

国立国語研究所学術情報リポジトリ

日本語日常会話コーパスを利用したコピュラ出現・非出現の研究

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2026-01-23 キーワード (Ja): ゼロコピュラ, だ抜き言葉, CEJC, 日常会話, 言語変異 キーワード (En): zero copula, da-deletion, CEJC, everyday conversation, language variation 作成者: 南部, 智史 メールアドレス: 所属: モナッシュ大学
URL	https://doi.org/10.15084/0002000594

This work is licensed under a Creative Commons Attribution 4.0 International License.



日本語日常会話コーパスを利用したコピュラ出現・非出現の研究

南部智史

モナッシュ大学／国立国語研究所 共同研究員

要旨

本稿では、変異理論の視点から「と思う」節内のコピュラ「だ」の出現・非出現を変異現象として扱い、日本語日常会話コーパスを用いた量的分析を行う。これまでに、日本語書き言葉均衡コーパスと日本語話し言葉コーパスを使った分析から、話し言葉の方が書き言葉より「だ」の使用が多いこと、女性の方が男性より「だ」の使用が多いこと、「だ」ナシから「だ」アリへの量的推移という言語変化の存在が報告されていた。本研究ではこれらの点について日常会話でも同様の傾向が見られるか検証を行った。分析の結果、日常会話においても、学会講演などのデータを含む話し言葉コーパスと同様に「だ」の高い使用率が観察され、書き言葉との差が顕著であることが明らかとなった。言語変化に関しては、見かけ上の時間の概念に基づいた年齢を用いた分析から、日常会話においても先行研究と同様に「だ」アリへの変化が確認された。一方で、男女差については先行研究とは反対の傾向が確認されたほか、話し相手（仕事関係 vs. 家族／親戚／友人・知人）に基づくフォーマリティの分析では、「だ」の使用に統計的に有意な影響は認められなかった*。

キーワード：ゼロコピュラ、だ抜き言葉、CEJC、日常会話、言語変異

1. はじめに

本稿では、コーパスを用いた言葉のバリエーション研究のプロジェクトの一環として、「だ抜き言葉」（塩田 2010）とも呼ばれる、「と」節内のコピュラ「だ」の出現・非出現の分析を行う¹。例えば、(1a) に挙げた文では名詞「問題」に続く位置で「時間の問題（だ）と思う」、(1b) では形容動詞語幹（以降「形容動詞」）「大事」の後に「大事（だ）と思う」と、それぞれ「だ」の挿入が可能である。

- (1) a. やはり時間の問題と思うとのことでした。（宮本百合子「獄中への手紙」1937 年）²
b. こういった活用、非常に大事と思うんですが、大臣、どのようにお考えなのか、お伺いしておきたいと思います。

（国会会議録，第 211 回国会参議院厚生労働委員会 2023 年 4 月 25 日，東徹）³

「と」節内の「だ」の出現・非出現については、(1) の動詞「思う」だけでなく、「考える」

* 本稿の一部は国立国語研究所の共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」（プロジェクトリーダー：小磯花絵）の研究成果である。本稿の内容は 2025 年 3 月 24 日のシンポジウム「日常会話コーパス」X での発表をもとに執筆したものである。

¹ 「だ」の代わりに「である」も同様の言語環境で生起可能だが、本稿で扱う日常会話のデータでは「である」の使用が確認されなかったため、本稿の議論では「だ」のみを対象とすることとした。

² 青空文庫：https://www.aozora.gr.jp/cards/000311/files/33186_15706.html（2025 年 11 月 22 日確認）

³ 国会会議録検索システム：https://kokkai.ndl.go.jp/（2025 年 11 月 22 日確認）

（「居眠りは問題（だ）と考える」）、「信じる」（「その条件を有利（だ）と信じる」）などの動詞においても同様の現象が観察される。一方で、（2）のように他の従属節では「だ」が必須である。また、（3）のように主節では「だ」が欠落する場合があるが、これはフォーマリティに関わる現象で、カジュアルな場面に観察されるコピュラの省略と見なされており（Yamaguchi 2007）、本稿で扱う「と」節内の「だ」出現・非出現の現象とは区別される。このことは、（1）に挙げた例が示すように、丁寧体が含まれる文であっても「と」節内で「だ」が現れない場合が観察されるという事実によって裏付けられる。

