テキスト関連属性と助詞選択：計量的アプローチに基づく探索的研究

―主語・主題を導く「は」と「が」をめぐって―

石川　慎一郎（神戸大学国際コミュニケーションセンター／国際文化学研究科）

Japanese Particles *Wa* and *Ga* As Topic/ Subject Markers: A Quantitative Analysis of Text-related Factors on the Particle Choice

Shin'ichiro Ishikawa (Kobe University, SOLAC/ GSICS)

1. はじめに

各種の助詞のうち「は」と「が」の選択は先行研究において様々な角度から問題にされてきた。一般に，「は」が「文で述べようとする事柄を『…について言えば』といった気持ちで話題としてとりたて，それについての説明を導く」のに対し，「が」は「文法的に主語となることを表す」ものとされる（『明鏡』2 版）。これにより，「象は鼻が長い」型の構文や，「僕はうなぎだ」型の構文についても一応の文法的説明が成り立つ。

ただ，実際には，「は」と「が」の選択はそれほど単純ではなく，様々な文法的・意味的・語用論的要因の関与が認められる。野田（1996）や堀川（2010）による先行研究のまとめを整理すると，(A)主格支配範囲（文末までかかれば「は」，節内にのみかかれば「が」），(B)陳述主観性（話し手の主観判断を表明する判断文は「は」，具体的現象をありのまま表現する現象文は「が」，話題を持つ題目的有題文は「は」，話題を持たない平説的無題文は「が」），(C)情報新旧性（既知・既定・不可変の旧情報ないしテーマは「は」，未知・未定・可変の新情報ないしレーマは「が」），(D)主述関係（主格要素の性質を述語が解説する内包・記述・措定の場合は「は」，主格要素と述語名詞が同一であることを表す外延・同定・指定の場合は「は」または「が」），(E)主格要素意味特性（対比的・対照的・単独物的意味を持てば「は」，排他的・総記的・集合物的・強調的意味を持てば「が」）などの観点が主として「は」と「が」の使い分けに関与しているとされる。

野田（1996）は以上の諸原理の妥当性を認めつつも，それら各々が「別々の視点」に基づいているとし，(A')主題を持てるか（持てなければ「が」，持てれば(B')へ），(B')主題を持つか（持たなければ「が」，持てば(C')へ），(C')何を主題にするか（格成分ならば「は」，述語ならば(D')へ），(D')主題を明示するか（明示すれば「は」，暗示すれば「が」）の 4 種と，(E')どう取り立てるか（対比なら「は」，排他なら「が」）からなる統一的な説明モデルを提唱している。

ただ，こうした選択モデルをもってしても，たとえば「これはペンです」と「これがペンです」にまつわる微妙な差のすべてを説明することは困難を極める。実際，日本語教育学などでは，母語話者や学習者を対象として，文中助詞を空欄にして「は」または「が」を埋めさせる実験が広く行われているが，母語話者であっても判断にぶれが見られること

が報告されている。ゲオルギエバ（2008）の実験では，「は」ないし「が」を埋めるべき 46 の項目中，18 項目において母語話者の判断が「は」と「が」に分かれ，母語話者であっても主格を導く助詞が「択一的に選択されているとは限らない」と結論されている。

こうした状況を踏まえると，言語記述においてまずもって明らかにすべきことは，「は」と「が」の優先性ないしデフォルト性であろう。中川（1996）は「これまで説明に用いられてきた概念の対立に準拠し，2 つの助詞を差異化しようとすれば，2 つの選択肢のどちらを選ぶにしてもそれらの概念に応じた動機が必要」になるが，実際には「それが確定できない状況，どちらともつかず選ぶことができない状況」があるとし，そうした無条件の選択に「零度」を認めることで助詞選択の問題を単純化できると指摘している。その上で，「私／先週／京都／行きました」に助詞を添えて文にする場合，ほとんどの日本人が「は」を選ぶことを根拠として，ここでの「は」の選択は何らかの文法規則による判断ではなく，「完全に無条件の選択」であり，「は」こそが選択の「零度」で，これを使ってはいけない場合にのみ「が」が選ばれるのだと結論している。このようなモデルを採用すれば，日本語教育においても「学習者には単純な条件と規則を示」すことができる。

