

IPAL用言例文への印象評定情報付与と代表義・典型用例の抽出

著者	加藤 祥, 浅原 正幸
雑誌名	計量国語学
巻	33
号	3
ページ	178-193
発行年	2021-12
URL	http://id.nii.ac.jp/1328/00003685/

doi: 10.24701/mathling.33.3_178

2021 年度特集・招待論文B

IPAL 用言例文への印象評定情報付与と代表義・典型用例の抽出

加藤 祥 (目白大学)

浅原 正幸 (国立国語研究所)

要旨

用例中の多義語の語義が、一般的な読み手によってどのように認定される傾向にあるのか、印象評定を用いた調査を試みた。計算機用日本語基本辞書 IPAL に含まれる用言（形容詞・形容動詞・最重要動詞、計 530 語）全文型の例文（5,125 例）に対し、文や用法に関する読み手の印象評定、分類語彙表番号と山崎・柏野（2017）の代表義情報を付与した。多義語の語義における代表義と印象評定を対照し、代表義の印象傾向を分析した。また、各用例の代表義を算出することで、見出し語における読み手が典型的と考える用例を抽出した。読み手の印象評定傾向から、助詞をはじめとする周辺語彙の差異により想起される文脈が異なること、文脈情報が用例の意味認定に関わる可能性を考察する。

キーワード：IPAL, 多義語, 代表義, 印象評定, 質問紙調査

1. はじめに

一般的に読み手が語の意味を判定するときは、典型的な意味との関係に基づき文脈情報や用法・文型などの様々な要素を用いていると考えられる。しかし多義語における中心的な意味の認識や、中心的な意味とその他の意味との関係性の認識は不明であり、読み手の用いる意味の認定根拠は明瞭でない。そこで、多数の一般的な読み手から、文の自然さや理解のしやすさ、用法の古さや新しさ、修辞性などの印象評定を収集し、語義認識を調査した。具体的には、計算機用日本語基本辞書 IPAL に含まれる用言（形容詞・形容動詞・最重要動詞）の全文型の例文について、クラウドソーシングを用いた印象評定を行った。さらに、例文に含まれる対象語の分類語彙表番号を付与し、山崎・柏野（2017）の代表義と対照した。本稿¹は、作成したデータについて、代表義と読み手の語義感覚との関わりから、代表義の認識傾向を調査するとともに、各用例の代表義度を読み手の語義感覚から線形回帰することで多義語の典型的な用例を調査し、読み手の意味判定に影響を及ぼすと考えられる要素について検討する。

2. 先行研究との関わり

2.1 多義語の中心的な意味

多義語の語義は当該言語において定着した意味であり、一般に辞書へ記載されるものと考えられる。瀬戸（2019）は、多義記述について、中心義の設定、意義の認定、意義関係

¹ 一部に言語処理学会第 27 回大会年次大会（オンライン）の発表内容を含む。

の明示、意義の配列を議論する。しかし、いわゆる多義語の語義認定には、関連語の対応による認定(国広 1982, 靱山 2002 など)、個別義認定の分離テスト・統合テスト(松本 2010 など)などの認定基準も検討されるが、語義数は辞書においてもまちまちである。語義の掲載順を見ても、出現年代順(例:古い順『広辞苑』, 現在通用している順『大辞林』など)や書き手による重要度判断順(例:意味の関連『ジーニアス』第3版まで)、頻度順(『COBUILD』, 『WISDOM』など)と統一されていない。語義数をはじめ、いずれの語義が中心的・代表的であるのか、複数語義がどのような派生関係にあるのかも明示されにくい。そもそも、多義語には何らかの中心的な意味が存在するとされる(瀬戸 2007 など)が、複数語義のうちいずれの語義が中心的(代表的, 基本的などとも)かという点についても議論が多い。

語義の派生関係を考える際には、歴史的な変化が参照される傾向にあるが、多義構造は再編成されるため、必ずしも中心的な意味が歴史的に古いとは限らない(Tyler & Evans 2001)。また、語義は基本的なものからそうでないものへと派生し、多義語の意味においては非対称性(派生的意味は基本的意味を想起させるがその逆は成り立たない)を有するとされ、典型性のある(最初に思いつく)語義が中心的と考えられるため、Gries (2006) は使用頻度との関連を考える。中心的な語義とは、最も確立されており、中立的なコンテキストで最も活性化されやすく、形式的な制約が少ない(靱山 2001, 2002, 2003 など)ともされる。なお、松本(2009)は、派生の方向性に基づく概念的な中心性と機能的中心性を認め、どちらも併せ持つ中心的意味が典型的な中心的意味であるとする。中心的な意味は一律に定められるものではなく、様々な要素によって成立しているのだといえよう。たとえば瀬戸(2019)は、「中心義」の設定にあたり、数多く体现すればするほど中心義としてふさわしい典型性を示す特性として、(i)文字通り、(ii)他意義の前提、(iii)具象性が高い、(iv)認知しやすい、(v)想起しやすい、(vi)用法上の制約を受けにくい、(vii)意義展開の起点(接点)となるのが最も多い、(viii)言語習得の早い段階で獲得される、(ix)使用頻度が高い、の9特性を挙げている。

