

Writing Style Changes by Disguising an Author's Gender

Wataru ZAITSU*, Mingzhe JIN**

(Received July 20, 2018)

This study examined writing style changes by conducting an experiment. A total of 48 participants (24 male and 24 female, ranging in age from their 20s to their 40s) rewrote blog posts with the goal of disguising the author's gender. Results indicated that when male participants disguised writing as that of a female author, the uses of "Kanji", "particle (pronominal)", and the characters "「僕」" and "「俺」" were decreased, while the uses of "Hiragana", "a verb (bound)", and the character "「私」" were increased. When female participants disguised their writing as that of a male, the use of "Kanji" was increased and the use of the character "「私」" was decreased. The rates of "parts-of-speech", "Katakana", "comma", small character of "Hiragana", "a sentence-final particle" and "character just before comma" were not significantly changed in the process of disguising an author's gender.

Key words : gender disguise, blogs, writing styles, chi-square tests, residual analysis

キーワード : 性別偽装, ブログ, 文体的特徴, χ^2 検定, 残差分析

性別を偽装した文章における文体的特徴の変化

財津亘, 金明哲

1. はじめに

1.1 性別を偽装した事件と犯罪捜査への影響

犯罪を敢行する際に、その過程で性別や年齢層を偽る者がいる。たとえば、児童買春にかかる犯罪では、成人男性が未成年の女子児童を誘い出す目的で、未成年の男子児童を偽って SNS や電子メールなどでやりとりをするといった事例が多くみられる。犯人が性別を偽装した有名な事件としては、1988 年から 1989 年に発生の「東京・埼玉連続幼女誘拐殺人事件」が挙げられる。この事件は、複数の女兒が誘拐された上に殺害されるといった痛ましい事件であったが、犯人は事件後に「今田勇子」という女性の偽名を使い、朝日新聞東京本社などに犯行声明文や告白文を郵送

している。その内容も、「子どもを亡くした母親」の立場として書かれたものであった。実際には、犯人の性別が男性であったにも関わらず、それらの文章が「情緒的」で、「陰湿」、「文章全体が長い」、「ネチネチしている」などの「女性らしさ」が見受けられるといった根拠を基に、メディアにおいて様々な識者が女性であると誤って性別を推定したとされている¹⁾。

このような性別偽装は、犯罪捜査をかく乱する目的で、意図的に行われることが多い。そのため、実際の犯罪捜査に影響を及ぼすこともありえる。そもそも、特定の被疑者が浮上していない上に、印字文書などしか存在しない場合、犯罪捜査は困難となる。そこで、印字文書などに記載の文章情報から著者の

* Toyama Prefectural Police Headquarters, Toyama E-mail: wataru0112csi@yahoo.co.jp

** Faculty of Culture and Information Science, Doshisha University, Kyoto E-mail: mjjin@mail.doshisha.ac.jp

特徴（性別や年齢層など）を推定する技術「著者プロファイリング」の登場となる。従来は、上記のマスメディアにおける識者の例のように、主観的に性別などを推定せざるを得なかったが、近年は文章の文体的特徴を機械学習などによって客観的に分析する技術が研究されている。我が国では、実際の犯罪捜査現場における応用例はほとんどみられないものの、今後はその技術が期待される。財津・金²⁾は、犯罪捜査に資する目的で、文章情報からの性別推定を目的として、ブログの文章を材料に用いて推定精度の検証を行っている。その研究では、100名のブログの文章をサンプル（1,000文字程度）として、まず男女間に違いがみられる文体的特徴を検討している。その結果によると、漢字やひらがな、カタカナ、名詞の使用率に加えて、品詞（動詞・形容詞・助詞・感動詞）や接続助詞「し」、助動詞「なかつ」、読点、文字（一人称代名詞「私」「僕」、小書き文字「っ」「ゃ」）の使用頻度といった文体的特徴に男女差がみられたとされている。この研究では続いて、これらの文体的特徴を実装した機械学習（ランダムフォレストとサポートベクターマシン）によって、新たな100名のブログの文章を用いて、著者の性別推定に関する精度を検証している。報告によれば、サポートベクターマシンに比べると、ランダムフォレストの推定成績の方が高く、最高で正解率86.0%（適合率：男性84.6%、女性87.5%）が得られたとされている。この研究では、ブログをサンプルとして用いているため、比較的自然的な状態で文章が記載されたものと推察される。しかし、財津・金²⁾でも述べられているとおり、犯罪現場では、犯罪の証拠を残さないようにするために、文章の表現方法などを作為的に変える可能性がある³⁾。常に犯人が文章を意図的に変容させるとは限らないが、性別を偽ることにともなう変化する文体的特徴に着目しては、性別を正確に推定することができなくなる可能性がある。そのため、財津・金²⁾で性別を推定する上で有効とされた文体的特徴の中でも、どの文体的特徴が変化しやすいのか把握する必要がある。

