

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@ninja1.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 **concreteness** 〈具象性〉、**abstractness** 〈抽象性〉、**obsoleteness** 〈古さ〉、**innovativeness** 〈新しさ〉、**polysemousness** 〈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。結果として、多義性は抽象性と弱い正の関連を示し、古さと新しさは互いに反対の傾向を示すことから、語の意味に対する通時的評価次元の両極として機能することが確認された。

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。英語については大規模で高品質な資源が存在する一方、日本語の整備状況は発展途上である。親密度（familiarity）やイメージ喚起性（imageability）には長い蓄積があり、抽象性も近年拡充されたが、現代的なモデリングと品質管理のもとで複数の主観規範を統合的かつ大規模に扱う単一の資源は、まだ存在しない。

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（**concreteness** 〈具象性〉、**abstractness** 〈抽

象性〉、**obsoleteness** 〈古さ〉、**innovativeness** 〈新しさ〉、**polysemousness** 〈多義性／語義の多様さ〉）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（polysemousness）〉と抽象性の体系的な結びつき、そして〈古さ—革新しさ〉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（KNOW 〈知る〉/WRITE 〈書く〉/READ 〈読む〉/SPEAK 〈話す〉/LISTEN 〈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；(iii) 強い不均衡下でも信頼できる項目推定を与える BLMM（ベイズ線形混合モデル）ワークフローを示し、評定プロジェクト間で再利用可能とする；(iv) 項目レベルのスコア、調査資料、解析コードを公開し、再現性と心理言語学・NLP における下流利用を支援する。

2 関連研究

本研究で検討する 具象性／抽象性および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の分野において、主観的評定（質問紙調査）に

よって大規模に収集されてきた。以下では、日本語の言語資源に焦点を当て、代表的なものを語彙数、尺度設定、および品質管理の観点から概説する。

2.1 具象性／抽象性

日本語において、AWD-J (Abstractness of Word Database for Japanese) [3] は、クラウドソーシングを通じて 15,220 語の 5 段階評価 (1 = 最も具体的~5 = 最も抽象的) を収集した。また、分布表現に基づいた拡張版 (AWD-J EX; 437,299 語) も公開されており、これは著名な英語の言語資源に匹敵する語彙数を実現している。

英語については、Brysbaert らによる約 40,000 語の 5 段階具象度評価 [4] が事実上の標準的な言語資源となっている。このデータは、信頼性を確保するために共通のキャリブレーション項目と品質管理項目を用いて、4,000 人以上の回答者からオンラインで収集された。これらの設定は、日本語の指示や品質管理を設計する際のテンプレートとして役立っている。

2.2 親密度

単語親密度 (word familiarity) は、日本語における中心的な主観的尺度の一つである。NTT データベースシリーズ『単語親密度』巻の平成版 (約 11 万語) [5] に続き、令和版 (16 万語以上) [6] では再評価と語彙数の拡充が行われた。尺度は 1~7 の 7 段階である。データへのアクセスは NTT 印刷を通じて提供されており、方法論の報告および使用ログの分析結果も公表されている。

Asahara は、『分類語彙表』[1] を基に単語親密度データベースを構築しました。このデータベースでは、単語の親密度について、「知っている (KNOW)」、「書く (WRITE)」、「読む (READ)」、「話す (SPEAK)」、「聞く (LISTEN)」の 5 つの観点を導入した。親密度は 1 から 5 までの評価に基づいており、100,830 語が含まれている。

上記のデータベースが主に日本語の内容語 (実質的な意味を持つ語) であるのに対し、Chen and Asahara (2023) は、日本語の機能語 (文法的な役割を果たす語) の語彙データベースである『Tsutsuji』[7] を基に単語親密度データベースを構築した。これには 6,396 の表層形が含まれている。

2.3 単語心象性

単語心像性 (imageability) は、主に NTT データベースシリーズの『単語心像性』巻に基づくものである。このデータベースでは、7 段階評価による主観的評価が用いられている。同形異義語や多義語については、各語義に対する判断の解釈を明確にするために、対象とする語義の選択と評価を組み合わせた調査手法が採用されている [8]。

