性・連続性と資料としての Twitter

### 2.3.1 デジタルデータの性質と Twitter のデータ

Twitter の仕組みの変化について本節で取り上げているのは、これが Twitter を言語資料として考える上でも厄介な問題ではないかと考えられるからである。

前提としてまず、デジタルデータそのものの継続性の問題がある。現在進められているデジタルアーカイブへの取り組みは、言語学はもちろんのこと人文系という大きな括りさえ超えた現代の研究活動の課題の 1 つと言えよう。資料のデジタル化の課題の 1 つとして逸村 (2016: 47) で「永続的な保存 (cold storage)」が挙げられており<sup>9</sup>、谷口 (2016) でもデジタル化のデメリットの 1 つとして「デジタルデータを扱うメディアを介してしか伝送できない」が挙げられている。資料をデジタル化した後にはその保存・継続が問題となるが、資料保存が人の手によって継続的に行われなくなってしまう場合にそのデータがある程度時間が経過した後に利用可能かどうかは、その後の技術的進展にもよるが、一般的には難しくなるように思われる。

何らかの目的の下に保存されるデジタルアーカイブと Twitter ではまた事情が異なる。しかし、1 つのプラットフォームとして国際的にその存在を確立している Twitter とはいえ、そのデータがいつまで、どのような形で利用可能かという心配はここ最近特に身近なものになっている。

このような Twitter への投稿をデジタルデータとして見た時の継続性の問題に加えて、Twitter の仕組みの変化が Twitter 上で行われる言語使用・コミュニケーションに何らかの影響をもたらす場合、そのデータとしての連続性も問題となる。日本語について言えば 2008 年以降の 15 年間の投稿を連続したものとしては取り扱えない場合もあるのではないかとことである。もちろんどれだけのデータを連続的に取り扱えるかというのは研究・調査のトピックによって異なる。たとえば、北村・佐々木・河井 (2016: 29) において各言語におけるリプライやリツイートなどの投稿行動の傾向につい

---

<sup>8</sup> たとえば、Twitter 上の英語を対象にした研究では、Pavalanathan and Eisenstein (2016) が Unicode による絵文字 (emoji) が使用できるようになったことによってそれまで用いられてきた顔文字 (emoticon) や非標準的な表記の使用が減少したことを明らかにしている。

<sup>9</sup> ただし、逸村 (2016) はその対象を「重要な資料」と限定している。



て 2010 年代前半に行われた研究がまとめられているが、これらのデータを 2015 年の引用リツイート採用後のデータと連続的なものとして扱って各行動の増減を議論できるかどうかについては検討が必要だろう。このように何らかの要因があることによるデータや資料を連続的に取り扱えないというのはデジタルデータに限らずデータ・資料一般に言えることだが、大きな変化が何回もあるとその分連続して取り扱えるデータの量が減る可能性が高まる。

### 2.3.2 ケータイメール研究との比較から

以上のことを踏まえ、Twitter 上の言語使用やコミュニケーションが日本語研究において重要な記述・分析の対象の 1 つであると考ええるなら、Twitter 上の投稿がデータとして参照可能なうちに研究や調査を行っておくというのが 1 つの対応策になる。たとえば Twitter がサービスを終了した後に過去の投稿を一部でもアーカイブ化したり研究用のデータとして提供したりするという可能性もゼロではないが、それを期待して研究・調査の優先度を下げるのは現実的ではない。

日本語研究において Twitter を対象にした研究を考える上で良い比較例になるのがケータイメールの研究である。ケータイメールは日本における CMC (Computer-Mediated Communication) 研究の対象として 2000 年代では代表的なものであると考えられ (落合 2023)、三宅 (2005) など三宅和子氏の一連の研究をはじめとして重要な記述・分析が存在し、数としてはそれほど多くない文献によって特有の表現、談話・コミュニケーション、文体、顔文字・絵文字までも含めた表記など、幅広くその性質が明らかにされている。また、「加藤安彦ケータイメールコーパス<sup>10</sup>」が整備されたことにより (宮寄他 2020)、2001 年～2010 年に収集されたケータイメールのデータが広く利用可能になっている。ケータイメールと Twitter や LINE といった現在使用されているサービス・ツール上における言語使用・コミュニケーションの比較といった研究も可能なのである (落合 2021)。

このコーパスは資料としての性質を考えずに単純にサイズだけを見た場合はそれほど大規模なものではないが、調査協力者を募ってデータを集めるしかないというケータイメールの性質を考えると、これほどの量のデータが時間をおいてコーパスとして利用可能になったことは僥倖としか言いようがない。なぜなら、このコーパスに収録されているような言語使用を行う話者は現在はおそらく (ほぼ) 存在しないと考

---

<sup>10</sup> <https://www.gsk.or.jp/catalog/gsk2023-b/>、2023/11/01 確認

れ、ケータイメールは高々10年ほど前のメディアであるにもかかわらずすでに「古典資料」とも言えるような存在になっているのである。文体的に見れば佐竹(2005)が明らかにしたように既存の文体の特徴を継承しており、そのほかの特徴についてもLINE上の日本語などに引き継がれているところがあるかもしれないが、2000年代に主要なコミュニケーション手段の1つとして用いられていたケータイメール上の言語使用のデータ、あるいはそれに準ずるものが今後新しく得られる可能性は限りなく低いと言えよう。

ケータイメールに関する日本語研究はある程度の研究とデータが残されたケースである。Twitterを対象にした研究そのものは1.3で触れた通りすでにかんがりの蓄積があるが、日本語研究という範囲で考えると、次節で述べるとおりまだ十分ではないのではないかと考える余地があり、今後も個別の研究の進展が望まれる。

### 3. Twitterを対象にした日本語研究はどのように公開されているか

本節ではTwitterを対象にした日本語研究がどのような媒体で発表されているかについて部分的な調査を行ったものをまとめる。先に述べておくと、調査の結果としての数値は掲載するが、計量的な調査としては予備調査に当たるかどうかという程度の段階にあり、あくまでも今後の調査や研究の方向性を定める際のきっかけとして位置付けられるものである。

#### 3.1 論文誌『日本語の研究』を対象にした調査

この調査は元々TwitterやLINEを対象にした日本語の研究が実際にどれくらい行われているかを調べるといった目的の下に行ったもので、田川(2021)の内容をまとめ直したものである。なお、この調査においてのみTwitterだけでなくLINEを対象にした研究についても調べている。

調査の概要は下記の通りである。『日本語の研究』を調査対象に選んだのは、現在日本国内で刊行されている日本語研究に特化した査読付き論文誌で、日本語研究者からの認知度・信頼度ともに高くかつ幅広い研究トピックを対象としていると判断したためである。