中川 (1996) の言う「は」の零度モデルは現象記述を単純化する上で有益なものだが，「は」の優先性が実際の言語使用においてどの程度一般的に再現されるかについては検証の必要がある。このとき，テキストに関連する言語的・非言語的属性を広く考慮に加えるべきであろう。たとえば，「私／先週／京都／行きました」の場合に「は」が優先的に選ばれるとして，「私」がその他の名詞であってもそうなのか，「私」の位置が文頭でなくてもそうなのか，テキストのタイプや時代に関わらずそうなのか，書き手の性別や年代に関わらずそうなのか，といった点についても確認の手順を踏むことが望ましい。

こうした問題を実証的に調査する場合，前述のように，従来は発話タスクや穴埋めタスクなどの誘引型手法が広く用いられてきたが，日本語における無条件・無意識の選択傾向を信頼できる形で抽出するには大規模な匿名データの解析が不可欠になる。本研究においては，「現代日本語書き言葉均衡コーパス」（以下 BCCWJ）を用い，「は」と「が」の選択におけるデフォルト性の問題を計量的観点から考察してみることとしたい。

2. リサーチデザイン

2. 1　目的と RQ

各種のコーパス研究により，言語の振る舞いは，テキストを取り巻く多様な要因によって複合的に影響されていることが示唆されている。本研究では，テキスト内・テキスト・テキスト外という 3 段階の階層要因モデルを仮定し，テキスト内レベルでは「は」と「が」を含む構文の言語特性として主格名詞の情報新旧性・位置・意味内容に，テキストレベルではテキストタイプ・年代・内容種別に，テキスト外レベルでは書き手の生年・性別にそれぞれ注目しつつ，「は」と「が」の頻度的優先性を多角的に検討してゆく。リサーチクエスチョンは下記の 5 点である。

RQ1 現代日本語の書き言葉全体で「は」と「が」はいずれが頻度上の標準であるか。

RQ2 テキスト内の言語的属性（(1)主格要素の情報新旧性，(2)位置，(3)意味内容）は助詞選択にどのような影響を与えるか。

RQ3 テキスト自身の非言語的属性（(1)テキストタイプ，(2)年代，(3)内容種別）は助詞選択にどのような影響を与えるか。

RQ4 テキスト外の非言語的属性（(1)書き手の生年，(2)性別）は助詞選択にどのような影響を与えるか。

RQ5 非言語的属性で助詞選択をモデル化することは可能か。

2．2 データと処理手順

データはBCCWJで，2013年7月に「中納言」インタフェースを介して調査を実施した。対象は主格要素（名詞・代名詞）の直後位置に出現する「は」と「が」で，検索は語彙素（ないし語彙素読み）単位で行った。このうち，「が」については，「水が飲みたい」や「画面が見にくい」のように他動詞と共起する例があるが，ここでは『明鏡』2版の定義に従い，これらも「主語を表す」用法とみなして他と同列に扱った。また，同辞書が「が」の別用法とする「名詞を修飾する」用法（例：我らが母校）と「同じ名詞をつなぐ」用法（例：親が親だから）については，BCCWJより無作為抽出した1,000例（500例×2回）を質的に検証した結果，当該用例の出現を認めなかったため，該当事例は無視できる程度に少ないものと判断し，検索で得られた値をそのまま使用することとした。このほか，BCCWJの形態素判定では，接続詞の「ところが」が「ところ」＋「が」と分析されている例があるなど（例：|ところ|(が)|、|私|の|職業|は|今|、|漫画|を|描く|こと|だ|。［LBr7_00041］），いくらか問題も認められるが，全体に占める比率は極小であるため，頻度修正は行っていない。なお，以下の議論において，「は」の零度性を検討する際には，「は」の頻度を「は」と「が」の合計頻度で割った「は」選択率を指標として使用する。