より古典的な研究としては、心像性、具象性、意味の明瞭さ (meaningfulness)、および学習容易性の 4 属性について 400 語を同時に評価し、これらの属性間の相関関係を示した研究がある [9]。

2.4 Japanese Lexical Decision Database

最近のデータとして、項目レベルの反応時間 (RT) / 正確度と参加者レベルの指標を備えた日本語初の大規模語彙判断データベースである Japanese Lexical Decision Database (JALEX) がある [10]。刺激は、『分類語彙表』から抽出した 5,736 語 (名詞・動詞・形容詞) で構成され、親密度、頻度、イメージ喚起性、抽象性など 8 つの主要資源でカバーされる項目に限定し、同数の非語を加えた。項目は 38 の均衡化リストに分割された。オンラインの語彙判断課題 (LDT; PsychoPy/Pavlov) では、日本語母語話者 1,652 名が、200ms のブランク、300ms の固視点、2,000ms の応答ウィンドウで実施した。ELP 方式の前処理の結果、項目平均の RT はおよそ 700-710ms、正確度はおよそ .96 となった。収束的妥当性として既知の効果 (頻度・親密度の促進、イメージ喚起性の優位、OLD20 の増大に伴う反応遅延) を再現し、親密度については $r \approx -0.42$ 、抽象性については小さなコスト ($r \approx .11$) を示した。文字近傍効果は表記体系に依存し、漢字では RT が短縮し、カタカナでは逆のパターンが報告された。

3 データ収集と倫理

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 Fig. 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

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: 脳活動・心・対象・感情 3, 3020

図 1 調査画面

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 名の一意な協力者から、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—abstract〈抽象性〉、concrete〈具象性〉、obsolete〈古さ〉、innovative〈新しさ〉、polysemous〈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として扱う。

4.1 データ構造

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 $y_i^{(d)} \in 0, 1, 2, 3, 4, 5$ は評定値、 $w_i \in 1, \dots, N_{\text{word}}$ 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。

4.2 モデル仕様

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。

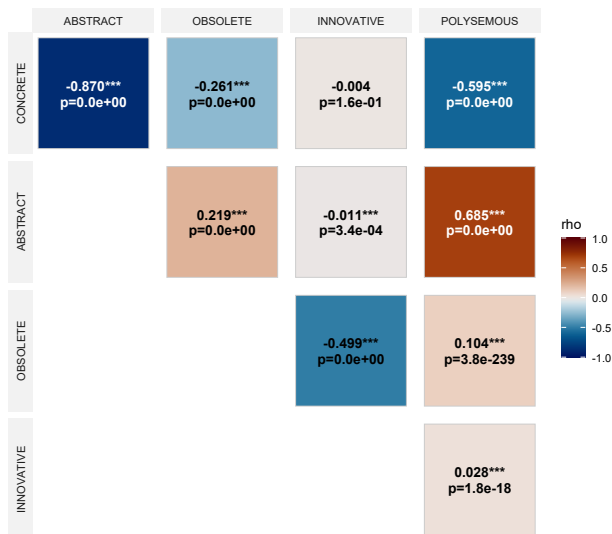
4.3 なぜリッカート評価に線形—ガウスモデルなのか？

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、ガウス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。

5 データ分析

5.1 5 つの評価次元間の相関

まず、全語を対象に、項目レベルで 5 つの評価次元の関係をペアごとの Spearman の順位相関



(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 ρ shown to 4 significant digits).

Var1	Var2	Spearman ρ
CONCRETE	ABSTRACT	-0.8700***
CONCRETE	OBSOLETE	-0.2610***
CONCRETE	INNOVATIVE	-0.0040
CONCRETE	POLYSEMOUS	-0.5950***
ABSTRACT	OBSOLETE	0.2190***
ABSTRACT	INNOVATIVE	-0.0110***
ABSTRACT	POLYSEMOUS	0.6850***
OBSOLETE	INNOVATIVE	-0.4990***
OBSOLETE	POLYSEMOUS	0.1040***
INNOVATIVE	POLYSEMOUS	0.0280***

