



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

中尾, 桂子

(Citation)

統計数理研究所共同研究レポート, 456:105-116

(Issue Date)

2022-03-03

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81013072>

(URL)

<https://hdl.handle.net/20.500.14094/81013072>



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—

中尾 桂子(大妻女子大学短期大学部)

nakao@otsuma.ac.jp

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

概要

一般に、創作された会話では、現実の自然な会話の場合より、ジェンダー的要素がデフォルメされ、また、その文体的特徴が、主に、女性語や男性語により表されると考えられているのではないか。この創作会話と自然会話の違いを、発話中に埋め込まれた性差という観点で比較してみるため、中尾(2021)を見直し、創作会話として『現代日本語書き言葉均衡コーパス』(BCCWJ)書籍レジスターの会話部分を、自然会話として『現日研・職場談話コーパス』(CWPC)を、各々に含有される母音、子音に基づいて判別分析を行った。結果、創作会話は自然会話より位相語が多いが、若干、判別精度が低めで、創作会話と自然会話では、判別に寄与する音素の影響範囲に違いがあると考えられた。このことから、音素による文体分析の可能性が示されると同時に、性別上の「らしさ」が統語レベルにもあり、深層部分でも表現上の社会的な性差の違いがあり得ることが考えられた。

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

日本語には位相語が多いが、中でも、女房詞に端を発する女性語と、そのカウンターパートとしての男性語は、その代表として認識されている。しかし、任(2010)が言うように、現代語では、最早、女性語、男性語という絶対的なことば遣いの差はなくなっており、「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、語用法の一つとなっている。

確かに、現実には女性語や男性語がさほど見られなくなった。しかし、創作の世界では、現在でも、対象読者年齢に関わらず、明示的に女性語、男性語が使い分けられている。おそらく、創作では、女性語や男性語を用いることで登場人物の性別やキャラクター描写を割愛させられるからだと考えられる。創作における登場人物のことば遣いは、現代社会の一般的な認識を反映するものだと考えられるが、このような機能的な面に着目し、金水(2003)は、「役割語」を次のように定義する。

ある特定の言葉遣い(語彙・語法・言い回し・イントネーション等)を聞くと特定の人物像(年齢、性別、職業、階層、時代、容姿・風貌、性格等)を思い浮かべることができる、あるいはある特定の人物像を提示されると、その人物がいかにも使用しそうな言葉遣いを思い浮かべることができる、その言葉遣いを「役割語」と呼ぶ。(金水 2003, 205 頁)

金水(2003)の言う役割語は、金水の「役割語からみた日本語とキャラ」によれば、「(実在と否とを問わず)社会的グループと結びつけられた話し方の(社会・地域)方言のステレオタイプなパターン」だとされる。このパターンは、「いかにも使用しそうな言葉遣いとしてイメージされる母語話者が共有する言語的知識で、いわば、社会的な意味の乗せられたパターン」である。役割語は「性別、年齢・世代、社会的階層や職業、居住あるいは出身地域」によって示される「社会的な共同体」の共有物であるとされており、共同体に着目する意識が強いものとして位置づけられている。

この定義や位置づけを見る限り、役割語は、とりたてて、性別を意識したものとは位置づけられていないと考えられるが、役割語は、その役割ということばで社会的な意味役割が与えられていることを指すため、実際には、社会的な性差を意味するジェンダーとも関連するものだと考えられる。創作において使用される性差が強調された語や表現は、母語話者の知識に働きかけることで描写を割愛することができ、また、その反面、性差が強調されるがゆえに、現実とは異なる創作として、わざとらしい女性らしさの強調、わざとらしい男性らしさの強調としても受け取られ得るからである。また、現在、自然会話では使用されなくなっている位相語が、創作会話の文体的特徴として顕著に使用されているのならば、このわざとらしさという印象も含めて、読み手に、社会的な性差を強調するがゆえに、ジェンダーという問題に抵触するものとなり得る可能性があり、反面、それが創作の文体的特徴になっている可能性も考えられる。