そこで、本研究は、意図的に文章表現を異性のものに偽装する場合に、どの文体的特徴に着目して変

化させるものなのかに関して実験による検証を行った。

1.2 性別偽装と文体的特徴の変化に関する先行研究

管見の限り、性別を偽装した文章を検証した研究は少ない。

菊池⁴⁾は、男女のデュエット曲「三年目の浮気（作詞・作曲：佐々木勉，歌：ヒロシ&キーボー）」の歌詞を題材に、男女のパートを交換して書き換えるといった課題を実施している。この報告によれば、書き換えられた特徴として、①人称代名詞、②終助詞、③命令・依頼表現、④その他の4つが見受けられたという。たとえば、①を例に挙げると、男性から女性の場合は「俺→私」、「お前→あなた」、女性から男性の場合は「あなた→お前」、「私→俺」といった書き換えが観察されている。②の終助詞については、男性から女性の場合で「よ→わ」、「ぜ→わ」など、女性から男性の場合で「わ→よ」、「だわ→だぜ」などの書き換えが多くみられている。③命令・依頼表現では、「みろよ（男性）→みてよ（女性）」、④その他に、「許してあげない（女性）→許してやらない（男性）」などがみられたという。このような書き換えによる変化が見受けられたものの、菊池⁴⁾については、①サンプル数が13名と少ないこと、②13名中11名が女子学生であったため、実験参加者の性別が偏っていること、③サンプルの年齢層も大学生に偏っていること、④性別偽装に用いた文章はこの歌詞一つのみであり、分析結果がこの歌詞に依存している可能性があること、⑤仮説検定などの統計手法による検証を行っていないといった問題点が挙げられる。

この他の性別偽装研究として、長谷川⁵⁾は、架空の人物に手紙を書く、もしくは同様の手紙を異性に偽装して書くといった2つの課題を行った後に、男性や女性の文章に対するイメージを自由に記述させるといった課題を行っている。報告によると、男性の文に関するイメージとして、①要件のみ、②事務的でかたい、③心配りがないといったものが見受けられたのに対して、女性の文に関するイメージは、①付加的な情報がある、②やわらかい印象、③心配りのあるといったものが挙げられている。性別を偽

装する際には、これらのイメージ（いわゆるジェンダーステレオタイプ）が参照されるために、男性／女性らしさへ偏る傾向があるという。つまり、性別を偽装すると、結果として過度に男性／女性らしさを演出してしまうことから、読み手により男性／女性らしい印象を与える文章を作成する傾向にあることが言及されている。

さらに、長谷川⁹⁾は、本研究の目的と同様に、偽装した文章における文体的特徴（読点の打ち方、漢字の使用率、一人称代名詞の使用頻度）の変化を検討している。実験参加者は32名（男性15名、女性17名）で、平均年齢は26.7歳（レンジ20から35歳）であった。課題は、自分の持っているCDを欲しがっている人に対して、それを譲ってもいいという趣旨の手紙を、①普通に書く場合と、②性別を偽装して書く場合について実験参加者内要因による検討を行っている。読点の打ち方、漢字の使用率、一人称代名詞の使用頻度のそれぞれについて分散分析を行ったところ、主効果・交互作用ともに有意差はみられなかったとしている。菊池⁴⁾で指摘された人称代名詞の変化は、長谷川⁹⁾では有意差はみられなかった。また、菊池⁴⁾で検討されていた終助詞に関しては、長谷川⁹⁾において検討していないようである。長谷川⁹⁾で扱った文字数をみると、85から125文字程度とかなり少なく、有意差がなかったのは文字数が少ないことによる検定力（有意差を正確に検出できる力）の低さが原因とも考えられる。通常の手紙を書くといった自然な状況と比べると、上記のような状況を想定して、白紙に文章を記載させるといった課題が実験参加者にとって困難であったために、記載された文字数が少なかった可能性がある。

