

二字漢語の品詞的用法に関するご質問への回答

太陽コーパス統合集計

2026 年 7 月 14 日

1. ご質問の整理

ご質問は次の六点到整理できる。

1. サ変動詞用法率と形容動詞用法率の散布図が L 字型になる理由。
2. 間淵（2018）の 5 段階「動詞比率レベル」と本調査のサ変動詞用法率が対応するという結果の意味、および博士論文の所在。
3. 変化幅図の「増加」「減少」が表す内容、語例の選定基準、数値の単位、図中の配置、増加語・減少語の総数。
4. 現代にも「思考する」が使われる中で、「思考」のサ変動詞用法率が大きく低下する理由。
5. 図中の略号と、各指標の分子・分母。
6. 名詞用法と形容動詞用法が同形になる用例の扱い、および分析上の限界。

2. 指標名・略号・計算式

図表と本文では「サ変動詞用法率」「形容動詞用法率」を用いる。内部集計の略号は再現処理の列名として残し、読者向けの図表には日本語名を示す。

表 1 主要な内部略号と操作的定義。

略号	名称	判定条件
V	隣接サ変動詞用法	対象語の直後に動詞「為る」または「出来る」が続く。活用形の「する・した・される・させる」も語彙素「為る」で検出する。
V_split	「を」を介するサ変動詞用法	対象語の直後に格助詞「を」、その次に動詞「為る」が続く。例：勉強をする。
A_na	連体形による形容動詞中核用法	断定助動詞の連体形が続く。現代語の「な」、文語の「なる・たる」を含む。
A_ni	連用形による形容動詞中核用法	断定助動詞の連用形「に」が続く。格助詞の「に」は含めない。
A_pred	名詞・形容動詞述語の曖昧群	その他の断定助動詞が続く。「健康だ」と「問題だ」の双方が入る。
A_no	「の」を伴う連体修飾の曖昧群	助詞「の」が続く。「健康の秘訣」と「研究の成果」の双方が入る。

語 w 、資料層 c におけるサ変動詞用法率は次式で求める。

$$R_{V(w,c)} = \frac{\text{層 } c \text{ における語 } w \text{ のサ変動詞用法数 (V + V_split)}}{\text{層 } c \text{ における語 } w \text{ の総出現数 (全フレーム合計)}} \quad (1)$$

形容動詞は、形態論情報から判定できる中核用法と、名詞との曖昧群を分ける。

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

$$R_{A+pred}^{proxy(w,c)} = \frac{\text{形容動詞中核用法数と名詞・形容動詞述語候補数 (A_na + A_ni + A_pred)}}{\text{層 } c \text{ における語 } w \text{ の総出現数 (全フレーム合計)}} \quad (3)$$

$$R_A^{upper(w,c)} = \frac{\text{中核用法数と全曖昧候補数 (A_na + A_ni + A_pred + A_no)}}{\text{層 } c \text{ における語 } w \text{ の総出現数 (全フレーム合計)}} \quad (4)$$

名詞側も同じ分母を用いる。

$$R_N^{core(w,c)} = \frac{\text{層 } c \text{ における語 } w \text{ の独立名詞用法数 (N)}}{\text{層 } c \text{ における語 } w \text{ の総出現数 (全フレーム合計)}} \quad (5)$$