まず，RQ1（全般頻度）では，BCCWJ全体で「は」と「が」の総頻度を比較する。

RQ2（テキスト内言語的属性）では，後続動詞は助詞選択にほとんど関与しないとされていることから（ヨフコバ，2007），議論を構文の主格要素に限定した上で，(1)情報新旧性については，BCCWJ全体を対象として主格要素が名詞の場合と代名詞の場合を，名詞については連体詞「その」（語彙素読み）が共起する場合としない場合を比較する。(2)位置については，構文の安定性が高い新聞・雑誌・白書データを用い，主格要素が文頭に来る（直前句点共起）場合と文中に来る（直前句点共起なし）場合を比較する。(3)内容については，まず，BCCWJの全データを用いて主要代名詞別に比較する。ついで，2000年代に刊行された書籍（図書館）に限定して，名詞＋「は」（全188,256件のうちダウンロード上限の10万件），名詞＋「が」（全209,181件中の10万件），代名詞＋「は」（36,056件），代名詞＋「が」（15,977件），合計252,033件の用例データをダウンロードし，両助詞と高頻度に共起する主格要素（名詞・代名詞）を抽出して質的に比較する。

RQ3（テキストの非言語的属性）では，(1)テキストタイプについては，BCCWJを構成する11種を比較する（コア・非コアは区別せず，書籍は生産母集団・図書館母集団・ベスト

セラーを統合)。(2)年代については，長期データを保有する国会会議録と白書を対象に，1970年代，1980年代，1990年代，2000年代を区分して比較する。(3)内容種別については，上述の書籍の25万件データを用い，日本十進分類別に比較する。

RQ4（テキスト外非言語的属性）では，上述の書籍の25万件データの中から，書き手が単独で生年・性別が明示され，かつ，日本十進法分類が明らかな158,780件を抽出した上で，(1)書き手の生年および (2)性別を分けて比較する。なお，(2)に関しては，BCCWJの生年帯（10年単位）ごとのサンプル数に大きなばらつきがあるため，サンプル数が1,000件を超える1890年代生まれ～1970年代生まれ（157,167件）のデータに限定して分析する。

最後に，RQ5（非言語的属性によるモデル化）では，RQ4で使用した約16万件の書籍データに対して Weka v3.77 を用いた決定木分析を実行し，テキストおよびテキスト外レベルの非言語的属性による助詞選択モデルの作成を試みる。これにより，先行研究で注目されることの少なかった非言語的属性の助詞選択に対する影響を計り，あわせて，はっきりした言語的選択動機が「確定できない状況」下での助詞選択傾向を観察する。分析アルゴリズムはJ48（Quinlan の C4.5 に基づき，データを反復的に分割し，情報利得が最大になるものを選ぶ）とする。剪定（pruning）のための confidence factor は0.25，葉（leaf）あたりの最低のインスタンス数は500，相互検証のための fold 数は10とする。なお，言語研究における決定木分析の適用過程の詳細については石川（2013）他を参照されたい。

3. 結果と考察

3．1　RQ1「は」と「が」の全体的標準性

「は」と「が」の出現状況を概観するため，BCCWJ全体で検索したところ，名詞・代名詞に後続する「は」の頻度は1,950,445回，「が」の頻度は2,007,682回で，「は」選択率は49.28％であった。助詞としての「は」と「が」の選択は拮抗しているが，差は有意であり（G^2=844.4, df=1, p<0.1％），「は」の頻度上の優先性は確認されなかった。このことは「は」の零度モデルを日本語の全体に無条件に適用することの問題点を示唆している。

3．2　RQ2 テキスト内言語的属性の影響

3．2．1　情報の新旧性

先行研究の多くは，旧情報ないしテーマは「は」，新情報ないしレーマは「が」とする分類基準を提唱している。では，実際のデータにおいて，情報の新旧性による違いは確認できるのだろうか。

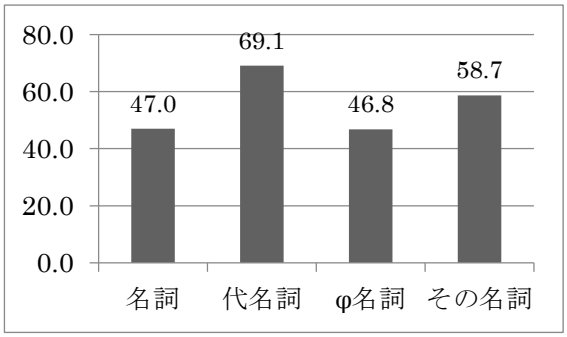

図1 主格要素タイプ別の「は」選択率

BCCWJ全体で名詞・代名詞別の検索を行い，次に，新聞・雑誌・白書に限って名詞前の「その」の共起の有無別に検索を行ったところ，それぞれの条件下での「は」選択率は図1のようになった。

