

# 政治的傾向と Twitter 投稿の誤字との関係について

大月 英明 \*

## 概要

特定の政治的傾向を持つ Twitter アカウントが、誤字を含むツイートを投稿しやすいかどうかについて統計的に調査した。ここでは Twitter アカウントをリベラル群と保守群にわけ、それぞれの群が特定の誤字を含むツイートを投稿する確率を推定する。この推定には、事前分布を一様分布とした事後分布を利用するベイズ統計的手法を用いた。誤字に関しては元内閣総理大臣の安倍晋三氏を「阿部首相」とするもの、百田尚樹氏の著書である「日本国紀」を「日本国記」とするものを調査した。この研究では、どちらの誤字に関してもリベラル群より保守群が誤字を投稿する確率が高いことが示された。

## 1 はじめに

日本語の誤字の一種として同音異義語に起因するものが挙げられる。コンピュータ上での文書作成においては漢字変換機能のアシストにより、同音異義語の誤字は手書きの文書よりも発生しにくい。しかし手軽な投稿が特徴の SNS では、十分推敲されることなく誤字を含む書き込みが行われることが多いと思われる。そのため Twitter で投稿される書き込みには誤字は少なくないと推測される。

近年、政治的な傾向が異なると、脳の構造や物事の認知、作業記憶の働きなども異なるという研究結果が発表されている。Kanai 達 [1] は、保守傾向のある人は脳の扁桃体が肥大しており、リベラル傾向のある人は脳の灰白質の重量が大きいことを示した。Buechner 達 [2] は、保守傾向のある人は既存の情報と矛盾する思考を抑制し、逆にリベラル傾向のある人は新しい情報に応じて認知を更新することに優れているという結果を得た。

我々は、このような保守とリベラルの脳機能の違いが、文章作成能力の違いにも現れるのではないかという仮説を立てた。特に、短文投稿 SNS の Twitter において、政治傾向の違いが文章作成時に生じる誤字に反映される可能性について検討する。

Twitter で投稿される同音異義語には特定のパターンが見受けられる。その例としては、元内閣総理大臣の安倍晋三氏を「阿部首相」とするもの、作家の百田尚樹氏の著書である「日本国紀」を「日本国記」とするものが挙げられる。ここに挙げた2名の人物は、保守層の人々からはポジティブな評価を、リベラル層からはネガティブな評価を受けている。八木 [3] は安倍晋三元首相を次世代のリーダーと評価し、また百田尚樹氏はその著作 [4] を通じて、保守層に大きな影響力を持って

---

\* 南山大学理工学部 ソフトウェア工学科

いる。一方、長谷川 [5] に見られるように、リベラル層からは安倍晋三元首相のナショナリズムに対する批判があり、西岡 [6] は「百田尚樹氏は保守のまがいもの」と強く批判している。このように、安倍首相と百田尚樹氏は、政治に関心のある人にとっては良くも悪くも非常に馴染み深い人物である。

この研究ではまず、先に挙げた誤字を含むツイートを収集する。次に政治的にリベラル傾向を持つアカウントと、保守傾向を持つアカウントを収集する。そしてそれらのアカウントのうち、誤字を含むツイートを投稿したことのあるアカウントを抽出する。これら2群のアカウント群について、誤字を含むツイートをを行ったアカウントの割合をそれぞれ求めて標本平均とする。最後にそれぞれの群に対して、一様分布を事前分布とする事後分布を求め、これら2群に差があるかどうかを推定する。

## 2 調査と分析手法

まず特定の語句を  $\text{Word}_n$  ( $n = 1, 2, \dots$ ) とし、そのツイートが投稿された期間を  $D_n$  とする。 $D_n$  期間内に投稿された  $\text{Word}_n$  を含むツイートの集合を  $T(n)$  とし、 $T(n)$  を投稿したユニークなアカウントの集合を  $U(n)$  とする。

次に特定の誤字を  $\text{Word}_e$  とする。また政治傾向を特徴付ける語句をリベラル傾向に関しては  $\text{Word}_L$ 、保守傾向に関しては  $\text{Word}_C$  とする。すなわち  $U(L)$  はリベラル傾向群であり、 $U(C)$  は保守傾向群である。ここで  $U(L) \cap U(C) = \emptyset$  とは限らないので、新たに  $A(L) \equiv U(L) \setminus U(C)$  をリベラル群、 $A(C) \equiv U(C) \setminus U(L)$  を保守群と定義する。ただし  $\setminus$  は差集合を表す。

