

二字漢語の品詞的用法の実態と通時変化

太陽コーパス統合版パイロット研究報告

2026 年 7 月 14 日

要旨 二字漢語のサ変動詞用法・形容動詞用法・名詞用法を前後文脈の形態論情報から分類し、近世から現代までの資料で用法率を測定する。太陽コーパス 5 時点を統合し、間淵（2018）の人手判定との対応、語別の増減、名詞・形容動詞間の曖昧性を検討した。

キーワード 二字漢語、サ変動詞、形容動詞、太陽コーパス、通時コーパス

1. 目的と方法

二字漢語が実際にどの程度サ変動詞（～する）・形容動詞（～な・～に・～だ）として使用されるかを、コーパス中の各トークンの前後文脈（語彙素・品詞・活用形）によるフレーム分類で測定した。トークンとはコーパス中に現れた個々の UniDic 短単位である。解析辞書の「サ変可能」等の潜在性ラベルは語彙的潜在性の仮説として実測値と照合する。フレーム判定の条件からは外す。指標は語彙素×時代層ごとのサ変動詞用法率・形容動詞用法率である。

$$R_{V(w,c)} = \frac{\text{サ変動詞用法数 (V + V_split)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (1)$$

$$R_A^{\text{core}(w,c)} = \frac{\text{形容動詞中核用法数 (A_na + A_ni)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (2)$$

$$R_{\text{res}(w,c)} = \frac{\text{総出現数} - \text{サ変動詞用法数} - \text{形容動詞中核用法数}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (3)$$

$$R_{A+\text{pred}}^{\text{proxy}(w,c)} = \frac{\text{形容動詞中核用法数と名詞・形容動詞述語候補数 (A_na + A_ni + A_pred)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (4)$$

$$R_A^{\text{upper}(w,c)} = \frac{\text{中核用法数と全曖昧候補数 (A_na + A_ni + A_pred + A_no)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (5)$$

$$R_N^{\text{core}(w,c)} = \frac{\text{独立名詞用法数 (N)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (6)$$

$$R_N^{\text{context}(w,c)} = \frac{\text{独立名詞用法数と複合語内用法数 (N + COMP)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (7)$$

$$R_N^{\text{upper}(w,c)} = \frac{\text{名詞的文脈数と全曖昧候補数 (N + COMP + A_pred + A_no)}}{\text{対象語の総出現数 (全フレーム合計)}} \quad (8)$$

残余用法率 R_{res} は、独立名詞、複合語内用法、 A_{pred} 、 A_{no} 、OTHER を含む。三軸表示では互いに排他的なフレームをまとめるため、各語・各層について

$$R_V + R_A^{\text{core}} + R_{\text{res}} = 1 \quad (9)$$

が成り立つ。

1. 1. 対象語と判定手順

語種が「漢」で、語彙素が漢字 2 字のトークンを対象とする。数詞を除外し、名詞、形状詞、副詞の各用例を含める。入力トークンは文書 ID・文書内番号順に保存する。同一文書内の直前 1 トークン

と直後2トークンまでを参照し、文書IDと文書内番号の連続性を各参照時に確認する。各対象トークンは下表の上から順に判定し、最初に合致した1フレームを付与する。

表1 文脈フレームの判定条件。上の行の条件を優先する。

フレーム	操作的定義	例	用法率での扱い
V	直後が動詞「為る」または「出来る」	勉強する・勉強できる	R_V に算入
V_split	直後が「を」、その次が動詞「為る」	勉強をする	R_V に算入
A_na	直後が断定助動詞の連体形	健康な・安全なる・厳然たる	中核用法に算入
A_ni	直後が断定助動詞の連用形「に」	健康に暮らす	中核用法に算入
A_pred	直後がその他の断定助動詞	健康だ・問題だ・安全なり	名詞・形容動詞の曖昧群
A_no	直後が助詞「の」	健康の秘訣・研究の成果	名詞・形容動詞の曖昧群
COMP	直前または直後が名詞・形状詞・接辞	思考実験・受験勉強・必要性	分母のみ
N	他の条件に入らない名詞	研究は進む・課題は研究。	分母のみ
OTHER	他の条件に入らない非名詞	堂々・一切	分母のみ

V は語彙素「為る」を見るため、「する」「した」「される」「させる」を同じ条件で検出する。「出来る」はサ変動詞の補充的な可能形として算入する。V_split は「対象語+を+為る」の連続だけを対象とし、「勉強をよくする」のような挿入要素を伴う例は現在の条件に入らない。

A_na には現代語の「な」と文語の「なる・たる」、A_ni には助動詞の連用形「に」を含める。「必要に応じる」の「に」は格助詞として解析されるため、A_ni に入らない。

