

# 日本語語彙評定データの構築： ベイズ線形混合モデルによる分析

浅原 正幸 国立国語研究所 masayu-a@ninjal.ac.jp

## 概要

本論文は、『分類語彙表』(WLSP; 100,827 語)に基づき、日本語語彙に対する新しい主観規範を提示する。母語話者を対象としたウェブ調査において、各語について具象性、抽象性、古さ、新しさ、多義性の5尺度(0-5)で評定を収集した。単語項目と評定者に対するランダム効果をもつベイズ線形混合モデル(BLMM)により、共通分散と項目・評定者起源の分散を分離し、信頼できる効果推定を得た。

## 1 はじめに

語彙評定データ—具象性・抽象性・多義性など、心理学的に重要な特性に対する人間の評定—は、心理言語学と自然言語処理における研究の基礎となる重要なデータである。これらは統制された刺激選定を可能にし、回帰分析において解釈可能な予測変数を提供し、記号接地の評価における参照層として機能する。

本論文は、日本語語彙に対する新しい主観評定データを、『分類語彙表』(WLSP; 100,827 語) [1]に基づいて構築する。日本語母語話者を対象としたクラウドソーシング調査では、各ターゲット語について、0-5のリッカート尺度による5つの側面(具象性、抽象性、古さ、新しさ、多義性)で評定を得た。元の語彙資源における語義の曖昧性に配慮し、評定はWLSPの語義ラベルで構造化した。これにより、必要に応じて語義ラベルを反映した項目レベルの推定が可能となる。

方法論的には、単語ごとおよび参加者ごとのランダム効果を含むベイズ線形混合モデル(Bayesian Linear Mixed Models; BLMM)で評定値を分析する。両方の群化因子にわたる分析により、(項目と参加者の異質性といった)極端な不均衡下でも推定が安定化し、項目分散と評定者分散を分離できる。

経験的には、5つの次元のあいだに明瞭な構造が確認された。すなわち、強い〈具象性—抽象性〉の

拮抗、〈多義性〉と抽象性の体系的な結びつき、そして〈古さ—新しさ〉が対極をなす通時的(時代的評価)軸である。さらに、本規範を外部的妥当化するため、既存の調査結果—たとえば親密度の5側面(〈知る〉/〈書く〉/〈読む〉/〈話す〉/〈聞く〉) [2]—との相関を検証した。

**貢献** 本研究の貢献は次のとおりである。(i) WLSP全語彙項目を対象に、5つの主観的規範を統合的かつ大規模にカバーする日本語規範を初めて提示する；(ii) 語義ラベル付き評定を提供し、語義の曖昧性を解消した語彙意味研究に適用可能とする。

## 2 関連研究

具象性・抽象性や多義性といった語彙的特性は、心理言語学および計算言語学において、主観的評定に基づく大規模データとして蓄積されてきた。本節では、日本語を中心とした代表的な言語資源を簡潔に概観する。

日本語では、AWD-J [3]が15,220語に対する5段階の抽象性評定を提供しており、分布表現に基づく拡張版(AWD-J EX; 約43万語)も公開されている。英語では、Brysbaertらによる約4万語の具象性評定 [4]が標準的資源として広く用いられており、本研究の設計上の参照点となっている。

単語親密度は日本語における主要な主観的尺度の一つであり、NTTデータベースシリーズ(平成版・令和版) [5, 6]により大規模に整備されている。さらに、Asahara [2]は『分類語彙表』に基づき、KNOW / WRITE / READ / SPEAK / LISTENの5側面からなる親密度評定を構築した。内容語に加え、機能語についてもTsutsujiに基づく親密度データが報告されている。

単語心像性については、NTTデータベースシリーズ『単語心像性』 [7]が代表的であり、多義語では語義選択と評定を組み合わせた調査手法が採用されている。古典的研究としては、心像性・具象性・有意味度・学習容易性の相関を示したOgawa and Inamura

次の単語の評価をお願いします。 Please rate the following word.