ただ、一方で、文章を心理学的に見ようとした波多野(1958)が考察したように、表面上、見えぬ印象があり得ること、また、それを示唆する川原(2017)で示されるような音象徴的観点での研究から、深層的な印象の存在が、文体的特徴として存在する可能性も考えられる。

そこで、創作会話と自然会話の違いを音素に基づく性差という観点で比較し、それにより、日本語の創作会話と自然会話の差について考えたい。女性語、男性語の語彙的要素は、創作会話には多いが、自然会話には少ないと考えられることから、本稿では、語彙ではなく、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

中尾(2021)は、文体の構成要素としての音素に着目し、川原(2017)の示した日本語音の音象徴的分析観点を援用して、波多野(1958)の著者性別推定を追認する形で、『現日研・職場談話コーパス』(CWPC)を用いて性別での違いがあるか調べた。

この川原(2017)の示す日本語における音象徴的分析観点とは、図 1, 2 様の関係の子音と母音である。川原は数多の音象徴における実証的研究を経て、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、共鳴音に分けた。また、母音は、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。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, 阻害音は角ばった男性的だという連想が働き, 共鳴音は丸っこい女性的な連想が働くのではないかとして, 図1のように整理している。また, 母音においても, 実証研究の結果から, 明示する印象に分けて図2のように整理している。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。

図1

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

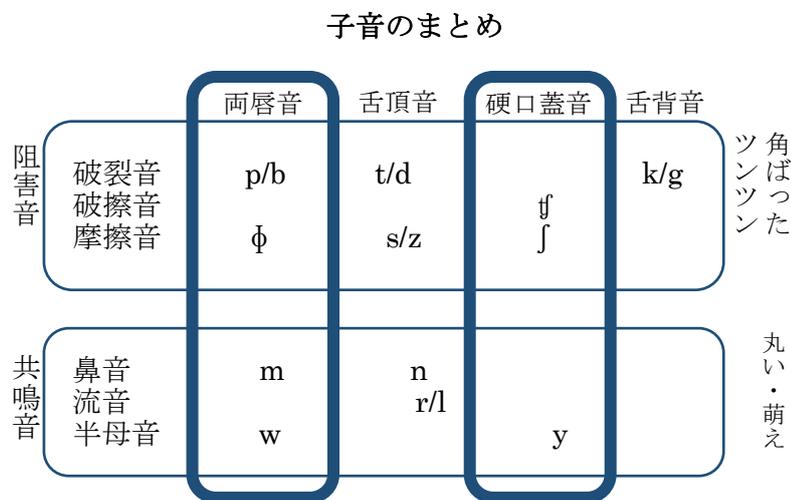
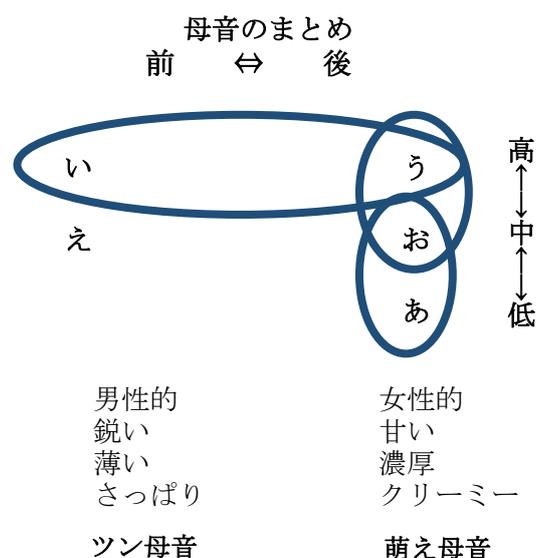