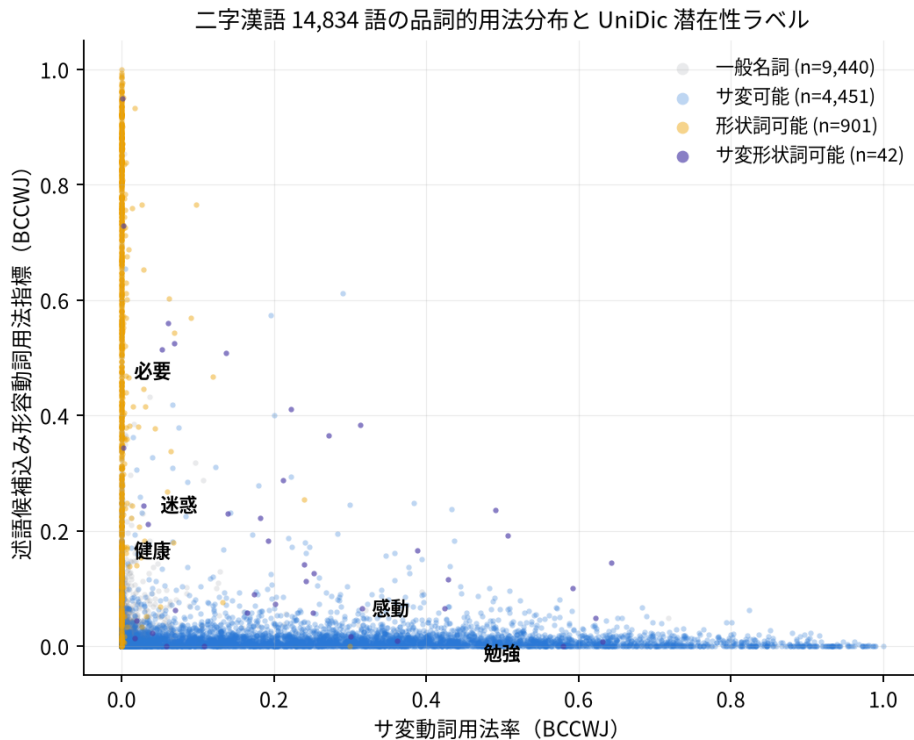
$$R_N^{context(w,c)} = \frac{\text{独立名詞用法数と複合語内用法数 (N + COMP)}}{\text{層 } c \text{ における語 } w \text{ の総出現数 (全フレーム合計)}} \quad (6)$$

$$R_N^{upper(w,c)} = \frac{\text{名詞的文脈数と全曖昧候補数 (N + COMP + A_pred + A_no)}}{\text{層 } c \text{ における語 } w \text{ の総出現数 (全フレーム合計)}} \quad (7)$$

3. L字型分布の意味

各用例は一つのフレームへ排他的に分類される。このため、サ変動詞用法率と形容動詞用法率の和は100%を超えず、散布図の右上が計算上制限される。名詞用法や複合語内用法も同じ分母に含まれるため、二つの率が同時に高くなる範囲は狭い。

BCCWJで頻度50以上の14,834語を調べると、サ変動詞用法率と述語候補込み形容動詞用法指標がともに10%以上の語は59語(0.4%)である。形容動詞中核用法率を用いる場合は14語となる。軸付近への集中には、各語がサ変動詞用法または形容動詞用法の一方へ偏る語彙的傾向も関与すると考えられる。この傾向は意味分類、語構成、頻度、潜在性ラベルを統制した分析で検証する。



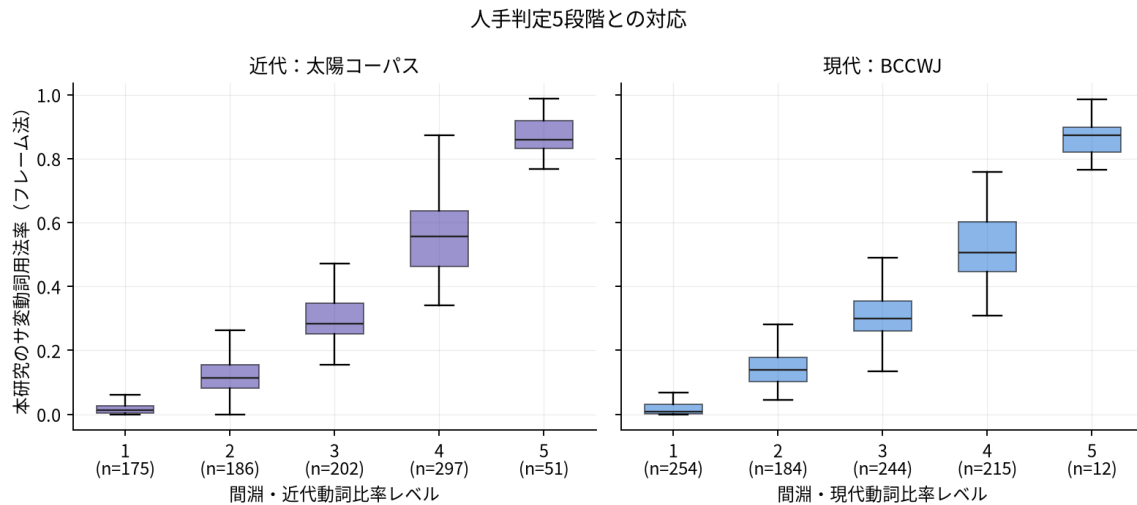
4. 間淵 (2018) の 5 段階との照合

間淵 (2018: 98–112) は、動詞用法頻度を出現総頻度で除した値を 5 段階に分類する。本調査の連続量を各段階別に集計した結果は次のとおりである。

表 2 間淵（2018）の動詞比率レベルと本調査のサ変動詞用法率。区間の上端は次段階に含む。

段階	間淵の範囲	太陽 n	太陽平均	BCCWJ n	BCCWJ 平均
1	5% 未満	175	2.0%	254	2.1%
2	5–20%	186	12.5%	184	14.8%
3	20–40%	202	29.5%	244	30.6%
4	40–80%	297	55.8%	215	52.7%
5	80% 以上	51	87.4%	12	86.3%