【不親切】 (ふしんせつ) Unkindness (fushinsetsu)

具体的なもの・ことを表していますか? Does it denote something concrete?

○1  
○2  
○3  
○4  
○5 そう思う

抽象的なもの・ことを表していますか? Does it denote something abstract?

○1  
○2  
○3  
○4  
○5 そう思う

古い言葉ですか? Is this an old word?

○1  
○2  
○3  
○4  
○5 そう思う

新しい言葉ですか? Is this a new word?

○1  
○2  
○3  
○4  
○5 そう思う

いろいろな異なる意味を表しますか? Does it have many different senses?

○1  
○2  
○3  
○4  
○5 そう思う

Scale: 0 = don't think so, 5 = strongly agree

095737 様 活動中心 研修 研修 3,3020

図 1 調査画面

[8] がある。

行動指標との対応関係を検討するための資源として、近年 Japanese Lexical Decision Database (JALEX) [9] が公開された。JALEX は 5,736 語を対象に、大規模オンライン語彙判断課題から反応時間と正確度を収集しており、親密度や抽象性など既存尺度との妥当な関連が確認されている。

### 3 データ収集

日本語母語話者を対象に、クラウドソーシングによる評価調査 (UI は図. 1 参照) を実施し、各刺激語を 0-5 のリッカート尺度により、具象性、抽象性、古さ、新しさ、多義性の 5 つの軸で評価を依頼した。本研究の手続きは所属機関の研究倫理委員会の審査を受け、承認を得た。データ収集は 2025 年 6 月 1-2 日に実施した。

参加者への謝礼は、15 件の回答ごとに 5 円相当のポイントとした。各参加者は 1 ページあたり最大

15 件まで回答でき、これを 5 ラウンド・11 セット繰り返したため、1 人あたりの理論上の最大回答数は  $15 \times 5 \times 11 = 825$  件となる。

最終的に、12,659 名の一意的協力者から、100,827 語について評価を得た。項目レベルの推定を安定させるため、各語につき独立した 20 名の評価者を目標とし、各語は実際に 20 名の異なる参加者から有効回答を受け取った。

## 4 ベイズ線形混合モデル

5 つの評価次元—〈抽象性〉、〈具象性〉、〈古さ〉、〈新しさ〉、〈多義性〉—について、各評価ごとのベイズ線形混合効果モデルを用いた。各モデルでは、0-5 のリッカート応答を連続量<sup>1)</sup>として扱う。

各対象次元  $d \in \{\text{abstract, concrete, obsolete, innovative, polysemous}\}$  について、

$$\mathcal{D}_d = \{(y_i^{(d)}, w_i, s_i)\}_{i=1}^{N_d}$$

を構成する。ここで  $y_i^{(d)} \in \{0, 1, 2, 3, 4, 5\}$  は評価値、 $w_i \in 1, \dots, N_{\text{word}}$  は語のインデックス、 $s_i \in 1, \dots, N_{\text{subj}}$  は参加者のインデックスである。総語数・参加者数を表す定数として  $N_{\text{word}} = 100,827$ 、 $N_{\text{subj}} = 12,659$  を用いた。