A_pred は断定助動詞が直結する述語候補である。この表層条件には「健康だ」と「問題だ」がともに合致する。A_no にも「健康の秘訣」と「研究の成果」が入る。短単位の形態論情報から両群を一意に分類することは難しい。形容動詞の主結果には A_na + A_ni の中核用法率を用いる。A_pred と A_no は曖昧群として個別に示す。述語候補込み代理値は A_pred を加え、操作的上限値は A_pred + A_no を加える。本報告の散布図と事例語の赤線は、旧集計との比較用に述語候補込み代理値を用いる。

1. 2. 名詞・形容動詞間の曖昧性

語彙素が持つ品詞資格と、個々の用例が置かれた構文を別々に記録する。「健康が第一だ」の「健康」は助詞「が」を伴う名詞的構文、「健康な生活」は形容動詞の連体修飾構文である。「健康だ」のコピュラは名詞述語と形容動詞述語に共通するため、A_pred に保留する。「健康の秘訣」も A_no に保留する。

出力には、格助詞が続く N、裸の N、複合語内用法、形容動詞中核用法、二つの曖昧群を区別する証拠区分と直後形式を保存する。形容動詞側は次の三つを併記する。

1. 形容動詞中核用法率 R_A^{core}
2. 述語候補込み代理値 R_A^{proxy}
3. 全曖昧候補を形容動詞側へ配分した操作的上限値 R_A^{upper}

名詞側は、独立名詞用法率 R_N^{core} 、複合語内用法を加えた名詞的文脈率 R_N^{context} 、曖昧群を名詞側へ配分した操作的上限値 R_N^{upper} を示す。

母集団における形容動詞用法率は、現在の候補集合と解析精度を前提として中核用法率と操作的上限値の間に置く。この識別区間は Wilson 信頼区間と役割が異なる。追加の用例判定では、時代・レジスター・頻度を層化して A_pred と A_no を抽出し、複数の判定者が「形容動詞」「名詞」「判定保留」の

3 値で注釈する。判定者間一致度と保留率を報告し、合意例から曖昧群の配分率を推定する。BCCWJ の長単位品詞も補助情報として照合する。

2. 資料

表 2 使用資料と規模。母集団は BCCWJ 頻度表の二字漢語 26,866 語（頻度 10 以上・数詞除外）。

時代層	資料	総トークン	分類対象（二字漢語）
1703–1720	近松浄瑠璃 4 作（活字翻刻・上方 UniDic 解析）	91,446	5,373
1802–1814	『東海道中膝栗毛』（同・江戸 UniDic）	203,531	7,941
1874–1888	明六雑誌・国民之友（人手修正済み形態論情報）	1,269,164	215,185
1895–1925	太陽コーパス 5 年・60 号・3,409 記事（近代文語 UniDic 解析）	9,697,509	1,535,512
1872–2000	青空文庫 新字新仮名 10,170 作品（形態素解析器 Vibrato）	115,638,298	9,151,995
1976–2008	現代日本語書き言葉均衡コーパス（BCCWJ・人手修正済み）	105,855,819	13,934,727
1876–2015	国立国会図書館（NDL）雑誌 n-gram（年次頻度・代理指標）	—	25,715,644 行

3. 結果

3.1. 潜在性と実態（共時, BCCWJ）

頻度 50 以上の 14,834 語について、UniDic の潜在性ラベルは集計レベルでは強く予測的である（サ変可能語の平均サ変動詞用法率 31.1%、非サ変語 0.40%；形状詞可能語の平均述語候補込み形容動詞用法指標 50.0%、非形状詞可能語 2.2%）。形容動詞中核用法率の平均は形状詞可能語 40.0%、非形状詞可能語 0.29% である。分布の裾には、サ変可能でもサ変動詞用法率 1% 未満の語が 221 語（労働・施設・科学・意見など）、ラベルなしで同用法率 5% 超が 180 語ある（無理・格好など、「～をする」型を含む）。図 1 では、サ変動詞用法率と述語候補込み指標がともに 10% 以上の語は 14,834 語中 59 語（0.4%）である。形容動詞中核用法率を使うと 14 語である。各用例を 1 つのフレームに分類するため、両率の和は 100% を超えず、右上の領域は計算上制限される。軸付近への集中には、語ごとにサ変動詞用法か形容動詞用法の一方が優勢になる語彙的な専門化も関与すると考えられる。後者は現段階の解釈であり、意味分類・語構成・頻度を統制した検証を要する。