##### (1) 『日本語の研究』の調査概要

- a. 雑誌：『日本語の研究』（日本語学会の論文誌）
- b. 対象：投稿論文、研究ノート、短信、資料・情報のみ（査読のあるもの）
- c. 除外したもの：特集、追悼、報告、書評
- d. 期間：6巻1号-16巻3号の計42号分（2010-2020年に刊行されたもの）
- e. 方法：TwitterあるいはLINEに言及があるものを題目や概要だけでなく本文も確認してチェックする。電子ファイルに対する検索と目視による確認の両方を行う

調査結果は表1の通りである。結論としては、Twitter/LINEを研究対象としているかどうかではなく言及しているかどうかという緩い条件で見ても、Twitter および LINE に明確に触れているものは1件もなかった。投稿論文の数が「(1)」となっているのは、論文中では明確な言及がないものの野田(2011)のGoogle検索の結果にTwitterへの投稿が含まれている可能性があるかと判断したことを示している。

表1 『日本語の研究』2010年-2020年の調査結果

カテゴリー	総数	Twitter/LINE への言及
投稿論文	110	(1)
研究ノート	6	0
短信	4	0
資料・情報	7	0

なお、調査対象外のものとしては2013年と2019年のシンポジウムの報告2件にTwitterへの言及が見つかった。2013年のものは石黒・橋本編(2014)所収の論文として発表されており、2019年のものの一部は岡田(2019, 2020)に反映されているようである。2件と少ない数ではあるがいずれも後述する査読付き論文誌以外の媒体に発表されているものとして数えることができる。

ほかの研究対象との比較がないため、この調査結果からTwitterあるいはLINEに関する研究であることが『日本語の研究』に掲載されない理由であるといった結論を導くことはできない。一口に「日本語」と言っても、まず通時的研究においては現代以

外の日本語も対象になるし、現代語についても琉球諸語はもちろんのこと地理的変種（いわゆる方言）もそれぞれ異なる体系を持った言語として取り扱われるのが一般的であろう。また、研究対象は資料や社会的変種などの条件でも別々に取り扱われることが多いため、日本語研究の対象となる「日本語」はかなりの数になる。このような事情を踏まえると上記の 130 件弱程度の研究の中に Twitter あるいは LINE が出てこなかったことが偶然であっても不思議ではない。

一方で、日本語に関する研究成果を探す際に優先度の高い査読付き論文誌に日本語話者が大量の発話・コミュニケーションを行っている Twitter あるいは LINE に関する研究が発表されていないというのは日本語研究に関わる一研究者としては惜しいと感じる。現在の『日本語の研究』は比較的時間を置かずその内容が J-STAGE 上で公開されるためアクセスしやすい研究の発表媒体の 1 つとなっており、『日本語の研究』に Twitter を対象にした日本語研究が発表される媒体としては書籍などの媒体よりもその成果を他分野からも参照しやすいと考えられるからである。

### 3.2 文献データベースを用いた調査

調査の次のステップとしては、『日本語の研究』以外の日本語研究の論文が掲載される査読付き論文誌、たとえば『日本語文法』（日本語文法学会）や『言語研究』（日本言語学会）などを対象とし、前述の調査と比較することが考えられる。

しかしまだそのような調査には取りかかることができていないので、分野を日本語研究にある程度限定した上で研究の掲載媒体についてより広く概観することのできる国立国語研究所の「日本語研究・日本語教育文献データベース<sup>11)</sup>」による検索結果を参考として以下に示す<sup>12)</sup>。

「データベース概要：「日本語研究・日本語教育文献データベース」について<sup>13)</sup>」によれば、この文献データベースは 1950 年から現在までの日本語学、日本語教育研究に関する研究文献のデータベースであり、収録件数は 2023 年 9 月で 283,661 件となっている。論文データ約 23 万件のうち 3 万件ほどが「図書としての論文集等の掲載論文」となっており、論文集掲載論文についても最近収集を始めてはいるもののまだ十分に

---

<sup>11)</sup> <https://bibdb.ninjal.ac.jp/bunken/ja/>

<sup>12)</sup> この調査を行うきっかけとしては茂木俊伸氏の助言が大きい。

<sup>13)</sup> <https://bibdb.ninjal.ac.jp/bunken/ja/help/about>、2023/10/01 確認

は収録が行われていない CiNii Research よりこの研究分野のものについては多くの文献を検索できるという強みがある。

検索結果は表 2 の通りである<sup>14</sup>。この結果は 2023 年 10 月 28 日に行った検索の結果で、同年 9 月 8 日に行った結果と違いがあったことから、2023 年 9 月 15 日のデータ追加が反映されていると判断できる。少しでも Twitter に言及のある文献をできるだけ広く収集するため、論文名、書名に加えてキーワード・章タイトルも同時に検索対象に含めることができる簡易検索を用いた。また、明らかに Twitter のサービス開始より前である 1996 年に刊行された図書の 1 件分を実際の実験結果から除いたものである。

表 2 日本語研究・日本語教育文献データベースの Twitter に関する検索結果

検索語	件数
Twitter／ツイッター／トゥイッター	34
Twitter	30
ツイッター	9
トゥイッター	0
マイクロブログ	2
参考：ケータイメール <sup>15</sup>	54

「Twitter／ツイッター／トゥイッター」はこのすべてを検索窓に入力し「いずれか」を指定して検索を行ったものである。この結果と「Twitter」「ツイッター」「トゥイッター」の合計が一致しないのは、一部重複する結果があるためである。簡易検索を用いているため、たとえば和文タイトルに「ツイッター」、英文タイトルに「Twitter」が用いられている研究は「ツイッター」と「Twitter」の両方の検索結果として出てくる。「マイクロブログ」は自然言語処理分野で Twitter を対象とする場合によく用いられる名称で、この 2 件とも言語処理学会発行の『自然言語処理』に掲載されているものであった。

<sup>14</sup> 岡田 (2014) を参考に「ツイッター」「トゥイッター」どちらの表記も検索語として採用した。

<sup>15</sup> 「ケータイメール」については表記のバリエーションが多いため、表中には個別の表記を載せなかった。実際に用いた検索語は「ケータイメール」「ケータイメール」「ケイタイメール」「ケイタイメール」「携帯メール」「携帯メール」で、このすべての語を検索窓に入れ「いずれか」を指定して検索を行った。