一般に名詞に比べて代名詞が旧情報を指示しやすいこと，また，同じ名詞でも「その」が共起することで旧情報を指しやすいことをふまえて上記のデータを検証すると，主格要素が旧情報となることで「は」の選択傾向が顕著に高まることが確認された。これは，先行研究で広く言われている「は」と既知（扱い）の旧情報，「が」と未知（扱い）の新情報の親和性をデータ面で裏付ける結果と言えよう。

3．2．2　主格名詞位置

野田（1996）のモデルには明示的に含まれていないが，一部の先行研究は「は」と「が」の選択に主格要素の位置が関係する可能性を指摘しており，たとえばヨフコバ（2007）は，小説や学術論文などの文頭文を調査した結果，「圧倒的に『は』が使われている」としている。そこで，主格名詞が文頭に出現する場合（句点＋名詞＋助詞）と，文中に出現する場合（句点なし＋名詞＋助詞）を比較したところ，図 2 の結果が得られた。

右図に明らかなように，主格が文頭に出現する場合，「は」の選択率が高まり，異なるテキストタイプでも傾向は不変である。これは先行研究を広く支持する結果である。談話では，初めに旧情報である話題を提示し，次いでそれに関する新たな情報を付加してゆくのが語用論上自然な展開であり，このことが文頭位置での「は」の選択に影響していると考えられる。

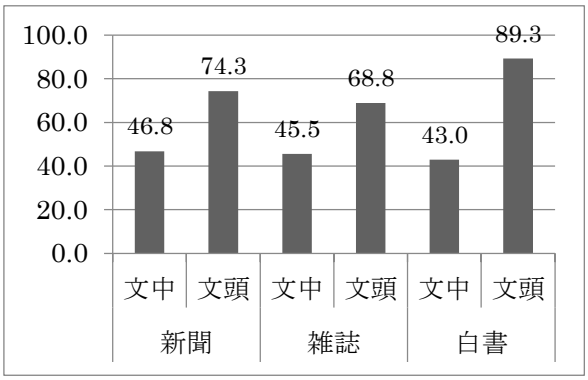

図 2　文頭・非文頭別の「は」選択率

3．2．3　主格要素意味内容

助詞選択に関して，後続する動詞タイプの影響はほとんどないとされているが，先行する名詞タイプの影響はどうであろうか。BCCWJ の全データを対象に，代表的な人称代名詞 3 種（「私」「彼」「彼女」）および非人称代名詞 3 種（「これ」「それ」「あれ」）をサンプルとして「は」選択率を調べたところ図 3 の結果を得た。

人称代名詞 3 種の「は」選択率の平均は 76.1%，非人称代名詞の場合は 73.7%で，主格要素が非人称化（モノ化）することで「は」選択率が低下する傾向が示唆されたようにも見えるが，個々の代名詞ごとに検証すると，傾向は可変的であり（「それ」を除くと，非人称であっても「は」選択率は高い），主格要素の人称・非人称性と助詞選択の間にはっきりした関係は検証されなかった。

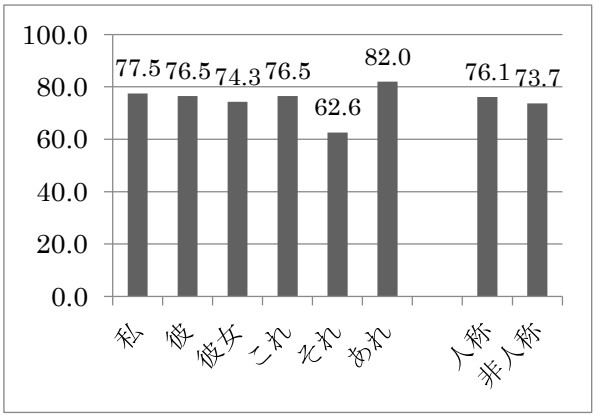

図 3　代名詞タイプ及び主要代名詞別の「は」選択率

続いて，書籍（図書館）の 25 万例の実例データより，「は」と「が」の主格要素のうち，高頻度語上位 30 語を抽出したところ，以下のようになった（「実は」「ほうが」「方が」などの非自立単位も含む）。

（は）私，それ，こと，これ，彼，わたし，彼女，人，僕，ぼく，もの，場合，あなた，とき，俺，実，男，おれ，あれ，ここ，われわれ，自分，今，今度，問題，人間，時，あたし，二人，日本

