

国立国語研究所学術情報リポジトリ

複数のコーパスを利用した言葉のバリエーションの研究：言語外的要因の探索

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2025-01-24 キーワード (Ja): コーパス, い抜き言葉, がの交替, がを交替, 言語外的要因 キーワード (En): corpus, i-deletion, nominative/genitive alternation, nominative/accusative alternation, language-external factors 作成者: 南部, 智史 メールアドレス: 所属:
URL	https://doi.org/10.15084/0002000454

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



複数のコーパスを利用した言葉のバリエーションの研究 ——言語外的要因の探索——

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、異なる特徴を持つコーパス間の比較とコーパスで利用可能な場面や話者（または書き手）の情報を活用することで、バリエーション現象における言語形式の選択に関わる言語外的要因についてどのような側面から分析可能か探索的に考察した。3種類のバリエーションを分析した結果、(i) コーパス間の比較と現代日本語書き言葉均衡コーパス (BCCWJ) のジャンル間の比較による書き言葉と話し言葉の差異、(ii) 年齢または生年情報を用いた言語変化および年齢差 (age-grading)、(iii) 日本語話し言葉コーパス (CSJ) の改まり度と日本語日常会話コーパス (CEJC) の発話場面を利用したフォーマリティ、(iv) 性差、について分析できる可能性を示した。また、CEJC の会話の録画映像の分析から、コーパスが持つ特性を活かした多角的な研究の可能性についても言及した*。

キーワード：コーパス、い抜き言葉、がの交替、がを交替、言語外的要因

1. はじめに

本稿では言葉のバリエーションの使用を取り巻く言語外的要因を探究するプロジェクトの一環として行った予備調査の結果を報告する。本稿の概念的枠組みには変異理論 (Variation Theory, Variationist Sociolinguistics, Labov 1966, 高野 2005, 2011, Tagliamonte 2012) を採用し、その定量的手法や言語変化に関する議論を参考にした。変異理論では、バリエーションにおける言語形式の選択に関わる言語内的要因だけでなく、(社会集団としての) 話者の属性などの社会的要因を含む言語外的要因も分析対象とし、また、進行中の言語変化と社会との関わりについても数多くの研究が行われている (Weinreich et al. 1968, Chambers 2013)。本稿では、分析の対象に進行中の言語変化が確認されているバリエーションを含めることで、言語変化の進行と社会の関係についても探索的に考察する。調査のためのデータには、現代日本語書き言葉均衡コーパス (BCCWJ, Maekawa et al. 2014), 日本語話し言葉コーパス (CSJ, Maekawa 2004), 日本語日常会話コーパス (CEJC, Koiso et al. 2022) の3つのコーパスを利用した。異なる特徴を持つコーパス間の比較やコーパスで利用可能な場面や話者（または書き手）などの情報を活用することで、複数の角度から言語外的要因の効果について検討する。また、CEJC で利用可能な会話の録画映像の分析から、コーパスが持つ特性を活かした多角的な研究の可能性についても言及する。

* 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) の研究成果であり、2024年3月18日に行われたプロジェクト発表会「日常会話コーパスを利用した言葉の変異と変化研究」での発表「日常会話コーパスを利用した言葉の変異と変化研究」を基にしている。

分析対象のバリエーション現象は、(1) に挙げた、「ている／てる」の交替（い抜き言葉）、主語につく助詞「が／の」の交替（「がの」交替）、目的語につく助詞「が／を」の交替（「がを」交替）の3種類とした。

- (1) a. い抜き言葉： ねこがご飯を食べ |ている／てる|。
- b. がの交替： 昨日部長 |が／の| 持ってきた土産
- c. がを交替： 母は英語 |が／を| 話せる。

これらのバリエーションにおける言語形式の選択に言語外的要因がどの程度関わっているかについては、これまであまり議論されてこなかった¹。しかし、近年の多様な大規模コーパスの登場により、言語使用を取り巻く様々な状況を定量的視点から多角的に捉えることが比較的容易になった。この点と関連して、本稿で3種類の現象を扱うことにした理由は、コーパスで利用可能な情報を用いることで、言語外的要因をどの程度そしてどのような側面から分析可能か探索的に考察することにある。

2. 本稿で扱うバリエーション現象について

い抜き言葉とも呼ばれる「ている／てる」のバリエーションについては、これまでに言語環境ごとの頻度調査が行われており（石川 1958, 坂梨 2002, 宮島 2006, 新井 2009），例えば動詞の語幹の長さ（短いほど「てる」使用率が上がる）や動詞の活用（促音便で「てる」使用率が上がる）といった要因の効果が明らかとなっている（新井 2009）。また、「てる」は江戸時代後期から使用が確認されているだけでなく、「浮世床」といった当時の作品において「てる」の使用は地の文ではなく会話文によく現れるといった、書き言葉と話し言葉の傾向についても指摘されている（坂梨 2002）。小磯（2022）では、「ている／てる」を含めた10種類の縮約形について、BCCWJ, CSJ, CEJC の3つのコーパスを用いて分析を行っており²、書き言葉より話し言葉で「てる」の使用率が高いこと、書き言葉では「てる」の使用率が、小説会話>ブログ>書籍>白書の順に高く、話し言葉では、日常会話>模擬講演>学会講演の順に高いという結果を報告している。次節で説明するが、本稿ではこの結果に加えて別の視点からも比較を行う。