「ケータイメール」の検索結果と数値上はそれほど差がないが、そもそもケータイメールを対象にした日本語研究もそれほど盛んに行われたものではなかったという点に注意が必要である。先に述べたように最近になってケータイメールコーパスが登場したものの、メディアの性質上ケータイメールの研究は基本的に調査協力者を募る形でデータを収集する必要がある研究に時間がかかる。データの収集に関してのみ言えば技術的なハードルがあったとしても Twitterの方がかかる時間は少ないと言えよう。さらに、Twitterの方の検索結果には言語処理学会の『自然言語処理』掲載論文が5件含まれているなど、日本語研究の範囲を厳しく考えればさらに件数は少なくなる。1節で述べたように「日本語研究」の定義や範囲を厳密にすることは有益ではなく本論文が目指すところでもないが、この文献データベースを対象にした調査からも Twitterを対象にした日本語研究はとても盛況であるとは言えない。

研究の発表媒体に目を向けると、「Twitter／ツイッター／トゥイッター」の検索結果のうち、いわゆる全国規模の学会発行の学会誌に掲載されたものは上述の言語処理学会『自然言語処理』の5件以外には、計量国語学会『計量国語学』から1件、表現学会『表現研究』1件のみである。ただし、この調査では『日本語の研究』を対象にした調査とは違い書籍・論文の本文までは確認していないので、実際には Twitterに言及している研究を拾い上げることができていない可能性は残されている。そのことを考慮しても、以下で述べるようにそのほかの媒体が Twitterを対象にした日本語研究の主な発表の場になっていると判断される。

### 3.3 そのほかの発表媒体

学会が発行する査読付き論文誌以外の媒体に目を向け、Twitter上の日本語に何らかの形で言及しているという条件で探すと、様々な研究を見つることができる。ここでは、学会発表、論文集などの書籍、商業誌、紀要論文、卒業論文における研究の発表についてそれぞれ簡単に触れる。

まず学会誌関連の発表媒体というところでは、先に検討した論文誌のほかに学会発表の内容が反映される発表論文集や学会誌もある。学会発表まで範囲を広げると Twitterを対象にした日本語研究はそれなりに行われていることが分かる。日本語研究・日本語教育文献データベースの検索結果に含まれていたのは日本認知言語学会の学会論文集掲載の西村(2022)くらいであるが、特に日本語用論学会の大会発表論文集

を調べると Twitter に特徴的な言語表現を対象にした研究や若者ことばのデータを得るために Twitter を調査対象にしている研究が複数見つかる。

論文集などの書籍では、たとえば Twitter への投稿を具体的に取り上げており Twitter の独話的性格を考察した先駆的研究である野田 (2014)をはじめ、さまざまなトピックの論文集・書籍で Twitter への言及を見つけることができる。しかし、日本語研究の分野では Twitter をトピックに据えた論文集・書籍というのは「ソーシャルメディア」や「SNS」までトピックの規模を大きくしてもあまり見受けられない。また最近では CiNii Research で一部が検索の対象になってきてはいるものの、論文集掲載の論文のような独立性が高いものでも和書では書籍の一部を文献として検索するのはまだ難しい。日本語研究に関するすべての書籍・論文集の内容に目を通すことも研究者個人としては限界があるため、このカテゴリーについては特に筆者が見逃している研究が多く残されている可能性がある。

日本語研究関連の商業誌として現在も定期的な刊行が行われている数少ない媒体の 1 つである明治書院刊行の『日本語学』では、Twitter への言及が比較的に見つかる。すでに紹介した Twitter を対象にした日本語研究における重要な文献の 1 つである岡田 (2019)や同著者による岡田 (2015)がそうであるし、そのほかにもたとえば 2021 年春号の特集「ポップカルチャーの日本語」では田中 (2021)が実際の用例として Twitter への投稿を取り上げている。『日本語学』はケータイメール研究についてもその発表の場として存在感がある。2000 年に「ケータイ・コミュニケーション」、2001 年に「ケータイ・メール」と立て続けにケータイメールに関わる特集を組んでいるし、三宅和子氏のケータイメールに関する研究はこの特集以外にも三宅 (2006, 2012)と複数掲載されている。Twitter についてはまだ個別のトピックとして特集を組まれるほどの扱いではないが、雑誌『日本語学』は日本語を対象にした CMC 研究にとって重要な媒体の 1 つであることは確かである。

Twitter を対象にした日本語研究は紀要論文としてもよく発表されている。なお、ここで言う「紀要論文」は個別の大学及びその組織と関連の強い雑誌に掲載されたものを広く指しているため、たとえば筑波大学日本語日文学会などのいわゆる学内学会発行の雑誌に掲載されたものも含む。すでに紹介した岡田 (2014, 2020)、岡田・西川 (2016)、植田 (2022a, b)のほか、Twitter を対象とした日本語研究の個別のトピックとしてよく言及される接尾辞「-み」に関する研究も宇野 (2015, 2018)、依田 (2016)、水野 (2017)と紀要論文として発表されているものが多い。新語などの新しい表現を対象に

した研究としてはさらに五味・辰巳・新田(2011)、岡田(2013)、高橋(2020)を挙げることができ、いずれも Twitter からデータを取るだけでなく Twitter のメディア・資料としての性質についても論じている点特徴的である。また、たとえば田川(2020)では新語「タピる」の例文を Twitter から挙げている。このような Twitter の研究であるとまでは言えないがデータや例文として Twitter に言及しているものまで含めると多くの紀要論文がさらに見つかるのではないかと予測される。

卒業論文の指導を担当している教員個人としての体感と他大学の卒業論文の題目一覧などを見る限りでは、Twitter を対象にした日本語研究はここまで見てきた媒体と同じかそれ以上に多く見つけることができる。ただし、卒業論文の web 公開は題目までという形が多く上述の接尾辞「-み」を取り上げている高橋(2017)のように概要が公開されていてある程度内容が確認できることもあるが<sup>16</sup>、多くの場合 web で直接論文本体を読むのは難しいため実際に Twitter から日本語のデータを得ている卒業論文はさらに存在するかもしれない。ちなみに筆者が 2023 年度に主指導を担当している学部生 6 名のうち当初は 3 名が Twitter への投稿から日本語のデータを採取する予定であったが、うち 1 名は Twitter から X への変更の影響が大きいと予測されたためテーマを大きく変更した。残り 2 名は Twitter を対象にした日本語研究を継続している<sup>17</sup>。また田川(2020)で取り扱っている「タピる」の用例も元々は指導学生の卒業論文で取り上げられたものである。すでに述べたように卒業論文は論文の本文全体が web に公開されることが少ないため、執筆者が大学院に進学するようなケースを除いては貴重な研究であってもアクセスが難しいのが一番の難点である。