（が）こと，それ，私，これ，人，もの，彼，気，ほう，ところ，自分，方，何，わたし，あなた，彼女，声，男，必要，音，人間，ぼく，誰，時間，僕，問題，目，俺，言葉，関係

括弧で示したように 30 語中 17 語が重複しており，主格要素の内容には一定の重複性が見られた。ただ，「は」の場合は，「おれ」や「あたし」といった口語的 1 人称代名詞や「あれ」や「ここ」といったくだけた仮名表記代名詞が頻出する。一方，「が」の場合は，「声」「必要」「時間」「言葉」「関係」といった固い文脈で多用される非主観的で抽象的な概念語や，一部の先行研究で指摘されている「何」や「誰」といった疑問詞などが頻出する。以上より，具体的で口語的な談話環境では「は」が，抽象的で書き言葉的な談話環境では「が」が選択されやすいという大まかな方向性が示唆される結果となった。これは，話し手の主観判断を表明する判断文では「は」が，現象をありのままに表現する現象文では「が」が選ばれるという先行研究の見解を異なる観点からサポートするデータと言える。

3．3　RQ3 テキストの非言語的属性の影響

3．3．1　テキストタイプ

助詞選択に関する先行研究の多くは日本語が不可分の実体であるという前提に基づいており，テキストタイプの影響はほとんど考慮されてこなかったわけであるが，実際にはどのような差が見られるのであろうか。テキストタイプ別の分析結果は図 4 のようになった。

コーパス全体で見た「は」選択率は 49.28%であるが（3.1 節参照），特殊性の高い韻文（59.8%）を除くと，すべてのテキストタイプにおいて「は」選択率は 42.5%～51.6%の範囲におさまっており，テキストタイプの影響は比較的限定的であることが明らかになった。

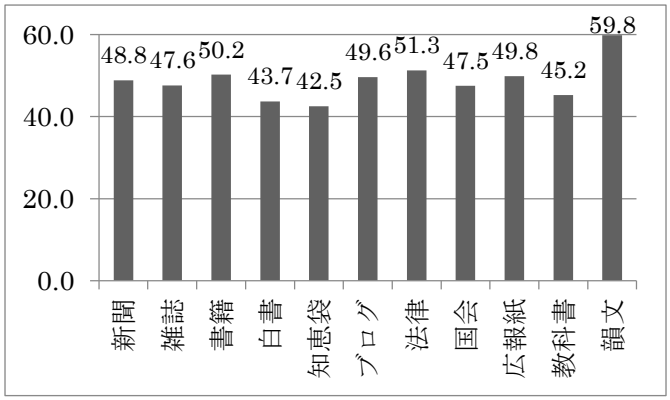

図 4 テキストタイプ別の「は」選択率

ただ，その範囲の中では一貫した差異の傾向も認められる。つまり，書籍・法律・広報紙・新聞など，単なる事実の羅列的伝達に終わらず，何らかの解釈・注

釈・判断が加わるテキストでは「は」の選択率が上昇し，知恵袋・白書・教科書のように，個人的・主観的な解釈が控えられ，具体的で客観的な情報や事実の透明な提示が主となるテキストでは「が」の選択が増えるのである。3.2.3 節で見たとおり，判断文では「は」が，現象文では「が」が選ばれるという傾向が反映された結果と言えるだろう。

3．3．2 テキスト年代

先行研究は年代についてほとんど考慮に入れていないが，「は」の零度性に年代は影響してないのであろうか。長期データが含まれる白書と国会に限って分析を行なったところ，図 5 の結果が得られた。

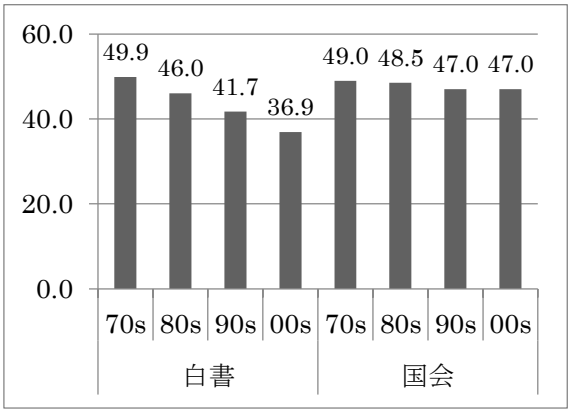

図 5 テキスト年代別の「は」選択率

「は」選択率は国会会議録においては微減しているだけだが，白書においては過去 40 年間に一貫して顕著に低下している。「は」の零度性を考える上でこれはきわめて興味深いデータである。仮に白書に見られる傾向が現代日本語全般の時系列変化を反映したものであるとするなら，40 年間の「は」の減少と「が」の増加は，日本語が，主題中心の伝統的な陳述形態から，主語中心の西欧言語型の陳述形態に接近しつつあることを示唆している可能性がある。