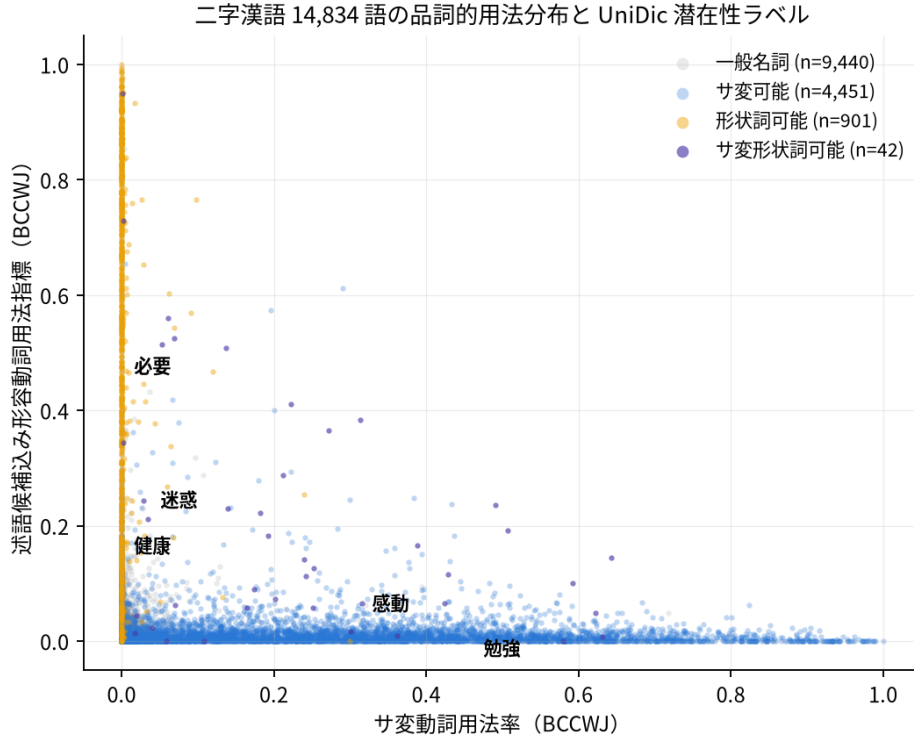


図 1 BCCWJ におけるサ変動詞用法率×述語候補込み形容動詞用法指標の分布。両方が 10% 以上の語は 14,834 語中 59 語である。

図 2 は同じ対象語を、サ変動詞用法率、形容動詞中核用法率、残余用法率の三軸に配置した透視図である。三率の合計が 100% になるため、全点は

$$R_V + R_A^{\text{core}} + R_{\text{res}} = 1 \quad (10)$$

の単体平面上に位置する。残余頂点への距離から、二つの主要用法率に現れなかった用例の量を確認できる。残余には名詞的文脈と曖昧群が含まれるため、その内部構成は語別集計と併読する。