### 3.4 背景にある事情への推測

筆者が目を通した限りでは、上述の調査で見つかったものには Twitter に投稿された日本語のデータを数理的・計量的に取り扱う研究はそれほど多くなく、個々の投稿を例文や発話のように扱うものが多い。

近年日本語研究においてもコーパス言語学的・数理的手法を用いた研究が盛んになってきた状況からするとこれは一見不思議なように思われる。たとえば、『日本語の

---

<sup>16</sup> human.hirosaki-u.ac.jp/faculty/wp-content/uploads/site/research28/03.pdf、2023/11/01 確認

<sup>17</sup> ただし筆者が担当している筑波大学人文・文化学群人文学類の言語学専攻応用言語学コースで日本語を対象とした研究を行う学生は、そもそもメディアと言語の関係など、社会言語学的なテーマに興味のある者が多いという偏りがあることには注意されたい。



研究』の特集「2018年・2019年における日本語学会の展望」の「数理的研究」領域について、次のような言及がある。

今や日本語研究のインフラとして定着した感のあるコーパス (間淵 2020: 114)

概説書や入門書についても、良質のものが数多く出版された (間淵 2020: 116)

しかしむしろ、このようにコーパスや数理的手法に関する各種のツール、概説書や入門書の整備が進んだことで、Twitter 上の日本語をテキストデータとして収集する必要がそれほどなくなっているのではないだろうか。テキストマイニングについても多くの優れたツールや入門書・概説書が提供されているとは言え、API を介してテキストを収集し、分析の準備となるテキストの処理・整形を行い、形態素解析をし、といった作業を行うのはまだすべての日本語研究者にとって基礎的な技術であるとまでは言い難い。研究の目的に沿うメタデータやアノテーションが十分に整備されたコーパスがあるのであれば、そのコーパスを用いる方が限られた時間や資源の中でより良い研究を行える可能性が高く、優先的に選択されるのは自然である。

その点では、せっかく植田 (2022a) という API の使用法に関する日本語研究者にとってもなじみやすい具体的なガイドとその実践例である植田 (2022b) が出た矢先に Twitter の方でその API そのものが研究には非常に使いにくい状況になってしまったことは大変残念である。

研究の掲載媒体の偏りに関して、特に紀要論文、卒業論文については岡田 (2019, 2020) が指摘するように Twitter が新しい言語現象の発見・観察に向くという特徴と新語・流行語や俗語などの特定の研究トピックとの相性が考えられる。使用されるようになってそれほど時間が経っていない表現について、まだ言語現象としては安定していないもののその存在と現時点での振る舞いを記録しておきたいというような場合には紀要論文がほかの媒体よりは投稿しやすいだろう。また、新語・流行語や俗語は研究対象としての扱いはそれほど簡単ではないが、卒業論文の研究テーマとして一定の人気があるように見受けられる。このことも、卒業論文が Twitter を対象にした日本語研究の発表媒体になることと関係がありそうである。

ここまで個別の事情についての考察を重ねた後に身も蓋もないことを述べるようであるが、Twitter を対象にした日本語研究が少ない大きな要因の 1 つとしてそもそも日

本語研究に携わる研究者が少ないことを挙げておきたい。たとえば接尾辞「-み」についての研究は上で触れたように複数の紀要論文が出ているだけでなく海外の論文誌にも権(2020)やSeraku(2021)といった研究が発表されているが、これは現時点ではむしろ例外的に研究が蓄積されているトピックである。このように Twitter で特徴的に観察される現象についてはうまく研究が繋がっていくことがあるとは言っても、前述したように日本語研究の対象が歴史、変種、資料によって実はかなり多岐に渡ることを踏まえると、1人の研究者が複数カバーできる組み合わせがあるということを考慮に入れても既存の研究トピックに対してでさえ研究者が足りているとはとても言えず、後発の研究トピックでもあり研究手法についても十分には確立されていない Twitter を対象にした日本語研究に手が回らないのは予想できるところである。

もちろん研究に関する文章は内容がまず重要なのであって、あらゆる人がどの種の媒体にも同じようにアクセスできるなら、ここまで述べてきた発表媒体による違いは大きな問題にはならないであろう。しかし実際には媒体によって探しやすさ、探してから手に入れるまでの手順、読むためにかかる料金、などの点で異なり、また発表媒体によって内容、特にどこまで専門的な議論が詳細に展開されるかといったことにも違いが出てくる。これらのことを踏まえると、研究がどのような媒体において発表されているのかというのは現在のその研究分野や研究トピックの状況を把握した今後の研究を盛況にする上で参考になる情報である。

## 4. CMC と Twitter の研究

### 4.1 言語使用、コミュニケーションの研究対象としての Twitter

前節までで Twitter を対象にした日本語研究を取り巻く状況やこれまでに行われてきた研究について概観し、Twitter が新語・流行語や俗語、あるいは若者ことばといったトピックに向いておりまた実際の研究もそのようなトピックについてのものが多いことを見た。

ここでは別の視点から例示として言語使用あるいはコミュニケーションに関する研究を取り上げる。言語使用、コミュニケーションに関する研究は、自然言語処理、社会心理学、社会学の分野でも行われており、Twitter については日本語研究に先駆けて行われていることが北村・佐々木・河井(2016)で挙げられている諸研究を見るだけでも分かる。一方言語学の強みは用いられた言語表現と談話を直接細かく分析する手法

にあり、本節で取り上げる会話の発生と引用という 2 つのトピックはその強みを生かせるものである。

たとえば対面会話が言語使用の最も基本的な形態 (の 1 つ) であるということには筆者も異存がない。しかし、そのような基本的な媒体・場における言語 (使用) を観察・記述・分析することで明らかになるものが言語の基本的な特徴であるかどうかというのは検証が必要な問題であり、たとえば Twitter におけるテキストを用いたコミュニケーションのような「新奇」なケースから言語の使用やコミュニケーション全般に関する新しい基本的な知見が得られる可能性もあるのではないだろうか。

#### 4.2 個別のトピック 1 : 会話の発生

Twitter への投稿を発話・談話と見た場合の面白い特徴の 1 つとして、会話・コミュニケーションの発生そのものが観察でき、それがデータとして取得できるという点が挙げられる。

そもそも Twitter に投稿する行為が当初 “tweet” と名付けられていたことに象徴されているように、Twitter そのものは必ずしも会話に相当する双方向的なやりとりが必須、あるいは主な使い方であるというわけではない。それは北村・佐々木・河井 (2016: 25-27) でも具体的に示されているし、野田 (2014) で指摘されているように個々の投稿を見ても独話的な特徴が観察される。その上で、独話的な “tweet” をきっかけにしてほかのユーザーとの会話・コミュニケーションが発生することがある。

