



音素に基づく性差判別の試み(再認) : BCCWJ の創作 会話と CWPC の自然会話を比べて

中尾, 桂子

(Citation)

統計数理研究所共同研究リポート, 465:105-117

(Issue Date)

2023-03-10

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/0100479391>

(URL)

<https://hdl.handle.net/20.500.14094/0100479391>



音素に基づく性差判別の試み(再認)

—BCCWJ の創作会話と CWPC の自然会話を比べて—

中尾 桂子(大妻女子大学短期大学部)

nakao@otsuma.ac.jp

A revalidation study of Discriminant Analysis of Gender Difference
based on Phonemes between Fictive Speech and Spontaneous Speech
NAKAO, Keiko (Otsuma Women's University)

概要

音象徴の観点で性差を分析した研究に触発され、波多野(1958)の著者推定追認(中尾, 2021)を経て、会話に含まれる音素に基づき、創作会話と自然会話に含有される母音と子音の数から性差判別を試みた(中尾, 2022)。中尾(2022)では『現日研・職場談話コーパス』(CWPC)からランダムに抜粋した文を自然会話とし、また、『現代日本語書き言葉均衡コーパス』(BCCWJ)の書籍レジスターの会話部分からランダムに抜粋した文を創作会話として、その含有音素数に基づいて判別分析を行ったが、その結果から、創作会話には使用音素の性質に基づいて性差が見られるとした。しかしながら、今回、中尾(2022)の調査方法を見直して再現したところ、性差判別では、明確な判別精度が得られるわけではないことが確認されるとともに、中尾(2022)の分析上の問題が考えられた。再認により、機能語に現れる母音、子音といった含有音素数のバランスでは、有意差、ならびに、文体的特徴の差を明示できるほどの精度を判別できるものではなく、音象徴的な印象には、単に含有数を見るだけではない別の方法を検討する必要があること、また、音素だけではなく、別の要因が関係していることを踏まえて対象データの再検討を行うべきだと考えられた。

キーワード

創作会話, 職場談話, 性差, 音素, 判別分析

1. はじめに

日本語の書きことばは、4 世紀頃から 7 世紀頃にかけての文字の借用が浸透する中から徐々にその形を発展させてきたものであるが、この書きことばの誕生で、言語としての日本語は変容し、話ことばにおいても、また、書きことばにおいても、社会的状況を背景に、関連し合ってダイナミックに変容、発展してきた。中でも、9 世紀頃の書きことばに男手、女手の区別が生まれ、その流れから、話ことばにも男性語、女性語と言われる性別の差が明確になり、この男性語、女性語の顕著な差が、近代までの日本語の特徴の一つでもあった。

ただ、現代では、最早、明確な男性語、女性語と言われるものはほぼ使用されなくなったと、広く、認識されている。言葉遣いにおける性差が消えつつある状況について、任(2010)は、女性語、男性語という絶対的なことば遣いの差がなくなったとしており、現代日本語のことば遣いにおける性差とは、「相対的に差のある語を、特定の状況で使いわけると」いう程度の、語用法の一つとして認められる差であると述べている。

しかしながら、一方で、表層の形態的な差だけでなく、心理的な印象を左右する差として、語用上の性差が存在するというのも、周知の事実である。いわゆる、ポライトネスや情意的表現の多寡等、表現方法における使用意識に、性差が見られるという社会言語学的な意味での差である。

また、廃れたと言われる女性語、男性語も、それは形式上のことで、依然として、コミュニケーション上の効果や、社会的な立場を明確にするために、性差があることも周知の事実である。日本語では立場の低い社会的位置づけの人物がより丁寧な表現を使うことが多いが、丁寧なことば遣いや敬語の使用割合が女性の方が多いということである。他にも、人物描写を省略する目的で、創作中の登場人物のキャラクターを立てるために、登場人物のセリフにステレオタイプの性差が表わされることがある。「わたくしが～ますわ」「君が～したまえ」など話者の性別が想起できる表現上のスタイル、すなわち、役割語(金水,2003)である。この役割語は、主に、フィクションで、その話し手のキャラクター設定が特定のものであることを端的に示す言葉遣いだと捉えられている。女性語や男性語が廃れたと言われる現在でも、創作中での登場人物には、場面設定上の社会的な意味役割が与えられており、2023 年現在、発売されている小説の会話やアニメのセリフにも見られるものである。このことは、実際には使用しない読者が、創作内での使用に対して違和感を覚えるわけではないということを示している。日常会話では性差を明示しないにも関わらず、創作では、使用されることが自然に受け入れられており、この状態が、言葉の性差が廃れたとはまだ言えないという状況を示していると考えられる。