「がの」交替および「がを」交替は、統語理論の観点から交替が可能な統語条件など言語環境に関する研究がこれまでに数多く行われている³。「がの」交替では、例えば「学生が来た。」のような主節では「が」の代わりに「の」が現れることがないように、「が」主語と比べて「の」主語の統語環境に制約があることが知られており、また、「学生 |が／の| 多いクラス」のような名詞修飾節が「の」主語の典型的な生起環境であることがこれまでのコーパス調査で明らかくなっている（南部 2007）。「がの」交替の言語変化に関しては、Harada (1971) が「の」主語か

¹ ただし、後述するように小磯（2022）の縮約形の調査がある。

² 小磯（2022）では本稿と異なり、い抜き言葉に関してはCSJのコアデータ、CEJCではモニター版に含まれる100時間分のデータを用いて調査を行っている。

³ 「がの」交替の統語研究の文献リストについては Nambu (2019), 「がを」交替については Nambu et al. (2018) に挙げられている文献リストを参照されたい。

ら「が」主語への変化の可能性を指摘して以来、コーパスを用いた定量的分析により「の」主語の使用が徐々に減少しているという進行中の変化が確認されている（南部 2007, Nambu 2019, Ogawa 2018）。

「がを」交替については、「が／を」目的語をとる述部の種類が助詞の選択に大きく影響することが知られている。Nambu et al. (2022) のコーパス調査では、「英語 {が／を} わかる」のような語彙述語では「が」目的語が好まれる強い傾向がある一方で、「会社 {が／を} 買収したい」のような「—たい」がつく願望を表す述部では他の述部の場合と比べて「を」目的語が好まれる傾向が明らかとなっている⁴。言語変化に関しては、可能の「れる／られる」と可能動詞において「が」目的語から「を」目的語へ変化している可能性に言及した渋谷 (1993) に基づいて Nambu et al. (2022) が行った調査では、可能を表す述部や語彙述部の場合には徐々に「が」から「を」へと推移している傾向を報告している。本稿では Nambu et al. (2022) で扱われた調査対象のうち、前述のように他の述部とは異なるパターンを示す「—たい」を除いて、可能を表す述部と語彙述部（「分かる」「出来る」「欲しい」「好き」「嫌い」の5つ）を調査対象とした。

3. 方法論

前述のように本稿のデータにはBCCWJ, CSJ, CEJCの3つのコーパスを利用した。BCCWJには、白書、新聞、雑誌やオンラインのブログなど様々なジャンルの書き言葉から約1億語が収録されている。CSJには661時間分の発話の書き起こしデータ約750万語が収録されており、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）がある。文章朗読のデータも含まれているが、本稿の分析対象からは除外した。また、CSJの発話には改まり度評定（5段階）のデータ（単独評定データ）も付与されている。CEJCには200時間におよぶ多様な状況での日常会話の書き起こしデータ約240万語が会話の録画映像とともに収録されている。今回の調査の規模などを考慮して、い抜き言葉と「がの」交替の調査にはBCCWJのコアデータに分類されている約110万語のデータを利用し、また、「がの」交替のCSJデータにはCSJのコアデータ約50万語を利用した。「がを」交替のBCCWJとCSJのデータは、Nambu et al. (2022) で用いられたデータセットを利用した。Nambu et al. (2022) では言語変化の分析に焦点を当てていたため、「がを」交替のBCCWJのデータには書き手の生年が判明している場合（書籍や雑誌など）のみが含まれている。

各現象のデータは、コーパス検索アプリケーション「中納言」(<https://chunagon.ninjal.ac.jp/>)から抽出した。抽出対象とした言語環境は、い抜き言葉のバリエーションでは動詞に後続する「ている」と「てる」、「がの」交替では「(格助詞の) が／の－動詞／形容詞－名詞」からなる名詞修飾節、「がを」交替では「(格助詞の) が／を－述部（述部の種類は2節の議論を参照）」である。データを抽出した後、「がの」交替と「がを」交替については項と述語の関係を形成していないなど当該現象にあたらぬものをデータから除外した。また、本稿では言語外的要因の分析を主

⁴ 願望表現の格助詞の歴史的変遷については山田（2018）を参照されたい。

眼としているため、言語内的要因については分析対象とせずになるべく統制することとした。まず、「がの」交替と「がを」交替では、それぞれ「が／の」主語・「が／を」目的語とその述語の間の介在要素が助詞の選択に影響を与えることがわかっているため (Harada 1971, Shibatani 1975, Nambu et al. 2018, Nambu and Nakatani 2023), 本稿の分析対象は「が／の」主語・「が／を」目的語が述語と隣接している場合のみとした。また、「がの」交替では「の」主語は述部が形容詞と動詞の場合に最も頻繁に見られることから (南部 2007), 量的に十分なデータが得られるこれらの述部を分析対象とした。

分析では、まず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用いて調査した。その後、各要因の効果をより正確に評価するため、言語内的および外的要因を含めたロジスティック回帰分析を行い、その際には個人差を考慮するため、ランダム効果としてランダム切片に話者（または書き手）を設定した混合効果モデルを構築した⁵。統計分析には R 環境 (R Core Team 2022) の lmerTest パッケージ (Kuznetsova et al. 2017) を使用した。今回の分析に用いた言語外的要因を以下にまとめた。