もっとも，BCCWJ は時系列分析用に開発されたコーパスではなく，年代ごとのデータ量も統制されていない。ゆえに，ここで得られた結果を過大に拡張解釈すべきではないが，たとえば複数の新聞社のデータベースなどで同一の経年パタンが確認されるならば，助詞選択の研究や，主語と主題の関係性をめぐる研究において，今後，時代や年代を組み込んだ分析が求められるようになろう。

3．3．3 テキスト内容種別

助詞選択におけるテキストタイプの影響はすでに検討したが（3.3.1 節），それでは，個々のテキストの具体的な内容種別は「は」と「が」の選択にどのように影響しているのであろうか。2000 年代に刊行された書籍（図書館）に絞って，内容種別ごとの「は」選択率を観察したところ，図 6 の結果を得た。

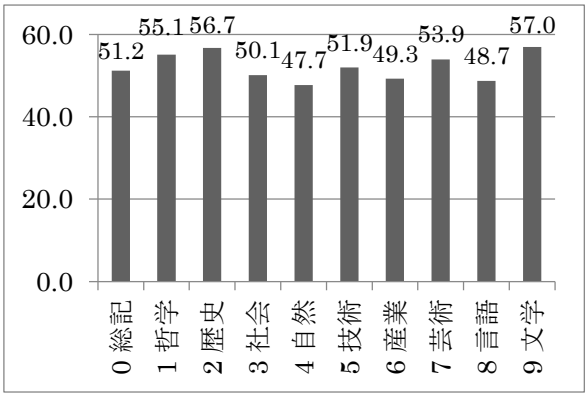

図 6 テキスト内容別の「は」選択率

コーパス全体での「は」選択率は 49.28%，テキストタイプ別では 42.5%～51.6%（韻文の 59.8%を除く）であったが，十進分類法に基づく内容種別では 47.7%～57.0%

となることがわかった。テキストタイプの場合より総じて高めであるものの，上下の幅は10%程度で，テキストタイプの場合とほぼ同等である。だが，ここでも，当該の範囲内においては一定の傾向性が看取できる。具体的には，歴史・文学・哲学など，書き手の解釈と判断に焦点を当てた叙述がなされる場合には「は」が選好され，自然・言語・産業のように。客観的に情報を提示する場合には「が」が選好されている。これは，すでに見たように，判断文と「は」，現象文と「が」の親和性を裏付ける結果と言える。

3．4　RQ4 テキスト外非言語的属性の影響

3．4．1　書き手の生年

過去 40 年間（1970 年代～2000 年代）を対象としたテキスト刊行年代の調査では「は」選択率の低下が示唆されたが（3.3.2 節），過去 90 年間（1890 年代生～1970 年代生）にまたがる書き手の生年の影響はどうであろうか。調査結果は図 7 の通りであった

単回帰による直線のあてはめを行うと，y = －0.7198x + 58.451 という一定の説明力を持った回帰式が得られる（R^2 = .35）。直線の傾きを示す係数は負であり，世代が進むにつれて「は」の選択率が若干ではあるが低下していることがわかる。なお，全体の中で相対的に低い値を示す1890年代と高い値を示す1900年代を除外して，1910年代以降のみで回帰分析をしても，回帰式はy = －0.7297x + 56.893（R^2 = .89）となり，やはり係数は負で，「は」選択率は低下しているように思われる。