以上のように考えてくれば、現代日本語には、形式的に顕著に現れるわけではなくても、依然、表現の中に埋め込まれた性別上の特徴というものの色濃く存在する可能性が想定される。たとえば、波多野(1958)は、読者の直感で作者の性別や作者そのものを特定できるか調査し、それにより、文体上に現れる印象や心理が存在することを明らかにしたいと考えた。また、川原(2017)は、日本語の音素にもより多くの表現上の情報がある例を多数示している。ポライトネスや情意表現の割合等、語彙だけではなく、音にも、性別を表す要素があり、それには、パターン化し得る規則性のある特徴が見られるのではないだろうか。特に、音調やフィラー等は、これまで言語の範疇外のものとして直接分析されることが少なかったが、音に関する日本語の特徴にも、性差等の法則性が見いだせる可能性もあるのではないだろうか。

そこで、中尾(2022)は、音象徴的観点での研究方法を援用し、創作会話と自然会話の内的な要素が無意識に知覚されていることについて、探索的ではあるが、何らかの考察ができるのではないかと考えて、深層部分に内在して、聞き手の印象を左右している文体的特徴を考えてみるために、試みに、音素に基づいて性別が判別できるか検討してみた。中尾(2022)は、創作会話と自然会話を、音素に基づいて性差という観点から判別できるかを調べようと、自然会話として『現日研・

職場談話コーパス』(以下, CWPC)を, また, 『現代日本語書き言葉均衡コーパス』(以下, BCCWJ)の書籍レジスターの会話部分から得られた文を創作会話と位置づけ, 女性話者と男性話者別に, BCCWJ, CWPC からランダムに抜き出した会話, 各 40 件に含有されている母音, 子音に基づいて性別の判別分析を行った。そして, 自然会話である CWPC に性別による判別精度の高いものがより多く含まれていること, ならびに, その判別精度が, 変数となる音素の種類により, 異なることを確認した。創作会話は, 性別ごとの「らしさ」を表す部分がデフォルメされていると考えられるが, 表層的な形式とは異なる, より深層的なものが音素にも存在する可能性が考えられた。波多野(1958)を追認した中尾(2021)同様, 中尾(2022)でも, 音素が文体的特徴を構成する一要素になり得る可能性, ならびに, 音素による文体分析の可能性が示されたのではないかと考えた。

しかし, 中尾(2022)の判別分析の計算手順に不備が認められたため, 本稿では, 中尾(2022)の分析手法を再確認し, 音素に基づいて創作会話と自然会話の文体的特徴の差を確かめる。

2. 先行研究

1990 年代中頃の「文章心理学」の試みの一つに, 波多野(1958)の著者性別推定がある。具体的には, 女性を主人公にした 12 人の作家の小説の冒頭から 3 文ずつ取り出した 14 種の実験用の文章をランダムに組み合わせて被験者に呈出し, その文章の筆者性別推定の的中率を調べたというものである。波多野(1958)の調査では, 調査対象全体で見ると, 的中率は高くないが, 特定の作品の適中率が, 特に被験者が女性の場合に高い点を指して, 作家性別推定の難易は, 作家の文体差に起因するだけではなく, 被験者の側にも環境に裏打ちされた「能力差」というようなものがあるからだと考察している。波多野(1958)はその能力差が具体的に何に起因するかまでは明らかにしていないが, 性差が, 印象に影響する可能性があるという点が指摘されていた。また, 同様に, 文章を心理学的な見地から明らかにしようとした研究の一つである安本(1965)は, それを語彙の分布に見出せるのではないかと考えており, 語彙の計量的調査で文体的特徴として実証的に明らかにしようとしていた。アプローチの違いはあるが, 文体差の分析に, 読者側の心理的な印象が影響する可能性を見い出そうとする観点は, 音象徴に関する研究に類似する点であると考えられ, 興味深い。

この音象徴とは, 乱暴に言ってみれば, 聞いた音によって何らかのイメージを喚起させられることである。『ブリタニカ国際大百科事典』の「音象徴」の項には, 「ある言語体系で, 特定の音や音連続に特定の意味が結びつく傾向があるとき, そのことを音象徴という」とある。また, 川原(2017)では, 「その名前に使われている音によって, その名前で表される物事のイメージに影響が出てしまう現象」だとされている。この音象徴という概念の研究は, 心理学者, ケーラーが行った実験に端を発する。心理学者のケーラーは, 角の尖った部分が多い図形と, 曲線で描かれた図形の 2 つを並べて, これらに名前を付けるとしたら, “kiki”と“bouba”のどちらの名前が適当か問い, 音象徴という現象を明らかにした。このときに実験では「とげとげ」した方の図に“kiki”, カーブの多い図に“bouba”をつけると応える人が 9 割方であったことから, 1929 年に, 言語音と図形の視覚的印象との間には連想的な関係があると報告した。これが, いわゆる, ブーバ／キキ効果であるが, その後, 1947 年

に『ゲシュタルト心理学』の中で、“takete” “maluma”という呼び方と2種の図形で、再度、検証し、音とイメージの関連性の傾向に、どのような音が関係するのかが示した。その後、1970年には、母語に関係なく、およそあらゆる言語に共通して、音とイメージに関連があることを指摘した(Köhler, 1929, 1947)。