- (2) a. それは必要 {だ／* \varnothing } から買った。
 b. それは必要 {だ／* \varnothing } が、買わなかった。
 (3) これは桜の花 {だ／ \varnothing }。

本稿では変異理論（variation theory, variationist sociolinguistics, Labov 1966, 高野 2005, 2011, Tagliamonte 2012）の枠組みから「だ」の出現・非出現を変異現象として捉え、日本語日常会話コーパス（CEJC, Koiso et al. 2022）を用いた量的分析を行い、日常会話における「と」節内の「だ」の出現・非出現という言語変異の選択に関わる要因について考察する。

2. 先行研究

1節で導入した「だ」の出現・非出現の評価に関して、塩田（2010）は、「行かなくても大丈夫だと思います」と「行かなくても大丈夫だと思います」という2つの文を用いた正しさの評価に関するオンライン調査（参加者763名）を行い、多数（77%）が「だ」ナシはおかしい（「だ」アリが正しい）と回答したと報告している。また、Nambu（2023）は、40文（「だ」アリ20文と「だ」ナシ20文のペア）を用いて5段階評価で「正しさ」（規範意識）についてのアンケート調査を日本語母語話者110名に対して行い、「だ」アリの方が統計的に有意に高く評価されることを示した。

言語使用に関して、田野村（2008）は1940年代から2000年代の国会会議録のデータを用い、漢字表記語（e.g., 「責任」）に続くコピュラ「だ／である」＋「と思う」のデータを分析し、「だ」ナシが経年的に徐々に減少し「だ」アリが増加するという進行中の言語変化を報告している。コピュラの使用頻度に関わる言語環境について、阿部（2015）は現代日本語書き言葉均衡コーパスを用い、国立国語研究所の検索システム「少納言」から「漢字表記語」＋「だ／である」＋「と思う。／と思われる。」という検索ワードでデータを抽出した結果、コピュラ非出現率は「～と思う。」（5.4%）よりも自発形の「～と思われる。」（52.2%）の方が高かったと報告している。また、阿部（2001）では「A{ が／を }B（だ）と思う」という言語環境におけるコピュラの有無と生成文法で扱われるECM構文（Chomsky 1981）との関係について議論しており、「太郎を天才と思う」のように対格「を」が付与された名詞句がコピュラ非出現と共起する場合の「と」節には時制がないと指摘している。

Nambu（2023）では、日本語書き言葉均衡コーパス（BCCWJ）と日本語話し言葉コーパス（CSJ）

を用いて、「[名詞／形容動詞] + 「だ／である／なり／たり」 + 「と」 + 「思う／考える／信じる」」のデータを抽出、分析した結果、書き言葉に比べて話し言葉の方がコンピュータを伴う形式がより多く出現する傾向があると指摘されている。表1に動詞「思う」の結果を示したが、カイ2乗検定の結果（表下）は書き言葉（BCCWJ）と話し言葉（CSJ）の間に有意な差があることを示している。

表1 「と思う」節内のコンピュータあり・なしの頻度（Nambu 2023）

	BCCWJ	CSJ
コンピュータあり	21,342 (76.6%)	2,559 (89.6%)
コンピュータなし	6,508 (23.4%)	297 (10.4%)

