

個人および生成 AI の「が」・「の」主語嗜好性: 文学作品と実験データに基づく初期研究*
Individual and generative AI's *ga/no* subject preference: A preliminary study based on literary works and experimental data

宮崎順大・牧秀樹
Jundai Miyazaki and Hideki Maki

岐阜大学・岐阜大学
Gifu University and Gifu University

本稿では、「が」・「の」交替に関して、各個人および生成 AI における嗜好性の有無を観察する。従来「が」・「の」交替に関わる研究に際して、コーパスを用いた研究あるいは母語話者集団を対象とした研究で、日本語母語話者全体としての言語変化を明らかにする試みが続けられてきた。かねてより、「の」主語の生起率は今日に至るまで低下していることが示唆されている一方、各個人に着目し、「が」・「の」交替に関する各々の嗜好性が現代に存在するか否かが不明瞭であった。また、今日飛躍的進歩を遂げる生成 AI について、その嗜好性を見ることには、生成 AI が日本語母語話者の特性を学習できるかを検証する意義がある。以上より本稿では、次の研究課題に取り組む。「現代において、関係節内における『が』主語および『の』主語嗜好性(以下、「が」・「の」嗜好性)を持つ日本語母語話者がそれぞれ存在しているのか。また、今日普及の進む諸生成 AI に、人間同様の嗜好性は認められるか？」関連するデータを吟味し、本稿は「『が』・『の』嗜好性は存在せず、『が』・『の』いずれも交替が自由に行われる傾向が現在においても存在する。一方 AI は『が』主語を好む傾向にあり、これは学習データに依拠するものであると考えられる。」と論ずる。

1. はじめに

従来、「が」・「の」交替に関わる研究では、コーパスを用いた研究、あるいは母語話者集団を対象とした研究を通じて、日本語母語話者全体としての言語変化を明らかにする試みが続けられてきた。「の」主語の生起率は今日に至るまで低下していることが示唆されている一方、各個人に着目し、「が」・「の」交替に関する各々の嗜好性が現代に存在するか否かが不明瞭であった。各個人において「の」主語生起率が低下しているのか、「の」主語嗜好者の減少により、日本語母語話者全体の「の」主語生起率の低下が起こっているのか、明らかにする必要がある。また、来る情報化社会において、大きくその活用が期待される生成 AI について、人間と同様の傾向がみられるかについても実験を行う。

(1) 研究課題

- a. 現代において、関係節内における「が」主語および「の」主語嗜好性(以下、「が」・「の」嗜好性)を持つ日本語母語話者がそれぞれ存在しているのか。
- b. 今日普及の進む諸生成 AI に人間と同様の嗜好性は認められるか。

2. 背景

南部 (2008) は、国会議事録(MJD)を用いて、個人内での「の」生起率について、加齢とともに減少・増加・横ばいの3種が存在することを示し、その変化の要因について研究を行った。まず、衆議院と参議院、また本会議と委員会における対象者の「の」生起率に有意な差がみられないことから、発話時の状況が「の」の生起率に影響を与えないことを論じた。また、国会議員か否かについては差が見られるとしつつも、国会議員発話時の(社会的)地位・役割が「が」および「の」の生起率に影響を与えないことを示した。

続く南部 (2014) では、MJD および日本語話し言葉コーパス(CSJ)を用いた定量観察を通じて、「が」・「の」交替に関わる言語変化の存在を明らかにし、またその使用は今日では安定に向かっている状態であることを示した。

また後述する実験に関連して、Maki et al. (2004) は、沖縄、鹿児島、岐阜の学部学生に主格・属格交替に関する調査を行い、2003 年調査当時において、学部学生年齢層では、「の」生起率が、統計的に有意に低いことを示した。

3. データ

3.1 文学作品の分析

まず、文学作品をもとに「が」・「の」嗜好性の調査を行った。なお、本稿では「が」主語嗜好性、「の」主語嗜好性の有無について論じるため、「が」目的語および「の」目的語と判断できたものは除外した。