この音とイメージの関連性は、言語を超えた共有意識が感じられることから、その後、多くの追認研究が盛んに行われた。たとえば、Perfors (2004) の報告がわかりやすい。詳しくは川原 (2017) に説明があるが、Perfors (2004) は、写真の中の男性と女性に対する魅力が、写真に添付された個人名の中の母音の違いに影響されることを示した。Perfors (2004) は、実験用に子音と母音の数を性別と合わせて調整した実在の名前を複数用意して 24 枚の友人の写真に添付し、「ユーザーが人の魅力を 10 点満点で獲得できる」というウェブサイトを投稿して魅力度判定を行った。そして、その判定結果から、名前の中の母音が、顔の魅力の知覚的判断に影響を与える可能性があることを明らかにした。男性の場合、前舌母音のある名前がついた写真が、後舌母音のある写真よりも一貫して魅力的であると認識され、女性の場合、その関係が逆であり ($p < 0.01$)、かつ、女性の場合には、子音も、魅力の違いに影響を与えている ($p = 0.01$) というものである。ここでいう前舌母音は英語のもので [i] [e] [æ]、後舌母音は [u] [o] [ʌ] である。

日本語でも、Kawahara & Shinohara (2012) が、同様に、音と印象に関係があることを報告している。丸みの多い図に口唇や歯茎等の「丸い」音、とげとげしい図に破裂や破擦等の「とげとげしい」音を当てはめる人が多いことから、共鳴する子音が「丸い」イメージを聞き手に持たせ、阻害的な子音が「とげとげしい」イメージを聞き手に持たせるとして、イメージの持たせ方から、子音を共鳴音と阻害音に分類できることを指摘している。

このほかにも、川原 (2017) には、日本語を音象徴的に分析した結果が多数紹介されているが、その分析の観点は、音の性質とそれに対する心理的印象である。そして、数多の音象徴における実証的研究を経て種類分けされた子音を、その調音点と調音法により、日本語における阻害音と共鳴音に分けた。また、母音を、その舌の位置に伴って生じる「男性的」「鋭い」「薄い」「さっぱり」というイメージに基づくものを「ツン母音」、「女性的」「甘い」「濃厚」「クリーミー」といったイメージに基づくものを「萌え母音」と呼び分けたが、この子音と母音の関係は、図 1, 2 のように図示されている。川原 (2017) では、この子音の阻害音と共鳴音を、「濁音が付けられるもの、もしくは濁音がついているもの＝阻害音」「そもそも濁点すら付けられないもの＝共鳴音」と端的に説明しているが、これは、阻害音により、角ばった男性的だという連想が働くこと、ならびに、共鳴音により、丸っこい女性的な連想が働くことからの分類としている。なお、この図 1, 2 の 2 図は、川原 (2017) のオリジナルの図の一部を割愛して筆者が改編したものである。

中尾 (2021) は、作家の性差が文体に表れるか考察した波多野 (1958) の〈読み手の印象〉と川原 (2017) の日本語の音素が与える〈読み手への印象〉を類似したものと捉え、また、計量的な調査による実証に基づいた安本 (1965) の姿勢に倣い、川原 (2017) の音象徴的調査で明らかにされた特性別の子音と母音を指標に波多野 (1958) の分析結果を追認した。文に内在する文体的要素として音素に着目するのは、波多野 (1958) の〈読み手の印象〉が表面には現れない隠

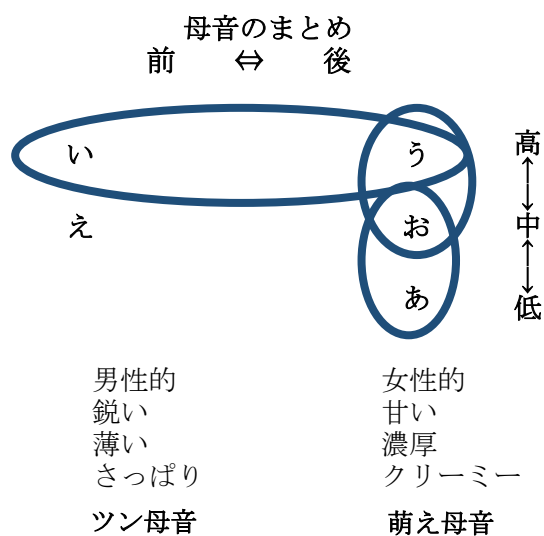
図 1

子音：阻害音・共鳴音と調音点・調音法（川原, 2017, p. 201 より）

		子音のまとめ				
		両唇音	舌頂音	硬口蓋音	舌背音	
阻害音	破裂音	p/b	t/d			ツン 角ばった ツン ツン
	k/g					
	破擦音 摩擦音	φ	s/z	tʃ ʃ		
共鳴音	鼻音	m	n			丸い・萌え
	流音		r/l			
	半母音	w		y		