$$\chi^2=251.81, df=1, p<.001$$

Nambu (2023) は田野村 (2008) が指摘した「だ」アリ増加という経年変化の傾向について検証している。Nambu (2023) では、日本語歴史コーパス (CHJ) を用いた分析結果から、元々は「と思う」の前にコンピュータを伴わない形式がデフォルトだったことを明らかにしている。さらに、BCCWJに含まれる書き手の生年が1900–1970年のデータを用いた統計的分析において、「と思う」はコンピュータを伴わない形式から伴う形式へ変化してきたことが報告されている。一方で、阿部 (2015) が指摘していたコンピュータの使用傾向が異なる自発形「と思われる」の場合には、そのような変化が見られず、コンピュータなしの形式が依然として優勢であることが示された。また、Nambu (2023) は阿部 (2001) のECM構文に関する議論を踏まえ、BCCWJのデータに含まれる「AがB(だ)と思う」と「AをB(だ)と思う」の頻度を比較し、前者の「が」の場合にコンピュータの出現頻度がより高い傾向があることを指摘している。その他の要因として、Nambu (2023) では、コンピュータ直前の名詞 (e.g., 「学生(だ)と思う」) と形容動詞 (e.g., 「重要(だ)と思う」) を比較すると、形容動詞の方がコンピュータの出現頻度がより高いこと (表2)、男女を比較すると、女性の方が男性よりも「だ」をより多く使用する傾向が報告されている (表3)。

表2 名詞／形容動詞 + 「と思う」におけるコンピュータあり・なしの頻度（Nambu 2023）

	BCCWJ		CSJ	
	名詞	形容動詞	名詞	形容動詞
コンピュータあり	20,555 (65.2%)	3,508 (76.5%)	2,507 (77.6%)	334 (88.6%)
コンピュータなし	10,960 (34.8%)	1,079 (23.5%)	561 (22.4%)	43 (11.4%)

$$\text{BCCWJ: } \chi^2=227.66, df=1, p<.001, \text{CSJ: } \chi^2=10.52, df=1, p<.01$$

表3 男女別, 「と思う」節内のコピュラあり・なしの頻度 (Nambu 2023)

	BCCWJ		CSJ	
	男性	女性	男性	女性
コピュラあり	5,851 (53.4%)	1,802 (68.9%)	1,647 (81.0%)	1,194 (84.7%)
コピュラなし	5,104 (46.6%)	813 (31.1%)	388 (19.0%)	216 (15.3%)

BCCWJ: $\chi^2=205.64$, $df=1$, $p<.001$, CSJ: $\chi^2=7.83$, $df=1$, $p<.01$

3. 方法論

本稿では, BCCWJ と CSJ を用いた Nambu (2023) の分析結果と比較するため, 日本語日常会話コーパス (CEJC, Koiso et al. 2022) を利用した。BCCWJ は新聞やオンラインブログなどの様々なジャンルから 1971 年から 2008 年にかけて収集された約 1 億語のテキストからなり (Mackawa et al. 2014), CSJ は 1999 年から 2003 年に収集された独話 (学会講演, 模擬講演など) および対話 (インタビューなど) の 661 時間の録音音声 (約 750 万語) で構成されている (Mackawa 2004)。一方, 本稿で用いた CEJC には, 家族との会話や職場での会議など多様な場面での会話 200 時間分の書き起こしデータ約 240 万語が収録されている。CSJ と同様に話し言葉のデータではあるが, CEJC では家族や友人・知人など親しい話者間の雑談が最も多くを占めていることから (小磯 2022), CSJ と比較するとカジュアルな場面での発話が中心となっている。また, CEJC には年齢や性別などの話者の属性に加え, 話者間の関係性といった情報も付与されており, 分析に活用することが可能である。

データは, コーパス検索アプリケーション「中納言」 (<https://chunagon.ninjal.ac.jp/>) を用いて抽出し, 対象とした言語環境は「[名詞/形容動詞] + 「だ/の」 + 「と思う」」とした。また, 「だ」と「思う」の異形態もデータに含めた。分析に用いた言語内的要因は, 「だ」の直前の要素 (名詞 vs. 形容動詞) であり, 言語外的要因は, 話者の性別と年齢, それから話者間の関係性 (仕事関係 vs. 家族/親戚/友人・知人) に基づいて想定したスタイル差 (フォーマル vs. カジュアル) の 3 つとした。