二字漢語の三軸用法空間（BCCWJ、頻度50以上）

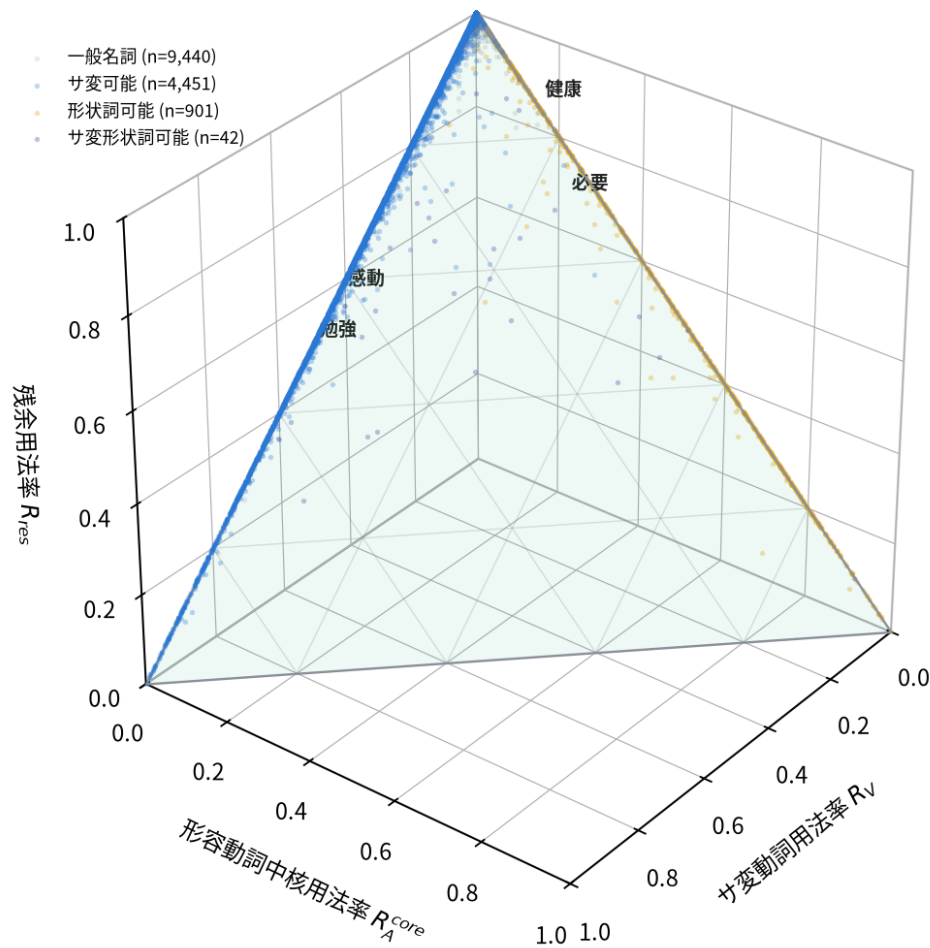


図2 BCCWJにおけるサ変動詞用法率、形容動詞中核用法率、残余用法率の三軸用法空間。三率は排他的なフレーム区分に基づき、合計100%となる。

3.2. 手法の妥当性検証

間淵（2018: 98–112）は、「動詞用法頻度／出現総頻度」を、5%未満・5–20%・20–40%・40–80%・80%以上の5段階に分類した。博士論文の資料1に掲載された2,202語との照合では、太陽コーパスで911語、BCCWJで909語が共通した。レベル1から5の本研究サ変動詞用法率の平均は、太陽で2.0%・12.5%・29.5%・55.8%・87.4%、BCCWJで2.1%・14.8%・30.6%・52.7%・86.3%となった（図3）。太陽側はCD-ROM版XMLを近代文語UniDicで解析した結果であり、間淵の形態論情報とは処理経路が異なる。同じ原資料から得た二つの集計が、近代・現代の双方で単調に対応した。

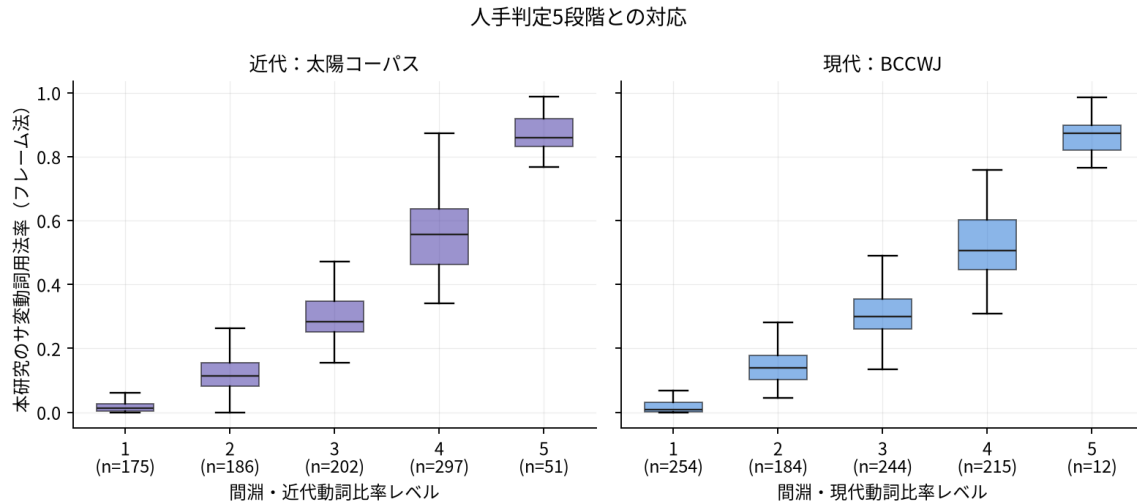


図3 間淵（2018）の動詞比率レベルにみた、本研究のサ変動詞用法率。左は太陽コーパスの近代値、右はBCCWJの現代値。箱の中央線は中央値、箱は第1-第3四分位数を示す。

3.3. 太陽コーパス内の変化（1895–1925）

全対象トークンを年ごとに合算すると、サ変動詞用法率は1895年9.9%、1901年10.9%、1909年10.4%、1917年10.7%、1925年10.1%である。全体比率は約10%で推移する。形容動詞中核用法率は3.6%・4.2%・4.2%・3.9%・3.8%、述語候補込み指標は5.6%・6.5%・7.4%・7.0%・7.9%となった。後者の上昇にはA_{pred}の増加や各年に出現する語の構成差も含まれる。形容動詞用法の増加を論じる際は、中核用法率、曖昧群の比率、同一語の変化を分けて確認する必要がある。

1895年と1925年の双方で30例以上ある1,032語について、サ変動詞用法率の差を求めた。絶対値15pt以上は95語で、減少54語、増加41語である（図4）。減少幅の上位には超過（−61.1pt）、執行（−55.7pt）、代表（−47.3pt）があり、増加幅の上位には比例（+38.2pt）、廃止（+33.5pt）、同情（+33.4pt）がある。図5は例数とWilson 95%区間を添えた5時点の系列である。語別系列には単調でない推移もあり、端点差と途中経過を併記する。