図 2

母音の下位分類（川原, 2017, p.204 より）



れた性差を示すものではないかと考えたからである。また、川原(2017)に整理された音素が文体上の見えない性差分析の可能性を期待したからである。その手順は次のようなものである。まず、波多野(1958)で用いられた14種のテキストを調査対象とし、それぞれのテキストをExcelの関数HEBONによってヘボン式のローマ字表記に変換することで「音素化」と見なし、ローマ字変換後のテキスト中の、共鳴音、阻害音、ツン母音、萌え母音(川原, 2017)に相当するローマ字の数を計上して、音素4組ごとの頻度表を作成することで基礎データとした。つぎに、基礎データの基本統計量に基づいて、テキストごとに、音素どうしを検定した。しかし、この結果からは、Perforce

(2004)のように、音素数による検定の結果に有意差が見られなかった。その理由は、おそらく、日本語が、基本的に、子音と母音の割合が同数になる音節単位の構成を持つからではないかと考えられた。念のため、中尾(2021)では、差のある部分をデフォルメするために、また差の意味の違いが分かりやすくなるように対応分析を行ったが、その結果では、子音 2 種と母音 2 種が各次元に分布し、波多野(1958)で「女性っぽい」文体、「男性っぽい」文体などという印象が持たれていたテキストが、それぞれの次元にそれぞれの子音、母音の特徴が顕著な作品として現れた。第 2 軸(縦軸)の正負で区分される特徴は、特に、「女性っぽい」印象と「男性っぽい」印象の作品の分別と一致していた。そこで、さらに、主成分分析を行い、第 1 主成分として特性が構成される(累積寄与率が集約)のが阻害音と萌え母音の組合せであること、そして、第 1 主成分に最も高い負荷量で寄与する変数が「萌え母音」であることを示した。このことから、「萌え母音」の多寡や、「萌え母音」に阻害音の影響が重なることが、文体上の特徴に何らかの影響を与え、それにより、著者推定の可能性が高まるのではないかと推定された。この結果は、4 種の音素の特徴でテキストが分類でき得ることを表しており、また、テキストには音素による何らかの傾向が反映されている可能性を指摘したことになるのではないかと考えた。ただし、中尾(2021)の調査対象にした波多野(1958)の 14 種のテキストは長さがまちまちであったこと、さらに、著者の性別が同数にそろえられていないなど、偏りのあるデータであることが否めないものであった。

そこで、性別と、作者、話者の条件をそろえ、ランダムに選んだデータで音素による判別を再検討することとした(中尾, 2022)。中尾(2022)では、自然会話として『現日研・職場談話コーパス』(以下, CWPC)を、また、『現代日本語書き言葉均衡コーパス』(以下, BCCWJ)の書籍レジスターの会話部分から得られた文を創作会話と位置づけ、女性話者と男性話者別にランダムに抜き出し、CWPC, BCCWJ の各 40 件に含有されている母音、子音に基づいて判別分析を行って、創作会話と自然会話を、音素に基づいて性差という観点から判別した。中尾(2021)で、僅差ではあるが、音素による性差の違いが考えられたことで、中尾(2022)においてデータを変えて音素の差を見ようとしたのだが、しかし、中尾(2022)の調査において、計量結果の分析時に計算方法に手違いがあったと考えられることから、本稿で、中尾(2022)の再検証を行うものである。

3. リサーチデザインと手法

3.1 研究目的と RQ

中尾(2022)の再検証として同内容を設定する。すなわち、創作会話と自然発話を、音象徴的印象差のある音素(川原, 2017)に基づいて判別した場合、男性の発話と女性の発話の判別が可能か(RQ1)そして、どのような音素がその判別に最も寄与し、分類に影響を与えているか(RQ2)。創作会話と自然会話は同様に判別が可能か(RQ3)、以上 3 点について考察するものである。

3.2 対象データ

調査対象は、CWPC と、BCCWJ から、話題やテーマによる影響を除くためにランダムに抜き出

した会話文を用いる。もちろん、BCCWJ は、サンプリングして集められた書きことばであるため、検索結果は、もともと、文脈に依存していないのだが、文脈のある CWPC に合わせて、「」でくくられた文からランダムに選び、そのうち、その会話の話者の性別が特定できる会話文を選んで、「の次から 20 語分の機能語部分を分析の対象とした。また、その検索対象を、「出版・書籍」「図書館・書籍」レジスターのうち、より会話らしい部分が抜き出されるように、「」でくくられていて「です。」で終わる文とした。その書籍での出版年をデータの全期間である「1970 年代～2000 年代」としている。

CWPC は、職場という特定の環境での談話が文字化されてコーパス化されたものである。女性談話は 1993 年 9 月から 10 月にかけて録音された、職場が異なる首都圏の有職者 19 名 (20 代～50 代) の自然談話からなるデータである。このデータは、朝、職場に着いてから 1 時間、会議・打ち合わせ 1 時間、休憩 1 時間を録音したうちから、それぞれ 10 分前後のまとまった談話が文字化されている。また、男性談話は、1999 年 10 月から 2000 年 12 月にかけて録音された、職場が異なる首都圏の有職者 21 名 (20 代～50 代) の自然談話のデータである。こちらも、職場に着いてから 1 時間、会議・打ち合わせ 1 時間、休憩 1 時間を録音したものから、それぞれ 10 分前後のまとまった談話が文字化されている。