このような独話的な発話をきっかけにした会話の発生は Twitter 以外の場でも起きており、それ自体が珍しいものではない。特に何らかの特定的话题や趣味がトピックになりやすい場、たとえば本屋における「あ、新しい本出たんだ」といった独話やスポーツの試合会場における「〇〇 (選手名) いいぞ!」といった独話を発端にして、近くにいる人、時にはまったく知らない人との会話が発生することはあり得る。

しかし、そのような会話の自然発生を研究用のデータとして得られるかという点が大変難しいのではないだろうか。筆者が知る限りでは対面会話をはじめほとんどの場合どのような手段を用いた会話・コミュニケーションのデータであってもあらかじめその参加者が会話・コミュニケーションを行うことが前提となっている。たとえば幼児と保護者の会話データを得るために特定の部屋を長時間撮影しているようなケースでは会話の発生に関するデータが得られるが、この場合も対象は特定の調査協力者に限られておりその間に会話・コミュニケーションが起きやすいことが前提となっている

であろう。このようなタイプの研究手法の規模を大きくしてたとえば1つの組織・建物や地域全体を対象に長時間撮影を行うことができればさまざまなタイプの会話の発生のデータが得られるかもしれないが、研究倫理をはじめとした調査のデザインが難しく実行の難易度は高そうである。

以上のことを踏まえると、会話の発生のデータが比較的容易に得られる可能性があることは新語や俗語などのデータを得やすいといったこととはまた別に Twitter のデータの貴重さを示している。Twitter における発話に独話的なものと対話的なものがあるという点についてはすでに岡田・西川 (2016)、岡田 (2019) にも指摘があるが、ここでは両者の間、特に独話から対話への移行に焦点を当てている。

Twitter における会話の発生のメカニズムや確率については、Twitter が形成する言語・コミュニケーションの場（以下「コミュニケーション空間」と呼ぶ）をモデル化することで捉えられそうである。残念ながら現在はまだそのモデル化が完成していないので、そのアイデアを示すに留める。

Twitter のコミュニケーション空間をモデル化すると物理的身体世界における対面会話など、既存のコミュニケーションの形態でほぼ同一と言えるものはなかなか思いつかないのだが、感覚的に捉えるために既存のものをを用いて比喩的に表現すると「物理的距離を Twitter 上の関係によって定義される距離で置き換えた立食パーティー会場」のようなもの（田川 2014）というのが1つの候補として考えられる。1つの場に黙っている人も会話をしている人もいて、新たに会話が発生することもある。パーティー会場ではたまたま近くにいることで会話になることもあれば、会話を目的にして目当ての人のところに移動することもあるだろう。Twitter のコミュニケーション空間でもこれらと同じようなことが起こるが、以下に見るようにその際の「近さ」や「移動」が物理的距離によらないということである。この「立食パーティー会場」モデルはあくまで比喩であって、重要なのはこのコミュニケーション空間がどのような要素によって構成され、それらが Twitter 上のコミュニケーションとどのような関係にあるかということである。

Twitter のコミュニケーション空間を会話の発生という観点から見ると、特に関わりそうなものは、1) 基本的な距離を決める要素、2) 個別に会話を誘発する要素、に大別でき、具体的には次のようにまとめることができる。

## (2) Twitter のコミュニケーション空間の基本的な距離を決める要素

- a. フォロー関係
  - b. リスト
  - c. アカウントの公開／非公開
  - d. 使用言語
- (3) Twitter のコミュニケーション空間で個別に会話を誘発する要素
- a. リツイート
  - b. いいね
  - c. ハッシュタグ
  - d. 検索
  - e. 使用言語における個別の言語要素

フォロー関係が Twitter におけるユーザー間の関係において最も基本的な要素であるというのは今でも変わりがないただろう。フォロー関係には方向性（一方向的か双方向的か）もあることから発話が誰に届き／見えやすいかということがフォロー関係に強く影響を受け、1つの会話・談話について物理的身体世界<sup>18</sup>における対面会話では起こりえないような話者と聞き手の複雑な関係が生じることもある（田川 2014、cf. 岡田・西川 2016）。

ほかの要素について見ると、フォロー関係になくても「リスト」を作成しユーザーを追加すればその発話を追うことができるし、フォロー関係にあるユーザーの中で特に注目しているユーザーをさらに「リスト」に入れるというような使い方もある。また、非公開のアカウントであっても公開アカウントのユーザーとのやりとりからほかのユーザーに影響を与えることもあり、コミュニケーションの観点からは完全に不可視としては扱えない。さらに「グローバルなサービスとはいえ、ツイッターは言語圏や文化圏で基本的には閉じた性質を持つ」（北村・佐々木・河井 2016: 29）と指摘されているように Twitter のコミュニケーションは使用言語に強く影響を受ける<sup>19</sup>。

---

<sup>18</sup> 単に我々の物理的身体が存在している世界（場）というくらいの意味である。専門的ではない用語として「ネット」と「リアル」が対置されることがあるが、仮想現実や拡張現実まで含めてネット（web）が言語やコミュニケーションの何にどこまで関わっているのかを考える上で、物理的身体を軸に考えると便利なので用いている。

<sup>19</sup> 一方、たとえば Lai, Toriumi and Yoshida (2023) の研究は COVID-19 についての誤った情報の伝播に複数の言語の関与が影響しているということを明らかにしており、トピックや言語の組み合わせによっては言語圏／文化圏をまたいで研究や対応を行わなければならない可能性がある。

個別に会話を誘発する要素はフォロー関係のあるユーザー間でも有効に働くが、フォロー関係のないユーザー間での会話の可能性も高める。Twitter に特徴的なケースは直接のフォロー関係のないユーザー同士間にリツイートされた発話を介して会話が発生するというものだろう。ハッシュタグはトピックを提示するようなもの以外にも「#~しよう」「#~募集」のように行動を促すものがあり、時には直接会話・コミュニケーションに誘導するものもある。さらに、それぞれの使用言語におけるコミュニケーションに関わる要素も影響してくる。たとえば日本語であればテキストにおけるスピーチレベルシフトの研究が示しているように（野田 2003）、丁寧体であるかどうかのほかのユーザーへの働きかけの強さに影響する可能性がある。

ほかにも会話の発生を考える上で Twitter のコミュニケーション空間を構成する要素の候補になるものはある。たとえば「おすすめ」として表示されやすいほどほかのユーザーとの会話が発生する可能性は増えそうである。また、現在そうであるようにインプレッション（投稿が表示された回数）が収益化において重要な値である場合、リプライなどのほかのユーザーや発話への関わりを促進する可能性がある。ただし、どちらも Twitter の機能としては最近の仕組みである上に、上で挙げた複数の要素からの影響を受ける二次的な要素とも考えられるので、ここではリストに含めなかった。