図2

母音の下位分類（川原, 2017, p.204 より）



波多野(1958)の著者性別推定とは、90年代中頃の「文章心理学」の試みの一つで、具体的には、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、的中率が高くはないが、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差し、作家性別推定の難易は、作家の文体差に起因するだけでなく、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、安本(1965)は、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〉に着目して、計量的調査による実証という安本(1965)の姿勢に倣い、川原(2017)の音象徴的調査観点から追認して、文体における性差印象に音素の影響があり得るか調査を試みた。その方法は、次のようなものである。まず、『現日研・職場談話コーパス』(CWPC)から、「」でくられ、末尾が「です」で終わる文のうち、文頭から20語を1文と仮定して抜き出し、その1文をExcelの関数HEBONによってヘボン式のローマ字表記に換えた。そして、文ごとに、川原(2017)の子音(阻害音・共鳴音)、母音(ツン母音・萌え母音)を数え、その音素の情報に基づき、僅差でも特徴を視覚化できるコレスポネンス分析で音素種による差があるかを確認し、ついで、判別分析、さらに、主成分分析を併用してデータの様子を確認した。

結果、特定の音素の組合せ(阻害音と萌え母音)には、対応関係が見られ、精度の高い判別結果が得られた。つまり、阻害音と萌え音の組合せと発話者の性差の違いには結びつきがあり得ることから、聞き手が、話し手の性差を印象として受け取る可能性が示唆されたと考えた。もちろん、中尾(2021)で用いたデータからだけでは、十分な資料が用いられとは言えないが、文字種、語種、語彙、表現の使用傾向といった従来からの文体形成要素以外にも、音素という、より深い部分に埋め込まれた要素も、文体の印象に影響を与える可能性があることが改めて確認されたと言えよう。ただ、中尾(2021)の方法は、プリミティブな試みでしかなく、十分な妥当性があるかを検討する必要がある。そのうえ、その分析からの考察がまだ十分ではない点もあった。

3. リサーチデザインと手法

3.1 研究目的とRQ

本稿では、先の中尾(2021)を見直しつつ、その成果を応用して、創作会話文と自然会話文における印象について性差からさらに考えるために、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、創作会話と自然会話のどちらの方が、性差の判別が可能だと言えそうか(RQ1)。そして、どのような音素がその判別に最も寄与、すなわち、分類に影響を与えているか(RQ2)。そこから、語彙的な違いと音素の違いのどちらの方が創作と自然会話の性差印象に影響していると予測できそうか(RQ3)、以上3点について考察する。

3.2 データ

話題やテーマによる影響を除くため、調査対象は、本稿でも、中尾(2021)で用いた『現日研・職場談話コーパス』(以下、CWPC)と『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)から、ランダムに抜き出した文を用いる。

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、文脈に依存しない。検索対象は、「出版・書籍」「図書館・書籍」レジスターとし、出版年は全期間「1970年代～2000年代」とした。

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)の自然談話である。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。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。

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、BCCWJから検索する創作の発表年(97年～2000年)から考えて、おおよそ、同時期の日本語が反映されたものと考えられること、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を優先してBCCWJの比較対象とした。

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。かつて対話であったことから、会話、すなわち、創作会話、自然会話と呼んでいる。

3.3 手法

3.3.1 データの事前処理

自然会話は、(1)～(3)の手順で事前に処理した。(1)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3,746件得た。このうち、男性発話が1,647件、女性発話が1,621件であったが、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、そこから、実質語である漢字表記の含まれる語を除いたものを(2)HEBON関数でローマ字に変換した。そして、(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データとした。

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。図3が40件の結果で、図4が全数の判別結果であるが、両方のデータ数は大きく異なるものの、その性差の分布や判別精度に大き

図 3

判別分析結果の散布図(40 件)