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

(ρ) で評価する。効果量と信頼性の双方を要約するため、 ρ 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し、図3では10組のペアについて線形トレンドを重ねた代表的散布図を示す。

主要なパターンは、具象性 Concreteness と抽象性 Abstractness 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 Polysemousness は抽象性 Abstractness と同方向に整列し、具象性 Concreteness とは反対方向に位置する。抽象性 Abstractness–多義性 Polysemousness では $\rho = 0.6850$ 、具象性 Concreteness–多義性 Polysemousness では $\rho = -0.5950$ *** である。これらの効果は、図3の該当パネルにおける細長い対角線状の点群として視覚的に

も確認できる。

次に、通時的評価を確認する。古さ Obsolescence と新しさ Innovativeness 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 Obsolescence は抽象性 Abstractness と小さな正の相関 ($\rho = 0.2190$)、具象性 Concreteness 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 Abstractness–新しさ Innovativeness $\rho = -0.0110$ ***)。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4は全体パターンを可視化し、表2は Spearman の順位相関 (ρ) 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 値だけでなく効果量 (実質的な重要性) を優先する。

具象性 具象性 Concreteness は親密度の5側面すべてで一貫して正の相関 ($\rho = .161-.325$) を示し、とりわけ KNOW (.325) と READ (.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 古さ Obsolescence はすべての側面と強固な負の相関 ($\rho = -.416-.446$) 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 多義性 Polysemousness は親密度と正の相関を示し、WRITE (.173), SPEAK (.155), LISTEN