3.1.1 芥川龍之介の文学作品に見る「が」・「の」嗜好性

芥川龍之介 (1892 年–1927 年)の『羅生門』(1915)において、「が」主語生起率が 24.0%、「の」主語生起率が 76.0%(観測数 25)であった。

これと、理論値である、「が」主語生起率 50.0%、「の」主語生起率 50%との母比率検定を行った結果、統計量(z)が 2.60 となり、これは有意水準(5%)の範囲である 1.96 を上回ることから、統計的に有意に「の」主語への偏りが認められる。

これにより、芥川は「の」主語を好んで用いる傾向にあり、「の」主語嗜好性を有している、と言える。(例(部分))

- | | |
|-----------------------------------|---|
| (2) a. 金銀の箱が [□] ついたりした木 | d. わしが [□] 今髪 [□] の毛を抜いた女 |
| b. 火の光が [□] 及ぶ範囲 | e. 仕方が [□] ない事 |
| c. 自分が [□] 盗人になる気でいた事 | f. 刻限が [□] 遅いせい |
| (3) a. 丹塗の [□] 剥げた大きな円柱 | d. 幅の [□] 広い梯子 |
| b. 崩れ目に長い草の [□] はえた石段 | e. 眶の [□] 赤くなった肉食鳥のような鋭い目 |
| c. 夕冷えの [□] する京都 | f. わしの [□] する事 |

3.1.2 星新一の文学作品に見る「が」・「の」嗜好性

次に、現代の作家である、星新一 (1926 年–1997 年)の『きまぐれロボット』(1999)に収録された 8 作品を対象に調査を行った。関係節内の「が」主語出現率は 26.1%、「の」主語出現率は 73.9% (観測数 23)であった。上記と同様に行った母比率検定から、統計的に有意に「の」主語への偏りが認められた($z \approx 2.29 > 1.96$)。よって、星も「の」嗜好性が高いと言える。(例(部分))

- | | |
|---|---|
| (4) a. ネズミたちが [□] とつぜん家から逃げ出したこと | d. 恐ろしい動物が [□] たくさんいたはず |
| b. 苦心した甲斐が [□] あったというもの | e. ラッパが [□] 出した音 |
| c. ハエやカが [□] 多いはず | f. 私が [□] 来たこと |
| (5) a. 人相の [□] よくない男 | d. パトカーは道の [□] あるところしか走れない |
| b. わたしの [□] つくった最も優秀なロボット | e. 文明の [□] おけている住民の [□] 住む星 |
| c. こんないい人たちの [□] 住む星 | f. 人通りの [□] 多い場所 |

3.1.3 吉本ばなの文学作品に見る「が」・「の」嗜好性

同じく現代作家である、吉本ばなの(1964年-)の『サンクチュアリ』(1988)の前半部において、関係節内の「が」主語出現率は50.0%、「の」主語出現率も50.0%(観測数20)であった。上記と同様に行った母比率検定から、統計的有意な主語嗜好性は見られなかった。 $(z=0 > 1.96)$ 。(例(部分))

- (6) a. 西陽が入って来る時刻
b. 電球が照らす夜店
c. ぴかっと音がするほど
d. 夫が突然死んだこと
e. 主婦同士が再会した時
f. 妹がいっぱいいたせい
- (7) a. 風の強い春の夜
b. 智明のいる小さなホテル
c. 人のいないがらんと広いホテル
d. わけのわからないところ
e. 希望のあるうち
f. 雲の浮かぶ青空