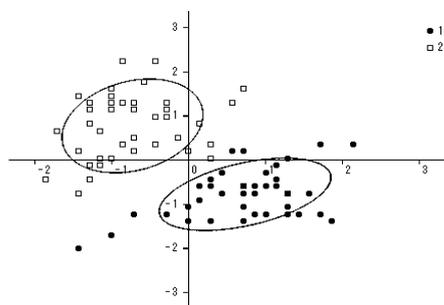
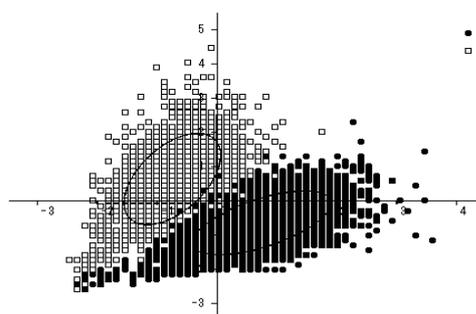


図 4

判別分析結果の散布図(116,562 件)



な違いが視覚的にも認められなかった。そのため、今回の比較は 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 件ずつランダムに抜き出して調査対象とした。

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。検索は、「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後、そこから、さらに「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、(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調査対象のデータとした。

以上、調査対象として 40 件ずつ取得後の事前処理は、いずれの場合も、次のような流れとなる。

- 1) 会話データを各コーパスから抽出
- 2) テキストを Excel 関数でローマ字に変換
 使用関数; HEBON
 例) =HEBON(C2)
- 3) 子音(阻害音・共鳴音), 母音(ツン母音・萌え母音:川原 2017)を計上
 使用関数; LEN
 例) =LEN(\$C2)-LEN(SUBSTITUTE(\$C2, "A", ""))/LEN("A")
- 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、この表に基づき、College Analysis 6.7 を用いて話者の性別の判別分析を行う。表 1 は頻度表から一部を抜粋した例である。

表 1

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 6.7 (福井)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、同時に出力される散布図で視覚的に判別精度を概観する。

4. 結果と考察

4.1 創作会話の判別分析の結果から

図 5~8 は正準判別分析結果の散布図である。この 4 図を概観すると、図 7 の判別精度の高さが窺える。そして、母音どうしや子音どうしの組合せでは判別精度が低い、子音と母音の組合せ、中でも、[萌え母音・共鳴音]の組合せで判別した場合の判別精度が特に高い。萌え母音を 1 群とした場合の判別関数は 1.062 で、対する 1 群の共鳴音が -0.171 である。また、2 群の判別関数は、萌え母音が -0.136 で、共鳴音が 0.889 である。p = 0.0000 で、マハラノビスの距離 12.432 で、誤判別確率実測値は 1 群→2 群 2.5%となり、逆が 1.3%と高精度になっている。

また、ツン母音と共鳴音もツン母音と阻害音(図 8)と同程度の判別結果を得ており、共鳴音と母音の組合せの中でも、特に、共鳴音と萌え母音の含有数が性別の違いを反映するのではないかと