ここまで取り上げてきた要素は現時点でも数理的にモデル化できるものが多い。しかしその中でも「使用言語における個別の言語要素」は（どのように）モデル化できるかは一律には決められず、言語あるいは言語現象ごとの検討が必要であろう。

#### 4.3 個別のトピック 2：引用

ここで取り上げる「引用」とは Twitter に機能として備わっている「引用ツイート」（現在は「引用」あるいは「引用リポスト」）などに関するもので、必ずしも従来日本語研究で引用を扱う際に主な対象となってきた「と」や「こと」などの引用された発話のマーカ―や「言う」などの引用動詞の研究との関わりがあるものとは限らない。しかし、引用を既存の発話を別の場で示す行為という観点から考えるのであれば、ここから先に示す Twitter における「引用」も引用の一種であり引用研究に何からの形で新しい知見をもたらすのではないかと期待される。

本論文では従来の引用研究との関わりについて詳しく論じることはできないが、メディアや機能の観点も考慮しながら引用に関する談話研究を行っており、また引用に関する丁寧な先行研究の整理を行っているものとしてメイナード (2023) の第 4 章を参

照されたい。ただし、たとえば文字言語である Twitter への投稿においてはメイナード (2023) で重視されている「声」に直接対応するものがないといった違いも存在する。

まず、かつて良く用いられていた Twitter における引用の方法として、自分の投稿の中にほかのユーザーの投稿をそのまま埋め込むという方法があった。下記に実際にほかのユーザに対して筆者が行った投稿を簡略化したものを例として示す。

- (4) 筑波大学は University of Tsukuba RT @abcd 大学の英語名で University of と地名の組み合わせなのは東京大学くらいかな

これは「abcd」というユーザーIDを持つユーザーの「大学の英語名で…くらいかな」という発話(下線部)を自分の投稿に含めており、そのことを「RT」という文字列が表している<sup>20</sup>。「筑波大学は University of Tsukuba」という部分(二重下線部)が自分(投稿者)のオリジナルの発話である。詳しい経緯は省くが、リツイートが公式の機能として登場するよりも早くから用いられていた投稿の方法であり、リツイートの登場後は「非公式リツイート」と呼ばれている。

やはり API に関する大幅な変更により現在は使えなくなってしまった公式以外のアプリケーションではこの形式の投稿が簡単に行える機能を備えているものも多く、公式に「リツイート」が Twitter の機能になってからもまだ引用を行う投稿方式として用いられていたが、公式の「引用ツイート」の機能が登場した後は用いられなくなっていったように見受けられる。

この「非公式リツイート」と公式の機能としての「引用ツイート」の大きな違いは、前者は引用したほかのユーザーの投稿内容に手を加えることができ、後者は仕組み上それが不可能であるということである。たとえば森山 (2012: 17) で触れられているような、デマなどの誤った情報が拡散していく過程で元の発話には付いていたモダリティ

---

<sup>20</sup> 「QT」という文字列が用いられることもあり、「Q」は“quote”の頭文字由来と言われていることからより引用であることを意識した形式であると推測されるが、こちらも現在は実際に使用されている例を見つけることが難しい。なお「RT」自体は今でも「リツイート」を指す頭文字語としては用いられており、最近では「リツイート」の新しい名称である「リポスト」に対応する頭文字語として「RP」が用いられることもある。

形式が省かれてしまう現象は、Twitter の非公式リツイートでまさに起こり得るものであった<sup>21</sup>。

このような元の発話の改変、特に短くすることは 1 投稿当たりの文字数制限があることから元の投稿が長かったり複数回の非公式リツイートを重ねたりする場合により起きやすく、Twitter というメディアの持つ特性が言語・コミュニケーションに影響している例の 1 つでもある。また、現在の公式の引用ツイートでは仕組み上このような引用された投稿内容への直接の干渉は不可能なので、大きく見ると「引用」を行っているものの、それが言語・コミュニケーションにどのように現れているかは非公式リツイートとは異なっている可能性がある。すなわち、2 節で取り上げた同じ Twitter のデータの連続性に関する問題としても考えることができるかもしれない。

メイナード (2023)でも整理されているように、引用が元の発話の(どの程度の)模倣なのかということは引用研究の大きなトピックの 1 つである。テキストによるコミュニケーションでは音声による情報がないこともあって元の発話の完全な模倣を実現しやすい。Twitter はさらにリツイートや引用ツイートなどの形で他者の発話をそのまま用いる機能を持っているため、引用がコミュニケーションの中でどのように用いられるかという課題に取り組む際に 1 つの面白い対象になるのではないだろうか。

## 5. まとめと課題

### 5.1 まとめ

本論文では、Twitter を対象にした日本語研究について、1) Twitter を日本語研究の対象とした場合のポイント、2) 実際にどのような形でどのくらいの研究が行われているか、3) コミュニケーションの観点から見た個別の研究トピック、の 3 点から大まかな整理を行った。

1)については Twitter の資料としての特性に注意が必要であるものの日本語研究の対象として貴重なメディア・媒体であること、2)については Twitter を対象にした日本語研究が査読付き論文誌掲載の論文としてはあまり発表されておらず、論文集をはじめとした書籍(の一部)や紀要論文としての公刊が多いこと、3)については会話の発生

---

<sup>21</sup> 森山(2012)ではNHKの番組で取り上げられた例として紹介されているが、発表の中で Twitter に言及したことが発表を聞いた際の筆者のメモに記録としてある。ただし、この事例そのものが Twitter のものなのか、関連のある話題として Twitter に言及があったのかは定かではない。



と引用を取り上げ、Twitter が新語や俗語といったトピックだけでなく言語によるコミュニケーション研究の対象としても興味深いものであることをそれぞれ確認した。

## 5.2 今後の展開

### 5.2.1 そのほかのソーシャルメディア、CMC 研究との関係

新しいメディア上の日本語を対象にした CMC 研究としては、Twitter よりも LINE が蓄積を重ねているようである (落合 2021)。その背景には、調査協力者を募るといった研究手法などの面でケータイメールの研究を受け継ぎやすかったといった事情があるのではないだろうか。

LINE の研究を中心にメディアと言語、コミュニケーションの研究を精力的に行っている研究者の 1 人である落合哉人氏による落合 (2021, 2023) は CMC 研究全体の優れたレビューにもなっている。これらの成果は現在は web で公開されているというような状態にはないが、今後広く公開され日本語を対象にした CMC 研究の隆盛につながることを期待される。