CWPC は、もともと、現代日本語研究会編(2011)の合本版としてひつじ書房から刊行された文字化テキストを、国立国語研究所が、MeCab + UniDic で解析した結果を合わせてコーパスとして公開したものである。BCCWJ に比べると、特定の話題に基づく自然談話が文脈ごと集められている向きが強いが、録音された時期が 1993 年から 2000 年までで、実際の制作時期がそれ以前だと考えれば、おおよそ、BCCWJ から検索する創作の発表年(1997 年～2000 年)と同時期の日本語が反映されたものと考えられる。また、CWPC は首都圏で収集され、基本的に書籍の日本語と共通点の多い語彙使用の談話だと考えられることから、BCCWJ と比較することとした。

以上をまとめたのが、表 1 と表 2 である。なお、CWPC はデータを談話と呼ぶが、本稿では文脈のある談話の一部を取り出した片方の発話を対象とすること、また、かつて対話であったことから、会話という呼び方でまとめて捉え、創作会話、自然会話と位置付けている。

表 1
調査対象(データ)

調査対象	コーパス	時代	男性発話 40 件	女性発話 40 件
創作 (会話)	現代日本語書き言葉均衡コーパス BCCWJ	全期間	男性筆者の創作した男性が発話した体で記述されている「」内冒頭の文字列 20 語分から実質語を除外したもの	女性筆者の創作した女性が発話した体で記述されている「」内冒頭の文字列 20 語分から実質語を除外したもの
自然 (談話)	現日研・職場談話コーパス CWPC	全期間	男性が発したもの	女性が発したもの

表 2

検索条件

調査対象	検索条件	
創作 (会話)	検索キーワード	カギかっこの開始記号”「”
	検索結果	小説内会話(116,562)
	対象	キーワード後件の出力からランダムに1行(20語)1件×40件×男女(男性筆者による男性発話, 女性筆者による女性発話に該当する)の会話文から実質語を除外したもの
談話	検索キーワード	書字形出現形「です」
	検索結果	3,746 件(男性発話 1647 件, 女性発話 1621 件)
	対象	キーワードの後件として出力された1行(20語)1件×40件×男女の発話から実質語を除外したもの

3.3 手法

3.3.1 データの事前処理

自然会話となる CWPC は, (1)～(3)の手順で事前に処理した。(1)として, まず, CWPC の全期間から「書字形出現形」が「です」, 「後方共起条件の追加」で“。”を加えて検索し, 文末が「です。」で終わる1文を得たが, その総数は 3,746 件であった。この 3,746 件は, 男性発話が 1,647 件, 女性発話が 1,621 件とであったため, ほぼ同数と考えて, そのまま, 男性, 女性別に冒頭の文字列 20 語分を1文(1件)として抜粋した。そして, そこから, 内容に左右されないように, 実質語である漢字表記の含まれる語を除いた機能語のみの部分を抜き出した。ついで, (2)機能語のみとなった抜粋部分を HEBON 関数でローマ字に変換した。そのうえで, (3) 調査対象の子音, 母音の数を数え, (4)4 種の音素ごとに集計した結果を判別分析のデータとして整理した。

図 3

判別分析結果の散布図(40 件)

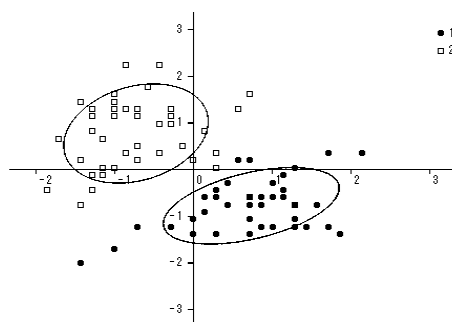
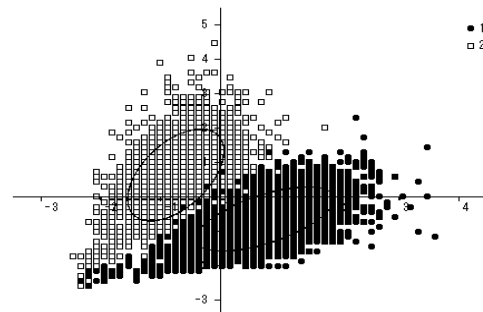


図 4

判別分析結果の散布図(116,562 件)



なお、自然会話から同様に抜き出した男性と女性の談話全数 3,746 件分で判別分析した結果と、そこから Excel のランダム関数を用いて抜き出した 40 件を、College Analysis 6.7 を用いて判別分析にかけ、それぞれで出力した散布図を比較してみた。その散布図が図 3, 4 である。図 3 が 40 件の結果で、図 4 が全数の判別結果であるが、両方は判別に用いたデータ数が大きく異なるものの、その性差の分布や判別精度に大きな違いが視覚的にも認められなかった。出力結果がそのデータ数に応じて分布を変えないことが確認できるため、本稿での調査においても、中尾(2022)同様、比較は 40 件のデータで行うこととし、自然会話も、創作会話も、性別ごとに、それぞれ 40 件ずつをランダムに抜き出して調査対象とする。

