

文章表現における男女の特徴の差

齋藤敦美[†][†] 筑波大学中山伸一[‡][‡] 筑波大学真栄城哲也[‡][‡] 筑波大学

図書館情報専門学群 図書館情報メディア研究科 図書館情報メディア研究科

1 はじめに

文章表現における男女の特徴を、計量情報学の手法を用いて調べた結果について報告する。文章の定量的な特徴を調べる様々な手法が発表されている [1]。これまでの研究から、執筆者が無意識に書く部分に個人のクセや特徴が出現する可能性が高いことが判っている。例えば、接続詞や助詞の使用頻度、また読点の付け方に個人の特徴が出やすい。本研究は、個人差ではなく、性差に焦点を当て、計量的な手法を用いて男女の性差について解析する。一方、文章についての心理的な側面についても様々な研究があり、例えば [2, 3, 4] がある。本研究では、被験者の考え方についてアンケート調査を実施し、計量的な手法と心理的な側面の両者から捉えることを目的としている。

文章の計量的な解析、特に書き手のクセの分析は真贋判定に関連しており、単語の長さ、文の長さ、特定の単語の出現頻度、接頭辞の出現頻度、読点の位置等が指標として用いられている [1]。ここでは、基本的な指標を使用した予備解析結果について述べる。

2 方法

男女それぞれ 20 人 (合計 40 人) の被験者 (大学生) に同じテーマについて、1 時間の制限時間内に 800 文字の文章を手書きで書いてもらう。ワープロではなく手書きとしたのは、その方がしっかりした文章を書くと考えられるからである。また、被験者にきっちり 800 文字書いてもらうことは過度な要求であり、書き方のクセの出現に影響を及ぼす可能性があるため、文字数を厳密に 800 文字に固定するのではなく、800 文字前後とした。文字数に関わる計測量の比較については後で正規化するために問題は生じない。

これらの手書きの文章をワープロで入力して電子ファ

イル化した後、形態素解析プログラム ChaSen* で形態素解析を行う。形態素解析の結果は手動でチェックし、間違いがあれば訂正する。このデータから、品詞毎の使用頻度を算出する。一方、電子ファイルから読点の数および 1 文当りの平均文字数を計算する。この作業を文章毎に行い、統計的な処理をする。

解析する品詞は、形容詞、助詞、助動詞、接続詞、動詞、副詞、名詞の計 7 種類である。

また、文章の男女差の特徴について調べるだけでなく、被験者がジェンダーというものに対してどのような意識を持っているかを調べた。文章を書いてもらった被験者に対して、ジェンダーに関する 20 項目から成るアンケート調査を同時に実施した。調査項目として、例えばジェンダーという言葉の捉え方や、文章を書く際に自己の性別を意識するか、等がある。

3 結果および考察

品詞の使用頻度については、ここでは助詞と動詞について詳細を調べた。

助詞については、個人のクセが出やすいことが知られている [1]。解析の結果、男女の間で使用頻度はほぼ同じであり、性別による明確な差はなかった (相対差 1.08%)。また、標準偏差についても正規化標準偏差が女性の場合 0.052、男性の場合 0.056 と、平均値同様、差はなかった。さらに、被験者間での使用頻度のヒストグラムを調べたところ、女性、男性どちらも正規分布に近い形状であった。正規分布は自然現象やランダムな現象に多く表れる分布である。このことは、助詞の使用については性差は存在せず、また文章を書く際の助詞の使用は意図的な要素は少ないと考えられ、この少なさが個人のクセの出やすさにつながっていると推測できる。

助詞と比較して、動詞の使用には個人差が出にくいと考えられるが、解析のコントロールデータ (対照データ) として解析した。動詞については、女性と男性で平均値の相対差は 1.6% とあまりないが、正規化標準偏差は男性の方が約 20% 高い。従って、男性の方が文の長さにつ

¹Differences in Properties of Text Expression Between Genders

²Atsumi Saito, Univ. of Tsukuba

²Shin-ichi Nakayama, Univ. of Tsukuba

²Tetsuya Maeshiro, Univ. of Tsukuba

*<http://chasen-legacy.sourceforge.jp/>

いてバラツキが大きい。さらに、使用頻度のヒストグラムを取ると、分布の形状も異なることが判った。男性の場合は正規分布に近い分布を示すが、女性の場合はピーク値が大きい値側にあり、その値より小さい値については分布が広く、大きい値については急速に頻度が下がる分布になっている。これは、正規分布が身長や体重などの自然現象やランダムな現象に多く表れる分布であることを考えると、女性の場合は何らかの要因によって男性よりも動詞の使用が制御されていると考えることができる。これは他の品詞の使用によって動詞の使用頻度が現象することも含まれる。男性は文章で表現する際に動詞の使用に女性ほど意識的に意味や意図を込めないとも言える。