前節で見た先行研究の結果を踏まえると, 「だ」の直前の要素については, 「形容動詞」の場合に「だ」の使用が多くなることが予測される。話者の性別に関しては, 女性の方が男性よりも「だ」を多く使用する傾向が予測される。年齢の効果については, 年齢差 (age-grading) と「見かけ上の時間」 (apparent time, Cukor-Avila and Bailey 2013) の 2 種類がある。年齢差は世代間で繰り返される, 個人が年齢の経過とともに言語使用を変える現象を指すのに対し, 見かけ上の時間は, 同時期に観察される異なる年齢層の言語使用は進行中の言語変化を反映していると推定する, 変異理論で提唱された概念である。本稿では, 田野村 (2008) と Nambu (2023) の言語変化の報告を踏まえ, 後者の見かけ上の時間に基づいて年齢の効果を捉えることとする。そのため, 年齢が若いほど「だ」がより多く使用される傾向が予測される。話者間の関係性については, CEJC にはより詳細なカテゴリーが付与されているが, 本稿ではスタイル差を分析するために, 「仕事関係」と「家族/親戚/友人・知人」の 2 つに大別したカテゴリーを比較することとした。スタ

イル差の効果については、現代では正しいと評価される「だ」の使用がフォーマルな場面でより多くなる傾向が予測される。また、ロジスティック回帰分析を行う際には、話者間の関係性には「その他」のカテゴリーも含めた。

ここで、本稿で扱わなかった要因について言及しておく。まず、この現象に関わる動詞として、先行研究では「思う」の他に「考える」と「信じる」も扱われていたが、CEJC で検索した結果、「考える」は「だ」アリが2件、「だ」ナシが2件、「信じる」はいずれも0件であったため、データには含めなかった。また、自発形「思われる」は「思う」と異なる振る舞いをするのが指摘されていたが、CEJC のデータには該当する環境で自発形「思われる」が使用された例が0件であったため、分析対象から除外した。それから、ECM 構文とされる「が」と「を」(「A{ が／を }B (だ) と思う」) の比較についても、収集したデータには該当する「が」は25件、「を」は3件のみであり、量的分析を行うには不十分と判断し、これも分析対象から除外した。言語外的要因については、話者の出身地を分析の際に考慮することも考えられたが、今回扱うデータは、主に関東圏、特に東京と神奈川出身の話者に集中していたため、分析には含めなかった。

分析では、まず上に挙げた要因ごとに各形式の頻度のクロス表を示し、必要に応じてカイ2乗検定またはロジスティック回帰分析を行った。それから、全ての要因を含めたロジスティック回帰分析を行った。全ての要因を同時に考慮することで、それぞれの要因が持つ独立した効果を評価した。また、分析の際に個人差を考慮するため、ランダム効果として話者をランダム切片に設定した混合効果モデルを構築した⁴。統計分析にはR環境(R Core Team 2022)を用い、混合効果モデルの構築にはlmerTestパッケージ(Kuznetsova et al. 2017)を使用した。

4. 分析結果

4.1 各要因の考察

表4にCEJCで観察された「と思う」節内の「だ」アリと「だ」ナシの件数、合計741件を示した。(4) から (7) はコーパスで観察された「だ」アリと「だ」ナシの事例である。

表4 CEJC, 「と思う」節内の「だ」
アリ・ナシの頻度

「だ」アリ	672 (90.7%)
「だ」ナシ	69 (9.3%)

(4) 名詞 + 「だ」アリ

- a. だから初めは商品開発だと思ってたんだけど (C001_003)⁵

⁴ ロジスティック回帰分析に先立ち、データのコーディングの際に「だ」アリを1, 「だ」ナシを0に変換した。また、年齢は全体の連続的な傾向を捉えるために、10歳代は10, 20歳代は20といった具合に連続変数に置き換えて分析を行った。