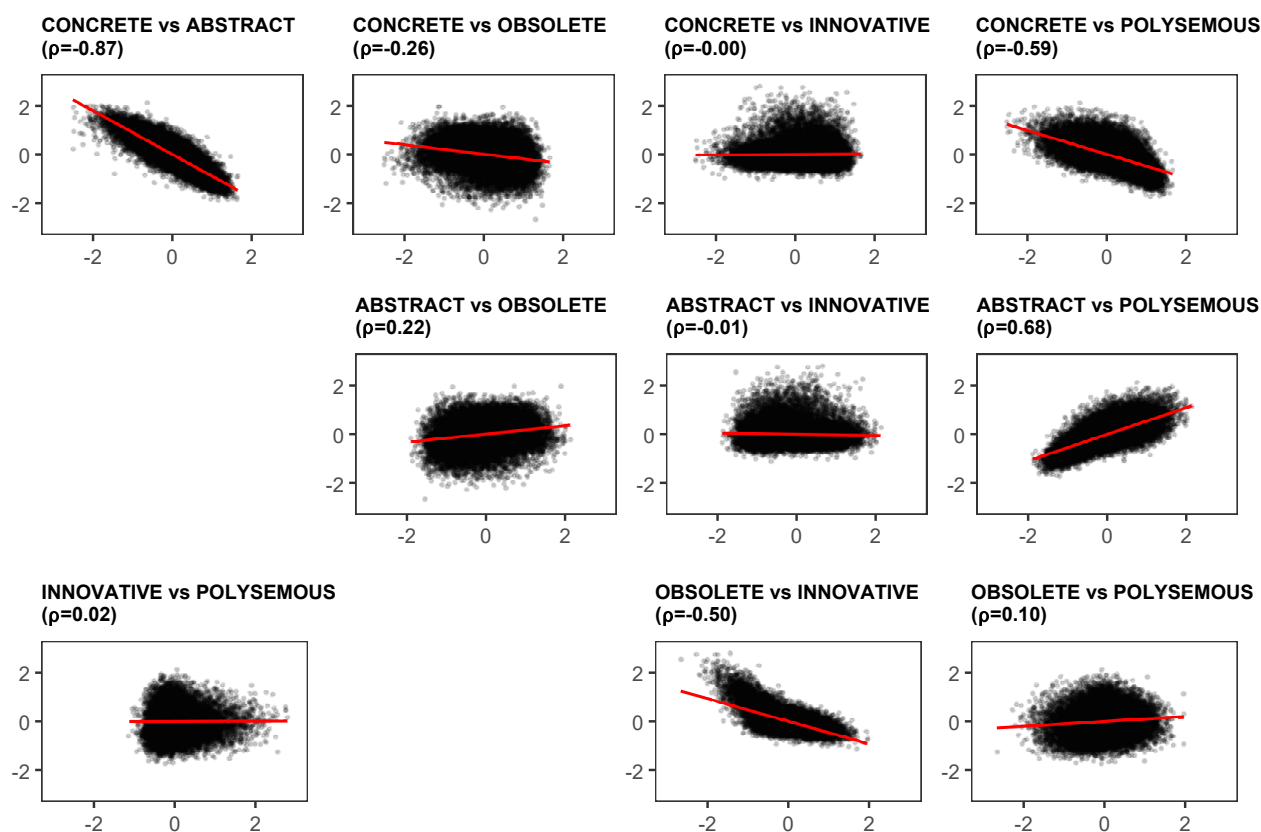


図3 Scatterplots among Subjective Norms

(.142), READ (.138) では控えめながら一貫した相関が見られる一方, KNOW (.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。

抽象性と新しさ 抽象性 Abstractness は符号が混在する小さな効果(例: KNOW -, .134, WRITE +.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 Innovativeness 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。

まとめ 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(Obsolescence と Innovativeness) 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

5.3.1 具象性-抽象性

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と、両尺度のあいだに鏡像的な構造が現れる。すなわち、体-生産物に結びつくラベルは CONCRETE (具象性) で高く、ABSTRACT では低い一方、談話機能・関係に関わるラベルはその逆のパターンを示す。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(Δ ; 0 = 平均的具象性)で表す。以下では各尺度について上位 5 件(値は Δ と項目数 N)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



図 4 Heatmap of Spearman correlations (ρ) between Subjective Norms and Familiarities.

表 2 Spearman correlations (ρ) 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).

(+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)

考察 (1) 人工物中心の体-生産物は、CONCRETE の具体側に集まり、ABSTRACT では負側に位置する。一方で、(2) 談話機能のおよびメタ言語的／論理的ラベル（感動詞、接続詞、呼びかけ、判断、操作）は抽象側の極を形成し、CONCRETE において大きな負の偏差を示す。

体-生産物-機械は具象性で最上位（+0.632）であると同時に ABSTRACT では最下位（-0.723）であり、これに対して他-動物の鳴き声は ABSTRACT で最も抽象的（+0.969）かつ CONCRETE で最も低い（-1.343）。