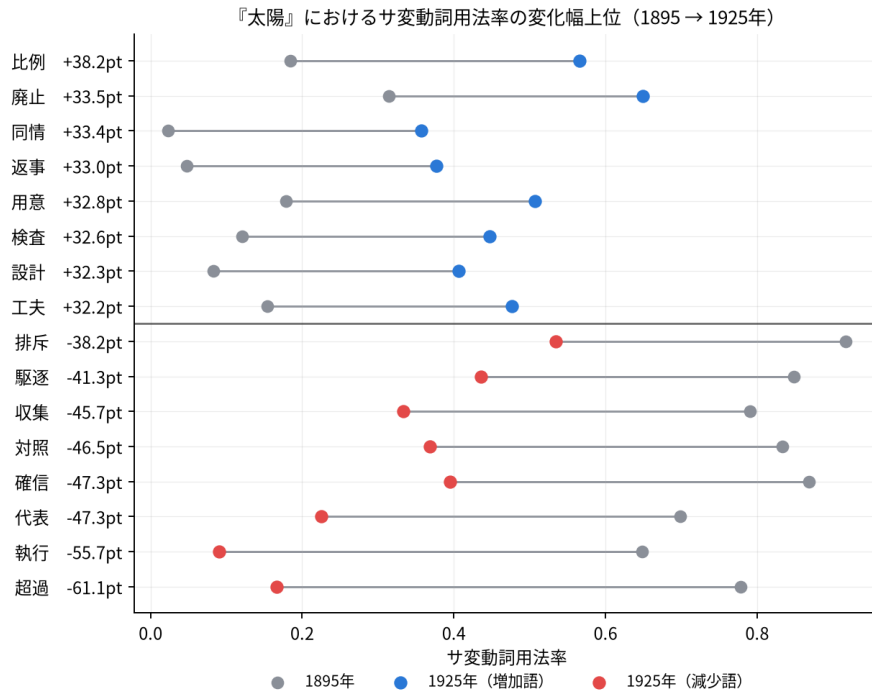


図 4 太陽コーパスの 1895 年 → 1925 年におけるサ変動詞用法率の変化幅上位各 8 語。両年 $n \geq 30$ 、|変化幅| ≥ 15 pt。語の後の数値は変化幅を示す。