現在は厳しい状況ではあるが、Twitter を対象にした日本語研究が進展させることができれば、LINE をはじめとした CMC 研究、あるいは言語・コミュニケーション研究全体に対しても新しいデータや知見をもたらすことができる。

### 5.2.2 研究・調査手法の今後

今後も API を用いたデータ収集が以前のように研究に広く利用できる状態に戻らなければ、積極的にそのほかの研究・調査手段を模索する必要がある。

1 つの可能性としては、LINE などと同じように調査協力者を募り、データを提供してもらうという方法がある。API を利用した収集ほどは短期間で大量のデータを得ることができないが、その代わりに言語研究にとって重要な話者の詳細な情報を得て研究を進めきることができる。実際に上で紹介した落合哉人氏による調査が進行中のようである。また、Twitter への投稿が web に公開されている状態が継続されれば、API を介してではなく web からテキストデータを得るという方法にも可能性があるかもしれない<sup>22</sup>。

---

<sup>22</sup> この手法についても落合哉人氏から情報と示唆を得た。

さらに、研究・調査の手法そのものとは少し話が違いますが、特に卒業論文で得られたデータのような公開されないまま死蔵あるいは散逸してしまう言語データを一部でも良いので保存・保持する仕組みを考えられないかという課題を挙げておきたい。これは Twitter を対象にした研究に限らず、ファッション誌や児童向け雑誌、あるいはマンガなどのコーパス・データベース化がされていない、あるいはしにくい資料・データについても同様の問題があるように思われる。

### 5.2.3 生成系 AI

すでに予定の分量を大幅に超過しており本論文では詳細に触れる余裕がないが、生成系 AI の発展と普及も Twitter を対象にした日本語研究への影響が予測される。

たとえば課題としては、Twitter に限らず web で得られる、あるいはコンピューターを介して生み出される言語すべてについて言えることではあるが、生成系 AI によって（部分的に）生み出された、あるいは修正された可能性のあるテキストをどのように言語研究の対象として扱うのかというものが挙げられる。

もちろん自動生成によって生み出された文章というものはこれまでも web に存在し、最近の生成系 AI の発展の前にも bot と呼ばれる自動応答を行うアカウントと人間の「会話」が話題になるようなこともあった。また、文章作成への影響ということであれば生成系 AI だけでなく IME のような基本的な入力そのものに関わるものをはじめさまざまなツール・サービス・アプリケーションに可能性がある。生成系 AI の影響にそれらを越えた大きなインパクトがあるかどうかは今後の調査・検討を待たねばならない。

### 5.3 課題など

冒頭でも述べたように、本論文は日本語研究の研究論文としては実際の言語現象の観察・記述・分析が少なく、レビュー論文としては日本語研究に限っても十分にはカバーできていないものがある。また、本論文で取り上げた日本語研究のトピックについて自然言語処理や心理学、社会学などの他研究分野で参考になる先行研究が多く存在しているにもかかわらずほとんど触れることができなかった。

Twitter はこれまで研究者コミュニティの形成・維持・拡大にも大きな役割を果たしてきた。筆者も Twitter を介して研究に関する情報を交換するだけでなく、研究会・勉強会や学会でのワークショップなどを企画したり、中には共同研究に発展したのもの

ある。田川 (2019)はそれが形になったものの1つである。そのような Twitter の側面も現在基本的には厳しい状況に直面しているように思われる。

これまでも繰り返し述べてきたように分野による違いがあることは前提とした上で、Twitter (X)の変化による研究への影響はさまざまな分野において小さくないと考えられる。本論文が日本語研究だけでなく他分野の研究にとっても何らかの参考になるようなことが少しでもあれば幸いである。

## 付記

本論文の一部は口頭発表である田川 (2014, 2021)の内容を発展させたものである。それぞれの発表の際に参加者の方々から貴重な質問・コメントをいただいた。また、本論文の構想を練っている段階で落合哉人氏、瀬楽亨氏、茂木俊伸氏から研究手法に関する助言や文献に関する情報を得ることができた。記して感謝したい。本論文に残された誤りや問題点はすべて筆者の責任によるものである。本研究の一部は JSPS 科研費 20K00641 「日本語とその変種における外来語・語彙層・借用の形式的研究」の助成を受けている。

## 参考文献

- 荒川歩 (2018) 「PythonとTwitterAPIによるビッグデータ事始め」『社会言語学会第42回大会発表論文集』, 101-104.
- 五味伸之・辰巳暢・新田優喜 (2011) 「Twitter を利用した言語形態の変化についての研究」『福井工業高等専門学校 研究紀要 人文・社会科学』45: 1-6. 福井工業高等専門学校.
- 石田基広 (2020) 『実践 Rによるテキストマイニング—センチメント分析・単語分散表現・機械学習・Pythonラッパー—』森北出版.
- 石黒圭・橋本行洋編 (2014) 『話し言葉と書き言葉の接点』, ひつじ書房.
- 逸村裕 (2016) 「第2章 記録を活かし、記録を循環させる—デジタルアーカイブの学術情報—」『デジタルアーカイブの資料基盤と開発技法—記録遺産学への視点—』水嶋英治・谷口知司・逸村裕編, 33-48. 晃洋書房.
- 桐村喬編 (2019) 『ツイッターの空間分析』, 古今書院.
- 北村智・佐々木裕一・河井大介 (2016) 『ツイッターの心理学』, 誠信書房.