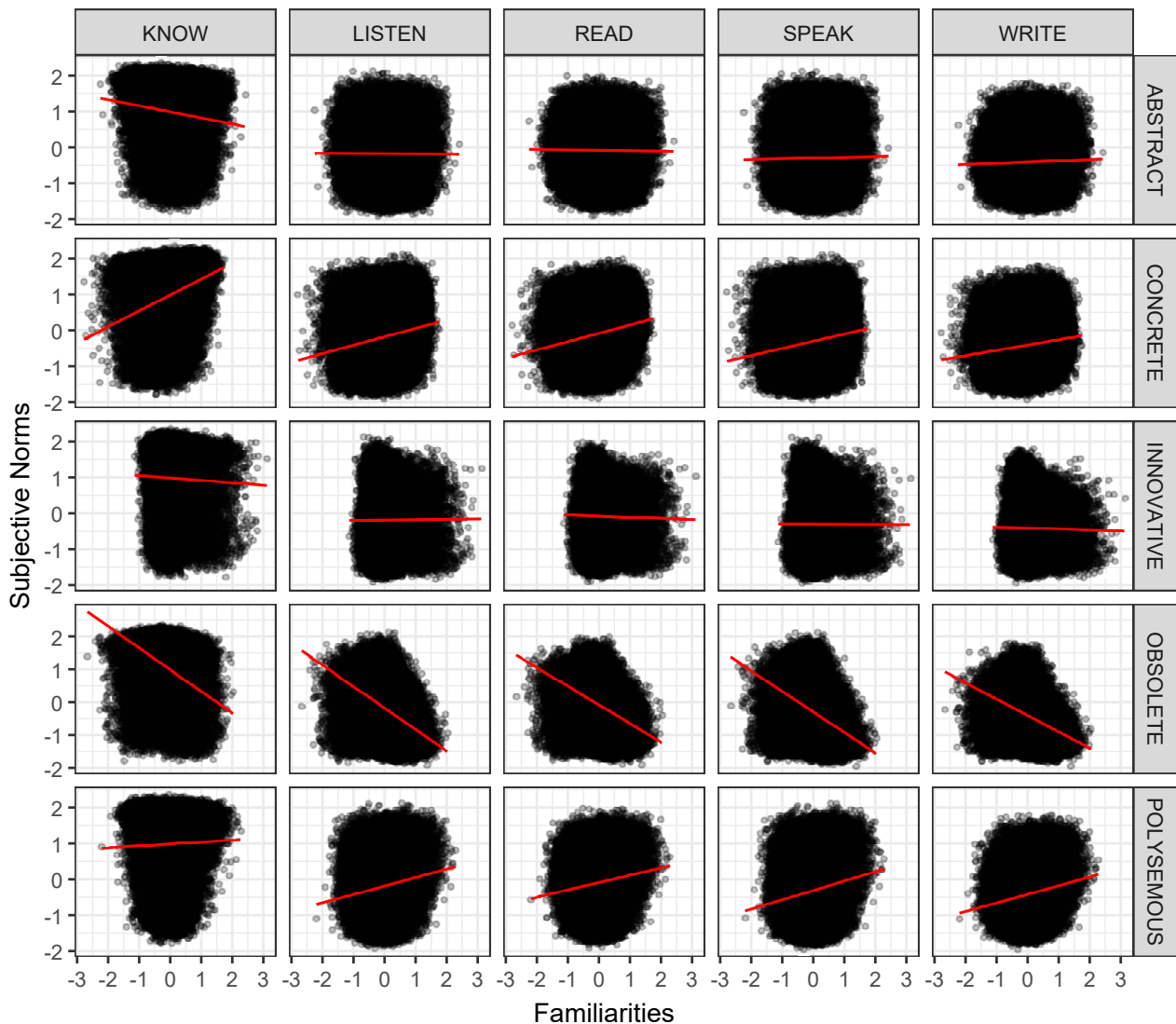


図 5 Scatterplots for Subjective Norms versus Familiarities.

同一のカテゴリ集合にわたるこのほぼ完全な反転は、カテゴリ水準における二つの尺度の対立的性質を示しており、項目水準で観察される強い負の相関とも整合的である。「他-」系のいくつかのラベル（例：動物の鳴き声）は N が小さいものの、両尺度での整合的な極端さは結果の頑健性を裏づける。

総じて、CONCRETE は物理的な指示可能性／知覚可能性を、ABSTRACT は談話・関係・論理・時間的／量的構造を捉えている。

5.3.2 古さ-新しさ

語義ラベルごとに BLMM の項目推定値を集約すると、通時的軸（OBSOLETE 対 INNOVATIVE）はカテゴリ水準では弱い構造しか示さない。多くのラベル平均はゼロ近傍（ $|\Delta|$ が小）に位置し、両尺度の

変動は小さく、しばしば逆方向である。これは、同一の語義ラベルの内部に「古い語」と「新しい語」が並存していることを意味する。言い換えれば、WLSP の分類は語が「何を扱うか」を組織化し、通時的判断は語が「いつ／どのように現代的に知覚されるか」を反映している、という分担が見られる。

小さい効果量ながら次の傾向が読み取れる。(i) ラベル内混在：主要ブランチの多くで、各尺度に対して正負のシフトを示す項目が同じラベル内に併存し、カテゴリ全体として一律に「古い／新しい」とはならない（カテゴリ間の分離よりも、カテゴリ内の語彙拡散が卓越）。(ii) 弱い非対称ドリフト：社会変化と結びつきやすい領域（例：技術・人工物・新しい語形成）は INNOVATIVE 側へ、定型表現・談話機能・親族／人称などの領域は OBSOLETE 側へ、

わずかな傾きが見られる。ただし大きさは小さく、領域全体で一様ではない。(iii) 構造語彙の中立性：関係・操作・量などの構造的／関係的語彙は、両尺度とも $\Delta \approx 0$ に集まり、平均的には通時評価が安定している。