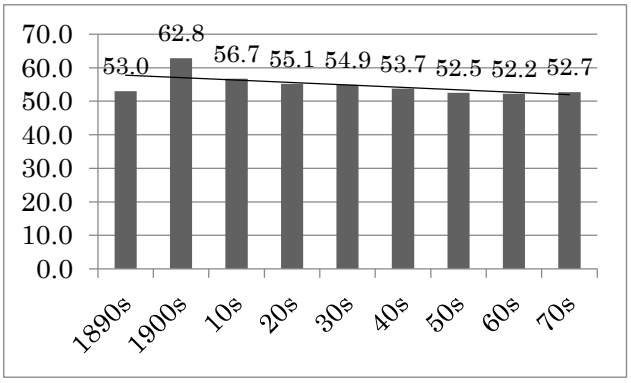

図 7 書き手の生年別の「は」選択率

3．4．2　書き手の性別

性差は言語使用の様々な局面に影響しているとされる。今回のデータで分析を行なった結果，「は」選択率は男性が 46.4%，女性が 53.8%で，男女間の「は」選択率の差は有意であった（G^2 =539.3, *df*=1, *p*<.001）。ただ，結果の再現性を確認するため，データセットを若干拡張し，書籍 25 万件中，単一著者の性別情報を含む 176,082 件全体で調べると，選択率は男性が 53.4%，女性が 53.6%となって選択率の差は有意でなかった（G^2 =0.56, *df*=1, *p*=.454）。このことを踏まえると，性差の影響は仮にあるとしてもごく限定的と思われる。

3．5　RQ5 非言語的属性による決定木モデル構築

以上の分析により，従来の研究でもっぱら分析対象になってきたテキスト内の言語的属性のみならず，テキスト自身およびテキスト外部の書き手に関わる非言語的属性も「は」と「が」の選択に一定の影響を及ぼしている可能性が示唆された。では，非言語的属性だけで「は」と「が」の選択はどの程度説明しうるのであろうか。

決定木分析を行なった結果，図 8 の結果が得られた。モデル中，ジャンル（G）の 0〜9 は書籍内容の十進分類を，書き手生年（B）の 1940 や 1950 はそれぞれ 1940 年代や 1950 年代を，書き手性別（S）の M と F はそれぞれ男性・女性を示す。分岐の結果はたとえば W（15320/6490）のように示されているが，これは当該条件においてモデルが「は」に 15,320 例を分類し，そのうち 6,490 例が誤分類であることを示す。

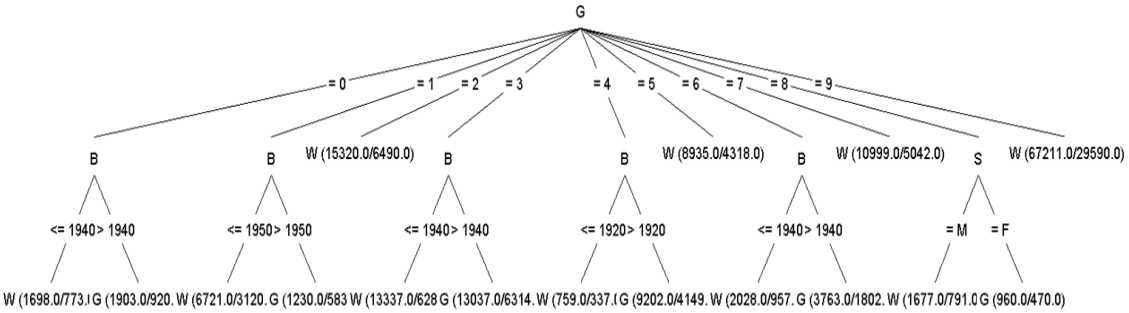

図 8 非言語的属性による「は」と「が」の選択の決定木モデル

モデルはデータの 79.4%を「は」に，20.5%を「が」に分類するもので，正判別率は 54.7%（*Kappa*=.053）である。非言語的属性（内容ジャンル，書き手の生年，書き手の性別）だけで助詞選択の過半が説明されたことは興味深い。また，使用したデータセット中，「は」は 85,223 件，「が」は 73,557 件で，「は」のほうが若干多くなっているが，モデルはより顕著に「は」を優先選択している。