(2) 分析に用いた言語外的要因

(i) 書き言葉 vs. 話し言葉

- a. コーパス間の比較 (BCCWJ vs. CSJ/CEJC)
- b. BCCWJ のジャンル間比較

(ii) フォーマリティ

- c. CSJ の改まり度 (5 段階)⁶
- d. CEJC の場面の比較 (「職場で会議」 vs. 「家で家族／親戚／友人と雑談」)
- e. CSJ の独話 vs. 対話

(iii) 性別

(iv) 年齢 (BCCWJ では生年)

(i) の BCCWJ のジャンル間比較では、話し言葉の特徴が入りやすいジャンル (Yahoo! 知恵袋およびブログ、雑誌) とそうではないジャンル (白書、新聞) を比較する。(ii) のフォーマリティはフォーマルな場面とカジュアルな場面の発話の比較になるが、言語使用に対するフォーマリティの影響には、話者が自身の発話により注意を向けているフォーマルな状況には規範に則した言葉遣いがより頻繁に見られるという点が挙げられる ('attention to speech', Labov 1972)。CEJC には多様な場面の会話が収録されているが、本稿ではその中でフォーマリティの差異が明確である 2 つの場面、「職場で仕事相手と会議」(フォーマル) と「家で家族／親戚／友人と雑談」(カジュアル) を比較することとした。CSJ の独話と対話には、対話者ありなしの違いの他に、学会講演

⁵ 分析に先立ち、データのコーディングとして「ている／てる」の場合は「ている」を 0、「てる」を 1、「がの」交替では「が」主語を 0、「の」主語を 1、「がを」交替では「が」目的語を 0、「を」目的語を 1 に変換した。

⁶ CSJ の対話データの印象評定は、「対話開始後 0 分から 3 分まで」、「4 分から 7 分まで」、「8 分以降」の 3 箇所に記録されているため、分析にはそれらの平均値を用いた。

のように内容をあらかじめ準備する独話ではインタビューのような対話よりも自身の発話への注意が高まると考えられるため、(ii) のフォーマリティに含めることとした。(iii) の性別は、変異理論におけるこれまでの研究において、女性が男性よりも規範形式を多く使用する傾向が様々な言語に観察されており（例えばアメリカ英語のINGバリエーション (-ing, -in), Labov 2001），また、言語変化の進行過程と性差の関連も指摘されていることから（Labov 1990, Meyerhoff 2014），分析に含めることとした。(iv) の年齢の効果には、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。見かけ上の時間は変異理論で考案された手法で、同時期に観察される異なる年齢層の言語使用からそれぞれの言語獲得期における言語変化の進行具合を推定するものであり、年齢差（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）とは区別される。年齢（BCCWJの場合生年）の効果の検証は、4.3節のロジスティック回帰分析で行う。

回帰分析の際にモデルに含めた言語内的要因は、い抜き言葉では動詞のモーラ数（長いほど「てる」率減少）と促音便あり／なし（促音ありで「てる」率上昇）（新井 2009）、「がの」交替では述部の種類（「が」主語率：動詞>形容詞）（南部 2007）、「がを」交替では述部の種類（「を」目的語率：可能動詞 / 可能「れる・られる」>動名詞 + 「できる」>語彙述部）（Nambu et al. 2022）である。

4. 分析結果

4.1 書き言葉と話し言葉の比較

コーパスから得られた各現象の言語形式ごとの件数を表1にまとめた。

表1 3つのコーパスで観察された各現象の言語形式の件数

	BCCWJ	CSJ	CEJC
い抜き言葉			
「ている」	88.8% (9,126/10,273)	47.0% (30,237/64,302)	2.7% (853/31,169)
「てる」	11.2% (1,147/10,273)	53.0% (34,065/64,302)	97.3% (30,316/31,169)
がの交替			
「が」	48.8% (1,233/2,526)	62.3% (308/494)	79.1% (1,413/1,787)
「の」	51.2% (1,293/2,526)	37.7% (186/494)	20.9% (374/1,787)
がを交替			
「が」	84.0% (4,037/4,805)	87.4% (2,314/2,648)	90.4% (1,144/1,266)
「を」	16.0% (768/4,805)	12.6% (334/2,648)	9.6% (122/1,266)

表1の全体の傾向を見ると、全てのバリエーション現象においてBCCWJ, CSJ, CEJCの順で出現頻度のきれいな傾向が確認できる。まず、い抜き言葉に関しては、小磯（2022）で確認されていたように書き言葉（BCCWJ）では「ている」が多く、反対に日常会話（CEJC）では「てる」がほとんどを占めている ($\chi^2(2, N=105,744)=29,932, p<.001$)。「がの」交替については、「が」主語は書き言葉より話し言葉で多くなることがわかった ($\chi^2(2, N=4,807)=401.71, p<.001$)。「がを」交替では、「を」目的語の使用が話し言葉より書き言葉が多い ($\chi^2(2, N=8,719)=39.85, p<.001$)。

次にBCCWJのデータをジャンル別に分けて表2に示した。3節でも述べたが、「がを」交替はデータ抽出の基準によりデータにジャンルの偏りがあったため、ここでの分析から除外した。

表2 BCCWJのジャンル間比較

	白書	新聞	書籍	雑誌	Yahoo! 知恵袋 およびブログ
い抜き言葉 「てる」率	0% (0/1,579)	1.1% (24/2,233)	4.5% (106/2,336)	10.2% (194/1,908)	37.1% (823/2,217)
がの交替 「が」率	43.1% (233/540)	59.7% (360/603)	41.3% (217/525)	45.0% (213/473)	54.5% (210/385)