リベラル群が誤字  $\text{Word}_e$  を投稿する確率を  $\mu_L(e)$  とする。ツイートの調査において、リベラル群のうち誤字  $\text{Word}_e$  を投稿した割合を  $x_L(e)$  とする。集合  $S$  の要素数を  $|S|$  で表すと  $x_L(e) = \frac{|A(L) \cap U(e)|}{|A(L)|}$  である。なお  $U(e)$  は誤字  $\text{Word}_e$  を投稿したユニークなアカウントである。これをツイートから得られた標本平均とみなし、事後分布  $P_L(e) \equiv P(\mu_L(e) | x_L(e))$  を求める。保守群についても同様に定義し、事後分布  $P_C(e) \equiv P(\mu_C(e) | x_C(e))$  を求める。なおこの調査では、事前分布として一様分布  $\text{Beta}(1, 1)$  を仮定する。ただし  $\text{Beta}(m, n)$  はパラメータ  $(m, n)$  のベータ分布を表す。したがって、 $P_L(e) = \text{Beta}(1 + |A(L) \cap U(e)|, 1 + |A(L)| - |A(L) \cap U(e)|)$  であり、 $P_C(e) = \text{Beta}(1 + |A(C) \cap U(e)|, 1 + |A(C)| - |A(C) \cap U(e)|)$  である。

### 2.1 リベラル群と保守群の分類について

この論文を通して、取り扱うツイートは「公式リツイート」を除いたものである。すなわち単に他人のツイートを編集することなく投稿し直すツイートは収集から除いている。ツイートを収集した日時は2021年2月14日であり、Twitter API v2. を用いて収集した。

リベラル傾向群を特徴付ける語句として、ハッシュタグ「#安倍やめろ」と「#安倍辞めろ」を仮定する。すなわち「#安倍やめろ」または「#安倍辞めろ」というハッシュタグをつけた投稿を投稿したアカウントをリベラル傾向群とみなす。ツイートの投稿期間は2020年8月27日から2020

年 9 月 16 日までとした。この期間にこのハッシュタグで投稿されたツイート数は 48214 件であり、ツイートしたユニークなアカウント数は 15084 であった。次に保守傾向群を特徴付ける語句として、「反日」を仮定する。この語句を含む投稿を投稿したアカウントを保守傾向群とみなす。ツイートの投稿期間は 2021 年 1 月 28 日から 2021 年 2 月 13 日までとした。この期間にこの語句を含むツイート数は 47068 件であり、ツイートしたユニークなアカウント数は 17873 であった。

ただしリベラル傾向群と保守傾向群の両方に含まれるアカウント数は 400 あった。これらのアカウントを 2 群からそれぞれ取り除いたものを、リベラル群  $A(L)$  と保守群  $A(C)$  とする。すなわち  $|A(L)| = 14684$  であり、 $|A(C)| = 17473$  である。

## 2.2 分類の妥当性について

上ではリベラル傾向群と保守傾向群を、その傾向を特徴付けるキーワードを含むツイートによって分類したが、この分類の妥当性について考察する。2015 年の安保法案採決時において、Twitter 上ではその態度表明として「#安保法案反対」と「#安保法案賛成」というハッシュタグが用いられた。前者はリベラル的、後者は保守的な態度であるとするのは妥当である。これらの語句をそれぞれ  $Word_a$  と  $Word_f$  とする。 $Word_a$  に関しては期間  $D_a$ : 2015 年 7 月 12 日から 2020 年 10 月 11 日までに投稿されたツイート、 $Word_f$  に関しては期間  $D_f$ : 2015 年 9 月 4 日から 2017 年 11 月 20 日までに投稿されたツイートを収集した (表 4)。それぞれのツイート数は  $|T(a)| = 3295$ 、 $|T(f)| = 2265$  であり、 $|U(a)| = 731$ 、 $|U(f)| = 725$  であった。