以上の先行研究を踏まえ、本研究では、性別や年齢層を統制すべく、20代から40代の男女で構成した6グループを設定し、それらのグループにおける実験参加者の数を等しく設定するとともに、菊池⁴⁾や長谷川⁹⁾に比べて、より多くの実験参加者のサンプルを取得して実験を行った。また、文字数の少なさの影響を避けるために、実験参加者と同じ性別年齢層の人物が書いた500文字程度の原文を用いて、性別を偽装してその原文の文章表現を書き換えると

いった課題を行い、原文と偽装文の文体的特徴の変化を検討した。さらに、菊池⁴⁾では、一つの歌詞のみを性別偽装に用いており、分析結果が歌詞内容に依存していた可能性がある。そこで、本研究では48名すべての実験参加者にそれぞれ異なる文章を用いた。分析対象とする文体的特徴については、性別推定に有効とされた財津・金²⁾における文体的特徴に加えて、菊池⁴⁾で指摘された終助詞などを検討した。また長谷川⁹⁾では、読点の打ち方（例、読点前の文字「は」など）も検討していることから、本研究においても読点前の文字「は」「が」などの使用率について原文から偽装文への変化を検討した。

2. 方法

2.1 実験参加者

警察職員48名（男性24名、女性24名、平均値34.7歳（標準偏差8.30）、中央値33.0歳、レンジ20-49歳）が実験に参加した。実験に先立って、すべての実験参加者には、本実験データを研究目的で利用することを説明し、同意書に署名することで同意を得た。

2.2 原文サンプル

原文のサンプルには、20代から40代の男女を自称する48名のブログの文章を用いた。ブログは、インターネットサイト「にほんブログ村（<http://diary.blogmura.com/>）」から選定したもので、ブログの選定方法として、性別および年齢層の影響を統制するために、3つの年齢層（20代から40代）×性別（男女）の6グループを設定し、1グループにつき8名のブログの文章を無作為に抽出した。文字数は、500文字を基準として、500文字以降最初の文末までの文章を用いた。サンプルは、テキストファイルに変換して用いた。

2.3 実験手続き

実験は、実験参加者が保有するパソコンを使用して行った。電子メールを介して原文のテキストファイルを送信し、実験参加者には「実験参加者と同年代でかつ同じ性別の人物が書いた文章が添付されているので、読み手が異性の文章であると誤って判断するように、つまり男性実験参加者は女性に（女性実

験参加者は男性に), その文章の表現や内容を書き換えて下さい」といった教示を行い, 課題に取り組みさせた. 所要時間はおよそ5分程度と教示したが, 文章の書き換えに要した時間は実験参加者の任意としたことから, 各実験参加者で異なっている. 課題終了後, 偽装した文章のテキストファイルを実験者に電子メールで送信するよう求めるとともに, 文章書き換え時に想定した異性の年齢および課題実施に際しての困難度を10段階(1:非常に簡単~5:どちらでもない~10:非常に困難)で評価させた.

2.4 分析手続き

文字数の増減を確認するために, 文字数を従属変数とした, 2(性別[男性, 女性], between 要因) × 2(文章種別[原文, 偽装文], within 要因)の2要因分散分析を実施した.

次に, 先行研究^{24,6)}を参考に, 以下の文体的特徴における性別偽装の影響を検討した. そこで, 原文と偽装文における文体的特徴の使用率の差を算出し, 男性実験参加者において増加(ないし減少)した人数と女性実験参加者において増加(ないし減少)した人数に関して χ^2 検定ならびに残差分析を行った. なお, 文字数について確認したところ, 後述の「結果」のとおり, 有意差がみられたことから原文と偽装文の文字数の違いを考慮し, 本研究では文章内における出現頻度ではなく, 文章内における割合(使用率)を分析対象とした.