創作会話は、次の(1)～(4)の手順で事前処理した。(1)BCCWJ の全期間の「書籍」レジスターから「中納言」を経て抜き出したが、検索には、「書字出現形」を“「」”とし、「後方共起条件の追加」で“です”、“。”、“”を加えて検索し、116,562 件の検索結果を得た。ついで、＜男性筆者が創作した作品＞から＜男性が発話した体で記述＞されている「」内の会話の冒頭の文字列 20 語分を 1 件として、また＜女性筆者が創作した作品＞から＜女性が発話した体で記述＞されている「」内冒頭の文字列 20 語分を 1 件として抽出していった。このとき、男性、女性の分類は、話者の性別が特定できる語の有無で判断し、話者の性別がわからないものは除外した。そのうえで、それぞれから、Excel 関数の RAND を用いて、簡易的ではあるが、ある程度のランダムさでもって、男性話者の会話文、ならびに、女性話者の会話文、それぞれ、40 文ずつ抜き出した。これらを、(2)HEBON 関数でローマ字に変換してから、実質語である漢字表記部分を除いて、(3)調査対象の子音、母音の数を数え、(4)4 種の音素ごとに集計して、その結果を調査対象のデータとした。

以上のように、調査対象として 40 件ずつ取得して、事前処理をおこなった手順は、いずれの場合も、次の流れである。

1)各コーパスから「」冒頭 20 語分の「会話」データを抽出

2)テキスト部分を Excel 関数(HEBON)でローマ字に変換

3)Excel 関数(LEN)で

子音(阻害音・共鳴音)、母音(ツン母音・萌え母音:川原 2017)毎に計上

4)4 種の音素特性(阻害音、共鳴音、ツン母音、萌え母音)ごとに集計

子音(川原, 2017, p.201)

阻害音:/p/ /t/ /k/ /b/ /d/ /g/ /s/ /z/ /h/

共鳴音:/m/ /n/ /y/ /r/ /w/

母音(川原, 2017, p.204)

ツン母音:/i/ /e/

萌え母音:/u/ /o/ /a/

5)頻度表にまとめる

3.3.2 分析の手順

創作会話と、自然会話に内包される音素は、3.3.1 節の 5 番目の手順により頻度表に整理した。表 3 がその例で、頻度表から一部を抜粋したものである。

表 3

分析用データ頻度表の一部(粗頻度):小説の会話の「」内の音素数(4 種ごとの集計結果)

	萌え母音	ツン母音	阻害音	共鳴音	グループ
BCCWJ-N-6	11	5	8	6	1
BCCWJ-N-7	16	6	16	10	1
BCCWJ-N-18	16	5	7	13	1
BCCWJ-N-22	15	3	10	7	1
(中略)					
BCCWJ-N-222	10	7	12	9	2
BCCWJ-N-229	6	6	12	2	2
BCCWJ-N-234	15	7	13	7	2
BCCWJ-N-239	7	1	7	3	2

表 1 のように整理した頻度表を用いて、College Analysis 8.0 (福井, 2021) で正準判別分析結果(標準化実測値)を求め、同時に、萌え母音、ツン母音、阻害音、共鳴音の 4 つの指標のうち、2 群ずつの組合せで判別した結果(判別関数、標準化係数、F 検定値、自由度、確率値、マハラノビスの距離、共分散間の差)を確認しつつ、散布図で視覚的に判別精度を概観した。

4. 結果と考察

4.1 RQ1 創作・自然会話での男性の発話と女性の発話の判別は可能か

まず、創作会話では、正規性の検定で 2 群とも正規性があり、また、等共分散の検定でも共分散に差があるとは言えないことから、判別分析が適用された。そして、2 群間の生起確率を同じとし、誤判別損失を等しいとすると、判別分析によって以下の判別関数が得られた。

$$y = 0.417 \text{ 萌え母音} + 0.152 \text{ ツン母音} - 0.162 \text{ 阻害音} - 0.290 \text{ 共鳴音} - 2.371$$

係数の有効性検定で、「萌え母音」に関しては $p=0.00106$ で有意に 0 でないことが示されたものの、他は有効性が低い。マハラノビスの距離が 0.534 とかなり低く、実測値から 1 群と 2 群を間違える割合は 38%、その逆も 38%であったことから、判別精度が低く、判別はうまく行われていないということになる。すなわち、音素の含有数に基づく判別は難しいということになる。

もう一方の自然会話の場合は判別分析が適用不可であった。以下のような判別関数が得られ、等共分散の検定で共分散に差があると言えるが、判別分析は適用不可と判断された。

$$y = -0.055 \text{ 萌え母音} - 0.072 \text{ ツン母音} + 0.132 \text{ 阻害音} - 0.199 \text{ 共鳴音} + 2.011$$