リベラル群について、 $Word_a$  と  $Word_f$  を含む投稿をしたアカウント数はそれぞれ  $|A(L) \cap U(a)| = 42$ 、 $|A(L) \cap U(f)| = 5$  であった。また保守群について、 $Word_a$  と  $Word_f$  を含む投稿をしたアカウント数はそれぞれ  $|A(C) \cap U(a)| = 15$ 、 $|A(C) \cap U(f)| = 48$  であった。それぞれの群に関して、その政治傾向を反映するアカウントの方が 3~8 倍程度多いことから、この分類が十分妥当であると判断できる。

## 3 調査と分析結果

### 3.1 安倍晋三氏を「阿部首相」とする誤字について

語句「阿部首相」を  $Word_1$  とする。 $D_1$ : 2015 年 2 月 2 日から 2021 年 2 月 14 日までに投稿されたこの語句を含むツイートを収集したところ、 $|T(1)| = 17962$  件であり、ユニークなアカウント数は  $|U(1)| = 11198$  であった。 $U(1)$  のうちリベラル群のアカウント数は 134 であり、保守群のアカウント数は 292 であった。すなわち  $x_L(1) = 134/|A(L)|$ 、 $x_C(1) = 292/|A(C)|$  である (表 1)。

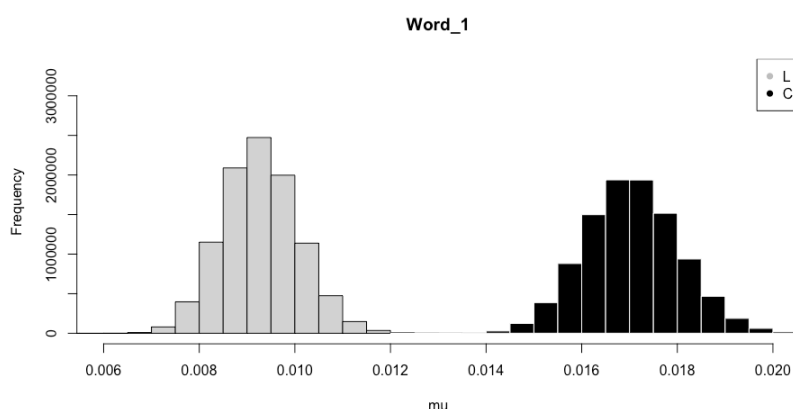
リベラル群と保守群について、誤字を含むツイートを行うアカウントの比率の確率分布を求める。ベイズ統計の手法を用い、一様分布を事前分布とした事後分布を求める。この場合、 $P_L(1) = \text{Beta}(135, 14551)$  である。保守群についても同様に  $P_C(1) = \text{Beta}(293, 17192)$  である。図 1 は事前分布として 10000000 個の一様分布に従う乱数を発生させたときのそれぞれの事後分

表 1 誤字ツイートを投稿したアカウント数

政治傾向	全数	「阿部首相」	「日本国記」
リベラル群 $A(L)$	14684	134(0.91%)	101(0.69%)
保守群 $A(C)$	17483	292(1.67%)	199(1.14%)