各次元  $d$  について、正規分布を仮定し、語および参加者ごとのランダム切片をもつモデルで表現する：

$$y_i^{(d)} \sim \text{Normal}(\mu_i^{(d)}, \sigma^{(d)}), \quad (1)$$

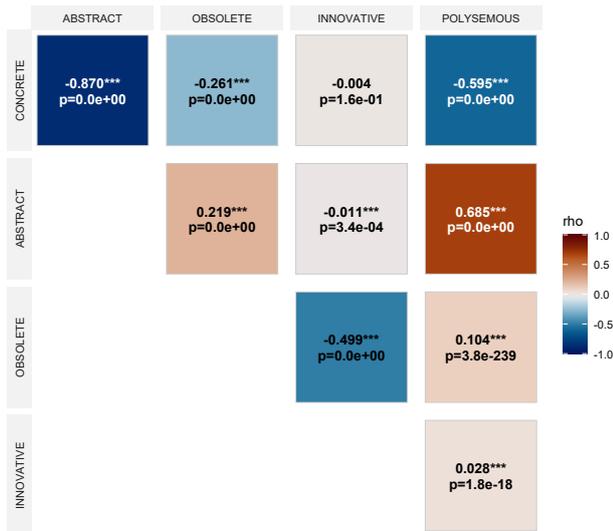
$$\mu_i^{(d)} = \alpha^{(d)} + u_{w_i}^{(d)} + v_{s_i}^{(d)}, \quad (2)$$

$$u_w^{(d)} \sim \text{Normal}(0, \sigma_{\text{word}}^{(d)}) \quad (w = 1, \dots, N_{\text{word}}), \quad (3)$$

$$v_s^{(d)} \sim \text{Normal}(0, \sigma_{\text{subj}}^{(d)}) \quad (s = 1, \dots, N_{\text{subj}}). \quad (4)$$

各次元について同一の MCMC 設定で推定を行った：3 chains, 各 chain 500 iterations (うち 50 iterations は warm-up)。監視したすべてのパラメータに対して Gelman-Rubin 統計量  $\hat{R}$  を算出し、受理基準として  $\hat{R} \leq 1.10$  を課した。

1) なぜリッカート評価に線形-ガウスモデルなのか?: 収集した評価は順序尺度 (0-5) ではあるが、(i) 尺度全体でカテゴリ使用が密であり、(ii) ランダム効果の推定が主たる目的である場合、線形混合モデルは標準的かつ実用的な選択である。本研究のパイロット分析では、この規模において線形モデルは順位に基づく推論で同等の結論を与えつつ、計算コストを大幅に抑えられることが示された。



(CVD-safe palette; white labels for  $|\rho| \geq 0.55$ )

図2 Heatmap of Spearman's correlations among Dimensions

表1 Pairwise Spearman correlations among subjective norms (Spearman  $\rho$  shown to 4 significant digits).

Var1	Var2	Spearman $\rho$
CONCRETE	ABSTRACT	-0.8700***
CONCRETE	OBSOLETE	-0.2610***
CONCRETE	INNOVATIVE	-0.0040
CONCRETE	POLYSEMIOUS	-0.5950***
ABSTRACT	OBSOLETE	0.2190***
ABSTRACT	INNOVATIVE	-0.0110***
ABSTRACT	POLYSEMIOUS	0.6850***
OBSOLETE	INNOVATIVE	-0.4990***
OBSOLETE	POLYSEMIOUS	0.1040***
INNOVATIVE	POLYSEMIOUS	0.0280***

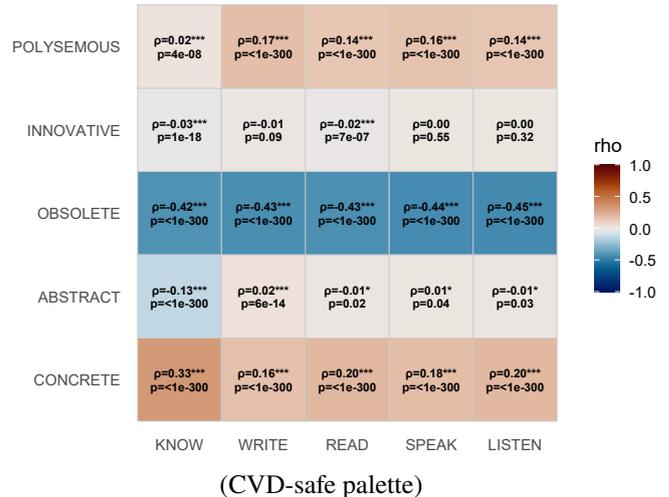
Stars indicate significance based on Benjamini-Hochberg adjusted  $p$ -values: \*\*\* < .001.