⁵ 各例文には、それぞれのCEJCの会話IDを括弧内に示した。

- b. あんな大変な仕事だと思わなかったから (C002_006b)
 - c. こっちがカメラだと思ってた (C002_004)
 - d. 部屋はたぶん広い部屋だと思うけど (K002_018)
 - e. これ取れない染みだと思うよ (K005_012)
- (5) 名詞+「だ」ナシ
- a. 変な名前と思って (S002_014)
 - b. 経費と思えばあれですけどね (K006_017)
 - c. そこを育てていくのが仕事と思えば (K001_004)
 - d. なんかこう友達と思ってたからみたいな (T001_003)
 - e. やっぱすごい先輩と思って (T011_015)
- (6) 形容動詞+「だ」アリ
- a. 今ちょっと大変だと思う (K002_016)
 - b. 俺ちょっとかわいそうだと思うよ (C001_004)
 - c. 絶対好きだと思うんだよね (K005_016)
 - d. 保護者の方が育った時代と大学が同じだと思わないでくださいって (T004_002)
 - e. 自分はまともだと思ってんだよ (T004_003)
- (7) 形容動詞+「だ」ナシ
- a. これは素敵と思ったけど (C001_001)
 - b. それを嫌と思う人もおるかもしれないから (W004_002)
 - c. すごい大変と思って (C001_003)
 - d. お酒とかたばことおんなじと思って取るんだったら (T002_015)
 - e. あれが入ってるから好きと思って (T014_018)

表4に示した通り, CEJC では「だ」ナシと比べて「だ」アリの方が圧倒的に多いことがわかる。また, その割合は2節の表1に示した Nambu (2023) で報告されている BCCWJ (書き言葉) と CSJ (話し言葉) を比べると, CSJ の割合に非常に近いことがわかる。前述の通り, CSJ に含まれるデータは CEJC と比較するとよりフォーマルな場面での発話が収録されているが, 今回の分析結果は, そのようなフォーマリティの違いによる発話スタイルの差は「だ」アリ・ナシに大きく影響しないことを示唆している。この点について検証するため, 表1のCSJと表5のCEJCのデータに対してカイ2乗検定を行ったところ, 予測通り有意な差は確認されなかった ($\chi^2=0.65$, $df=1$, $p=.42$)。

(4)～(7)に挙げたCEJCで観察された事例は, (4)と(5)は「だ」の直前が名詞の場合, (6)と(7)は形容動詞の場合である⁶。表5に, これらに対応する名詞と形容動詞の「だ」アリ・ナシの件数を示した。2節表2に示したように Nambu (2023) では名詞と比べて形容動詞の場合にコピュラがより多く現れると報告されていたが, カイ2乗検定の結果, CEJCのデータでは名

⁶ 品詞については, コーパスのアノテーションに従った。

詞と形容動詞において有意な差が確認されなかった ($\chi^2=0.33$, $df=1$, $p=.57$)。

表5 CEJC, 名詞・形容動詞別「と思う」節内の「だ」アリ・ナシの頻度

	名詞	形容動詞
「だ」アリ	530 (90.3%)	142 (92.2%)
「だ」ナシ	57 (9.7%)	12 (7.8%)

性差に関しては、2節表3に示した Nambu (2023) の傾向とは反対に、CEJC では男性の方が女性よりも「だ」をより多く使用することがわかった ($\chi^2=8.31$, $df=1$, $p<.01$) (表6)。また、スタイル差の影響を検討するために用いた話者間の関係性については、「仕事関係の相手」(フォーマル)と「家族／親戚／友人・知人」(カジュアル)では「だ」の使用に統計的に有意な差が見られなかった ($\chi^2=0.006$, $df=1$, $p=.94$) (表7)。この結果は、フォーマリティの影響を検討するために CSJ と CEJC を比較した際に差が見られなかったことと一致している。

表6 CEJC, 男女別, 「と思う」の前の「だ」アリ・ナシの頻度

	男性	女性
「だ」アリ	302 (94.4%)	370 (87.9%)
「だ」ナシ	18 (5.6%)	51 (12.1%)

表7 CEJC, 話者間の関係性と「と思う」節内の「だ」アリ・ナシの頻度

	仕事関係	家族／親戚／友人・知人
「だ」アリ	149 (90.9%)	446 (90.7%)
「だ」ナシ	15 (9.1%)	46 (9.3%)

最後に、年齢との関係について表8に示した。年齢の分析では、コーパスに年齢データが含まれていなかった3名を除いた計738名のデータを使用した。全体的には年齢が若いほど「だ」アリが多くなる傾向が見受けられるが、40歳代(または50歳代)と70–80歳代はこの傾向に当てはまらない。70–80歳代の割合は30歳代と同程度であるが、この年代での出現頻度が少ないため、傾向を正確に把握するのが難しい可能性がある。また、全体の傾向として年齢と「だ」の使用の相関を捉えるために、年齢を主効果としたロジスティック回帰分析を行ったところ、その傾向に統計的に有意な相関が確認された(推定値 $=-0.02$, 標準誤差 $=0.008$, $z=-2.99$, $p<.01$)。これは田野村(2008)と Nambu (2023) で報告されていた、コピュラなしからコピュラありへという言語変化の方向性と一致している。