- ・ 漢字, ひらがな, カタカナの使用率
- ・ 品詞の使用率: 名詞, 動詞(自立), 動詞(非自立), 形容詞(自立), 助詞(接続助詞), 助詞(連体化), 感動詞
- ・ 読点の使用率
- ・ 小書き文字の使用率:
「っ」, 「ゃ」
- ・ 一人称代名詞の使用率:
「私」, 「僕」, 「俺」
- ・ 終助詞の使用率:
「か」, 「ね」, 「よ」, 「な」, 「なあ」, 「わ」, 「の」
- ・ 読点の打ち方, 読点前の文字の使用率:
「は」, 「が」, 「で」, 「て」

3. 結果

3.1 各文章内の文字数

原文は500文字を基準としたものであるが, 性別を偽装することによって文章内の文字数に変化がみられたか否か確認した. 男性実験参加者の原文は平均521.4文字, 偽装文は平均519.1文字であったことからあまり違いがみられなかったものの, 女性実験参加者の原文が平均516.5文字であったのに対し, 偽装文で472.0文字と違いがみられた.

そこで, 各文章の文字数について, 2(性別[男性, 女性], between 要因) × 2(文章種別[原文, 偽装文], within 要因)の2要因分散分析を実施したところ, 5%水準で性別および文章種別の主効果のみ有意差がみられた($F(1, 46) = 5.05, p < .05, \eta_p^2 = .10$; $F(1, 46) = 4.48, p < .05, \eta_p^2 = .09$). このように, 偽装にともない文字数が有意に減少するといった現象がみられたことから, 以降の文体的特徴に関する分析では, 文字数の多寡による影響がないように文章内の割合(使用率)に着目することとした.

3.2 漢字, ひらがな, カタカナの使用率

Fig. 1に, 漢字の使用率の変化を示す.

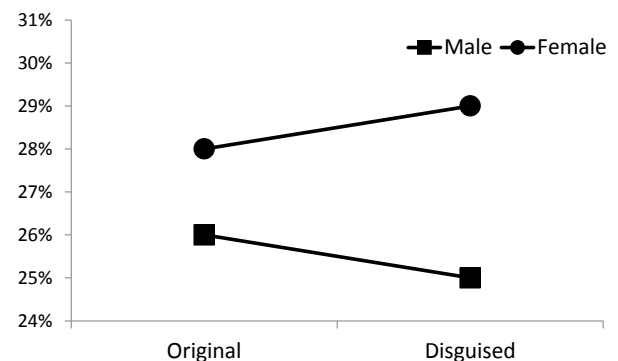


Fig. 1. Change rate of the uses of “Kanji”.

先行研究²⁷⁾では, 女性の文章に比べて, 男性の文章で漢字の使用率が高いとされているものの, Fig. 1の原文における漢字の使用率は, 女性の方で高いように見える. この点については有意差がみられなかった. 女性に偽装した男性実験参加者の66.7%(24名中16名)において, 漢字の使用率が減少したのに対して, 男性に偽装した女性実験参加者の66.7%(24名中16名)で漢字の使用率が増加した. これに関し

て χ^2 検定を実施したところ、有意差がみられるとともに ($\chi^2(1) = 4.08, p < .05, \phi = .29$)、続く残差分析において、男性実験参加者が女性に偽装することで漢字の使用率が有意に減少する一方で ($p < .05$)、女性実験参加者が男性に偽装する際には漢字の使用率が有意に増加することを示唆した ($p < .05$)。

Fig. 2 は、ひらがなの使用率に関する変化を示している。偽装にともなう変化として、66.7% (24 名中 16 名) の男性実験参加者でひらがなの使用率が増加し、75.0% (24 名中 18 名) の女性実験参加者においてひらがなの使用率が減少した。 χ^2 検定を実施したところ、有意差がみられた ($\chi^2(1) = 6.80, p < .01, \phi = .38$)。残差分析を実施した結果、漢字の使用率とは逆の方向、つまりは男性実験参加者が女性として偽装する場合にひらがなの使用率が有意に増加し ($p < .01$)、女性実験参加者が男性として偽装する場合にひらがなの使用率が有意に減少することがわかった ($p < .01$)。

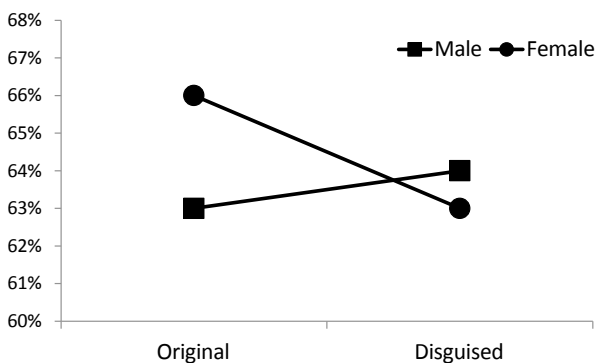


Fig. 2. Change rate of the uses of "Hiragana".