表2では、い抜き言葉の方は書き言葉の中でも話し言葉の特徴が反映されやすいジャンルにおいて「てる」使用率が高くなる傾向が見られるが、「がの」交替ではそういった傾向は見られない。

4.2 フォーマリティと性差について

まず、CSJに付与されている発話の改まり度ごとのデータを表3に示した。

表3 CSJの発話の改まり度に基づく比較

	くだけた	ややくだけた	普通	やや改まった	改まった
い抜き言葉 「てる」率	80.1% (2,613/3,262)	71.5% (10,237/14,310)	50.9% (16,106/31,645)	36.4% (4,250/11,680)	25.2% (859/3,405)
がの交替 「が」率	76.9% (10/13)	55.4% (36/65)	62.8% (118/188)	64.1% (82/128)	64.1% (25/39)
がを交替 「を」率	2.9% (2/69)	11.9% (46/386)	13.6% (159/1,166)	14.1% (81/576)	13.7% (21/153)

表3から、い抜き言葉ではくだけた発話ほど「てる」率が高くなる傾向が明らかとなった。その傾向について、改まり度を説明要因に含むロジスティック回帰分析を行ったところ、統計的に有意な相関が確認された ($z = -69.01, p < .001$)。また、「がを」交替の場合にも、改まった発話ほど「を」目的語の使用が増えるという傾向が明らかとなった ($z = -2.54, p < .05$)。一方で、「がの」交替ではそのような相関は見られなかった ($z = -0.43, p = .67$)。

次に、CEJCに収録された多様な状況での会話の中から、フォーマリティの差異が明確な「職場で仕事相手と会議」(フォーマル)と「家で家族／親戚／友人と雑談」(カジュアル)の2つの場面を表4に示した。

表4 CEJCのフォーマル・カジュアルな場面の比較

	職場で仕事相手と会議 [フォーマル]	家で家族／親戚／友人と雑談 [カジュアル]
い抜き言葉 「てる」率	92.0% (752/817)	98.6% (9,683/9,818)
がの交替 「が」率	88.5% (46/52)	77.4% (380/491)
がを交替 「を」率	14.6% (7/48)	7.1% (23/324)

表4では、い抜き言葉についてはカジュアルな場面で「てる」使用率がより高くなることが示され、統計的有意差も確認された ($\chi^2(1, N=10,635)=173.48, p<.001$)。一方、「がの」交替と「がを」交替ではフォーマルな場面でそれぞれ「が」主語、「を」目的語がより多く現れる傾向が見られたが、2つの場面に有意差は確認されなかった（「がの」： $\chi^2(1, N=543)=2.78, p=.10$ ；「がを」： $\chi^2(1, N=372)=2.23, p=.14$ ）。

表5に、CSJデータを独話と対話に分けて示した。

表5 CSJの独話と対話の比較

	独話	対話
い抜き言葉 「てる」率	52.1% (32,692/62,738)	87.8% (1,373/1,564)
がの交替 「が」率	63.1% (292/463)	51.6% (16/31)
がを交替 「を」率	12.7% (332/2,616)	6.3% (2/32)

い抜き言葉では対話において「てる」がより多く観察されたのに対して ($\chi^2(1, N=64,302)=778.34, p<.001$)、「がの」交替と「がを」交替ではそのような差は見られなかった（「がの」： $\chi^2(1, N=494)=1.17, p=.28$ ；「がを」： $\chi^2(1, N=2,648)=0.68, p=.41$ ）。

表6～8にコーパスごとの性別のデータを示した。BCCWJのデータについては、複数の著者の場合や著者の性別情報が付与されていない場合を除外してある。

表6 BCCWJの性差

BCCWJ	女性	男性
い抜き言葉 「てる」率	10.1% (118/1,162)	2.2% (41/1,830)
がの交替 「が」率	38.9% (96/247)	43.9% (187/426)
がを交替 「を」率	15.8% (147/929)	15.9% (607/3,808)

表7 CSJの性差

CSJ	女性	男性
い抜き言葉 「てる」率	48.4% (13,454/27,822)	56.5% (20,611/36,480)
がの交替 「が」率	57.8% (134/232)	66.4% (174/262)
がを交替 「を」率	11.6% (109/943)	13.2% (225/1,705)

表8 CEJCの性差

CEJC	女性	男性
い抜き言葉 「てる」率	97.1% (16,783/17,280)	97.4% (13,533/13,889)
がの交替 「が」率	80.8% (723/895)	77.4% (690/892)
がを交替 「を」率	9.0% (57/635)	10.3% (65/631)

表6～8で有意差が確認されたのは、い抜き言葉のBCCWJ ($\chi^2(1, N=2,992)=86.91, p<.001$) とCSJ ($\chi^2(1, N=64,302)=419.71, p<.001$) であり、「がの」交替と「がを」交替ではどのコーパスにおいても性差が確認されなかった。また、い抜き言葉のBCCWJデータでは女性の方が「てる」率が高いのに対して、CSJでは男性の方が「てる」率が高いという結果になっており、性差の効果にねじれが生じている。この点については次節の回帰モデルの結果も合わせて5節で議論する。