しかし、一般的な読み手は、文脈上多義語に相対するとき、概ねそれほどの労なく語義を読み取ることが可能であり、当該語の中心的な意味や語義の派生関係は一般に共有されているものとも考えられる。

また、多義語の代表義や基本義、中心的意味については、例語を挙げた詳細な分析が数多くなされている。このうち、網羅的に認定を行ったデータとして、山崎・柏野が、『分類語彙表』増補改訂版(国立国語研究所 2004)の多義語に人手で代表義を付与する作業を行っている。『分類語彙表』の項目となっている多義語のすべて(14,102語;山崎・柏野)について、代表義と判定された語義の情報を付与したものである。複数人の判断によって行われ、また、恣意的にならないように留意するための作業基準も設けられており、BCCWJ-WLSP(加藤・浅原・山崎 2019)を用いた語義別の頻度調査も可能である。但し、概ね専門性を有した作業者の内省的判定であるため、一般的な読み手の判断とは差のある可能性も考えられた。一般的な読み手の中心的な意味の認識を調査し代表義認定を検証すること、代表義がどのような用例において代表的な語義と認識されるのかを調査することが期待される。

2.2 語義・語義関係の調査

読み手が語義の判定を行う作業として、『現代日本語書き言葉均衡コーパス』に対する分類語彙表番号の付与 (BCCWJ-WLSP, 加藤・浅原・山崎 2019) がある。BCCWJ-WLSP では、多義語の意味を付与する際、作業者が当該文脈によって意味の判定を行い、掛詞等の特例以外は、一つの語に対し一つの語義のみを付与するとした。しかし、語義情報の付与作業においては、作業仲間あるいは一作業者でも文脈によって揺れの生じる場合がある。特に、派生関係があると考えられるような複数語義の選択にあたり、どちらとも読めると判断される場合がある。読み手によって意味の揺れが生じやすい語については、用例間の類似度判定を試みている (西内ほか 2020 など) が、意味的に類似度が高いと判断されやすい用例では、文脈 (周辺語句) の類似のほか、用法や文型の影響も考えられた。このほか、研究者により様々な多義語について、基本義の設定や中心的な意味の認定、語義関係の調査が行われている (国語研究所『基本動詞ハンドブック』, 靱山 2003 など多数)。一般的な読み手が多義語の語義を判定するにあたっては、典型的な意味や高頻度などの影響可能性が指摘されるが、何に基づいて当該語の意味を決定しているのかは明確ではない。

読み手の語義判定の根拠を調査するためには、用法や文型を検証可能な用例を網羅的に用いる必要があろう。

3. 印象評定調査

3.1 調査方法

一般的な読み手の語義判定根拠としての印象や語義感覚を調査するため、先行研究で指摘される多義語の複数語義間の派生関係に関わると考えられる要素を含む評定項目を、回答しやすいと考えられる項目として設定することにした。

表 1: 本研究の設定項目と瀬戸 (2019) の中心義特性の関連

瀬戸 (2019) \ 本研究	自然さ	わかりやすさ	古さ	新しさ	比喩性
(i) 文字通り			+		
(ii) 他意義の前提				-	
(iii) 具象性が高い		+			
(iv) 認知しやすい		+			
(v) 想起しやすい		+			
(vi) 用法上の制約を受けにくい	+				
(vii) 意義展開の起点 (接点) となることが最も多い			+		-
(viii) 言語習得の早い段階で獲得される		+			
(ix) 使用頻度が高い				-	

中心的と考えられる語義には歴史性や思い出しやすさ、一般的に知られていることなどが関わるとすれば、「古い」「自然」「わかりやすい」印象のあることが期待される。なお、

文法性の判断は、一般的な読み手に対して問いにくいのが、当該用例が「自然」であるかどうかの判定が文法性の認識に関わると考えられる。また、派生義が増えるにあたっては、定着までは「新しい」印象が期待されるほか、頻度が低いことなどにより獲得度が低く「わかりやすい」印象が若干低い可能性もある。そして、派生義に換喩や提喩をはじめ比喩的な転換が感じられる場合には「比喩性」の印象の残ることが期待される。表 1 に、瀬戸 (2019) の中心義特性との関連を示す。本稿の設定した項目で中心義特性が大まかに確認できると考えられる。よって、各用例の印象評定の観点は、「自然さ」「わかりやすさ」「古さ」「新しさ」「比喩性」の 5 種類とした。

調査にあたっては、Yahoo! クラウドソーシング (<https://crowdsourcing.yahoo.co.jp/>) (2021 年 3 月確認) により、Yahoo! 日本語 ID を有する 20 歳以上の実験協力者を、IPAL の各例文 (調査対象データについては次 3.2 節参照) に対し 20 名ずつ募集した。