布のヒストグラムである。この結果、 $Word_1$  の誤字ツイートを投稿する確率に関して、リベラル群が保守群を上回る確率はほぼゼロであることがわかる。

図 1 誤字「阿部首相」を投稿する確率分布



### 3.2 「日本国紀」を「日本国記」とする誤字について

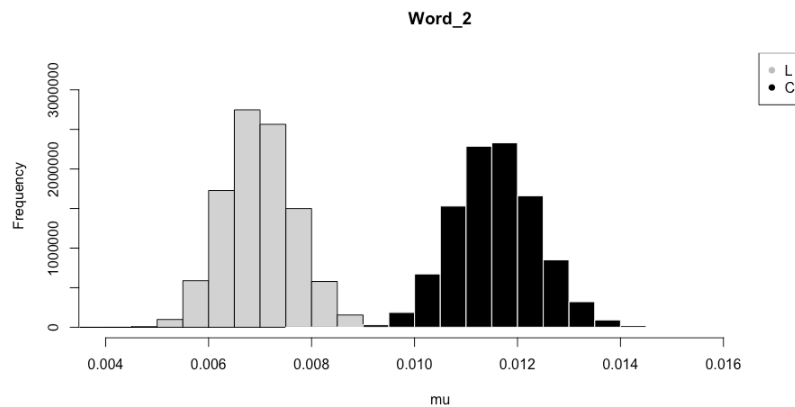
語句「日本国記」を  $Word_2$  とする。  $D_2$  : 2018 年 11 月 1 日から 2019 年 1 月 1 日までに投稿されたこの語句を含むツイートを収集したところ、 $|T(2)| = 3557$  件であり、ユニークなアカウント数は  $|U(2)| = 2621$  であった。このうちリベラル群のアカウント数は 101 であり、保守群のアカウント数は 199 であった。すなわち  $x_L(2) = 101/|A(L)|$  ,  $x_C(2) = 199/|A(C)|$  である (表 1)。

$Word_2$  についても  $Word_1$  と同様の分析を行う。この場合  $P_L(2) = \text{Beta}(102, 14584)$  であり、 $P_C(2) = \text{Beta}(200, 17275)$  である。図 2 は事前分布として 10000000 個の一様分布に従う乱数を発生させたときのそれぞれの事後分布のヒストグラムである。この結果、 $Word_2$  の誤字ツイートを投稿する確率に関して、リベラル群が保守群を上回る確率は 0.001% 以下であることがわかる。

## 4 議論

ここでは誤字を含んだ Twitter の投稿について、政治的な傾向と関係があるかについて 2 つの誤字を選んで調査した。一見すると、ポジティブな対象の表記に関しては、ネガティブな対象よりも誤字が少なくなるという仮説は自然に思われる。しかしこの調査では、保守的傾向がある人に

図 2 誤字「日本国記」を投稿する確率分布



としては親しみ深いと思われる語句に関しても、誤字を投稿する確率がリベラル傾向がある人よりも高いことが示された。これは政治的な傾向が異なる場合、認知機能や作業記憶の働きも異なっている可能性を示唆するとも考えられる。

ここではツイート投稿期間として、それぞれの特徴を持つツイート数およびアカウント数を十分に確保するための期間を選んでいる（表 2, 3）。本研究では統計的解析に十分な数のサンプルを確保しているが、ツイート投稿期間や、誤字として選ぶ語句を変えることによって、本研究と異なる結果が得られる可能性については排除できない。

表 2 ツイート投稿期間（政治傾向）

安倍やめろ OR 安倍辞めろ $D_L$	反日 $D_C$
2020/8/27-2020/9/16	2021/1/28-2021/2/13

表 3 ツイート投稿期間（誤字）

阿部首相 $D_1$	日本国記 $D_2$
2015/2/2-2021/2/14	2018/11/1-2019/1/1

表 4 ツイート投稿期間（安保）

安保反対 $D_a$	安保賛成 $D_f$
2015/7/12-2020/10/11	2015/9/7-2017/11/20

## 参考文献

- [1] Kanai R, Feilden T, Firth C, and Rees G. Political orientations are correlated with brain structure in young adults. *Curr Biol.*, Vol. 21, No. 8, pp. 677–680, 2011.
- [2] Bryan M. Buechner, Joshua J. Clarkson, Ashley S. Otto, Edward R. Hirt, and M. Cony Ho. Political ideology and executive functioning: The effect of conservatism and liberalism on cognitive flexibility and working memory performance. *Social Psychological and Personality Science*, Vol. 12, No. 2, pp. 237–247, 2021.
- [3] 八木秀次. { 保守主義革命 } の象徴たれ (特集安倍晋三総理待望論) – (安倍幹事長に期待する–若い世代のリーダーとして「強い日本」をつくれるか). PHP 研究所, 2003.
- [4] 百田尚樹. 百田尚樹、憂国の大演説 奇跡の国・日本が危ない! 天皇と憲法改正. 飛鳥新社, 2020.
- [5] 長谷川裕. 安倍晋三における保守主義・ナショナリズムと教育 (特集 「安倍教育改革」批判). かもがわ出版, 2013.
- [6] 西岡研介. 保守のまがいもの・百田尚樹 (櫻井よしこ・百田尚樹 大批判). K&K プレス, 2016.