4.3 ロジスティック回帰分析

4.2節で議論した言語外的要因に加えて話者の年齢（BCCWJの場合は書き手の生年）、それから3節に挙げた各現象の言語内的要因（い抜き言葉：動詞のモーラ数（長いほど「てる」率減少）と促音便あり／なし（促音ありで「てる」率上昇）（新井2009）；「がの」交替：述部の種類（「が」主語率：動詞>形容詞）（南部2007）；「がを」交替：述部の種類（「を」目的語率：可能動詞／可能「れる・られる」>動名詞+「できる」>語彙述部）（Nambu et al. 2022））をモデルに含めてロジスティック回帰分析を行った⁷。その際、個人差の影響を考慮するため、ランダム効果としてランダム切片に話者（または書き手）を設定した混合効果モデルを構築した。BCCWJではジャンルによって利用できる書き手の情報（年齢、性別）が偏っていたため、ジャンルは分析から除外した。

回帰分析の結果の詳細は煩雑にならぬよう付録に掲載し、以下に有意な効果が認められた要因を現象ごとにまとめた。

表9 い抜き言葉で「てる」使用率が有意に高かった要因⁸

BCCWJ	CSJ	CEJC
●性別：女性>男性	●年齢：高い>低い ●性別：男性>女性 ●発話：対話>独話 ●改まり度：低い>高い ●語幹のモーラ数：少ない>多い ●促音：あり>なし	●会話相手： 「家族／親戚／友人」>「仕事関係」 ●語幹のモーラ数：少ない>多い ●促音：あり>なし

⁷回帰分析の際は、分析に使用できるデータ数を考慮してCEJCの場面比較には会話場所は含まず会話相手のみを要因に含めることとした。

⁸CSJデータにおいて、い抜き言葉は学会講演よりも模擬講演でより多く観察される（小磯2022）で報告されている。また、学会講演では年齢が高い発話者が極端に少ない。これらが交絡要因として年齢差の分析に影響を与えていないか確認するため、独話を学会講演と模擬講演に細分化した要因を組み込んだ分析も行った。その結果、付録表Bで観察された全ての要因の傾向は変わらず、年齢の効果も有意であった（推定値=0.04, z値=12.06, p値<.001）。

表10 「がの」交替で「が」使用率が有意に高かった要因

BCCWJ	CSJ	CEJC
●生年：若いほど「が」	●年齢：低い>高い	●性別：女性>男性
●性別：男性>女性	●述部：動詞>形容詞	●述部：動詞>形容詞
●述部：動詞>形容詞		

表11 「がを」交替で「を」使用率が有意に高かった要因

BCCWJ	CSJ	CEJC
●生年：若いほど「を」	●述部：動名詞>可能>語彙	●年齢：若いほど「を」
●性別：女性>男性		●述部：動名詞 / 可能>語彙
●述部：動名詞>可能>語彙		

結果は前節でのクロス表による分析と概ね一致しているが、「がを」交替において CSJ の改まり度の効果が回帰分析では確認できなかった点と、「がの」交替と「がを」交替において回帰分析では性差が確認された点に相違が見られた。ただし、性差についてはどの現象でもコーパス間で一様な効果は見られなかった。

5. 議論

4 節のクロス表による分析とロジスティック回帰分析の結果から 3 つの現象を比較すると、い抜き言葉は「がの」交替や「がを」交替と比較して言語外的要因の影響を受けやすいと言える。以下では 3 つの現象についてそれぞれ見ていく。まず、い抜き言葉の傾向を（3）にまとめた。

(3) い抜き言葉の傾向

- 書き言葉と話し言葉（特に日常会話）で真逆の傾向
- フォーマルな発話では「ている」、カジュアルでは「てる」
- 年齢が上がると「てる」使用率が上がる
- 性別の効果はあるが一定ではない

江戸時代の作品で確認されていたように（坂梨 2002），い抜き言葉には書き言葉と話し言葉に大きな差があり、コーパスデータに見られる極端な傾向から、書き言葉では「ている」、話し言葉（特に日常会話）では「てる」という 2 種類のスタイルがステレオタイプとして確立されているようと思われる。フォーマリティに関しても強い傾向が見られるものの、日常会話においてはフォーマルな場面でも「てる」使用率が 90% を超えており、その圧倒的な優位性は変わらなかった。CSJ のデータにおいて年齢の効果が観察されたが、い抜き言葉に対するフォーマリティの効果を考慮すると、「見かけ上の時間」で考えられるような「てる」から「ている」へという言語変化を反映しているというよりも、例えば年齢が上がるとともに、規範に則した言語使用が求められるという社会的抑圧が小さくなることによる年齢差（age-grading）の可能性が考えられるが、今後の調査で検証する必要がある。性別の効果については、BCCWJ のデータでは女性が男性より「てる」率が高いのに対して、CSJ では男性の方が「てる」率が高いといった性差の効果のねじれが観察された。コーパス間で見られた性差の異なる効果については、今後さらに検証していく必要がある。

次に「がの」交替と「がを」交替の傾向についてまとめて見ていく。

(4) 「がの」交替の傾向

- a. フォーマリティの効果なし
- b. 年齢（生年）が若いほど「が」主語
- c. 書き言葉では「の」主語、話し言葉では「が」主語使用の傾向
- d. 性別の効果はあるが一定ではない