実験協力者は、各例文を読み、各項目について、0 (まったく違う) ~5 (そう思う) の 6 段階から評定値を付与した (図 1 参照)。

以下の事例について判定してください。

日本は物価が高い

1. 自然な表現ですか。	<input type="radio"/> 0	<input type="radio"/> 1
<input type="radio"/> 2	<input type="radio"/> 3	
<input type="radio"/> 4	<input type="radio"/> 5: そう思う	
2. わかりやすい表現ですか。	<input type="radio"/> 0	<input type="radio"/> 1
<input type="radio"/> 2	<input type="radio"/> 3	
<input type="radio"/> 4	<input type="radio"/> 5: そう思う	
3. 古い表現ですか。	<input type="radio"/> 0	<input type="radio"/> 1
<input type="radio"/> 2	<input type="radio"/> 3	
<input type="radio"/> 4	<input type="radio"/> 5: そう思う	
4. 新しい表現ですか。	<input type="radio"/> 0	<input type="radio"/> 1
<input type="radio"/> 2	<input type="radio"/> 3	
<input type="radio"/> 4	<input type="radio"/> 5: そう思う	
5. 同様の他の事例でたとえ (比喩) ていますか。	<input type="radio"/> 0	<input type="radio"/> 1
<input type="radio"/> 2	<input type="radio"/> 3	
<input type="radio"/> 4	<input type="radio"/> 5: そう思う	

図 1: 判定実験画面の例

3.2 調査対象としたデータ

BCCWJ-WLSP では実例の収集が可能であるが、多義語の語義 (複数の分類語彙表番号) が必ずしも全て収集できるのではない。実際には使用されにくい語義もあり、用法や文型の点においては同種の文型に偏りが見られるなど体系立った用例収集は難しい。といて、網羅的な語義、用法、文型という点において、一般的な辞書からは十分な用例が収集しにくい。そのため、計算機用日本語基本辞書 IPAL の例文を用例として使用することにした。IPAL は、意味記述のほか統語情報として詳細な用法が記述され、各々の用法と文型についてそれぞれの文例が記されている。調査にあたっては、IPAL に含まれた用言²の最重要

² 「IPAL は一つの言語を辞書と文法とで相補的に記述する立場にたっているが、特定の文法理論に依存していない。たとえば学校文法では、「きれい」を「形容動詞語幹」としているが、IPAL は品詞論にとらわれず、「ナ」を介して他の体言を修飾する用法を名詞の一用法として捉え、名詞辞書の枠組みの中でこの用法を持つ名詞約 290 語を記述としている。なお、動詞辞書では漢語サ変動詞

動詞 365 語, 形容詞 135 語, 形容動詞 30 語について, 掲載されていた文例全てを使用した。最重要動詞の 2,806 例と形容詞類の 2,319 例を対象とする。また, 用例における各対象語の語義として, 分類語彙表番号を付与した。IPAL と分類語彙表は語の認定と語義の認定が異なる場合がある。同音異義語の認定については他データとの重ね合わせを目途とし UniDic の短単位を基準とした。なお, IPAL の動詞類には分類語彙表番号が付与されているが, ほとんどの番号が改訂版において該当しないため, 手作業で全例の確認と修正を行った。また, IPAL 語義は分類語彙表の分類と重ならない場合が多いため, IPAL が多義もしくは複数用法や文型を認める場合に分類語彙表に当該分類がないと判断されれば, 該当する分類語彙表番号を付与した。反対に, 分類語彙表が多義と認める場合に IPAL に該当する用例の掲載がなければ, 備考を付した。これらの分類語彙表番号の付与は, 作業員 2 名が IPAL の語義を参照しながら手作業で進めた。但し, 印象評定の収集時には, 語義を提示せず用例のみを提示するため, 用例において一意に定められないと判断された場合, 複数の分類語彙表番号を付与した³。

4. データの分析と考察

4.1 印象評定調査の基礎統計

表 2 に本調査で収集した印象評定の平均値を示す。名詞的な形容動詞でのみ「比喩性」が若干下がる傾向が見られるが, 動詞形容詞の別に関わらず, 類似した評定値が得られている。「自然さ (Max:5.00, Min:1.15, 分散 0.27)」と「わかりやすさ (Max:5.00, Min:0.95, 分散 0.28)」は平均 4 以上の高い評定値が得られた。「古さ (Max:4.20, Min:0.55, 分散 0.24)」「比喩性 (Max:3.90, Min:0.05, 分散 0.29)」「新しさ (Max:2.90, Min:0.35, 分散 0.11)」は平均的に低めの評定となっており, 「新しさ」については, 最大値が付与された用例でも, 評定値が 3 を下回った。

表 2 : 印象評定調査結果 (評定は平均値)

品詞	語数	自然さ	わかりやすさ	古さ	新しさ	比喩性
形容詞	2,122	4.19	4.21	1.99	1.37	1.17
形容動詞	197	4.29	4.31	1.97	1.37	1.04
最重要動詞	2,806	4.04	4.04	2.08	1.36	1.12
計	5,125	4.11	4.12	2.03	1.37	1.14