3.1.4 鈴木ふみの文学作品に見る「が」・「の」嗜好性

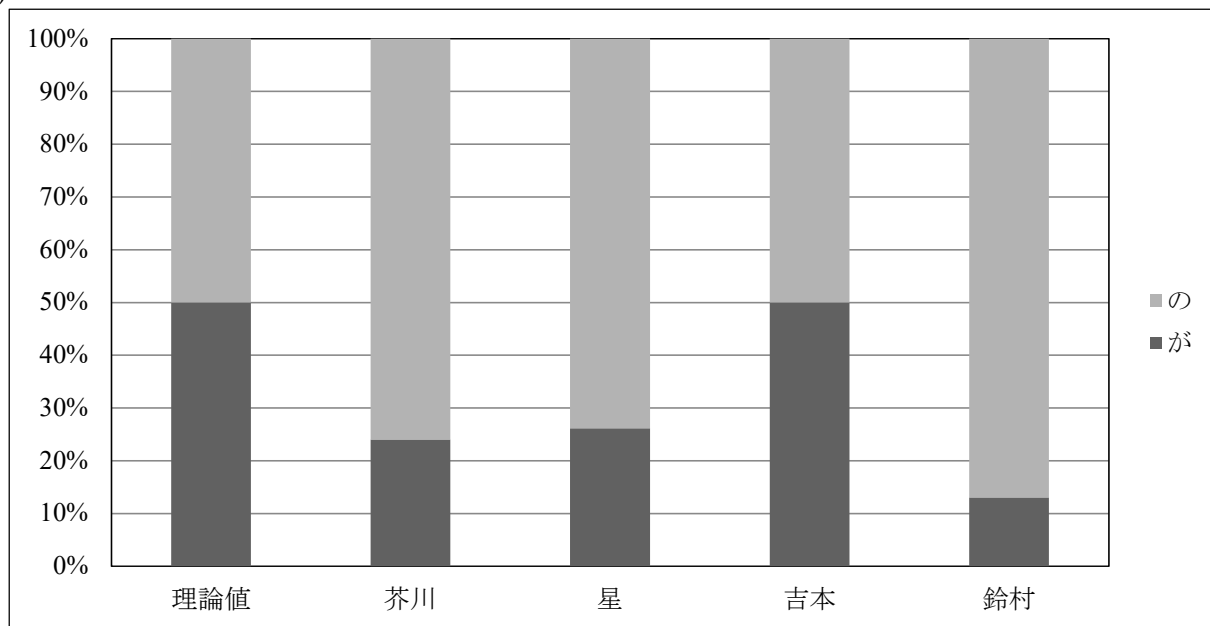
より若年の現代の作家の分析を行うために、鈴木ふみ(1995年-)の2021年小説すばる新人賞作品である『櫓太鼓がきこえる』(2021)の内、名古屋場所の章を対象として、関係節内の「が」・「の」嗜好性を調査した。関係節内の「が」主語出現率は87.0%、「の」主語出現率は13.0%(観測数54)であった。これも上記と同様に行った母比率検定より、統計的有意に「が」主語への偏りが認められる $(z=5.44 > 1.96)$ 。このことから、鈴木は「が」嗜好性が高いと言える。(例(部分))

- (8) a. 気が遠くなるような目標
b. 取組が組まれること
c. 色黒で眼光が鋭い親方
d. 本場所と巡業がどちらもない二月と六月
e. 宮川さんがそう思っていること
f. 達樹が言った話
- (9) a. 気の抜けた返事
b. 冷房のががん効いた部屋
c. 頼りがいのある直之さん
d. 土俵経験のある力士
e. 迷いのないすっきりとした表情
f. 体の細いざんばら髪の新弟子

3.1.5 まとめと考察

以上、文学作品の分析による調査の結果について、芥川の時代から鈴木ふみの時代へ経るにつれて、「が」主語を嗜好する傾向が高まっており、これは南部の観察に適うものである。また、Maki et al. (2004)の調査の通り、若年である鈴木ふみについて、「が」主語の嗜好が見られた。しかし、一方で吉本は、観察範囲の限りでは理論値と同値で、嗜好性が見られず、これについて、吉本の年齢に起因するものか、或いは嗜好性の無い人の存在がみられるのかの検討が必要である。

(10)