段階が上がるにつれて、本調査の平均サ変動詞用法率も単調に上がる。共通語は太陽コーパス 911 語、BCCWJ 909 語である。太陽側は同じ原資料を別の形態論解析と集計手順で処理している。近代・現代の双方で対応が確認できる。



参照文献は、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）である。定義と 5 段階は 98–112 頁、2,202 語の一覧は資料 1（169–199 頁）にある。明治大学学術成果リポジトリで書誌情報と本文を確認できる。

5. 変化幅図の読み方

「増加」と「減少」は用法率の差を表す。用法数が 0 になった場合だけを減少と呼ぶ基準や、0 から初出した場合だけを増加と呼ぶ基準は採用していない。図の名称には「サ変動詞用法率の増加」「サ変動詞用法率の減少」を用いる。

$$\Delta R_{V(w)} = R_{V(w, \text{比較後の資料})} - R_{V(w, \text{比較前の資料})} \quad (8)$$

単位はパーセントポイント (pt) である。20% から 50% への変化は +30pt となる。明治期から BCCWJ までの図は、明治期 40 例以上、BCCWJ 100 例以上、変化幅の絶対値 20pt 以上を条件とし、増加幅と減少幅の上位を 9 語ずつ示す。各語に変化幅を付し、増加群と減少群の間に仕切り線を置く。

絶対値 20pt 以上の語は 840 語で、減少 682 語、増加 158 語である。頻度条件も満たす 442 語では、減少 364 語、増加 78 語となる。

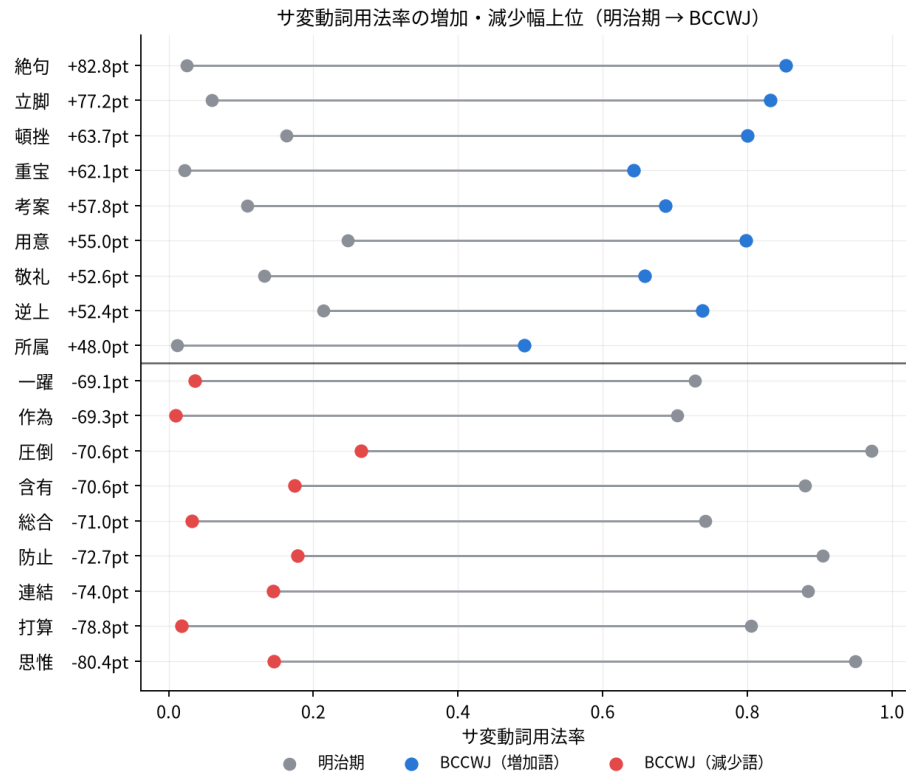


図 3 明治期から BCCWJ までのサ変動詞用法率の変化幅上位。

6. 太陽コーパス内の変化

太陽コーパスの 1895・1901・1909・1917・1925 年、全 60 号 3,409 記事を同じ条件で解析した。総トークン数は 9,697,509、分類対象の二字漢語は 1,535,512 トークンである。