4.2 代表義情報との対照

次に, 代表義情報と印象評定情報の対照を行った。各例文に対象語の語義 (分類語彙表番号) を付与することとし, 多義語については山崎・柏野の代表義情報を付与した。多義語, 複数語義の認定については, IPAL と分類語彙表で異なるため, 一部に代表義情報が存在しない語がある。また, IPAL の例文に対し, 分類語彙表に当該見出し語が分類に掲載の

を扱わず, その後動詞辞書の枠組みを使った 50 語の試作例を提供していただけであったが, これも名詞の一用法として捉え, サ変動詞用法のある名詞約 650 語を名詞辞書の中で記述している。(IPAL)」

³ 揺れの生じた用例については, 追加作業員による判定作業を進めている。今後, 3 名の作業員の判定結果に基づき修正を行う場合がある。

ない場合でも、該当する分類番号を付与したため、分類語彙表に当該語の語義として該当する分類番号の掲載がない場合についても、代表義情報は無い。

分析は代表義情報を 5~1 の代表義度として数値化した⁴うえで、印象評定値を固定効果とし、用例をランダム効果とした次式による一般化線形混合モデルにより回帰して行った。

$$\text{代表義度} \sim \text{自然さ} + \text{わかりやすさ} + \text{古さ} + \text{新しさ} + \text{比喩性} + (1|\text{用例})$$

表 3 に分析結果について示す。表中カッコ内は標準誤差を表す。また、有意差 $p < 0.05$ の場合 ** を、有意差 $p < 0.01$ の場合 *** を固定効果に付記する。「わかりやすさ」が高く判定される場合、代表義度が高い傾向にある。また、「比喩性」が高く判定される場合、代表義度は低い傾向にある。自然でわかりやすい語義が代表義であり、比喩性が感じられる語義は派生義である可能性が高いと考えられる。また、「古さ」と「新しさ」については強い傾向ではないが、古いと代表義度が低くなり、新しいと代表義度が高くなる傾向がみられた。一般にいわゆる歴史性や原義の意識は薄く、現在も一般的に使用されている語義が代表義に近いと考えられている可能性がある。最後に「自然さ」は代表義との関連性は確認できなかったが、「わかりやすさ」と相関係数が高く (0.8 以上 ($p < 0.01$)) 多重線形性の問題があったと考える。

表 3 : 代表義度の回帰分析による各評定値の固定効果推定値

	形容詞・形容動詞		動詞	
	係数	標準誤差	係数	標準誤差
自然さ	-0.002	(0.008)	0.012	(0.008)
わかりやすさ	*** 0.025	(0.008)	*** 0.033	(0.008)
古さ	** -0.008	(0.003)	*** -0.015	(0.004)
新しさ	*** 0.018	(0.004)	*** 0.018	(0.004)
比喩性	*** -0.040	(0.004)	*** -0.024	(0.004)
切片	*** 2.133	(0.115)	*** 1.965	(0.071)
データポイント数	46,400		56,120	

また、表 3 の結果から固定効果のパラメタのみを使って定義される回帰直線により、推定代表義値を以下の通り定義する：

$$\begin{aligned} \text{推定代表義度(形容詞)} := & -0.002 \times \text{自然さ} + 0.025 \times \text{わかりやすさ} - 0.008 \times \text{古さ} \\ & + 0.018 \times \text{新しさ} - 0.040 \times \text{比喩性} + 2.133 \end{aligned}$$

$$\begin{aligned} \text{推定代表義度(動詞)} := & 0.012 \times \text{自然さ} + 0.033 \times \text{わかりやすさ} - 0.015 \times \text{古さ} \\ & + 0.018 \times \text{新しさ} - 0.024 \times \text{比喩性} + 1.965 \end{aligned}$$

⁴ 山崎・柏野のデータにおける記号を、「●」:5, 「○」:4, 「△」:3, 「×」:2, その他(「?」など):1 とした。

代表義度の高い用例は、ある語における典型的な用例であることが期待される。推定代表義度により、語ごとの用例群において最も代表義度の高い用例を抽出することで、最も典型的と考えられる用例を得た。表4に抽出例を示す。代表義度の高い用例は、代表義と判定された語義の用例の中でも、その用法・文型において典型的である可能性が高い。分類語彙表で多義語の扱いとなっていない語であっても、本調査で付与した代表義度を用いることにより、代表義の推定も可能である。たとえば、表4の形容詞「あたらしい」「うつくしい」、形容動詞「しずか」「りっぱ」、動詞「かこむ」は、分類語彙表で多義が定義されていなかったが、IPAL では複数語義が定義されていた例である。代表義度の高い語義から、代表義を推定することが可能であった。また、形容詞「あつい」や動詞「のこる」のように、山崎・柏野において「●」認定等のない語についても、代表義度の高い用例から、一般的な読み手の判断を追加することが可能となった。