## 5 データ分析

### 5.1 5つの評定次元間の相関

まず、全語を対象に、項目レベルで5つの評定次元の関係性をペアごとの Spearman の順位相関 ( $\rho$ ) で評価する。効果量と信頼性の双方を要約するため、 $\rho$  は有効数字4桁で示し、有意性は Benjamini-Hochberg (BH) で調整した  $p$  値で評価する (表1)。図2には、上三角の相関構造を可視化する。

主要なパターンは、具象性と抽象性の非常に強い対立 ( $\rho = -0.8700$ ) であり、両尺度が単一の支配的な意味軸を捉えていることを裏づける。多義性は抽象性と同方向に整列し、具象性とは反対方向に位置する。抽象性-多義性では  $\rho = 0.6850$ 、具象性-多義



(CVD-safe palette)

図3 Heatmap of Spearman correlations ( $\rho$ ) between Subjective Norms and Familiarities.

性では  $\rho = -0.5950^{***}$  である。

次に、通時的評価を確認する。古さ obsolete と新しさ innovative は中程度の負の相関 ( $\rho = -0.4990$ ) を示し、時間的両極として設計された役割と整合的である。さらに、古さ obsolete は抽象性 abstract と小さな正の相関 ( $\rho = 0.2190$ )、具象性 concrete と小さな負の相関 ( $\rho = -0.2610^{***}$ ) を示し、古く感じられる語ほど平均して具象性が低く、やや抽象的である傾向を示唆する。

具象性と新しさの相関は BH 補正後には統計的に非有意 ( $\rho = -0.0040$ ,  $p_{BH} = 0.1639$ ) であり、具体-抽象軸が知覚される新しさによって自明に説明されないことを示す。さらに、いくつかの組み合わせは効果量が極めて小さいにもかかわらず有意に達する (例: 抽象性 abstract-新しさ innovative  $\rho = -0.0110^{***}$ )。これは大きな標本による帰結であるため、結果の解釈では  $p$  値よりも効果量の大きさを重視する。

総合すると、この構造はほぼ直交する二つの成分-具象性-抽象性/多義性の軸と、古さ-新しさの軸を示唆する。こうした分解は、これらの次元が親密度に基づく指標や語彙判断行動とどのように関係するかを検討する後続分析で活用する。

### 5.2 親密度との相関

次に、5つの主観規範を、親密度の5側面 KNOW, WRITE, READ, SPEAK, LISTEN と対応づけて検討する。図3は全体パターンを可視化し、表2は Spearman の順位相関 ( $\rho$ ) と BH 補正後の有意性を報告する。標本数が大きいことを踏まえ、解釈では  $p$  値だけで

表2 Spearman correlations ( $\rho$ ) between Subjective Norms and Familiarities.

	KNOW	WRITE	READ	SPEAK	LISTEN
CONCRETE	0.325***	0.161***	0.200***	0.179***	0.195***
ABSTRACT	-0.134***	0.024***	-0.008*	0.006*	-0.007*
OBSOLETE	-0.416***	-0.430***	-0.427***	-0.441***	-0.446***
INNOVATIVE	-0.028***	-0.005	-0.016***	0.002	0.003
POLYSEMOUS	0.017***	0.173***	0.138***	0.155***	0.142***

Stars denote significance (\* < .05, \*\* < .01, \*\*\* < .001).

なく効果量（実質的な重要性）を優先する。

具象性 concrete は親密度の5側面すべてと一貫して正の相関 ( $\rho = 0.161-0.325$ ) を示し、とりわけ KNOW (0.325) と READ (0.200) が大きい。これは、知覚的基盤の強い語ほど、理解課題・産出課題を問わずより親密だと評定されるという古典的知見の再現である。

古さ obsolete はすべての側面と強固な負の相関 ( $\rho = -0.416 \sim -0.446$ ) を示し、とりわけ LISTEN と SPEAK で最も強い。古風だと知覚される語は、受容・産出の双方で親密度が低く評価されやすく、この通時的軸がモダリティを超えて一般化する使用の稀少性を捉えていることを示す。