図 5

萌え母音-ツン母音

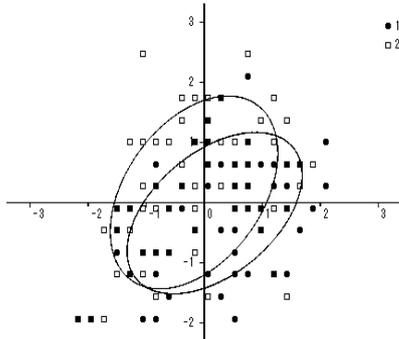


図 6

共鳴音-阻害音

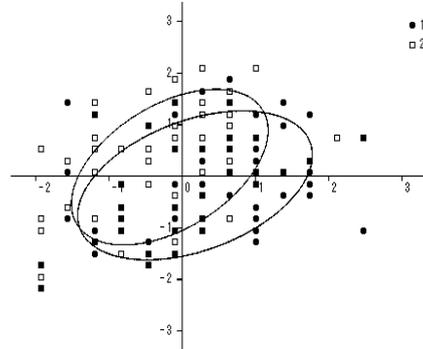


図 7

萌え母音-共鳴音

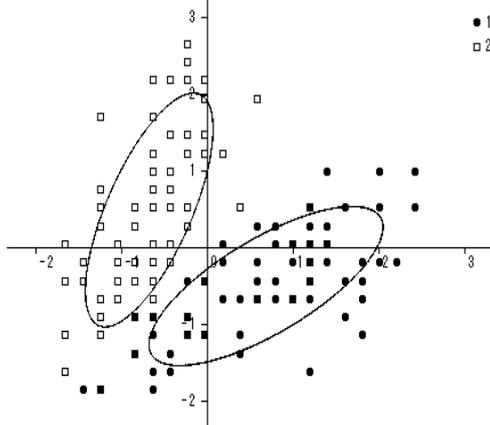
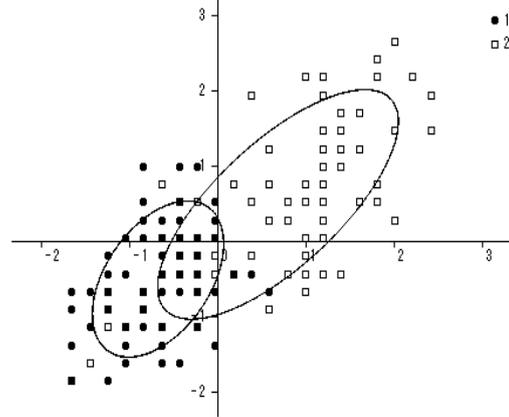


図 8

ツン母音-阻害音



考えられる。

この分析に用いた BCCWJ のデータサンプルと CWPC のデータサンプルをそれぞれ 3 例ずつ下に抜粋する。サンプルの下線部は性差の違いが特に現れる語だと考えられるものである。なお、この下線は、本稿の例示のために筆者が付与したが、本稿の調査で計上する対象ではない。音素数を数えたのは太字の部分のみで、形状の際にはローマ字に変換されている。この下線の数を見ても、自然会話 CWPC に比べて、創作 BCCWJ の女性会話文に、女性語の影響が色濃く見える。しかし、図 9 で創作会話に内包される音素の内訳を比べると、単一の音素数の性差はさほど明確なものでもないことが窺える。

つまり、明示的な性差のある語数に差があっても、音素数での性差の判別には大きく影響しないこと、一方で、[萌え母音-共鳴音]の相乗効果により、性差の判別精度が高くなることが窺える。