表4：推定代表義度による各語の典型的用例（例）

語	用例	自然さ	わかりやすさ	古さ	新しさ	比喩性	分類番号	代表義	推定代表義度	分類項目
あおい	この木は実の色が青い	4.20	4.25	1.45	0.75	0.15	3.5020	●	2.227	色
あたらしい	この靴はまだ新しい	4.65	4.65	1.60	1.60	0.75	3.1660	-	2.226	新旧・遅速
あつい	夏は暑い	4.55	4.50	1.65	1.30	0.55	3.1915	○	2.225	寒暖
あまい	この柿は甘い	4.80	4.75	1.65	0.70	0.30	3.5050	●	2.230	味
あらい	彼は気性が荒い	4.50	4.60	2.25	1.20	0.95	3.3430	●	2.205	行為・活動
うつくしい	バラの花は美しい	4.75	4.65	2.40	1.50	0.60	3.1345	-	2.224	美醜
うまい	あの店はラーメンがうまい	4.85	4.95	1.80	1.15	0.60	3.5050	●	2.229	味
おおきい	この学校のグラウンドは大きい	4.80	4.85	1.95	1.45	0.50	3.1912	-	2.235	広狭・大小
おもい	太郎は体重が重い	4.95	4.95	2.40	1.15	0.55	3.1914	●	2.226	軽重
しずか	早朝は近所の子供も静かだ	4.35	4.55	1.20	1.30	0.55	3.5030	-	2.230	音
だめ	予約が入っているのでその会議室はまだめです	4.45	4.30	1.15	0.85	0.10	3.1332	●	2.234	良不良・適不適
りっぱ	あの先生は立派だ	4.90	4.90	1.35	0.90	0.55	3.1302	-	2.229	趣・調子
うしなう	彼は友人を交通事故で失った	4.65	4.70	1.40	1.45	0.40	2.1250	●	2.171	消滅
かこむ	日本は福に囲まれている	4.80	4.80	1.45	1.00	0.85	2.1535	-	2.157	包み・覆いなど
したがう	弟は兄の命令に従った	4.75	4.75	1.15	0.80	0.45	2.3670	●	2.165	命令・制約・服従
すむ	仕事が済んだら出掛けよう	4.70	4.75	2.25	1.50	0.35	2.1503	●	2.163	終了・中止・停止
できる	道路に水溜りが出来た	4.70	4.80	1.10	0.80	0.55	2.1220	●	2.165	成立
のこる	金が財布に残っている	4.70	4.80	1.55	1.20	0.45	2.1240	○	2.167	保存
はずす	監督は彼を先発メンバーから外した	4.80	4.80	1.30	1.05	0.80	2.1251	●	2.161	除去
ようする	この患者は手術を要する	4.75	4.85	2.40	1.25	0.10	2.3711	●	2.166	受給

4.3 一般的な読み手の代表義認識

前節で見た代表義度から、概ね代表義と典型的な用例が抽出可能であったといえる。しかし、必ずしも容易に代表義が抽出可能ということではない。一般的な読み手が代表義と考える語義は、いわゆる代表義ではない場合が散見されるためである。

例として、表 5 に、動詞「とく（語彙素：解く）」の調査結果を示す。語彙素「解く」の語義は、BCCWJ-WLSP において、分類番号 2.1552（分類項目：分割・分裂。以降同様に分類項目のみ示す）が 2 件、分類番号 2.3062（注意・認知）が 2 件と同数例得られ、語義による頻度差は見られない。但し、BCCWJ-WLSP における分類番号 2.1552 は、「緊張を解く（PB26_00004）」「解いた髪（PB29_00003）」に付されていた。「緊張を解く」は比喩的な表現であり、「解いた髪」は「ほどいた」と読まれた可能性のあることから、典型的な用例とは見にくい。分類番号 2.3062 の用例が「文章題を解く（PB43_00001）／解き方（PM25_00084）」という類似の用例であったとはいえ件数が少なく、頻度のみでは比較しにくい。

さて、山崎・柏野において代表義と認められている語義は、分類番号 2.1552 であり、IPAL における代表義の用例は「彼女が包みを解いた」「彼女は帯を解いた」である。用例により印象評定が異なるが、「自然さ」「わかりやすさ」が高く、「比喩性」が低いという傾向からすれば、代表義における典型例は「彼女は帯を解いた」であるといえる。

しかし、同傾向で語彙素「解く」の全用例を見ると、「研究者が暗号を解いた」が最も代表的な用例であり、語義は分類番号 2.3062 に当たる。同様に分類番号 2.3062 の用例は「誤解を解く」「問題を解いている」も挙げられており、代表義とされた 2 例よりも代表義的な印象の評定が為されている。一般的な読み手の認識においては、「包みを解く」「帯を解く」のような分類番号 2.1552 が代表義なのではなく、「暗号を解く」を典型例とする分類番号 2.3062 が代表義となっている可能性がある。なお、代表義における「包みを解く」は「古さ⁵⁾」の評定は 1.65 と取り立てて高くないものの「自然さ」と「わかりやすさ」が低めの評定値であり、古さや新しさに関わらず、読み手にとって見聞きしにくい印象として評定されたと考えられる。「自然さ」の評定は、文法性の判断よりも読み手が見聞きするかどうかという判断による可能性がある。「自然さ」と「わかりやすさ」がそれぞれ最も低い用例を(1)(2)に示す。「古さ」は若干高めの評定ではあるが、取り立てて高いということはなく、「自然さ」「わかりやすさ」が低い値となっていることがわかる。前節で見た代表義の傾向として「古さ」の影響が少なかったことも、一般的な読み手の評定が用例の「古さ」を問わないことと関係すると考えられる。