(5) 「がを」交替の傾向

- a. フォーマリティの効果なし
- b. 年齢（生年）が若いほど「を」目的語
- c. 書き言葉では「を」目的語、話し言葉では「が」目的語使用の傾向
- d. 性別の効果は限定的ではあるが、女性の方が男性より「を」目的語を多く使用

い抜き言葉と異なり、「がの」交替と「がを」交替では CSJ の改まり度と CEJC の日常会話の場面比較においてフォーマリティの効果が見られなかった。年齢または生年の効果は、「がの」交替 (BCCWJ, CSJ) と「がを」交替 (BCCWJ, CEJC) それぞれで進行中の言語変化の方向性と一致していることから、観察された結果は年齢差 (age-grading) ではなく、「の」主語から「が」主語へ、「が」目的語から「を」目的語へという言語変化の過程を反映していると考えられる。「がの」交替と「がを」交替でも書き言葉と話し言葉に違いが見られたが、い抜き言葉のように真逆の傾向とまではならなかった。「がの」交替の場合、話し言葉では「が」主語が 8 割ほどであるのに対して、書き言葉では話し言葉と比べて「が」主語が抑えられ「の」主語がより多く現れている。進行中の言語変化（「の」から「が」への変化）を考慮すると、この傾向は書き言葉は「がの」交替の変化に保守的であることを意味する。一方、「がを」交替の場合は進行中の変化（「が」から「を」への変化）と書き言葉・話し言葉の関係は反対となり、書き言葉で「を」目的語が話し言葉より多く使用されている。2 種類の言語変化における書き言葉・話し言葉の振る舞いの違いについては、それぞれのバリエーションにおける助詞の選択に対する話者または書き手の規範意識やスタイルの選好など、様々な要因を今後検証する必要がある。性別については、「がの」交替の場合は BCCWJ で男性が女性より「が」率が高いのに対して、CEJC では女性の方が「が」率が高いという効果のねじれが観察され、「がを」交替では、BCCWJ においてのみ女性が男性より「を」率が高いという傾向が確認された。

以上の観察から、3 つのコーパスで利用可能な情報を用いることで、言語外的要因には以下の情報がバリエーション研究に利用できる可能性を示した。

- (6)
- i. コーパス間の比較、BCCWJ のジャンル間比較による書き言葉と話し言葉の差異の検証
 - ii. 年齢または生年情報を用いた言語変化および年齢差 (age-grading) の検証
 - iii. CSJ の改まり度、CEJC の発話場面を利用したフォーマリティの検証
 - iv. 性差の検証

性差に関しては、コーパス間で性差の効果にはらつきが見られたことから、バリエーション研究にコーパスを用いる際には複数のコーパスを使って検討した方がよいかもしれない。今回観察された性差の効果については、女性が言語変化をリードする状況 (Labov 1990) や女性が男性より規範とされる形式を多く使用する状況 (Labov 2001) などの先行研究での一般的傾向に基づいて、言語変化やフォーマリティの視点から何らかの説明を与えることは可能かもしれない。しかし、今回扱った3つのバリエーション現象では今までに性差の効果が考察されていなかったことを踏まえると、まずは別の手法を用いて性差の効果について検証した方がよいと思われる。

6. CEJC の会話の録画映像活用の可能性

本節では前節までとは異なり、CEJC の会話の録画映像活用の可能性について述べる。4節では、い抜き言葉の「てる」使用率はカジュアルな発話でより高くなることを示したが、表4のCEJCのデータでは1.4%と非常に少ないながらも、カジュアルな場面（家で家族/親戚/友人と雑談）においても「ている」の使用が確認されている。ここでは、「てる」が圧倒的に多く使用される中で稀に現れるカジュアルな場面での「ている」の使用がどのような役割を果たしているのかについて、録画映像を活用することで検討したい。

まず、カジュアルな場面で「ている」が使用された例を(7)に挙げる。

- (7) なんか すげえ落ちる 冷蔵庫の床が割れかけている 気がしたけど
(CEJC, 自宅, 雑談, 恋人と料理しながらの会話)

(7) は恋人と一緒に自宅で料理をしている際の会話という非常にカジュアルな場面だが、「てる」ではなく「ている」が使用されている。この会話部分を録画映像で確認してみると、この発話の前後も合わせてこの発話者が話し相手である恋人に対して改まったりより丁寧に話そうとしている様子は見られなかった。つまり、フォーマリティという尺度からはこの「ている」の説明が難しい。そこで、他の説明の可能性を探るため、映像で確認できる会話状況から「ている」の機能について検討した。映像では(7)の発話の際に話者は聞き手と顔を突き合させて話している状況ではなく、話者が観察した状況を描写している中の「ている」の使用であった。この状況から、「ている」の選択は話し相手に対してフォーマルさや丁寧さを表現しようとしているわけではなく、別の要因が働いている可能性が考えられる。

他の「ている」の例として、CEJCで確認された、10歳未満の子供による「ている」の使用全2件を以下に挙げる（10歳未満の子供の「てる」の使用は231件であった）。

- (8) a. 今4分撮影しています (CEJC, 5–9歳, 自宅, 家族との会話)
b. あれでお前はもう死んでいるって (CEJC, 5–9歳, 弘前城で観光中, 家族との会話)

(8b) はアニメのセリフからの直接引用だったため、ここでは議論から除外し、(8a)についてのみ見ていく。(7)の場合と同様に録画映像で発話部分を確認すると、自宅において5–9歳の子供が、会話の録画に使われているカメラの方を向いて、「今4分撮影しています」と家族に対して