表8 CEJC, 年齢と「と思う」節内の「だ」アリ・ナシの頻度

	10 歳代	20 歳代	30 歳代	40 歳代	50 歳代	60 歳代	70–80 歳代
「だ」アリ	41 (100%)	122 (96.8%)	119 (90.2%)	123 (87.9%)	130 (89.7%)	105 (86.1%)	29 (90.6%)
「だ」ナシ	0 (0%)	4 (3.2%)	13 (9.8%)	17 (12.1%)	15 (10.3%)	17 (13.9%)	3 (9.4%)

4.2 ロジスティック回帰分析

次に、4.1 節で議論した要因を全てモデルに含めたロジスティック回帰分析を行い、個人差の影響を考慮するためにランダム効果としてランダム切片に話者を設定した混合効果モデルを構築した。その結果を表9にまとめた。

表9 CEJC, 「だ」アリ・ナシについてのロジスティック回帰分析の結果

	推定値	標準誤差	z 値	p 値
切片	4.18	0.83	5.04	<.001
年齢	-0.03	0.01	-2.07	<.05
直前要素				
名詞	-0.29	0.41	-0.69	.49
性別				
男性	1.27	0.46	2.76	<.01
会話相手				
その他	-0.23	0.56	-0.41	.68
仕事関係	0.03	0.42	0.07	.94

表9が示すように、回帰分析の結果は前節でのクロス表による分析を裏付けるものとなった。具体的には、年齢の効果と男女差に統計的に有意な差が確認された。年齢については、若いほど「だ」の使用がより多くなることが示され、男女差では男性の方が女性より「だ」の使用が多い傾向が支持される結果となった。

5. 議論

「(だ) と思う」について、先行研究での報告と4節のCEJCの分析結果から、以下の傾向が指摘できる。

(8) 「(だ) と思う」の傾向

- 経年的な変化として、「だ」ナシから「だ」アリへと推移している可能性がある
- 話し言葉において、発話場面のフォーマリティは「だ」の使用に影響しない
- 書き言葉と話し言葉には大きな差があり、話し言葉の方で「だ」がより多く使用される
- 日常会話では、「だ」の使用は直前の要素（名詞・形容動詞）の違いによる影響を受けない

e. 日常会話では、男性の方が女性より「だ」を多く使用する

まず、該当する言語環境での「だ」の使用については、「だ」ナシから「だ」アリへの経年的な変化が見られることが、先行研究での報告に加えて、見かけ上の時間の概念を用いた今回の年齢による分析でも示された。田野村（2008）は、「「問題だと思う」のように明示的なコピュラを伴う表現が元来の形であり、「問題ゆと思う」はそこにあったコピュラが端折られた表現であろうと解釈した。」（p. 61）と、田野村自身の分析前の直感を述べている。この直感を支持するデータとして、2 節に挙げた、塩田（2010）と Nambu（2023）の評価に関する調査では、「だ」ナシ（「問題と思う」）より「だ」アリ（「問題だと思う」）の方がより正しいと現代では認識されていることがわかっている。一方で、田野村（2008）に続いて行われた Nambu（2023）の CHJ と BCCWJ を用いた通時的分析では、かつては「だ」ナシ（「問題と思う」）の方が多く使用されていたという事実が示されており、現代の規範意識と過去の実際の用法との間にはねじれが生じていると言える⁷。また、「だ」アリが正しいという現代の規範意識に基づいて今回行ったフォーマリティの分析では、話し言葉の CSJ と CEJC の比較および CEJC の話し相手との関係性（仕事関係（フォーマル）vs. 家族／親戚／友人・知人（カジュアル））の比較から、発話場面のフォーマリティは「だ」アリ・ナシに影響を与えないことが示された。そのため、正しくないと評価される「だ」ナシの形式が、フォーマリティ以外の要因によってどのような場面でより多く現れるのかについて、今後調査を進める必要がある。