(1) 彼は偽小判を吹いた（自然さ：1.15，わかりやすさ：1.20，古さ：2.15）

(2) 霜が置く（自然さ：1.20，わかりやすさ：0.95，古さ：1.85）

以上により、一般的な読み手の代表義の認識は、原義というよりも「自然さ」「わかりや

⁵⁾ 分類語彙表においては別語扱いとなるが、「髪をくしでとく」は「古さ」が 3.25 と高い評定値であるため代表義度が低くなっていた。「髪をとく」は古い印象によって代表義度が低い例であるといえ、「古さ」が強く影響していると判じられる用例も見られる。

すさ」が高い典型的用例として現れる用法であり、派生義であっても「比喩性」の印象は薄れているものと考えられる。

この他「任を解く(2.3630)」のような例は分類語彙表に掲載がなく(「印綬を解く」は掲載がある)、分類語彙表の「解く」にある分類番号 2.3670(命令・制約・服従)の例が IPAL にないなど、用例や調査語義の不足していることが課題として残った。

表 5: 「とく」

用例	自然さ	わかりやすさ	古さ	新しさ	比喩性	代表義 ⁶	推定代表義度	分類番号	分類項目
彼女が包みを解いた	3.70	3.60	1.65	1.15	1.15	●	2.10	2.1552	分割・分裂・分散
彼女は帯を解いた	4.30	4.30	2.30	1.10	1.00	●	2.12	2.1552	分割・分裂・分散
当局は戒厳令を解いた	3.45	3.60	2.35	1.25	1.10	×	2.09	2.3670	命令・制約・服従
自衛隊は包囲を解いた	3.80	4.05	2.05	1.35	1.85	×	2.09	2.3670	命令・制約・服従
彼女の説明で彼は誤解を解いた	4.30	4.10	2.20	1.75	1.05	×	2.13	2.3062	注意・認知・了解
彼の説明が彼女の誤解を解いた	4.10	4.05	1.85	1.40	1.15	×	2.12	2.3062	注意・認知・了解
学生が問題を解いている	4.15	4.15	1.75	1.45	1.00	×	2.13	2.3062	注意・認知・了解
研究者が暗号を解いた	4.20	4.55	1.55	1.40	0.70	×	2.15	2.3062	注意・認知・了解
社長は営業部長の任を解いた	4.15	4.10	2.60	1.55	1.15	なし	2.11	2.3630	人事

4.4 データの傾向に見る用法と文型の影響

個別の用例を取り上げ、データの傾向と本調査に残る課題を確認する。表 6・表 7 に形容詞「うるさい(語彙素: 煩い)」「くさい(語彙素: 臭い)」、表 8 に形容動詞「きれい(語彙素: 奇麗)」、表 9 に動詞「さす(語彙素: さす)」の例を挙げ、付与情報に見られる傾向を示す。

表 6 の形容詞「うるさい」用例を見ると、最も代表義度の高い用例は「プリンターの音がうるさかった(分類番号: 3.5030)」であり、代表義とされる分類番号 3.5030 に関する「プリンターの音」「(犬の) 鳴き声」などにおいて「古さ」・「新しさ」・「比喩性」が低い傾向にある。また、「自然さ」と「わかりやすさ」が高く評定される傾向も見られる。但し、「隣の犬がうるさい」のように、文脈的に音であることが明示されない場合は、「比喩性」が高く(2.15) 評定される。しかし、「蚊がうるさくて、眠れなかった」のように理由節に当該語が含まれる場合には、「眠れなかった」原因が音であることが連想されるためもあるが、「比喩性」は低く(0.35) 評定されることになる。分類語彙表番号付与作業における作業員間の認定揺れも、文脈の影響によるものと考えられるため、文型と文脈的な情報量の影響を整理する必要があるといえる。なお、「近所づきあいがるさい」の語義「邪魔あるいは面倒で、できれば避けたいと思う。(IPAL)」(分類番号: 3.3014) は、自然さとわ

⁶ 表 5・9 の「なし」は分類語彙表に当該語の語義として掲載がなかったが、作業員が語義として付与した分類番号であることを示し、同様に表 6・7・9 における「0」は山崎・柏野のデータにおいて記号記入のなかった語義を示す。

かりやすさが低く、「蚊がうるさくて、眠れなかった」が同義の用例とは読まれていなかったことがわかる. このように分類番号 3.3014 のような一般的に読み取られにくい語義は、はたして対象語の語義と認めるべきかという問題もあろう. 音に関する用例に次いで「自然さ」と「わかりやすさ」が高い用例は「あの会社はあれこれと注文がうるさい」と「僕は紅茶にうるさい」(分類番号: 3.3100) であるが, IPAL ではそれぞれ「注文や要求などが、いやになるほど細かくしつこい。」「いろいろと細かい文句や注文をつけて厳しい。」という別語義の例文とされている. どちらも語義として認められるのかという問題については、本調査から判定し難い.