多義性 polysemous は親密度と正の相関を示し、WRITE (0.173), SPEAK (0.155), LISTEN (0.142), READ (0.138) では控えめながら一貫した相関が見られる一方、KNOW (0.017) はゼロ近傍である。このパターンは、語義が豊かな語ほど、とりわけ産出的モダリティにおいて日常言語使用に強く定着しがちであるという見解と整合的である。

抽象性 abstract は符号が混在する小さな効果（例：KNOW -0.134, WRITE +0.024, その他はゼロ近傍）を示し、具象性と多義性を考慮すると抽象性それ自体の親密度への寄与は小さい。新しさ innovative は本質的に親密度と直交しており（すべて  $|\rho| \leq .028$ ）、知覚される新しさは日常的な親密度の観点で古さの単純な裏返しではないことを示唆する。

全体として、解釈しやすい二つの傾向が見られる。第一に、高い具象性・より高い多義性を持つ語は、諸側面にわたって親密度の高さと整合する。第二に、通時的成分 (obsolete と innovative) は非対称であり、知覚される古さは親密度を一貫して低下させる一方、知覚される新しさそれ自体の寄与は小さい。これらの規則性は、意味評定と語彙判断行動を結びつける後続の分析に示唆を与える。

### 5.3 『分類語彙表』 語義ラベルによる評価

付録参照。

## 6 おわりに

日本語語彙の評定データとして、『分類語彙表』(100,827語)を対象に、具象性、抽象性、古さ、新しさ、多義性の5尺度を新たに整備した。母語話者による0~5のリッカート評定を収集し、単語と参加者の両群にまたがる部分プーリングを備えたベイズ線形混合モデルで分析した。この枠組により、項目(単語)由来の分散と評定者由来の分散を分離し、大規模条件下でも安定した項目水準推定を得た。

実証的には、二つの強固な規則性が得られた。第一に、語彙意味空間は強い具象性-抽象性の拮抗で構造化され、多様性は抽象性に整列し、具象性とは逆方向に位置づく。第二に、ほぼ直交する通時的成分があり、古さと新しさが対極をなす。5つの親密度側面 (Know/Write/Read/Speak/Listen) との外的妥当化では、解釈しやすい連関が確認された：具象性と多義性は親密度と正に相関し、古さはすべてのモダリティで負に相関し、新しさはほぼゼロに近い相関しか示さない。総合すると、知覚的基盤/語彙的定着と通時的評価という二つの成分が、日本語の意味変異の補完的な源泉であることが示唆される。

**貢献** (i) 『分類語彙表』全語彙を対象に、5つの主観的規範について語義に紐づく項目レベル評定を提供；(ii) 項目・参加者ランダム効果を備え、極度に不均衡な評定データに適した再利用可能な BLMM ワークフロー；(iii) 次元間および次元×親密度の関係に関する包括的な相関要約；(iv) 心理言語学および NLP における再利用を促進する、スコアなどの公開パッケージ<sup>2)</sup>。

2) <https://github.com/masayu-a/WLSP-norms/>

## 謝辞

本研究は JSPS 科研費 JP25K00459, JP23K21935, JP22K18483 および国立国語研究所共同研究プロジェクトの助成を受けたものです。