3.2 愛知大学における「が」・「の」嗜好性の調査

3.2.1 調査の目的および内容

先項の結果を受けて、年齢を鈴木と同年代の若年層に調査対象に絞り、「が」嗜好性が顕著に見られるか否かの調査を行うこととした。「が」嗜好性を有する話者が多数であると見込み、愛知大学にて 10 代から 20 代にかけての日本語を母語とする若年層の学生 81 名を対象として、「が」・「の」嗜好性の有無を調査した。

(11) 調査方法

羅生門より引用した 6 文 ((11a)–(11f))と、同様に「が」或いは「の」いずれも非文とならない 4 文の合計 10 文を使用し、ターゲットとする格助詞部分を空欄としたうえで、穴埋め作業として回答させる方式を採用した。調査に使用した 10 文は以下の通りである。

(12) 使用した 10 問

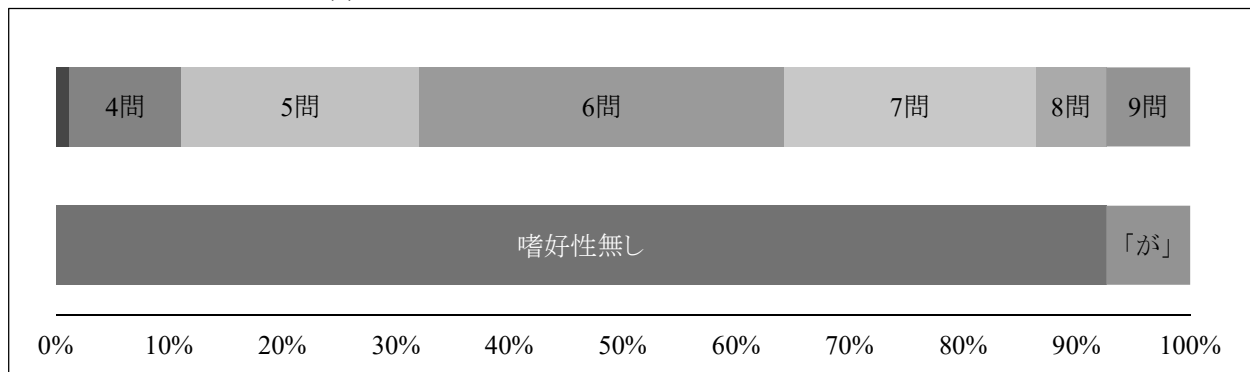
- a. 老婆()床に挿した木切れ
- b. がまがえる()つぶやくような声
- c. この女()売る干魚
- d. この女()した事
- e. わし()した事
- f. わし()する事
- g. ぼく()作った料理
- h. ぼくたち()作った料理
- i. 優子()作った料理
- j. りんご()実る時期

3.2.2 調査結果

調査の結果、統計的有意にいずれかの嗜好性があるといえる回答(10 問中、いずれかが 9 問以上の場合、有意水準を 5%としてどちらかに嗜好性があると認められる。(9 問の場合、 $z \asymp$

2.52>1.96))をした人は、81名中わずか6名であり、残りの75名はいずれの嗜好性も認められなかった。これは、嗜好性を有する者と有さない者を1:1とした場合の理論値との母比率検定から、統計的有意に嗜好性を持たない日本語母語話者が多数である事がわかった。(z=8.11>1.96)