全対象トークンを年ごとに合算したサ変動詞用法率は 9.9%・10.9%・10.4%・10.7%・10.1% であり、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% となる。後者には A_pred の比率と 各年の語彙構成が影響する。全語合算値と同一語の変化を分けて解釈する。

1895 年と 1925 年の双方で 30 例以上ある 1,032 語のうち、変化幅の絶対値 15pt 以上は 95 語で、減少 54 語、増加 41 語である。減少幅の上位は超過 (−61.1pt)、執行 (−55.7pt)、代表 (−47.3pt)、増加幅の上位は比例 (+38.2pt)、廃止 (+33.5pt)、同情 (+33.4pt) である。

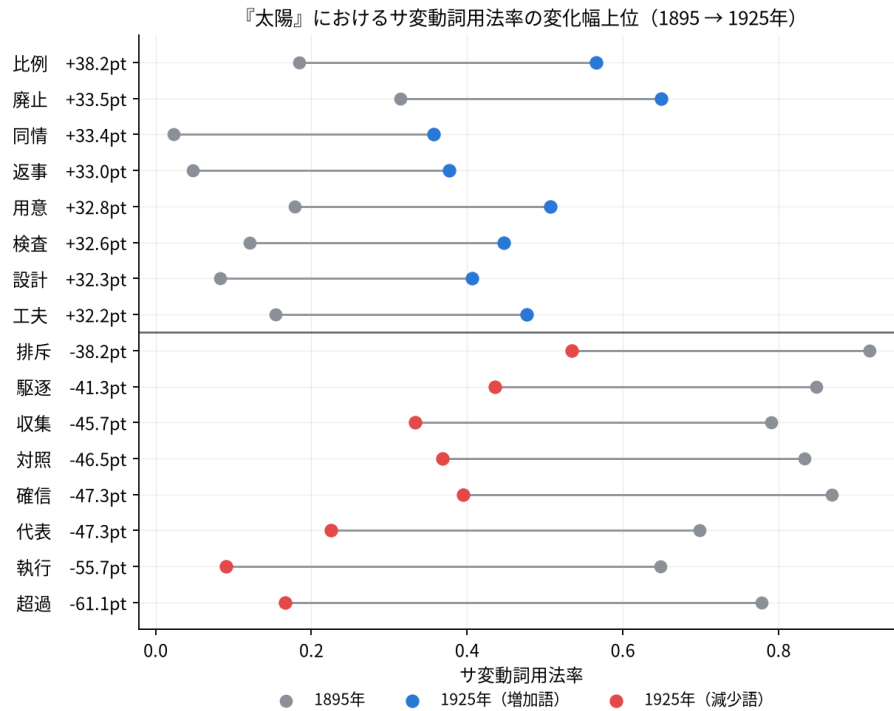


図 4 太陽コーパスの 1895 年から 1925 年までの変化幅上位。両年 $n \geq 30$ 、|変化幅| $\geq 15\text{pt}$ 。

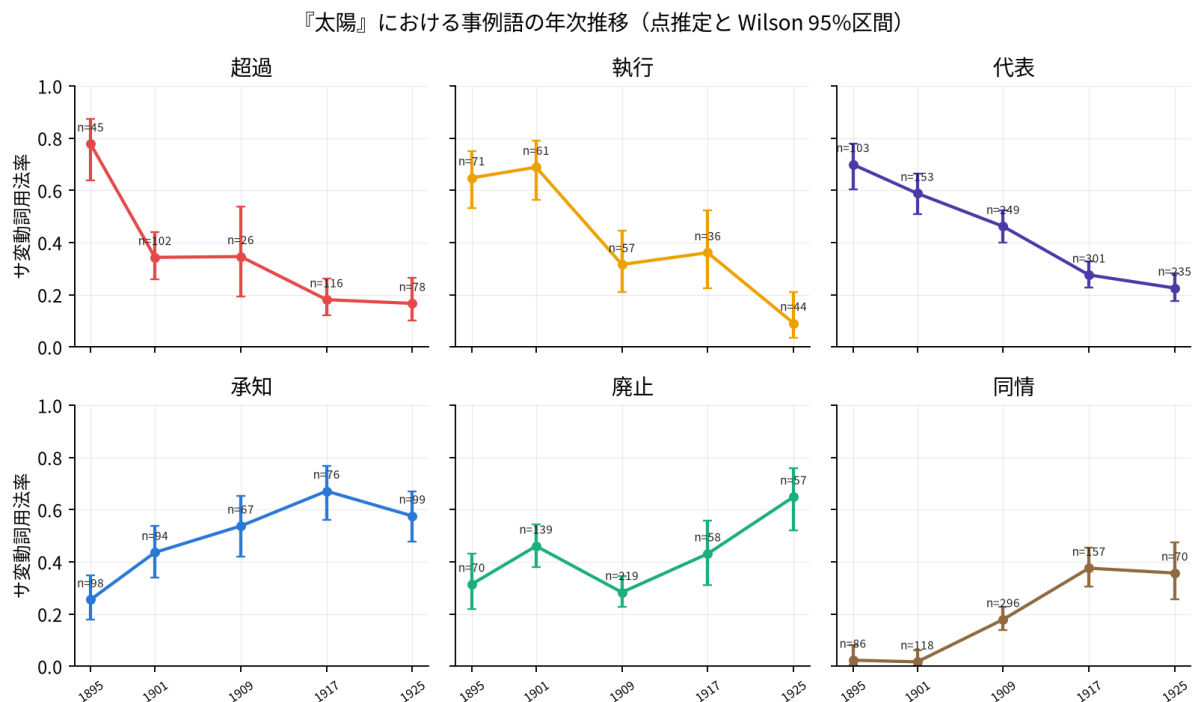


図 5 事例語の 5 時点系列。縦線は Wilson 95% 区間、ラベルは各年の総出現数を示す。

7. 「思考」の解釈

「思考」のサ変動詞用法率は、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。現代にもサ変動詞用法が 130 例あり、「思考する」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、「の」を伴う連体修飾用法 425 例などが分母に入る。「思考過程」「思考実験」「思考の方法」のような用法が多いため、サ変動詞用法率は低くなる。

太陽コーパス各年の内訳は 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 時点の例数は少ない。層別系列の低下には、サ変動詞用法数、名詞・複合語内用法数、資料のレジスター構成が関与する。NDL 雑誌 n-gram では 1945–1960 年の低下が目立ち、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。分母、OCR、収録雑誌、形態論判定の差を調べ、同一レジスターの用例列で時期を再検証する。