表 6: 「うるさい」

用例	自然さ	わかりやすさ	古さ	新しさ	比喻性	代表義	推定代表義度	分類番号	分類項目
私にはプリンターの音がうるさかった	4.55	4.65	1.30	1.20	0.40	●	2.24	3.5030	音
プリンターの音がうるさい	4.70	4.60	1.50	0.75	0.85	●	2.21	3.5030	音
隣の犬がうるさい	4.05	4.10	2.00	2.45	2.15	●	2.17	3.5030	音
室内がうるさい	4.10	4.05	1.50	1.05	1.25	●	2.18	3.5030	音
この犬は鳴き声がうるさい	4.30	4.35	1.75	0.85	0.50	●	2.21	3.5030	音
この部屋はプリンターの音がうるさい	4.65	4.60	1.50	1.15	0.90	●	2.21	3.5030	音
この通りは車がうるさい	4.10	3.80	1.95	1.60	1.20	●	2.19	3.5030	音
僕は紅茶にうるさい	4.65	4.50	2.70	1.35	1.75	0	2.17	3.3100	言語活動
うちの親は外出にうるさい	3.95	3.85	2.05	1.95	1.65	0	2.17	3.3100	言語活動
その労組は労働条件にうるさい	3.95	4.05	1.35	1.05	1.30	0	2.18	3.3100	言語活動
親の注文があれこれとうるさいので、おもしろくない	4.10	4.15	2.25	1.80	1.55	0	2.18	3.3100	言語活動
あの会社の注文はうるさい	3.90	3.90	2.40	1.65	2.55	0	2.13	3.3100	言語活動
あの先生はうるさい	4.15	4.30	2.45	1.25	1.45	0	2.18	3.3100	言語活動
あの人は注文がうるさい	4.25	4.30	2.00	1.30	1.45	0	2.18	3.3100	言語活動
あの会社はあれこれと注文がうるさい	4.75	4.70	1.80	1.15	1.30	0	2.20	3.3100	言語活動
あの老人はこごとがうるさい	4.35	4.10	2.45	1.45	1.20	0	2.19	3.3100	言語活動
私には蚊がうるさくて、眠れなかった	4.35	4.55	1.45	0.40	0.35	0	2.22	3.3014	苦悩・悲哀
近所づきあいがうるさい	2.65	2.60	2.10	1.90	2.15	0	2.12	3.3014	苦悩・悲哀
消費税についての議論がうるさい	4.20	4.10	2.30	1.70	1.50	0	2.18	3.3100	言語活動
世間では、原発反対の声がうるさい	4.30	4.20	3.00	1.55	1.20	0	2.19	3.3100	言語活動

同様に、表 7 の形容詞「くさい」例では、最も代表義度の高い用例は「(ブルーチーズは)においが臭い」であり、代表義とされる分類番号 3.5040 (におい) に関する用例で「比喻性」の評定値が低く、特に「におい」の語が例文中に明示される場合において低い(「煙草の」「ブルーチーズは、」「においが臭い) 傾向にある. 但し、「にんにくのにおいは臭い」のみ若干「比喻性」が高くなっていた (1.10). 同文型の「足は臭い」において類似した「比喻性」の評定値 (1.20) が得られていることから、「A が臭い」と「A は臭い」の差異が生じているものと考えられる. 係助詞「は」によって対照するものが想起されるという文型によって、「比喻性」の印象が異なる傾向があり得る.

なお、分類項目「におい (分類番号: 3.5040)」の例とされていた「何か臭い」は新しさ

が高く、比喩性も高く評定されており、「証言が臭い」「このあたりが臭い」同様に「重要な情報が隠されていると感じさせる。(IPAL)」(分類番号: 3.3068)として読まれた可能性があろう。語義に対応した例文であっても、文脈的に対象が明示されない場合には、情報量が少ないために連想される余地が多く比喩性が上がるものと考えられる。「隣の犬がうるさい」「何か臭い」のように曖昧性の高い用例については新しさの印象が強まるため、4.2 で見た「新しさ」が代表義度に影響している可能性があろう。