次に、書き言葉と話し言葉を比較すると、話し言葉の方で「だ」がより多く使用されるという結果が示された。これは、レジスターの違いを反映している可能性もあるが、書き言葉のデータとして使用された BCCWJ には書き手の生年が 1900 年代からのデータが含まれており（Nambu 2023）、話し言葉の CSJ や CEJC と比較するとかなり古い。したがって、「だ」ナシから「だ」アリへの経年変化がこれらコーパス間の違いを生み出している可能性も考えられることから、書き言葉と話し言葉というレジスターの違いが与える影響については、今後より詳細な分析を行う必要がある。

最後に、「だ」の直前の要素（名詞・形容動詞）と男女差に関して、CEJC のデータでは先行研究での報告とは異なる結果が確認された。具体的には、名詞と形容動詞の場合には想定されていた有意な差は見られず、男女差については女性の方が「だ」をより多く使用するという BCCWJ と CSJ を用いた Nambu（2023）の報告とは逆の傾向が見られた。今後、これらの結果に差異が生じた原因について検討する必要がある。名詞と形容動詞という言語環境における差異については、それぞれのコーパスに現れた名詞と形容動詞の語彙の種類を分析し、「だ」の使用に偏りが見られるか検討することが挙げられる。また、心理言語学的手法を用いた容認度実験を行ってその効果を検証することも考えられる。男女差に関しては、なぜ今回異なる結果が観察さ

⁷ この点について、Nambu（2023）は、過去にはコピュラがない形式がデフォルトであったが、その後コピュラが現れるようになり、さらに現代ではそのコピュラが省略される場合がある、という言語変化の仮説を述べている（p. 12、詳細は Nambu（2023）を参照のこと）。

れたのか、特に性別がどのような言語使用の指標となっているのかについて検討する必要がある。例えば、男女差をジェンダーという視点から捉えたと、ジェンダーは、話者が所属するさまざまな実践共同体 (communities of practice) における多面的な実践を通じて形成される社会的産物とされる (Eckert and McConnell-Ginet 1992, Meyerhoff 2014)。したがって、今回観察されたジェンダーの影響を解釈するには、その背景にある言語の規範意識や、ジェンダーと関わる社会的実践が言語行動にどのように影響しているのかなどについて検討する必要がある。

6. おわりに

本稿では、日本語日常会話コーパスを用いて、「と思う」に先行する「だ」の出現・非出現という変異現象の量的分析を行った。その結果、(i) 「だ」ナシから「だ」アリへと推移する言語変化が観察されたこと、(ii) 発話場面のフォーマリティは「だ」の使用に影響しないこと、(iii) 書き言葉と比べて話し言葉の方で「だ」がより多く使用されること、(iv) 日常会話では「だ」の使用は直前の要素 (名詞・形容動詞) の違いに影響されないこと、(v) 日常会話では男性の方が女性より「だ」を多く使用すること、が明らかになった。これらの結果には先行研究での報告とは異なる点も含まれるため、今後、コーパス分析に加えて実験など多様な手法を用いることで、今回検証した要因についてより詳細に考察する必要がある。