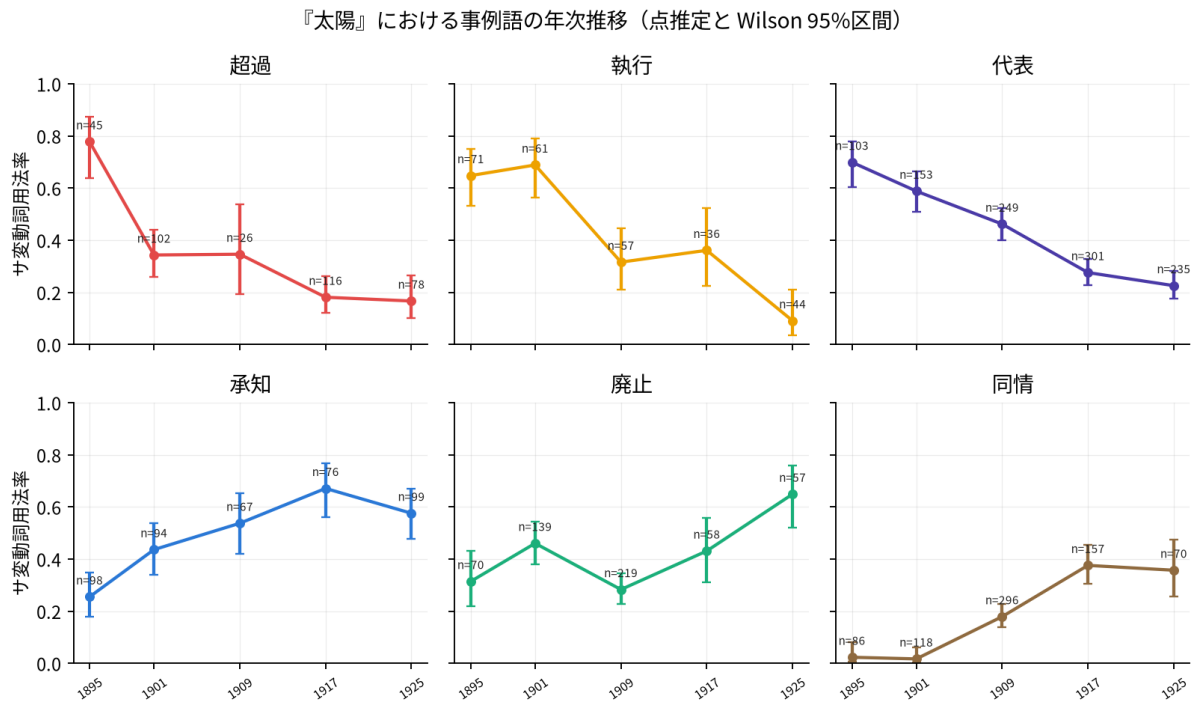


図 5 太陽コーパスにおける事例 6 語のサ変動詞用法率。縦線は Wilson 95% 区間、ラベルは各年の総出現数を示す。

3. 4. 通時変化（1703–2008）

図 6 に事例 6 語の層別推移を示す。感動のサ変動詞用法率は 1874–1889 年の 74.4% から BCCWJ の 39.7% へ低下する。健康の述語候補込み形容動詞用法指標は 15.8% から戦後層の 39.8% まで 上がり、

BCCWJでは19.7%となる。案内のサ変動詞用法率は膝栗毛の12.5%から戦後層の53.9%まで上がり、BCCWJでは37.8%となる。層別比較には資料構成の差が含まれる。

変化幅は「BCCWJのサ変動詞用法率－明治期のサ変動詞用法率」とし、パーセントポイント（pt）で示す。絶対値20pt以上の840語は、減少682語、増加158語である。