カタカナの使用率 (男性実験参加者の原文 7.8%、偽装文 7.8%、女性実験参加者の原文 6.1%、偽装文 6.2%) に関しては、 χ^2 検定において有意差はみられなかったため、性別偽装による影響がみられなかったといえる。

3.3 品詞の使用率

動詞 (非自立) の使用率に関する変化を Fig. 3 に示す。動詞 (非自立) とは、具体的には「(して) いる」、「(行って) もらう」、「(終わって) い(ない)」などが該当する。性別の偽装によって、動詞 (非自立) の使用率が、男性実験参加者の 70.8% (24 名中

17 名) で増加し、残り 7 名で減少した。他方、女性実験参加者では、70.8% (24 名中 17 名) で減少し、6 名が増加した (1 名のみ変化なし)。 χ^2 検定を実施した結果によると、これらの人数に有意差がみられた ($\chi^2(2) = 10.4, p < .01, \text{Cramer's } V = .47$)。残差分析によると、性別を偽装することにより、男性実験参加者では動詞 (非自立) の使用率が有意に増加し ($p < .01$)、女性実験参加者では有意に減少することを示唆した ($p < .01$)。

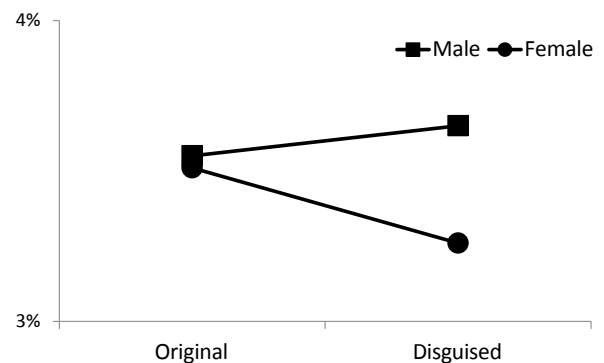


Fig. 3. Change rate of the uses of "a verb (bound)".

助詞 (連体化) の使用率に関する変化を Fig. 4 に示す。

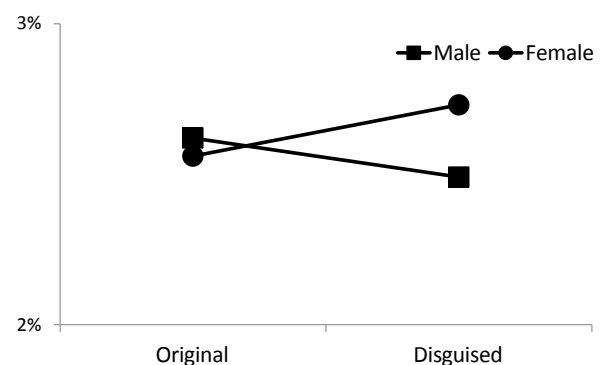


Fig. 4. Change rate of the uses of "particle (pronominal)".

助詞 (連体化) の例を挙げると、「(今回) の」、「(そのため) の」などが該当する。性別を偽装することで、助詞 (連体化) の使用率が、男性実験参加者の 66.7% (24 名中 16 名) で減少し、29.2% (24 名中 7 名) で増加した (1 名のみ変化なし)。女性実験参加者については、助詞 (連体化) の使用率が 66.7% (24 名中

16名)で増加し,残り8名で減少した. χ^2 検定を実施したところ,有意差がみられた($\chi^2(2) = 7.19, p < .05, \text{Cramer's } V = .39$).残差分析の結果によると,性別を偽装することにより,男性実験参加者では助詞(連体化)の使用率が有意に減少し($p < .05$),女性実験参加者では有意に増加することを示唆した($p < .01$).

その他,名詞(自立),形容詞(自立),助詞(接続助詞),感動詞の使用率については χ^2 検定の結果で有意差がみられず,性別偽装による影響はみられなかった.

3.4 読点の使用率

読点の使用率は,文章内の全文字数における読点の割合を意味する.この文体的特徴に関して, χ^2 検定を行ったものの,有意差はみられず,性別偽装による読点の使用率への影響は確認できなかった.