実測値から 1 群と 2 群を間違える誤判別確率が 60%, 逆は 75%とかなり高く、また、4 つの中では[阻害音]が強めに影響しており、確率値も 0.2525 と高い。マハラノビス距離も 0.719 で、うまく判別できたとは言えそうにない。正準判別分析結果散布図（図 4）でも視覚的に概観してみるが、判別されたと言い難い様子がうかがえる。

図 3

創作会話における [萌え母音] [ツン母音] [共鳴音] [阻害音] の散布図

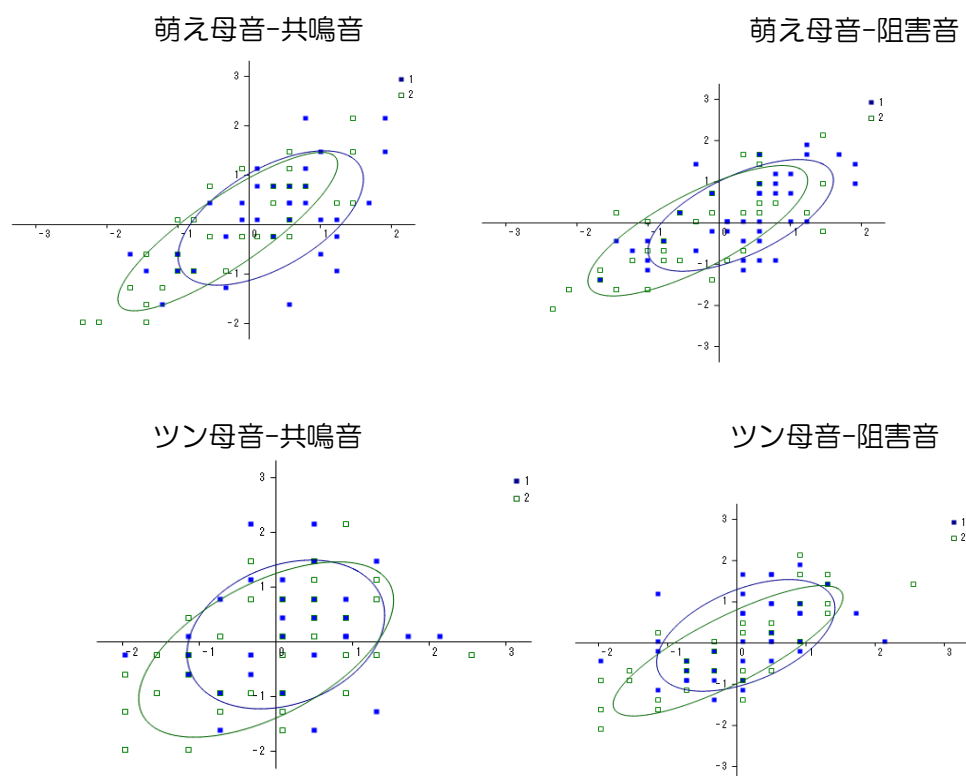
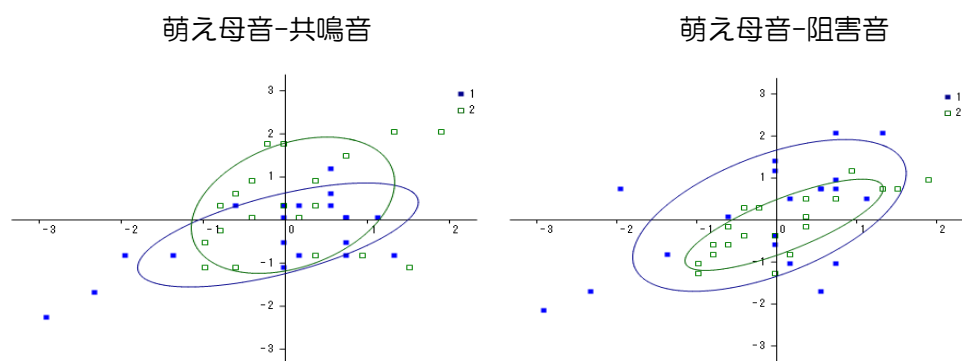
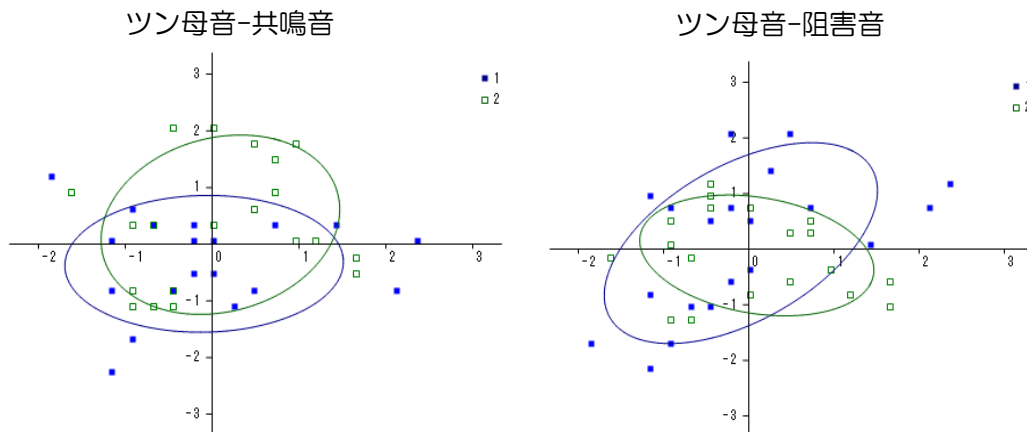