報告しているという状況での「ている」の使用であった。このことから、(7) と同様に「ている」の使用は聞き手である家族に対して改まっている状況ではないことがわかる。また、その直後の場面で同じ子供が母親に向かって「ママ、映ってるの？」と尋ねる際には「ている」ではなく「てる」を使用していたことが観察された。ここで、「ている」と「てる」の使用場面を映像で比較すると、「ている」の際は対話者に向かってではなくカメラに向かって発話しているのに対して、「てる」の場合は対話者に向かって発話しているという違いがある。(7) と (8) の観察のみで結論づけることは無論不可能であり、作業仮説を立てる前段階でしかないが、例えばこのようなカジュアルな場面で使用される「ている」は、報告という行為が典型的に行われる新聞などの書き言葉やフォーマルな話し言葉で使用される「ている」が報告のスタイルとして派生した結果、発話者が観察した事象を描写し報告していることを相手に伝えるためのマーカーとしての機能があるかもしれない。今後きちんと仮説を立てて検証する必要があるが、このように、CEJC の録画映像は、フォーマリティによる「ている」の使用ではなく別の要因が働いている可能性など、テキストのみでは得られない新たな分析の視座を提供してくれる可能性を指摘しておきたい。

7. おわりに

本稿では、3つのバリエーションの分析を通して、異なる特徴を持つ複数のコーパスで利用可能な情報から、言語外的要因についてどのような側面から分析可能か探索的に考察した。その結果、(i) コーパス間の比較と BCCWJ のジャンル間の比較による書き言葉と話し言葉の差異、(ii) 年齢または生年情報を用いた言語変化および年齢差 (age-grading)、(iii) CSJ の改まり度と CEJC の発話場面を利用したフォーマリティ、(iv) 性差、について分析できる可能性を示した。また、CEJC の会話の録画映像の分析から、コーパスが持つ特性を活かした多角的な研究の可能性についても言及した。

参考文献

- 新井文人 (2009) 「「い抜き」に関する言語内の制約条件」『第 23 回社会言語科学会発表論文集』20–23.
- Chambers, J.K. (2013) Patterns of variation including change. In: Chambers and Schilling (eds.) (2013), 297–323.
- Chambers, J.K. and Natalie Schilling (eds.) (2013) *The handbook of language variation and change*. Oxford: Wiley-Blackwell.
- Cukor-Avila, Patricia and Guy Bailey (2013) Real time and apparent time. In: Chambers and Schilling (eds.) (2013), 237–262.
- Harada, Shin-ichi (1971) Ga-no conversion and idiolectal variations in Japanese. *Gengo Kenkyu* 60: 25–38. https://doi.org/10.11435/gengo1939.1971.60_25
- 石川禎紀 (1958) 「「ている」と「てる」について一話しことにみられる現代語の一特色一」『解釈』4(3): 22–24.
- 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—」窪薙晴夫・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, DC: Center for Applied Linguistics.
- Labov, William (1972) Some principles of linguistic methodology. *Language in Society* 1: 97–120. <https://doi.org/10.1017/S0047404500006576>
- Labov, William (1990) The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2: 205–254. <https://doi.org/10.1017/S0954394500000338>
- Labov, William (2001) *Principles of linguistic change. Vol. 2. Social factors*. Oxford: Blackwell.
- 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.
- 宮島達夫 (2006) 「「テイル」と「テル」」土岐哲先生還暦記念論文集編集委員会（編）『日本語の教育から研究へ』239–252. 東京：くろしお出版。
- 南部智史 (2007) 「定量的分析に基づく「が / の」交替再考」『言語研究』131: 115–149. https://doi.org/10.11435/gengo.131.0_115
- Nambu, Satoshi (2019) Japanese subject markers in linguistic change: A quantitative analysis of data spanning 90 years and its theoretical implications. *Linguistics* 57(5): 1217–1238. <https://doi.org/10.1515/ling-2019-0018>
- Nambu, Satoshi, Hyun Kyung Hwang, David Y. Oshima and Masashi Nomura (2018) The nominative/accusative alternation in Japanese and information structure. *Journal of East Asian Linguistics* 27(2): 141–171. <https://doi.org/10.1007/s10831-018-9169-1>
- Nambu, Satoshi and Kentaro Nakatani (2023) An experimental study of the adjacency constraint on the genitive subject in Japanese. *Glossa: A Journal of General Linguistics* 8(1): 1–30. <https://doi.org/10.16995/glossa.9509>
- Nambu, Satoshi, David Y. Oshima and Shin-ichiro Sano (2022) The nominative-to-accusative shift in Japanese: Diachronic and synchronic considerations. *Journal of Japanese Linguistics* 38(2): 161–191. <https://doi.org/10.1515/jjl-2022-2057>
- Ogawa, Yoshiki (2018) Diachronic syntactic change and language acquisition: A view from nominative/genitive conversion in Japanese. *Interdisciplinary Information Sciences* 24(2): 91–179. <https://doi.org/10.4036/iis.2018.R.01>
- R Core Team (2022) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria. <https://www.R-project.org/>
- 坂梨隆三 (2002) 「『浮世床』『浮世風呂』のテルとテイル」東京大学国語国文学会（編）『国語と国文学』79(8): 38–48. 至文堂。
- Shibatani, Masayoshi (1975) Perceptual strategies and the phenomena of particle conversion in Japanese. In: Robin E. Grossman (ed.) *Papers from the parasession on functionalism*, 469–480. Chicago: Chicago Linguistic Society.
- 渋谷勝己 (1993) 「日本語可能表現の諸相と発展」『大阪大学文学部紀要』33(1): i–262.
- 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.
- Weinreich, Uriel, William Labov and Marvin Herzog (1968) Empirical foundations for a theory of language change. In: Winfred P. Lehmann and Yakov Malkiel (eds.) *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.
- 山田昌裕 (2018) 「格助詞「ガ」の用法拡大の様相—ヲ格名詞（対象）に下接する用法を中心として—」東京大学国語国文学会（編）『国語と国文学』95(1): 51–62. 明治書院。