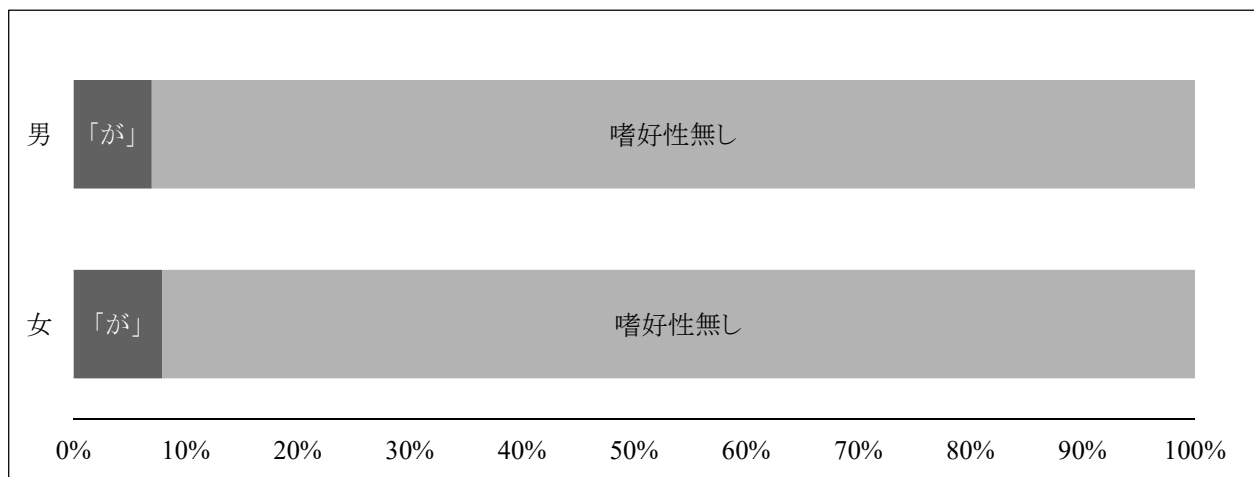
(13) 3.2.1 の調査の結果 (a)



なお、調査にあたって性別も収集しており、性差による嗜好性の有無の傾向も併せて調査した。なお、今回の調査では「の」嗜好者は見られなかったため、性差による「が」嗜好性・「の」嗜好性の偏りを調査することはできなかった。

女性の総数は38名で、うち「が」嗜好性のある人は3名、残る35名は嗜好性がなかった。男性の総数は43名で、うち「が」嗜好性のある人は3名、残る40名は嗜好性がなかった。比率にして、男性で「が」嗜好性のある人は約6.98%、女性で「が」嗜好性のある人は約7.90%であった。両性の差は非常にわずかで、したがって男女間で嗜好の有無に明確な差は見られないと認められる。

(14) 3.2.1 の調査の結果 (b)



3.3 諸生成 AI の「が」・「の」嗜好性の調査

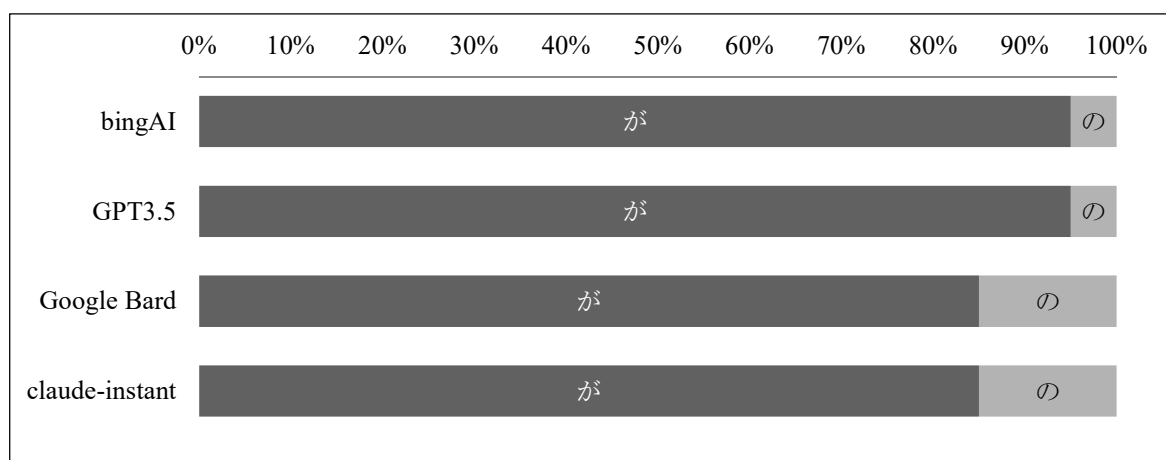
「Bing AI」(現 Microsoft Copilot)、「Google Bard」(現 Gemini)、「GPT3.5」、「Claude-instant」を用いて、3.2と同様の実験を行った。(例(部分))