- 権裕羅 (2020) 「感情・感覚形容詞の新しいミ形について—Twitter上の使用に注目して—」 『日本語教育』 92: 65–77. 韓国日語日文學會.
- Lai, Cameron, Fujio Toriumi, and Mitsuo Yoshida (2023) A cross-lingual analysis on the spread of misinformation using the case of Ivermectin as a treatment for Covid-19. *Scientific Reports* 13.
- 間淵洋子 (2020) 「数理的研究」 『日本語の研究』 16: 114–121.
- メイナード泉子・K (2023) 『ミステリードラマの日本語 発話と記号の演出を探る』, くろしお出版.
- 三宅和子 (2005) 「携帯メールの話しことばと書きことば 電子メディア時代のビジュアル・コミュニケーション」 『メディアとことば2』 三宅和子・岡本能里子・佐藤彰編, 234–261. ひつじ書房.
- 三宅和子 (2006) 「携帯メールに現れる方言—「親しき志向」をキーワードに」 『日本語学』 25: 18–31. 明治書院.
- 三宅和子 (2012) 「ケータイの絵文字: ヴィジュアル志向と対人配慮」 『日本語学』 31: 14–24. 明治書院.
- 宮下萌編 (2022) 『テクノロジーと差別 ネットヘイトから「AIによる差別」まで』, 解放出版社.
- 宮寄由美・林直樹・田中ゆかり・三宅和子 (2020) 「一般公開を視野に入れた「携帯メールコーパス整備の試み—加藤安彦氏の遺志を受けて—」 『社会言語科学会第44回大会発表原稿集』, 262–265.
- 水野みのり (2017) 「ネット集団語における接尾辞「-み」の語基拡張」 『思言: 東京外国語大学記述言語学論集』 13: 167–174. 東京外国語大学地域文化研究科・外国語学部記述言語学研究室.
- 森山卓郎 (2012) 「社会的言語習得論—グローバルな「国語力」にむけて—」 『日本語学会2012年度春季大会予稿集』, 15–18.
- 那須川哲哉編. (2020) 『テキストアナリティクス2 テキストマイニングの基礎技術と応用』, 岩波書店.
- Neubig, Graham, and Kevin Duh (2013) How much is said in a tweet? A multilingual, information-theoretic perspective. *AAAI Spring Symposium: Analyzing Microtext*.
- 西村綾夏 (2022) 「Twitter投稿中の絵文字使用に見られる記号的流動性の考察」 『日本認知言語学会論文集』 22: 150–161.
- 野田春美 (2014) 「疑似独話と読み手意識」 『話し言葉と書き言葉の接点』 石黒圭・橋本行洋編, 57–74. ひつじ書房.
- 野田尚史 (2003) 「テキスト・ディスコースを敬語から見る」 『朝倉日本語講座8 敬語』 菊地康人編, 73–92. 朝倉書店.

- 野田大志 (2011) 「[他動詞連用形+具体名詞]型複合名詞の意味形成」『日本語の研究』7: 1-16.
- 落合哉人 (2021) 「「打ちことば」の基盤的研究」筑波大学博士論文.
- 落合哉人 (2023) 「「打ちことば」とは何か—言語学周辺におけるCMC研究のこれまでとこれから」  
口頭発表, 大東文化大学言語学勉強会, 大東文化大学.
- 岡田祥平 (2013) 「Twitterを利用した新語・流行語研究の可能性: アイドルグループ「Sexy Zone」の略語を例に」『新潟大学教育学部研究紀要 人文・社会科学編』6: 49-74. 新潟大学教育学部.
- 岡田祥平 (2014) 「「Twitter」のカタカナ表記は「トゥイッター」か「ツイッター」か: 外来語受容における「原音主義」と「慣用主義」の相克」『新潟大学教育学部研究紀要 人文・社会科学編』7: 49-68. 新潟大学教育学部.
- 岡田祥平 (2015) 「インターネット上に観察される略語: ツイッターと「質問サイト」を対象とした研究の可能性」『日本語学』34: 4-16. 明治書院.
- 岡田祥平 (2019) 「一語からはじめるSNSのことばの研究: SNSの「特性」と先行研究から、その可能性を考える」『日本語学』38: 56-66. 明治書院.
- 岡田祥平 (2020) 「SNSを日本語研究資料として利用するための覚書: 『日本語学』2019年4月号掲載の拙論に対する補遺として」『新潟大学教育学部研究紀要 人文・社会科学編』12: 163-180. 新潟大学教育学部.
- 岡田祥平・西川由樹 (2016) 「日本語研究資料としてのTwitter: コミュニケーション構造の観点から」『新潟大学教育学部研究紀要 人文・社会科学編』9: 93-111. 新潟大学教育学部.
- Pavalanathan, Umashanthi, and Jacob Eisenstein (2016) More emojis, less:) The competition for paralinguistic function in microblog writing. *First Monday* 21(11).
- Russell, Matthew A. (2011) *Mining the Social Web*. O'Reilly.
- 佐々木裕一 (2018) 『ソーシャルメディア四半世紀』, 日本経済新聞出版社.
- 佐竹秀雄 (2005) 「メール文体とそれを支えるもの」『講座社会言語科学2 メディア』, 橋元良明編, 56-68. ひつじ書房.
- Seraku, Tohru (2021) *Mi-nominalizations in Japanese Wakamono Kotoba 'youth language.'* *Pragmatics* 31: 278-302.
- 田川拓海 (2014) 「リプライするかないか: Twitterと言語(学)研究」口頭発表, 第8回言語学×自然言語処理合同勉強会, 国立情報学研究所.
- 田川拓海 (2019) 「独話に現れる愚痴命令文と反事実性」『日本語文法』19: 126-134.
- 田川拓海 (2020) 「外来語の形態論研究: 外来語系接辞と新語形成」『日本語と日本文学』, 25-38. 筑波大学日本語日本文学会.

- 田川拓海 (2021) 「TwitterやLINEはどれくらい日本語研究の対象になっているか」 口頭発表, 言語学フェス2021, オンライン (Remo) .
- 高橋快征 (2020) 「インターネット上にある新語・新表現について: Twitter・Yahoo! 知恵袋での最古例を中心に」 『言語の研究』 7: 15-32. 東京都立大学言語研究会.
- 田中ゆかり (2021) 「方言とポップカルチャー: 「方言萌えマンガ」から探る両者の関係」 『日本語学』 40: 70-82. 明治書院.
- 谷口知司 (2016) 「第4章 デジタルおよび情報記録 (取得) の基礎」 『デジタルアーカイブの資料基盤と開発技法—記録遺産学への視点—』 水嶋英治・谷口知司・逸村裕編, 71-97. 晃洋書房.
- 鳥海不二夫 (2015) 「Twitter上のビッグデータ収集と分析」 『組織科学』 48: 47-59.
- 植田麦 (2022a) 「研究資源としてのTwitter」 『明治大学教養論集』 564: 17-41. 明治大学教養論集刊行会.
- 植田麦 (2022b) 「「就活」のつぶやき」 『明治大学教養論集』 566: 1-31. 明治大学教養論集刊行会.
- 宇野和 (2015) 「Twitterにおける「新しいミ形」」 『国文』 123: 106-94. お茶の水女子大学国語国文学会.
- 宇野和 (2018) 「Twitterで見られる名詞に後接する接尾辞ミ: 「ぼさ」「らしさ」と比較して」 『人間文化創成科学論叢』 20: 29-37. お茶の水女子大学大学院人間文化創成科学研究科.
- 依田綾乃 (2016) 「ツイッターに用いられる「-み」の用法」 『信大国語教育』 26: 1-12. 信州大学国語教育学会.

たがわ たくみ/人文社会系  
(2023年12月21日受理)