<BCCWJ 女性>

PB39_00134 九月でも、ずいぶんお暑いのね。一どう、お国では皆様

PB49_00594 本当は気が進まないのに、ここに連れてきてくれたでしょ。わかっ

PB39_00144 男の子なんか用なんかないわよ、行っちまいなさいってば！

<CPWJ 女性>

CWPC_F01A041 やだなあ。もうねえ、この状況はもう、ほとんど追加公演状況

CWPC_F03K011 がよくないですねえ。ええ、ええそうですね。発色が。高い紙なん

CWPC_F04K011 からの連絡は以上だそうで。ちょっと休憩の、これは。はいあのゲスト

<BCCWJ 男性>

PB19_00248 われら松平党は、こたびの出陣において栄えある先鋒をうけたまわった！

PB1n_00020 おれの調査員にそこからの足どりを追わせるか、それともあんたがやる

PB19_00287 あのころはみんなが尾羽打ち枯らしてたもんよ。交易はさっぱりだし、工場は

<CPWJ 男性>

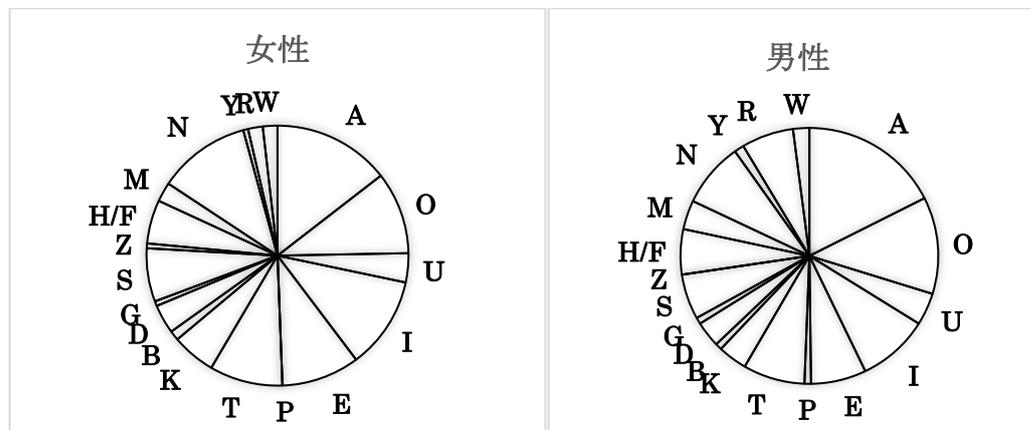
CWPC_M01A071 強力ですか、普通のですか。普通でいい。845円

CWPC_M01K0112個頼む、2回、あの一、1枚の伝票で2回頼んじゃったん

CWPC_M02K011 はい。周辺環境とゆうのは、なんかわかるようでわからないんです。

図 9

創作会話のサンプル内含有音素の内訳



注: 萌え母音(A,O,U),ツン母音(I,E),阻害音(P,T,K,B,D,G,S,Z,H/F),共鳴音(M,N,Y,R,W)

4.2 職場の会話の判別分析の結果から

図 10～13 は正準判別分析結果の散布図である。この 4 図を概観すると、図 10～12, 特に、図 12 の精度の高さが窺え、[萌え母音と共鳴音]の組合せでの性差判別精度が高そうに見える。図 12 においては誤判別確率が 0.025, すなわち、97.5%の精度で判別された。また、 $y = -0.874$ 共鳴音 + 0.694 萌え母音 + 1.648 で、判別の分点を 0 とし、2 群に分けられる。ただし、 $p = 0.0000$ のため、マハラノビス距離 15.705 から、理論的な誤判別確率として $p = 0.024$ が予想される。高精度のため、等共分散の検定結果を見ると、[萌え母音-共鳴音]では、自由度 3, χ^2 統計値 6.133, 片側確率 $P < 0.105$, 有意水準 $\alpha < 0.05$, $P \geq \alpha$ より、共分散間に差があるといえない(等共分散とみなす)。

以上から、判別分析が有効に適用可能であると判断する。

図 10 の[ツン母音-共鳴音]は、 $y=0.313$ 共鳴音 $+0.172$ ツン母音 -8.999 で、 $p=0.0000$ のため、マハラノビスの距離 5.008 から、理論的な誤判別確率として $p=0.132$ が予想される。自由度 1, χ^2 統計値 3.625, 片側確率 $P=0.057$, 有意水準 $\alpha=0.05$, $P \geq \alpha$ より、共分散間に差があるといえず(等共分散とみなす), 判別分析が有効に適用可能であると判断する。しかし、図 11 の[萌え母音-阻害音]は、 $y=0.059$ 阻害音 $+0.192$ 萌え母音 -4.770 で、 $p=0.0000$ のため、マハラノビス距離 7.980 から、理論的な誤判別確率として $p=0.025$ が予想されるが、自由度 3, χ^2 統計値 9.781, 片側確率 $P=0.021$, 有意水準 $\alpha=0.05$, $P < \alpha$ より、共分散間に差があるといえ(異共分散), 判別分析が有効に適用可能ではないとされた。さらに、子音どうしの組合せ[阻害音-共鳴音]は正準判別分析ができなかった。

そこで、図 10 の[ツン母音-共鳴音]と図 12 の[萌え母音-共鳴音]の散布図で考察を進めるが、自然会話でも、[萌え母音-共鳴音]と[ツン母音-共鳴音]の組合せで精度の高さが出ると考えれば、自然会話では、共鳴音の含有数に性差が反映される可能性があるのではないかと考えられる。