$$\Delta R_{V(w)} = R_{V(w, \text{BCCWJ})} - R_{V(w, \text{明治期})} \quad (11)$$

図7は、さらに明治期40例以上・BCCWJ100例以上という頻度条件を満たす442語（減少364語、増加78語）から、増加幅と減少幅の上位を9語ずつ示す。

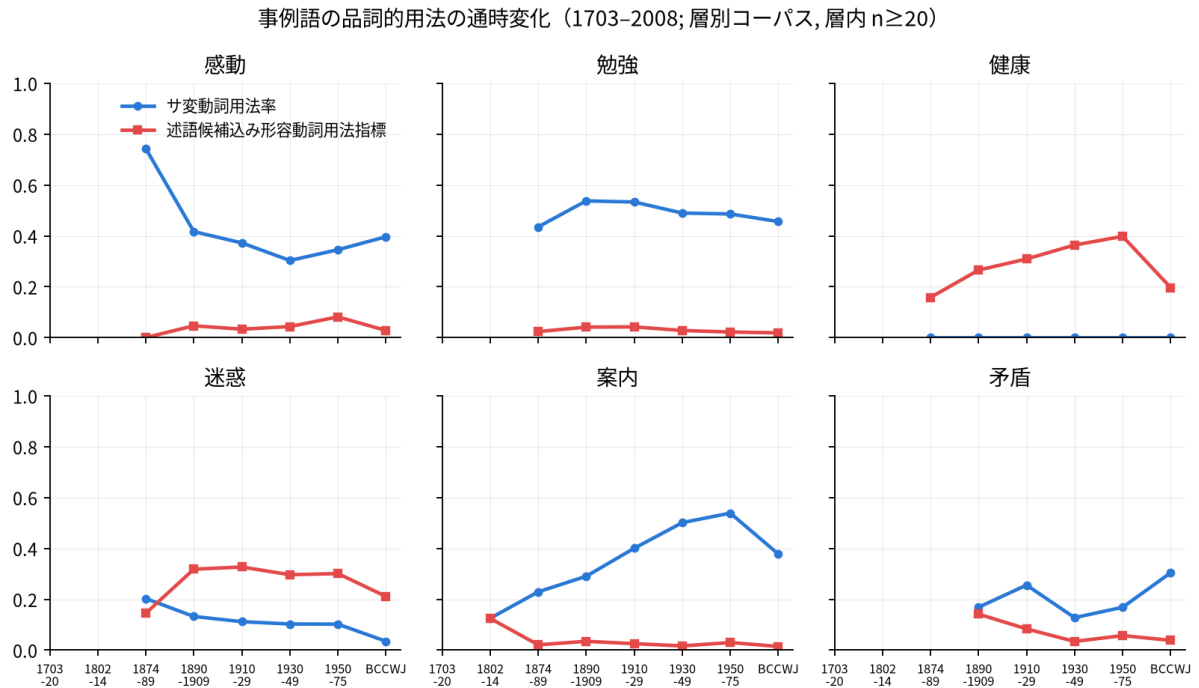


図6 事例語のサ変動詞用法率（青）・述語候補込み形容動詞用法指標（赤）の層別推移。層内頻度20未満は非表示。

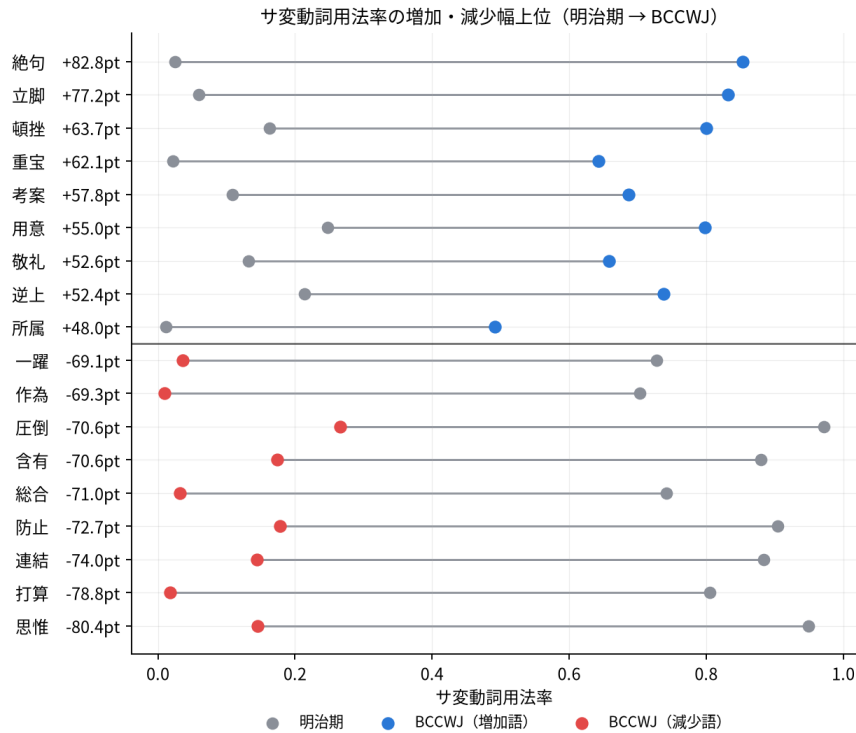


図7 明治期 → BCCWJ のサ変動詞用法率の増加幅・減少幅上位各9語。語の後の数値は変化幅 (pt) を示す (明治期 $n \geq 40$ ・BCCWJ $n \geq 100$)。

「思考」のサ変動詞用法は、1874–1889年 46/58 (79.3%)、1890–1909年 60/91 (65.9%)、1910–1929年 20/99 (20.2%)、1930–1949年 38/321 (11.8%)、1950–1975年 6/78 (7.7%)、BCCWJ 130/2,763 (4.7%)と推移する。現代にも130例があり、用法自体は存続している。BCCWJでは名詞用法1,022例、複合語内用法1,147例が分母の大部分を占める。太陽コーパスの5時点だけを見ると、1895年 6/12、1901年 27/38、1909年 24/27、1917年 5/8、1925年 3/9である。後半2時点は例数が少ない。層別系列の低下には、サ変動詞用法数の変化、名詞・複合語内用法数の変化、資料のレジスター構成差が関与する。

3.5. 変化時期の同定 (年次解像度)

NDL雑誌 n-gram (1876–2015, 年次) によるサ変動詞用法率の代理系列 (図8) は、年ごとの表層形頻度から変化時期の候補を得る資料である。「実感」は1975年以降に上昇が加速し、「誕生」は1935–1945年に急伸する。「思考」はNDL代理系列で1945–1960年の低下が目立つ一方、形態論情報を使う層別コーパスでは1910–1929年までに大幅な低下が現れる。二系列は分母、OCR、収録雑誌、形態論判定が異なる。変化点の確定には、同一レジスターの用例列と複数資料の一致を用いる。