3.5 小書き文字の使用率

小書き文字「っ」および「ゃ」についても, χ^2 検定による有意差はみられなかったことから,性別の偽装による影響は見受けられなかった.

3.6 一人称代名詞の使用率

「私」の使用率に関する変化をFig. 5に示す.58.3%(24名中14名)の男性実験参加者が女性に偽装することで,「私」の使用率が増加したのに対して,1名のみ減少した(9名は変化なし).54.2%(24名中13名)の女性実験参加者は,男性に偽装する際に「私」の使用率が減少し,3名のみは増加した(8名は変化なし).これら使用率の増減した実験参加者の割合には,有意差がみられた($\chi^2(2) = 17.5, p < .01, \text{Cramer's } V = .60$).残差分析を実施したところ,性別を偽装することによって,男性実験参加者では「私」の使用率が有意に増加し($p < .01$),女性実験参加者では有意に減少することを示した($p < .01$).

「僕」の使用率に関する変化はFig. 6のとおりである.男性実験参加者の24名中7名(29.2%)が女性に偽る場合に,「僕」の使用率が減少しており,残りの17名においては増減がみられなかった.逆に,女性実験参加者の24名中3名(12.5%)において増加する傾向がみられ,残りの21名については増減がみられなかった.これらの実験参加者における増減の

割合に関して, χ^2 検定を実施したところ,有意差がみられた($\chi^2(2) = 10.4, p < .01, \text{Cramer's } V = .47$).残差分析によると,性別を偽装することにより,男性実験参加者では「僕」の使用率が有意に減少し($p < .01$),女性実験参加者については「僕」の使用率の増加に関して有意傾向がみられた($p < .10$).

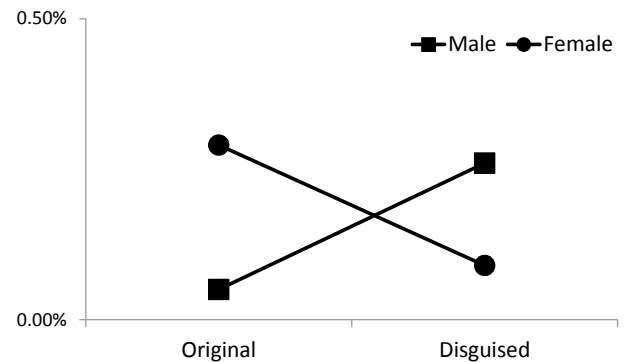


Fig. 5. Change rate of the uses of “the character 「私」”.

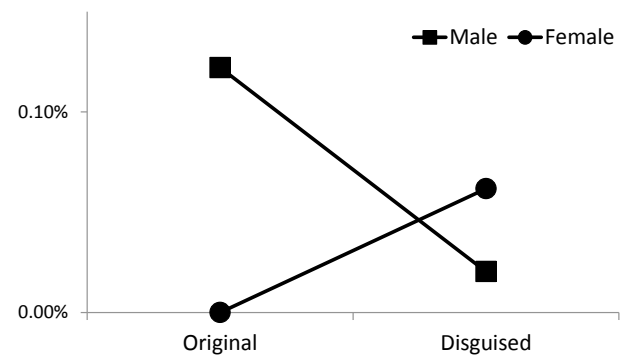


Fig. 6. Change rate of the uses of “the character 「僕」”.

Fig. 7は,「俺」の使用率の偽装による変化を示している.男性実験参加者(24名)の内2名において,女性に偽装した結果として「僕」の使用率が減少したものの,残りの22名については変化がなかった.一方,女性実験参加者の29.2%(24名中7名)において,「俺」の使用率が増加する傾向がみられた一方で,残りの17名には増減がみられなかった.これに関しては, χ^2 検定において有意差がみられたことに加えて($\chi^2(2) = 9.64, p < .01, \text{Cramer's } V = .45$),残差分析で,女性実験参加者において有意に増加することが示唆された($p < .01$).男性実験参加者における「俺」の使用率の減少については,残差分析による有

意差はみられなかった。

一人称代名詞「私」の使用率に関しては、 χ^2 検定において有意差がみられた上に、効果量の値が非常に高かったことから、性別を偽装する際に積極的に書き換えられる特徴であることが示された。同様に、「僕」の使用率も男女ともに性別偽装の影響を受けやすい特徴であることが示唆された。