【付録】

表 A い抜き言葉、BCCWJ の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	– 44.08	25.56	– 1.72	.08
生年	0.02	0.01	1.62	.11
性別	男性	– 1.59	0.57	– 2.78
モーラ数		– 0.22	0.11	– 1.90
促音あり		0.18	0.21	0.88
				.38

表 B い抜き言葉、CSJ の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	1.48	0.20	7.24	<.001
年齢	0.05	0.004	13.97	<.001
性別	男性	0.51	0.10	5.19
発話	独話	– 2.33	0.13	– 18.42
改まり度		– 0.41	0.02	– 17.91
モーラ数		– 0.12	0.01	– 10.14
促音あり		0.18	0.02	8.73
				<.001

表 C い抜き言葉、CEJC の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	4.86	0.27	17.99	<.001
年齢	0.002	0.005	0.50	.61
性別	男性	0.01	0.18	0.07
会話相手	その他	– 0.60	0.13	– 4.60
	仕事関係	– 0.88	0.22	– 3.97
モーラ数		– 0.16	0.05	– 3.29
促音あり		0.25	0.08	3.30
				<.001

表 D 「がの」交替、BCCWJ の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	41.18	2.39	17.21	<.001
生年	– 0.02	0.001	– 16.71	<.001
性別	男性	– 0.62	0.29	– 2.10
ジャンル	雑誌	0.09	0.27	0.32
	新聞	0.10	0.63	0.16
述部	動詞	– 1.08	0.21	– 5.10
				<.001

表 E 「がの」交替, CSJ の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	− 0.004	0.77	− 0.005	.996
年齢	0.03	0.01	2.18	<.05
性別	男性	− 0.24	0.29	− 0.82
発話	独話	− 0.66	0.52	− 1.27
改まり度		− 0.09	0.15	− 0.61
述部	動詞	− 0.77	0.24	− 3.24

表 F 「がの」交替, CEJC の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	− 0.86	0.21	− 4.04	<.001
年齢	0.004	0.004	0.98	.33
性別	男性	0.28	0.14	2.05
会話相手	その他	− 0.005	0.15	− 0.03
	仕事関係	− 0.48	0.27	− 1.75
述部	動詞	− 1.01	0.14	− 7.43

表 G 「がを」交替, BCCWJ の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	− 17.41	4.88	− 3.57	<.001
生年	0.009	0.003	3.46	<.001
性別	男性	− 0.55	0.14	− 3.94
述部	語彙	− 2.60	0.14	− 18.47
	動名詞 + できる	1.63	0.12	13.29

表 H 「がを」交替, CSJ の混合効果モデルの要約

	推定値	標準誤差	z 値	p 値
切片	− 0.66	0.96	− 0.69	.49
年齢	− 0.006	0.006	− 0.92	.36
性別	男性	− 0.33	0.19	− 1.73
発話	独話	− 0.43	0.90	− 0.48
改まり度		0.03	0.11	0.25
述部	語彙	− 2.66	0.23	− 11.82
	動名詞 + できる	1.21	0.20	5.93

表 I 「がを」交替, CEJC の混合効果モデルの要約

		推定値	標準誤差	z 値	p 値
切片		− 0.54	0.33	− 1.65	.10
年齢		− 0.01	0.007	− 2.13	<.05
性別	男性	0.07	0.23	0.30	.76
会話相手	その他	0.02	0.25	0.08	.94
	仕事関係	0.12	0.37	0.34	.73
述部	語彙	− 2.54	0.24	− 10.64	<.001
	動名詞 + できる	0.30	0.48	0.64	.52

Corpus-based Study on Language Variation: Exploring Language-external Factors

NAMBU Satoshi

Monash University / Project Collaborator, NINJAL

Abstract

This article explores how the effects of extra-linguistic factors on the choice of linguistic forms in variation phenomena can be analyzed by comparing corpora with different characteristics and utilizing the information about contexts and speakers (or authors) they contain. Analyzing three variation phenomena, this article demonstrates the potential of such analysis in the following respects: (i) written vs. spoken language by comparing the corpora and genre differences within the Balanced Corpus of Contemporary Written Japanese (BCCWJ); (ii) language change and age-grading using age or year of birth information; (iii) formality using the Corpus of Spontaneous Japanese (CSJ) and the conversational situations in the Corpus of Everyday Japanese Conversations (CEJC); and (iv) gender differences. In addition, it highlights the potential of analyses using video recordings of conversations in the CEJC.

Keywords: corpus, *i*-deletion, nominative/genitive alternation, nominative/accusative alternation, language-external factors