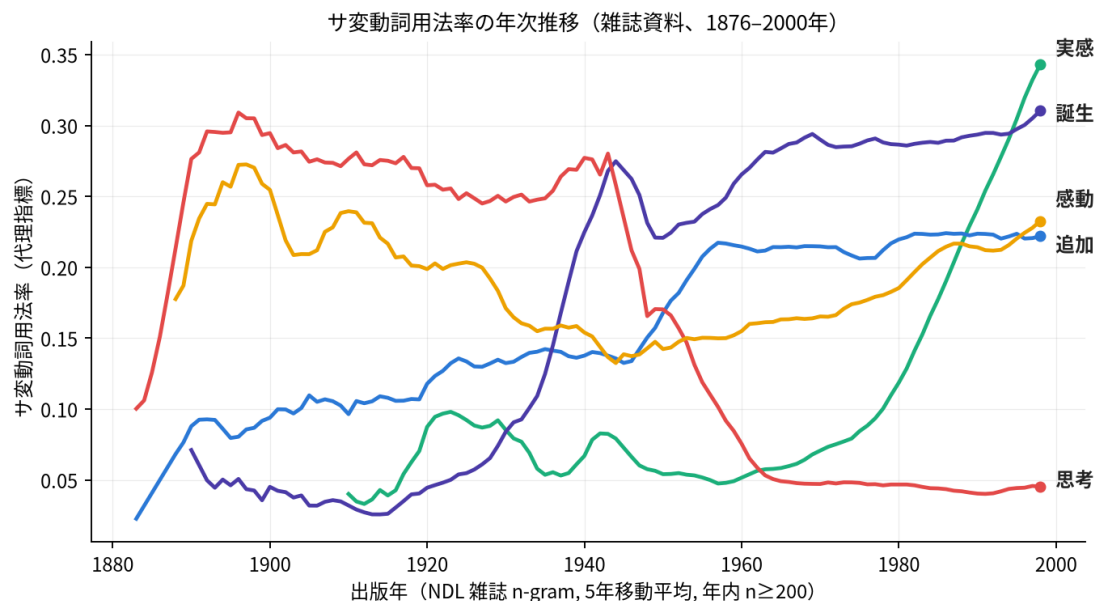


図 8 雑誌資料の年次サ変動詞用法率の代理系列（5 年移動平均・年内頻度 200 以上）。

4. 限界と留意点

(i) n-gram 系列は表層形に基づく代理指標であり、分母に複合語内用法を含む。(ii) 青空文庫は底本・初出年の混在と文学偏重、雑誌 n-gram は OCR ノイズを含む。(iii) 太陽コーパスは 1895・1901・1909・1917・1925 年の標本であり、年次連続資料ではない。(iv) 時代層とレジスターが交絡する。レジスター内通時比較には帝国議会・国会会議録を用いる。(v) 近世層は約 30 万トークンのサンプルである。(vi) A_{pred} と A_{no} は名詞・形容動詞間の曖昧群であり、中核用法率と操作的上限値を併記する。(vii) 太陽コーパスの原文・派生データは購入製品の利用条件に従い、本文・トークン列を公開しない。二次的著作物の公開前に許可条件を確認する。(viii) 太陽の形態論解析では 17,867 トークン (0.18%) の語彙素が空欄となり、その多くは固有名詞・普通名詞である。これらは二字漢語のフレーム集計から除外される。

5. 今後

太陽コーパスの号・文体・ジャンル別分析、SHC による 1925 年以後の補完、国会・帝国議会会議録の統合、変化点検出の統計化を行う。意味クラス（分類語彙表）・語構成・アスペクト型は回帰分析の説明変数に用いる。曖昧群 A_{pred}・A_{no} は時代・レジスター・頻度で層化抽出し、複数判定者による配分率を推定する。

6. 参考文献・資料

- 間淵洋子 (2016) 「近現代漢語におけるサ変動詞用法の変化」『国際日本学論集』4, 17-36.
- 間淵洋子 (2018) 『近代漢語における表記・語法の多様性とその変化に関する計量的研究』明治大学博士論文.
- 永澤済 (2010) 「変化パターンからみる近現代漢語の品詞用法」『東京大学言語学論集』30, 115-168.
- 永澤済 (2011) 「漢語『-な』型形容詞の伸張」『東京大学言語学論集』31, 135-164.
- 塚脇幸代 (2018) 「日本語の名詞と接辞『な』『に』『の』」『言語処理学会第 24 回年次大会発表論文集』P6-5.
- 国立国語研究所 (編) (2005) 『太陽コーパス—雑誌『太陽』日本語データベース』博文館新社.

7. 用語

表 3 本報告で用いる主要用語。

用語	本報告での意味
UniDic	国立国語研究所が整備する電子化辞書。短単位の語彙素、品詞、活用型、活用形、語種を提供する。
Vibrato	UniDic の情報を付与する形態素解析器。青空文庫の処理に用いた。
BCCWJ	『現代日本語書き言葉均衡コーパス』。
NDL	国立国会図書館。
OCR	紙面画像の文字を機械的に読み取ってテキスト化する光学文字認識。
語彙素	活用形や表記変異をまとめる UniDic の辞書見出し形。「し」「する」の語彙素は「為る」。
文脈フレーム	対象トークンの前後に現れる形態素列による操作的な用法分類。
層	年代、資料、レジスターを揃えた集計区分。
レジスター	雑誌、書籍、会議録などの媒体・使用場面の種類。
潜在性ラベル	UniDic の「サ変可能」「形状詞可能」等が示す辞書上の用法可能性。
代理指標	直接の品詞判定を伴わない頻度表から近似的に求める値。
n-gram	連続する文字または単語列とその出現頻度。
5 年移動平均	各年と前後 2 年の値の平均。年次変動を平滑化する。
Wilson 信頼区間	二項比率の標本誤差を表す区間推定。
パーセントポイント (pt)	2 つの割合の差。20% から 50% への変化は 30 pt。

再現 : analysis/2026-07-10_pilot/make_figures.py (データは data/out/rates.parquet, rates_taiyo.parquet, rates_taiyo_year.parquet, ngram_frames_zasshi-all.parquet, mabuchi2018_appendix.tsv)。