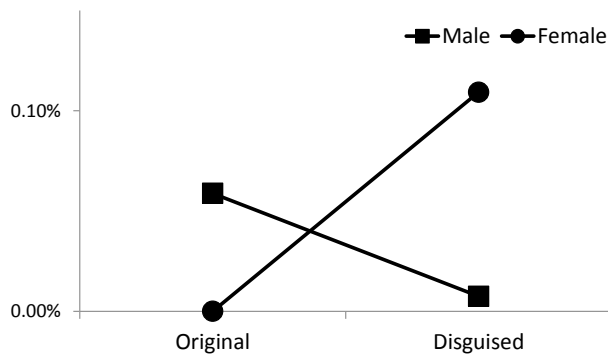


Fig.7. Change rate of the uses of “the character 「俺」”.

3.7 終助詞の使用率

本研究で扱った文章（原文・偽装文）内には、終助詞「か」、「ね」、「よ」、「な」、「なあ」、「わ」、「の」がみられた。これらの終助詞の使用率に関して、原文から偽装文への変化を検討したものの、 χ^2 検定において有意差はみられなかった。したがって、これらの終助詞に対する性別偽装の影響はあまりないものと推察された。

3.8 読点前の文字（読点の打ち方）の使用率

χ^2 検定を実施したところ、有意差がみられた読点前の文字の使用率は皆無であった。したがって、読点の打ち方は、性別偽装による影響を受けない可能性がある。

3.9 想定した異性の年齢層および課題の困難度

性別偽装時に想定した異性の年代については、実験参加者48名中46名で、同年代あるいは10代上（ないし下）を想定したと回答した。2名については、40代男性実験参加者1名が10代女性を想定し、20代女性実験参加者が50代男性を想定したと回答を得た。

性別偽装の課題に対する困難度（10段階評価）に

関しては、男性実験参加者で平均6.92（標準偏差2.00）、女性実験参加者で平均6.17（標準偏差1.95）と、多少課題が難しかったと感じる実験参加者が多かった。男女間で困難度評価に有意差はみられなかった（ $t(46) = 1.32, p = .20$ ）。

4. 考察

本研究では、ブログから抽出した文章を原文とし、読み手が異性の文章であると誤って認識するようにその原文を書き換えるといった実験を行い、性別偽装によって変化する文体的特徴を検討した。その結果によると、男性実験参加者が女性に偽って文章を書き換えることで、「漢字」や「助詞（連体化）」、一人称代名詞「僕」の使用率が減少する一方で、「ひらがな」や「動詞（非自立）」、一人称代名詞「私」の使用率が増加するといった傾向がみられた。それとは反対に、女性実験参加者が男性として偽った場合は、「漢字」の使用率が増加するなど、男性実験参加者と反対の変化がみられた。また、動詞（非自立）や助詞（連体化）を除く「品詞」や「カタカナ」、「読点」、「小書き文字」、「終助詞」、「読点前の文字」の使用率については、性別偽装による有意な変化はみられなかった。

効果量 ϕ 係数やCramer's V には、その評価基準として、「小（.10）」、「中（.30）」、「大（.50）」といったものがある⁸⁾。それらの評価基準に照らし合わせると、「漢字」の使用率は「小」に該当し、「ひらがな」や「動詞（非自立）」、「助詞（連体化）」の使用率については、「中」程度の効果量であった。それら文体的特徴に比べて、一人称代名詞「私」「僕」「俺」の使用率に関しては、「大」程度の効果量に相当し、非常に高い効果量を示した。このことから、一人称代名詞は、性別偽装の際に積極的に書き換えられる特徴であることがわかった。一人称代名詞に関する結果は、菊池²⁾と類似の結果といえる。また、「漢字」や「ひらがな」の使用率が性別偽装の影響を受けるといった新たな知見も得られた。長谷川⁵⁾によると、男性の文章に対するイメージとして「事務的でかたい」などが挙げられ、女性の文章については「やわらかい印象」といったものがあった。確かに、全体的に「漢字」の割合が多ければ「かたい」イメージとなり、「ひらが