## 参考文献

- [1] 国立国語研究所. 分類語彙表 増補改訂版. 大日本図書, 東京, 2004.
- [2] Masayuki Asahara. Word familiarity rate estimation using a Bayesian linear mixed model. In **Proceedings of the First Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP**, pp. 6–14, Hong Kong, November 2019. Association for Computational Linguistics.
- [3] 日本語抽象度辞書「awd-j: Abstractness of word database for japanese common words」. AWD-J core (15,220 語) / AWD-J EX (437,300 語) ; データ更新日: 2019-04-04; ライセンス: CC BY 4.0.
- [4] Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness ratings for 40 thousand generally known english word lemmas. **Behavior Research Methods**, Vol. 46, No. 3, pp. 904–911, 2014.
- [5] 天野成昭, 近藤公久 (編). 日本語の語彙特性. 三省堂, 東京, 1999.
- [6] 藤田早苗, 奥村優子, 小林哲生, 服部正嗣. 絵本と幼児向けの発話に出現する語の多様性比較. 言語処理学会第 24 回年次大会 発表論文集, pp. 1264–1267, 岡山, 3 2018. 言語処理学会.
- [7] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中正之, 天野成昭, 近藤公久 (編). 日本語の語彙特性 第 8 巻: 単語心像性. NTT データベースシリーズ. 三省堂, 東京, 2005. 2 冊構成 (単語心像性 1・2)、付属 CD-ROM1 枚. 監修: 東京都老人総合研究所・NTT コミュニケーション科学基礎研究所.
- [8] 小川嗣夫, 稲村義貞. 言語材料の諸属性の検討: 名詞の心像性・具象性・有意味度および学習容易性. 心理学研究, Vol. 44, No. 6, pp. 317–327, 1974.
- [9] 太田直斗, 望月正哉. Jalex: 日本語版語彙判断課題データベース. **Frontiers in Language Sciences**, Vol. 3, p. 1506509, 2025. Data Report; 公開日: 2025-01-13; 付随データ: OSF (<https://osf.io/qr2sg>).

## 『分類語彙表』 語義ラベルによる評価

語義ラベル(中項目相当)ごとにBLMMの項目推定値を集約する。要約のため、カテゴリー平均をコーパス全体平均からの偏差( $\Delta$ )で表す。以下では具象性・抽象性・古さ・新しさについて上位5件(値は $\Delta$ と項目数 $N$ )、多義性について上位3件・下位3件を列挙する。

### 具象性: Top 5

- 1.46: 体-生産物-機械  
( $\Delta = +0.632, N = 1149$ )
- 1.43: 体-生産物-食料  
( $+0.609, 1941$ )
- 1.41: 体-生産物-資材  
( $+0.486, 976$ )
- 1.45: 体-生産物-道具  
( $+0.485, 2172$ )
- 1.44: 体-生産物-住居  
( $+0.466, 1216$ )

### 抽象性: Top 5

- 4.50: 他-動物の鳴き声  
( $\Delta = +0.969, N = 34$ )
- 4.31: 他-判断  
( $+0.925, 225$ )
- 4.30: 他-感動  
( $+0.898, 141$ )
- 4.11: 他-接続  
( $+0.885, 209$ )
- 3.15: 相-関係-作用  
( $+0.838, 479$ )

### 古さ: Top 5

- 1.21: 体-主体-家族  
( $\Delta = +0.368, N = 654$ )
- 3.52: 相-自然-天地  
( $+0.335, 37$ )
- 1.20: 体-主体-人間  
( $+0.298, 895$ )
- 4.33: 他-挨拶  
( $+0.267, 171$ )
- 1.22: 体-主体-仲間  
( $+0.163, 269$ )

### 新しさ: Top 5

- 4.50: 他-動物の鳴き声  
( $\Delta = +0.317, N = 34$ )
- 1.46: 体-生産物-機械  
( $+0.311, 1149$ )
- 1.53: 体-自然-生物  
( $+0.298, 170$ )
- 3.53: 相-自然-生物  
( $+0.211, 13$ )
- 3.17: 相-関係-空間  
( $+0.137, 38$ )

### 多義性: Top 3

- 4.32: 他-呼び掛け  
( $\Delta = +0.624, N = 126$ )
- 4.30: 他-感動  
( $+0.551, 141$ )
- 4.50: 他-動物の鳴き声  
( $+0.495, 34$ )

### 多義性: Bottom 3

- 1.43: 体-生産物-食料  
( $\Delta = -0.477, N = 1941$ )
- 1.46: 体-生産物-機械  
( $-0.464, 1149$ )
- 1.42: 体-生産物-衣料  
( $-0.353, 1179$ )