(15) 3.3 結果 (a)

- a. この女がした事(「が」100%)
- b. わしがした事(「が」100%)
- c. 老婆の床に挿した木切れ(「が」50%、「の」50%)
- d. りんごの実る時期(「が」75%、「の」25%)

調査の結果、4つすべてにおいて、「が」嗜好性が認められた。但し、別日に再度試行をしたところ、回答にばらつきが見られた。しかし、ほぼ全ての試行において、「が」嗜好性を有していた。(8回中 6回) (16)は別日の試行を含めたうえでの嗜好性を表しており、累計 20 問のうち、いずれかが 15 問以上の場合、有意水準を 5%としてどちらかに嗜好性があると認められる。(15 問の場合、 $z=2.24>1.96$)この結果は、AI が学習済みのデータは、「が」主語が優勢であることを示唆している。

(16) 3.3 結果 (b) (※名称はいずれも調査当時のもの)



4. 議論

本稿では、以下 2 つの研究課題について考察を行った。

(1) 研究課題(再掲・部分)

- a. 現代において、関係節内における「が」主語および「の」主語嗜好性(以下、「が」・「の」嗜好性)を持つ日本語母語話者がそれぞれ存在しているのか。
- b. 今日普及の進む諸生成 AI に人間と同様の嗜好性は認められるか。

まず、a.について、文学作品内における「が」・「の」嗜好性は、芥川から鈴木に至る時間内では、「の」嗜好性から「が」嗜好性に傾向が移っているように見える。ところが、10 代から 20 代にかけての若年層日本語母語話者に関しては、実際に、「が」と「の」を選択させるという課題においては、「が」・「の」嗜好性を持たない話者が統計的有意に多数である。これは、「が」・「の」いずれも、交替が自由に行われることを許容する傾向が、2024 年 3 月現在においても、存在していることを示唆している。

次に、b.について、現時点での諸生成 AI は、統計的有意に「が」嗜好性が強いと言える。生成 AI は、学習データによる生成に依存していることから、本調査で使用した生成 AI は、どれもその学習データ自体に、「が」嗜好性があることが示唆される。

5. 参考文献および資料

5.1 参考文献

Maki, Hideki and Tamami Morishima (2004) “A Statistical Analysis of the Nominative/Genitive Alternation in Japanese: A Preliminary Study,” *Bulletin of the Faculty of Regional Studies, Gifu University* 14, 87–119.

南部智史 (2008) 「「が/の」交換における個人内変化の研究」, 松田謙次郎 (編) 『国会会議録を使った日本語研究』, 135–157, ひつじ書房, 東京.

南部智史 (2014) 『コーパス言語学および実験言語学に基づく格助詞交替の分析』, 博士論文, 大阪大学.

5.2 資料

芥川龍之介 (1915) 「羅生門」, 『芥川龍之介龍之介全集 1』, 52–62, 筑摩書房, 東京.

鈴木ふみ (2021) 『櫓太鼓がきこえる』, 集英社, 東京.

星新一 (2002) 『気まぐれロボット』, 角川書店, 東京.

吉本ばなな (1988) 『うたかた/サンクチュアリ』, 福武書店, 東京.

鈴木、星、吉本については、電子書籍版のためページ数は不定。

*本稿のすべての不備は執筆者自身のものである。3.2 におけるデータ収集・提供に関し、斎藤正高氏に感謝する。

連絡先メール住所: MIYAZAKI Jundai <miyazaki.jundai.d4@s.gifu-u.ac.jp>