図 10
ツン母音-共鳴音

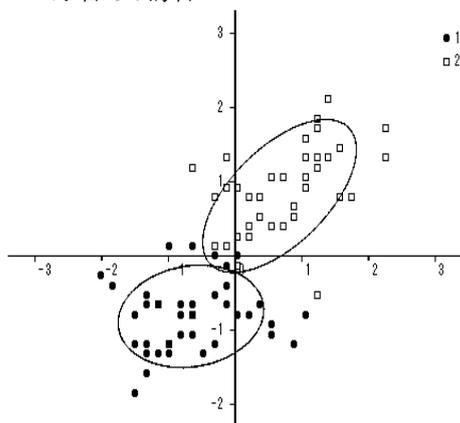


図 11
萌え母音-阻害音

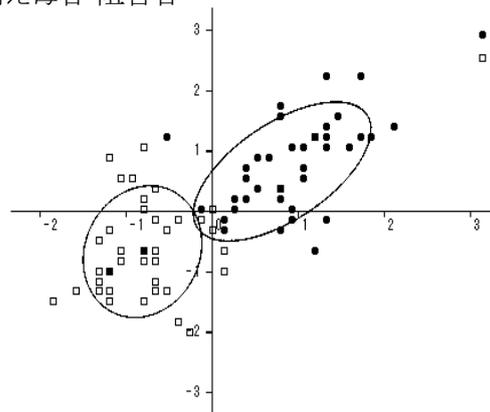


図 12
萌え母音と共鳴音

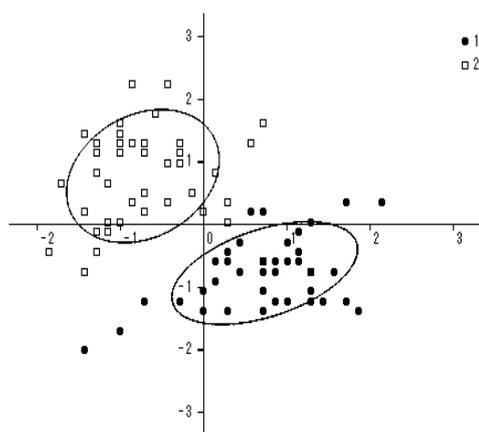
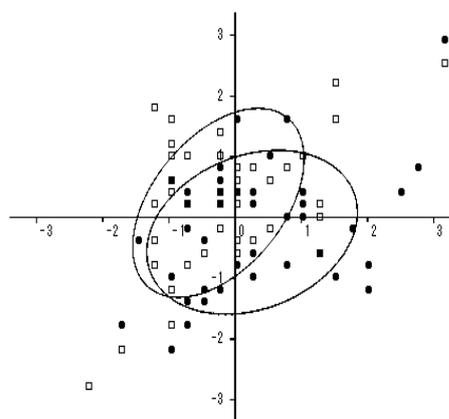


図 13
ツン母音-阻害音



4.3 RQ について

今回の RQ1) は、音象徴的印象差のある音素で男性と女性の会話文の判別が可能か。それは、創作会話と自然会話で異なるかというものであった。また、判別に寄与するのはどのような音素かというのが RQ2) で、語彙的な違いと音素の違いのどちらの方が創作と自然会話の性差印象に影響していると予測できそうかというのが RQ 3) であった。

まず、RQ1) について言えば、4.1, 4.2 の結果を見る限り、どちらの会話でも音素による判別は可能だと言えそうである。ただし、判別に影響する音素は特定の音素、すなわち、共鳴音だと考えられたが、創作会話と自然会話ではその影響力が異なる可能性が示唆された(RQ2)。それは、創作会話と自然会話の判別において、創作会話では、萌え母音が 1 群、共鳴音が 2 群となる組合せにおいて、判別精度が高くなるものの、一方の自然会話では、創作会話とは異なり、母音と子音の組合せではなく、共鳴音という子音のみの含有数に影響を受ける可能性が示唆されたことによる。つまり、判別に寄与する音素として、自然会話では共鳴音、創作会話では共鳴音と萌え母音の組合せによる影響が強いのではないかと、また、その判別精度は自然会話の方が高いのではないかと考えられる(RQ2)。さらに、RQ3) については、創作会話と自然会話では共鳴音の影響力が異なるのではないかと考えられること、また、4.1 節でみたように、女性語の影響が色濃く見える創作会話であっても、含有音素での性差判別には大きく影響することはなく、また、自然会話の方が、共鳴音に基づく判別精度が高そうだと考えられることから、使用語彙よりも、内包音素のほうが、性差印象には影響があるのではないかと。