8. 名詞用法と形容動詞用法の曖昧性

「健康な生活」は形容動詞の連体修飾、「健康が第一だ」は格助詞を伴う名詞的構文として判定できる。「健康に暮らす」の「に」が断定助動詞の連用形なら形容動詞中核用法となり、「必要に応じる」の「に」が格助詞なら名詞的構文となる。

「健康だ」と「問題だ」はコピュラが同形であり、短単位の形態論情報だけでは用例ごとの所属を一意に決められない。「健康の秘訣」と「研究の成果」に現れる「の」も同じ問題を持つ。この二群は A_pred と A_no に保留し、中核用法率、曖昧群の比率、操作的上限值を併記する。名詞側にも中核値、複合語込み値、操作的上限值を示す。

追加の用例判定では、時代・レジスター・頻度で曖昧例を層化抽出する。複数の判定者が「形容動詞」「名詞」「判定保留」の 3 値で注釈し、判定者間一致度、保留率、合意例から推定した曖昧群の配分率を報告する。BCCWJ の長単位品詞、連体形への言い換え、程度副詞との共起、格成分の取り方、文脈内の意味関係を補助資料に用いる。

9. 分析仕様と図表に必要な項目

1. 指標名を「サ変動詞用法率」「形容動詞用法率」に統一し、略号を凡例で定義する。
2. 全指標の分子、分母、対象フレームを数式で示す。
3. L 字型分布について、計算上の制約と語彙的傾向を分けて検証する。
4. 間淵 (2018) の段階範囲、共通語数、平均値、箱ひげ図の読み方を図の近くに置く。
5. 変化幅の式、単位、頻度条件、掲載順位、増加語数、減少語数を図注に示す。
6. 各語に変化幅を付し、増加群と減少群の間に仕切り線を置く。
7. 主要例は各時期の分子・分母、用法別内訳、Wilson 95% 区間を示す。
8. 太陽コーパスの端点差と 5 時点系列を併記する。
9. 全語合算値と同一語の変化を分け、語彙構成の影響を記す。
10. A_pred と A_no を曖昧群として保存し、人手判定の一致度と保留率を報告する。

10. 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」『国際日本学研究論集』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.

数値は太陽コーパス統合後の同一フレーム集計による。原文と派生トークン列の扱いは製品の利用条件に従う。