参考文献

- 阿部二郎 (2001) 「「A ヲ B ダト思ウ」と「A ヲ B ト思ウ」」『日本語と日本文学』33: 14–24.
- 阿部二郎 (2015) 「引用句内におけるコピュラの非出現について：「～だと思う」と「～と思う」」阿部二郎・庵功雄・佐藤琢三 (編) 『文法・談話研究と日本語教育の接点』57–78. 東京：くろしお出版.
- Chomsky, Noam (1981) *Lectures on government and binding theory*. Dordrecht: Foris Publications.
- Cukor-Avila, Patricia and Guy Bailey (2013) Real time and apparent time. In: J. K. Chambers and Natalie Schilling (eds.) *The handbook of language variation and change*, 237–262. Oxford: Wiley-Blackwell.
- Eckert, Penelope and Sally McConnell-Ginet (1992) Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21: 461–490.
- 小磯花絵 (2022) 「書き言葉・話し言葉における縮約形の実態—コーパスに基づく分析を通して—」窪蘭晴夫・朝日祥之 (編) 『言語コミュニケーションの多様性』61–78. 東京：くろしお出版.
- Koiso, Hanae, Haruka Amatani, Yasuharu Den, Yuriko Iseki, Yuichi Ishimoto, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, Yuka Watanabe and Yasuyuki Usuda (2022) Design and evaluation of the Corpus of Everyday Japanese Conversation. *Proceedings of LREC2022*, 5587–5594.
- Kuznetsova, Alexandra, Per B. Brockhoff and Rune H. B. Christensen (2017) lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1–26.
- Labov, William (1966) *The social stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.
- Maekawa, Kikuo (2004) Design, compilation, and some preliminary analyses of the Corpus of Spontaneous Japanese. In: Kikuo Maekawa and Kyoko Yoneyama (eds.) *Spontaneous speech: Data and analysis*, 87–108. Tokyo: The National Institute of Japanese Language.
- Maekawa, Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka and Yasuharu Den (2014) Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation* 48(2): 345–371.
- Meyerhoff, Miriam (2014) Variation and gender. In: Susan Ehrlich, Miriam Meyerhoff and Janet Holmes (eds.) *The handbook of language, gender, and sexuality*, 87–102. Oxford: Wiley-Blackwell.
- Nambu, Satoshi (2023) A quantitative study on zero copula in Japanese. *Language Sciences* 96. Online first. <https://doi.org/10.1016/j.langsci.2023.105000>.

org/10.1016/j.langsci.2022.101534

R Core Team (2022) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

塩田雄大 (2010) 「「だ抜きことば」？」『最近気になる放送用語』NHK 放送文化研究所. <http://www.nhk.or.jp/bunken/summary/kotoba/term/140.html> (2024 年 12 月 1 日アクセス)

Tagliamonte, Sali (2012) *Variationist sociolinguistics: Change, observation, interpretation*. Oxford: Wiley-Blackwell.

高野照司 (2005) 「言語変異理論再考：「相関主義」という誤解と説明能力」『北星論集』42(2): 37–54.

高野照司 (2011) 「バリエーション研究の新たな展開」『日本語学』30(14): 256–275.

田野村忠温 (2008) 「大規模な電子資料に見る現代日本語の動態」『待兼山論叢』42: 55–77.

Yamaguchi, Toshiko (2007) *Japanese linguistics: An introduction*. London: Continuum.

関連 Web サイト

国立国語研究所 (2024) 『日本語日常会話コーパス』(バージョン 2023.3, 中納言バージョン 2.7.2) <https://clrd.ninjal.ac.jp/cejc/> (2024 年 11 月 21 日確認)

Use/Non-use of Copula Using the Corpus of Everyday Japanese Conversation

NAMBU Satoshi

Monash University / Project Collaborator, NINJAL

Abstract

From the perspective of variation theory, this study examines the presence and absence of the copula *da* within *to omou* “think that” clauses as a variation phenomenon, conducting a quantitative analysis using the Corpus of Everyday Japanese Conversation. Previous studies utilizing the Balanced Corpus of Contemporary Written Japanese and the Corpus of Spontaneous Japanese have shown that spoken Japanese exhibits a higher frequency of *da* usage than written Japanese, that females use *da* more frequently than males, and that a quantitative shift from the absence to the presence of *da* indicates a linguistic change. This study investigates whether these trends are similarly observed in everyday conversations. The results revealed that, similar to the Corpus of Spontaneous Japanese, which includes data from academic lectures, everyday conversations indicate a high frequency of *da* usage, thus highlighting a marked difference from written Japanese. Regarding language change, the analysis using age as a variable based on the apparent-time concept confirmed a shift toward the use of *da* in everyday conversations, which is consistent with previous studies. By contrast, a trend opposite to those of previous studies is observed for gender differences, and the analysis of formality based on conversation partners (i.e., work relations vs. family/relatives/friends) indicate no statistically significant influence on the use of *da*.

Keywords: zero copula, *da*-deletion, CEJC, everyday conversation, language variation