総じて、通時的評価は分類語彙表の語義ラベルを横断し、ラベル間の差というよりラベル内の不均質性（継続的な語彙の追加と衰退）を捉えている。これは、項目水準で観察された OBSOLETE と INNOVATIVE の負の関連を補完しつつ、カテゴリ平均がゼロ近傍に留まる理由を説明する。

5.3.3 多義性

語義ラベル別に BLMM 推定を集約すると、多義性 POLYSEMOUS は 談話機能・関係／論理・自然現象系で高く、人工物（生産物）系で低いというパターンが現れる。これは先に示した「多義性は抽象性と正に、具象性と負に結び付く」という全体相関と整合的である。以下にラベル平均（コーパス平均からの偏差 Δ ）と項目数 N を添えて、上位／下位の代表例（各 3 件）を示す。

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)

なお、相の類でも高値が見られ（例：相-自然-自然 $+0.444$, $N = 741$ ；相-関係-形 $+0.418$, 135；相-関係-量 $+0.345$, 1127）、関係・論理・制度的な作用／属性が語義の広がりを持ちやすいことが示唆される。

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)

考察 談話標識や間投詞・呼格などの機能語的／用法横断的資源は、文脈依存の意味機能が多層に重なりやすく、多義性が高くなる。これに対し、具体的な人工物クラス（食品・機械・衣料）は参照対象が

明確で語義の幅が狭まりやすく、単義に寄る。さらに、関係・量・形・自然現象といった構造的／体系的ドメインでも比較的高い値が観察され、語の用いられ方（関係づけ・分類・叙述枠組み）そのものが意味の拡張を促すことが示唆される。全体として、POLYSEMOUS は (i) 談話機能・関係語で高く、(ii) 人工物名詞で低く、(iii) 抽象的構造語彙で中～高水準という三層構造を示し、CONCRETE/ABSTRACT の軸に沿った理論的一貫性を裏づける。

6 おわりに

日本語語彙の評価データとして、『分類語彙表』（100,827 語）を対象に、具象性、抽象性、古さ、新しさ、多様性の 5 尺度を新たに整備した。母語話者による 0～5 の Likert 評価を収集し、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）由来の分散と評価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、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／語彙の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価を提供；(ii) 項目・参加者ランダム効果を備え、極度に不均衡な評価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¹⁾。

限界と今後の方向性 本研究で序数応答を線形ガウス（連続）として扱ったのは実用上の近似である。評価者の属性や文体・レジスターの効果は本分析に含めておらず、層化解析の対象となる。今後は、機能語を含む複単語表現や専門領域へのカバレッジ

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

拡張、頻度・表記・近隣指標などの共変量を組み込んだ多変量 BLMM、さらに行動ベンチマーク（例：語彙性判断）での予測力検証が有望である。本資源は、分布表現との整合を図りつつ定期的に更新される「生きた」規範集合として、日本語の語彙処理と意味変異に関する累積的研究を支えることを目指す。

日本語は空白による単語分割を前提としないため、頻度推定や語彙単位は採用する形態素解析・辞書・表記揺れ・複合語規則などの単語分割に依存する。これにより他資源との比較も難しく、トークン定義が一致しないため一対一対応は一般に困難である。本研究では影響を最小化するため『分類語彙表』の単位（見出し語／語義ラベル）に統一し、外部比較は『分類語彙表』と整合的な範囲に限定した。その結果、提示する頻度や相関は『分類語彙表』の単位に依存する。今後は複数のトークン化方式をまたぐアラインメント層を整備し、資源横断で再現可能な比較を実現したい。

利益相反の開示

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。企業などとの共同研究および企業などからの助成は受けていない。

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] Suguru Matsuyoshi, Satoshi Sato, and Takehito Utsuro. Compilation of a dictionary of japanese functional ex-

pressions with hierarchical organization. In Yuji Matsumoto, Richard W. Sproat, Kam-Fai Wong, and Min Zhang, editors, **Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead**, pp. 395–402, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

- [8] 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.
- [9] 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.
- [10] 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>).