1文当りの長さについて、男性の方が女性よりも平均で8%長かった。正規化した標準偏差については男女差がなかった(男性:0.195, 女性:0.199)ことから、文の長さのバラツキに性差は見られないが、女性の方が短い文を書く傾向があると考えられる。この結果は、被験者全体に対する正規化標準偏差がほぼ同じ値であったことから裏付けられる(0.202)。一方、1文当りの文字数のヒストグラムについては、動詞の使用頻度と同様、分布の形状も異なることが判った。男性は正規分布に近い分布を示すが、女性は正規分布ではなく、ある値よりも短い文は同じような頻度で存在するが、長い文は少ないというステップ状に類似した分布である。このことは、男性の場合、女性ほど文の長さについて左右する要因が少ないか弱いことを示唆していると思われる。動詞の使用頻度の場合のように、正規分布は自然現象やランダムな現象に多く表れる分布だからである。女性の場合、文の長さに意識的または無意識的に意味や意図を込めている可能性もある。

1文当りの読点の使用頻度についても解析した。平均値、標準偏差ともに女性と男性の間で差が見られ、いずれの値も男性の方が大きく、平均値と正規化標準偏差の相対差はそれぞれ11.5%と25.6%であった。文章心理学の観点からは、高い読点の使用頻度は文章の工夫と関連付けられ、低い使用頻度はありのままの文章を書く性格に由来する[2]。しかし、ここで得られた平均値の相対差の値(11.5%)から、このような性格の差が導けるかは詳細な検討を要する。正規化標準偏差の値が異なることは分布の形状の違いを意味している。ヒストグラムを調べたところ、女性、男性のどちらもピークの左右の幅が非対称な三角形に近かったが、女性の場合はピークの値よりも読点の使用頻度が少ない側は急速に現象し、多い側は緩やかに減少する。一方、男性の場合は女性とは反対の形状を示し、ピークの値よりも読点の使用頻度が少

ない側が緩やかな減少を示した。このことから、読点の使用頻度については女性と男性は反対の性質を持っていることが言える。これは個人差ではなく、同じ性別の集団間の差である。

アンケートについては、20項目の設問のうち、文章表現の性別差に深く関連するのが、「レポートを書く際に自己の性別を意識するか」と「手紙を書く際に自己の性別を意識するか」である。どちらも回答は「意識する」から「意識しない」の5段階である。レポートについては意識しない被験者が大半であった。しかし、手紙については男女ともに半数が意識しないが、残りの分布には性別差が表れ、女性は均等に分布していたが、男性については意識するという回答が得られた。よし詳細な解析が必要だが、今回の実験で被験者に書いてもらった文章は、手紙ではなくレポートの部類に属し、被験者は自己の性別を意識せずに書いたと仮定できる。従って、前述の計量解析で得られた性別による差は、文章を書く際の性別に依存する特性であると見なすことが妥当性だと考えられる。

これまで、本研究で用いた計量的な指標は個人差の検出に用いられ、その有効性が確認されてきた。例えば、文の長さは新約聖書や源氏物語の真贋判定に使われている[5, 6]。一方、ここで対象とした性差は、個人の集団の間での判定である。平均値が同じでもヒストグラムの形状が大きく異なる場合があり、この性別差は分布の違いに大きく出る結果が得られた。今後、アンケート結果との関連も含めて、より精密な解析を行っていく予定である。

参考文献

- [1] 村上征勝,「真贋の科学」, 朝倉書店, 1994
- [2] 波多野完治,「文章心理学」, 小学館, 1990
- [3] 波多野完治,「ことばの心理と教育」, 小学館, 1990
- [4] 波多野完治,「文章診断学」, 小学館, 1990
- [5] A.Q. Morton, "The authorship of Greek prose", *J. Royal Stat. Soc.*, A128, 1965.
- [6] 安本美典, "文体統計による筆者推定 — 源氏物語・宇治十帖の作者について —", *心理学評論*, 2, 1958.