性差印象には、共鳴音(/m//n//y//r//w/)が影響するのではないかと考えられるが、一方で、創作会話では共鳴音(/m//n//y//r//w/)と萌え母音(/a//o//u/)の両方によって性差が醸し出される可能性が考えられることから、創作会話が、共鳴音だけでなく、萌え母音の持つ音象徴的印象によっても、それらしくデフォルメされていると言えるのかもしれない。

さらに、この含有音素数の違いは、音素を用いることで、位相語の使用に影響を受けずに、性別が区別し得ることを示唆すると考えられるが、音素数で性差があるということは、語を組み立てる際の方法、すなわち、統語的な文体差も、ある程度、社会的な習慣やイメージに影響されており、性別の差になって表れているという可能性を示唆すると考えられる。日本語での表現上のジェンダー意識が、位相語だけでなく、文体の差にも現れている可能性が改めて確認されたと言えるのではないかと。

5. まとめ

本稿では、創作会話と自然会話の違いを性差という観点で比較してみるため、自然会話として『現日研・職場談話コーパス』(CWPC)を、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得られた文を創作会話と位置づけ、サンプルとして抜き出した各 40 件の文に含有されている母音、子音に基づいて話者の性別で判別分析した。

結果、創作会話は、自然会話よりも判別精度が、若干、低めで、自然会話の方が性別による判別精度の高いものがあつた。その判別精度は、どの音素が変数になるかで異なっていた。このこと

から、創作会話は、その含有される音素の特徴からも創作らしさが確認されたとと言えるのではない。また、この結果は、創作会話での性別の「らしさ」を表す部分が、女性語や男性語といった表面的な語の使用方法で示されるもの以外にもあることを示しており、現実の会話では、女性語や男性語といった目に見えるものではなく、より深層的な統語方法や語用方法に見られるジェンダー的な傾向で示されるものである可能性が考えられた。したがって、中尾(2021)同様、本稿の再考においても、音素は、文体的特徴を構成する要素の一つとして位置づけられ、また、音素による文体分析の可能性も、ある程度、示されたのではないかと考える。

ただし、本稿で、会話のサンプルとして用いた日本語は、BCCWJ と CWPC からのものである。これらは、現在から 20 年ないし、50 年前の日本語である。そのため、本稿で用いたサンプル日本語に含有される音素の多寡は、厳密には 2022 年現在の語を反映しているわけではない。また、今回のデータでは、調査対象の待遇レベルをそろえていない。調査対象を変えて繰り返し、検証する必要があるが、それは今後の課題とする。

但

本研究は、2021 年 7 月 15 日開催の第 102 回大妻国文学会例会にて報告した「創作会話は象徴的にも『らしい』のか—子音と母音にみる性差—」を文章に起こしたものである。

引用文献

College Analysis 6.7 (福井正康): <http://www.heisei-u.ac.jp/ba/fukui/analysis.html>

波多野完治(1958)『ことばと文章の心理学』新潮文庫。

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。

金水敏「役割語からみた日本語とキャラ」より(2022.2.9 訪問)

<http://skinsui.cocolog-nifty.com/sklab/files/yakuwarigokaramitanihongotokyaranihongobunpogakkai.pdf>

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.

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。

任利(2010)『女ことばは女が使うのかしら』ひつじ書房。

Perfors Amy (2004) What's in a name?: The effect of sound symbolism on perception of facial attractiveness, *Proceedings of CogSci2004*.

<https://escholarship.org/uc/item/9bq5v5c7>

安本美典(1965)『文章心理学入門』誠信書房。