表 1 分類結果（confusion matrix）

	は（分類）	が（分類）
は	69,710	15,513
が	56,404	17,153

非言語的な諸属性のうち，選択に最も影響しているのはジャンルである。書き手の主張に重点が置かれやすい 2（歴史）・5（技術）・7（芸術）・9（文学）類では他の属性の介在なしに「は」が選択される。一方，客観的な事実の紹介が主となる 0（総記）・1（哲学）・3（社会）・4（自然）・6（産業）類では書き手の生年情報が，8（言語）類では書き手の性別が介在的に影響を及ぼして助詞決定がなされる。なお，書き手の生年は 1840 年代から 1980 年代に及んでいるが，分岐の多くは 1940 年代ないし 1950 年代の以前か以後かで起こっており，いわゆる「戦後生まれ」であれば「が」が選択されると言える。この結果はテキスト刊行年代や著者生年の時系列分析（3.3.2 節，3.4.1 節）の結果を精緻化するものである。

4. まとめ

以上，主語・主題標識としての「は」と「が」の選択について，テキストを取り巻く 3 段階の階層要因モデルをふまえた検証を行ってきた。得られた結果についてまとめておく。

まず, RQ1（現代日本語書き言葉全体における標準性）については,「は」選択率は 49.28%で，日本語全般において無条件に「は」が優先されるわけではないことが確認された。

RQ2（テキスト内言語的属性影響）については,「は」の選択は主格要素の (1)情報新旧性に関しては旧情報と，(2)位置に関しては文頭と，(3)意味内容に関しては口語性・具体性・1 人称性と親和性を持つことが示唆された。これらは，話し手の解釈や判断を表明する判断

文で「は」が選ばれやすいという先行研究の主張を支持するものと言える。

RQ3（テキストの非言語的属性影響）については，「は」の選択が，(1)テキストタイプに関しては解釈・注釈・判断を含むテキスト（書籍・法律・広報紙・新聞）と，(2)年代（1970年代〜2000 年代）に関してはより古い年代と，(3)内容種別に関しては解釈・判断が加わるもの（歴史・文学・哲学）と親和性を持つことが示された。判断文と「は」の結合が再確認されたことに加え，テキストの刊行年代が下がるにつれて「は」の優先性が低下する興味深い事実が示された。

RQ4（テキスト外非言語的属性影響）については，「は」の選択は，(1)書き手の生年に関しては古い年代と親和性を持つものの，(2)性別について確定的な結果は得られなかった。

RQ5（非言語的属性によるモデル化）については，6 割弱の正分類を行う決定木モデルが得られ，モデルは「は」を優先するものであった。非言語的属性が助詞決定に関与しうること，いずれかの助詞選択を促す明確な言語的動機を欠く状況では「は」が優先される可能性があること，内容ジャンルが相対的に強く影響しうること，生年が戦前か戦後かで助詞選択が「は」から「が」に変化しうることなどが示唆された。

本研究の結果は，いわゆる「は」の零度性が，頻度の上で，日本語全体に無条件に当てはまるわけではないものの，非言語的要因に基づくモデルにおいては，そうした傾向性が一定程度認められることを示すものであった。ただし，本研究で使用したコーパスは時系列分析や書き手の属性分析を前提として構築されたものではなく，結果の解釈に慎重さが求められるのは言うまでもない。今後，年代や書き手属性ごとのデータ量をより厳密に統制したコーパスの整備がなされ，量的観点に基づく助詞研究が進展することが期待される。

謝辞：草稿に対し，前川喜久雄氏より「が」の用法識別について，森秀明氏より BCCWJ の著者生年の扱いについて貴重な指摘を賜り，一部加筆を施した。記して感謝申し上げる。

文　献

ゲオルギエバ，ベロニカ・トドロバ(2008)『「が」と「は」の使いわけとその理由：母語話者と非母語話者の実態調査を比較して』早稲田大学修士論文.

堀川智也(2010)「日本語の「主題」をめぐる基礎論」『大阪大学世界言語研究センター論集』*4*, 103-117.

石川慎一郎(2013)「語彙難度・語彙多様性・文構成度：母語話者と学習者の区分基準は何か―決定木を用いた学習者コーパス分析―」『統計数理研究所共同研究レポート』*290*, 107-124.

中川正弘(1996)「『は／が』と助詞選択の零度」『広島大学留学生日本語教育』*8* ,11-23.

野田尚史(1996)『新日本語文法選書 1「は」と「が」』東京：くろしお出版.

ヨフコバ四位，エレオノラ (2007)「「は」と「が」に関する一考察 : 外国語としての日本語教育との関連」『横浜国立大学留学生センター教育研究論集』*14*, 159 – 189.