図 4

自然会話における [萌え母音] [ツン母音] [共鳴音] [阻害音] の正準判別分析結果散布図





以上の結果を見る限りでは、男性発話と女性発話を、音素を指標に、精度よく判別することはできなかったと言える。

4.2 RQ2 判別に寄与する音素はどれか

RQ1 の結果から、創作会話の判別はかろうじて可能であったものの、その精度は低く、自然会話では判別不可であったことから、判別に影響を与えると考えられる音素を明確なものとして特定することはできないと考えられた。

ただ、「萌え母音」の影響が若干見られる点は、中尾（2021）でも見られたことであるから、判別に寄与するとは言えない参考程度の情報ではあるが、「萌え母音」には、なんらかの影響力のようなものがあるとも考えられる。

4.3 RQ3 創作会話と自然会話は同様に判別が可能か

RQ1 の結果から見る限りでは、創作会話と自然会話を同様に判別することはできず、また、創作会話の方が、若干、判別精度が高かった。これは、創作された会話が、その「らしさ」をより強調して使用しているということを表すのかもしれない。

ただ、いずれにしても、やはり、日本語の機能語部分という骨格に乗っている音素には、性差に関しての情報量は少ないと考えられる点が再確認されたと言える。

5. まとめ

今回、音象徴的観点を反映すると考えられた子音と母音で、日本語の骨格部分に特化して文体的特徴を統計的に概観しようと試みたが、性差という観点からの判別精度は低かった。このことから、今回の方法では、発話者の性別が十分に意味のある状態で判別しきれないと考えられた。

ただし、＜やわらかい音＞だと認識される「萌え母音」の有無が、創作会話では、若干、性別の「らしさ」に影響している可能性が考えられた。これは、創作が、自然談話より、性差がデフォルメされていることを表している可能性があり、もしそうなのであれば、心象を左右するも音もあるという可

能性が残る。

また、「萌え母音」がその判別に影響している可能性が否定できないことが、中尾(2021)で行った波多野(1958)の追認結果とも類似する。これを考えれば、ある種の母音の働きに焦点を当てた分析方法を検討することで、音素が文体に影響する可能性を検討することもあり得ると考えられる。

さらに、前回の中尾(2022)、今回の再認ともに、談話内容に影響を受けないことを優先して調査対象から実質語を除外している。しかし、実は、語彙的な部分の選択に性差がより現れる可能性もあることから、語彙と音素と性差の関係を、再度、検討する必要もある。

以上を考慮するなら、今回の試みだけで結論を出すには尚早だとも考えられるとともに、今回用いたデータの妥当性、ならびに、事前準備からの手順や分析手法も、さらに、検討し直す必要がある。話者を調整した実際の自然会話で、また、同方法、ならびに、別の観点と角度から、調査方法を検討し直し、性差という観点での文体分析をもう少し検討してみたい。今後の課題である。

引用文献

- 川原繁人(2017)『「あ」は「い」より大きい!?—音象徴で学ぶ音声学入門』ひつじ書房。
- 金水敏(2003)『ヴァーチャル日本語 役割語の謎』岩波書店。
- 現代日本語研究会編(2011)『合本 女性のことば・男性のことば(職場編)』ひつじ書房。
- 中尾桂子(2021)「音素からみる文体—文体形成要素としての音素の影響を考える①—」『大妻国文』52, 1-20.
- 中尾桂子(2022)「音素に基づく性差判別の試み—BCCWJの創作会話とCWPCの自然会話を比べて—」2021年度統計数理研究所共同研究レポート No. 456.
- 任利(2010)『女ことばは女が使うのかしら』ひつじ書房。
- 波多野完治(1958)『ことばと文章の心理学』新潮文庫。
- 福井正康(2020) College Analysis 8.0: <http://www.heisei-u.ac.jp/ba/fukui/analysis.html>
- ブリタニカ・ジャパン(2014)『ブリタニカ国際大百科事典 小項目電子辞書版』Britannica.
- 安本美典(1965)『文章心理学入門』誠信書房。
- Kawahara, Shigeto and Kazuko Shinohara (2012), A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation. *In. N. Miyake, D. Peebles and R. P. Cooper (eds.) The Proceedings of the 34th Annual Meeting of the Cognitive Science Society. Austin: Cognitive Science Society. pp. 569-574.*
- Köhler Wolfgang (1947), Gestalt psychology. *New York: Mentor Books.*
- Köhler, Wolfgang (1970), Gestalt Psychology: The Definitive Statement of the Gestalt Theory Paperback, Liveright; 2nd Revised edition.