な」の割合が多ければ「やわらかい」イメージとなりそうである。したがって、実験参加者の意識的か無意識的かは不明であるが、性別を偽装する際に、このようなジェンダーステレオタイプが影響して、「漢字」や「ひらがな」の使用率が変化したことが考えられる。さらに、長谷川⁵⁾は、性別を偽装する場合に、ジェンダーステレオタイプを参照することで、過度に男性や女性らしさを演出する可能性があると言及していた。このことを示すように、Fig. 7では、女性実験参加者における「俺」の使用率が偽装条件において増加したことに加えて、男性実験参加者の原文よりも高い使用率を示している。この他の、「動詞（非自立）」や「助詞（連体化）」の使用率の変化については、先行研究においても指摘されておらず、今後も検討する余地があろう。

菊池⁴⁾では終助詞の変化が指摘されていたことから、本研究においてもいくつかの終助詞の使用率について検討した。しかしながら、本研究ではどの終助詞においても有意差がみられなかった。そもそも、この先行研究では、統計的仮説検定を実施していなかったため、有意な変化があったかは不明である。たとえ、有意な変化がみられたとしても、その終助詞の変化は、男女のデュエット曲「三年目の浮気」のみを題材とした研究から得られた結果であり、その歌詞に依存していた可能性がある。この曲における歌詞の特徴として、男性パートで「よ」、「ぜ」など、女性パートで「わ」、「ね」などが多用されており、男性らしさ女性らしさが過度に表現されている。このような終助詞が多用されていたため、性別偽装の際に異性らしい終助詞に書き換えやすかったとも考えられる。これに関連して、菊池⁴⁾は、女性特有の終助詞とされる「かしら」を、日常会話の中で耳にすることはほとんどないなどとも言及している。したがって、本研究で用いたブログの文章には、男性や女性特有の終助詞が元々あまりなかったことで、原文から偽装文への変化がみられなかったものと推察される。

「漢字」や「ひらがな」、「動詞（非自立）」、「助詞（連体化）」、一人称代名詞「私」「僕」の使用率が性別偽装によって影響を受けたという事実は、財津・金²⁾の

性別推定の精度に対しても影響があることが予想される。これらの文体的特徴が変化するのであれば、最高で正解率 86%であった推定精度も下がる可能性がある。一方で、本研究では、「動詞（非自立）」や「助詞（連体化）」を除く「品詞」の使用率や「カタカナ」の使用率、「読点」の使用率、小書き文字「っ」「ゃ」の使用率に関する性別偽装の影響はみられなかった。このことから、偽装が疑われる場合には、これらの偽装の影響を受けない（かつ先行研究で性差が確認済みの）文体的特徴に着目して性別を推定する必要があるかもしれない。

本研究では、ブログに書かれた文章を原文とし、その原文を書き換えるといった課題を実施したものであるが、今後は長谷川^{5,6)}のように、白紙欄に対して自由記述を求めるなど課題も検討する必要がある。ただし、その場合は、ある程度の文字数を確保するために、課題を工夫することが求められる。

参考文献

- 1) 武田春子, “言語性差のステレオタイプ—「今田勇子」への「識者」のコメントを読む—”, 女性学年報, **11**, 28-39(1990).
- 2) 財津亘, 金明哲, “ランダムフォレストによる著者の性別推定—犯罪者プロファイリング実現に向けた検討—”, 情報知識学会誌, **27**[3], 261-274 (2017).
- 3) P. Juola, “Authorship attribution”, *Foundations and Trends in Information Retrieval*, **1**[3], 233-334 (2006).
- 4) 菊池悟, “現代大学生における言語の性差意識—男女の歌詞書き替えの課題から—”, 岩大語文, **12**, 109-101 (2007).
- 5) 長谷川直宏, “擬装文書の検査法—ジェンダーステレオタイプに注目して—”, 日本法科学技術学会誌, **11**[別冊号], 178 (2006).
- 6) 長谷川直宏, “擬装文書の検査法 2—文章計量学的アプローチによる検証—”, 日本法科学技術学会誌, **13**[別冊号], 188 (2008).
- 7) 島崎洵子, “新聞投書の文体分析—性差を中心に—”, 武庫川女子大学言語文化研究所年報, **19**, 5-35 (2007).
- 8) 水本篤, 竹内理, “研究論文における効果量の報告のために—基礎的概念と注意点—”, 英語教育研究, **31**, 57-66 (2008).