表7:「くさい」

用例	自然さ	わかりやすさ	古さ	新しさ	比喩性	代表義	推定代表義度	分類番号	分類項目
私には煙草のおいが臭い	3.85	4.20	2.30	1.40	0.45	●	2.22	3.5040	におい
あいつの足は臭いなあ	4.70	4.50	2.10	0.95	1.20	●	2.19	3.5040	におい
あ、何か臭いな!	4.10	4.30	1.55	1.90	1.60	●	2.19	3.5040	におい
にんにくのおいしは臭い	4.55	4.60	1.55	1.45	1.10	●	2.21	3.5040	におい
たくあんは臭い	4.25	4.40	1.90	1.10	1.00	●	2.20	3.5040	におい
この部屋は煙草で臭い	4.55	4.65	1.10	1.25	0.95	●	2.22	3.5040	におい
ブルーチーズはにおいが臭い	4.00	4.25	0.75	0.95	0.40	●	2.23	3.5040	におい
あいつは足が臭い	4.60	4.70	2.05	1.20	0.75	●	2.22	3.5040	におい
うちの子は汗で体が臭い	4.20	4.45	1.90	1.65	0.70	●	2.22	3.5040	におい
酔っぱらいは息が臭い	4.30	4.40	1.60	1.40	0.55	●	2.22	3.5040	におい
目撃者の証言が臭い	3.20	3.35	2.30	1.60	2.10	0	2.14	3.3068	詳細・正確・不思議
このあたりが臭い	4.60	4.50	2.30	1.05	1.55	0	2.17	3.3068	詳細・正確・不思議

表8は形容動詞「きれい」の例である。代表義は「目や耳を通して人に美しいと感じさせる。(IPAL)」(分類番号: 3.1345)とされている。しかし、本稿の調査では「私の部屋はきれいだ(分類番号: 3.5060)」の例が最も代表義度が高い評定値となっており、「彼の仕事は仕上がりがきれいだ(分類番号: 3.1340)」が次ぐ。これらの例は、「ゴミや汚れや不純物がなく、清潔である。(IPAL)」(分類番号: 3.5060)や「外見や形式が整っていて、気持ちのよい印象を与える。(IPAL)」(分類番号: 3.1340)の用例とされていたが、分類番号3.1340と3.1345は隣接分類であり、分類番号3.5060を根拠として、中項目3.13(様相)が判定されるという点において、一般的な語義認識に大差のない可能性が考えられる。

別途用例の類似度評定の調査を進めているため、今後印象評定値と類似度評定値を組み合わせた検討により、代表義と派生義の調査を目指す必要がある。

表 8: 「きれい」

用例	自然さ	わかりやすさ	古さ	新しさ	比喩性	代表義	推定代表度	分類番号	分類項目
桜の花はきれいだ	4.65	4.60	1.70	0.60	0.35	○	2.22	3.1345	美醜
彼女の髪はきれいだ	4.70	4.45	1.80	0.60	0.70	○	2.20	3.1345	美醜
夜空の星がきれいだ	4.50	4.45	1.80	1.65	1.05	○	2.21	3.1345	美醜
彼女は声がきれいだ	3.90	3.75	1.65	1.70	1.30	○	2.18	3.1345	美醜
この絵は色がきれいだ	4.70	4.70	1.90	0.95	0.65	○	2.22	3.1345	美醜
彼女は髪がきれいだ	4.45	4.45	1.60	1.00	0.65	○	2.21	3.1345	美醜
その鳥は羽根がきれいだ	4.20	4.35	2.00	1.05	0.65	○	2.21	3.1345	美醜
この服は柄がきれいだ	4.40	4.60	1.70	0.80	1.05	○	2.20	3.1345	美醜
私の部屋はきれいだ	4.95	4.95	1.80	1.10	0.45	△	2.23	3.5060	材質
その海の水はきれいだ	4.50	4.50	1.65	1.40	1.10	△	2.20	3.5060	材質
ここの台所は流しが特にきれいだ	4.60	4.75	2.40	1.65	0.95	△	2.22	3.5060	材質
その書類の字はとてきれいだ	4.35	4.25	1.45	0.70	0.95	×	2.19	3.1340	調和・混乱
この製品の仕上がりがきれいだった	4.25	4.25	2.15	1.60	1.40	×	2.19	3.1340	調和・混乱
彼女は言葉づかみがきれいだ	4.45	4.70	2.05	1.10	0.70	×	2.22	3.1340	調和・混乱
その書類は字がきれいだ	3.60	3.70	1.85	1.75	0.60	×	2.21	3.1340	調和・混乱
彼の仕事は仕上がりがきれいだ	4.55	4.80	1.65	0.95	0.40	×	2.23	3.1340	調和・混乱
彼女の心はきれいだ	4.65	4.65	3.65	1.40	1.80	×	2.16	3.1340	調和・混乱
あの政治家はやり方がきれいでない	3.75	3.90	2.30	2.05	1.75	×	2.17	3.1340	調和・混乱

表 9 の動詞「さす」は、UniDic の語彙素「さす」に対応する見出し語が IPAL では 2 語に分けて扱われていたため、IPAL の最重要動詞 2 と 229 の用例をまとめて示してある(表 9 では二重線の上下で IPAL の見出し語の区別を示す)。分類語彙表では、分類番号 2.1532 (入り・入れ): 「挿す」「差す」「刺す」、分類番号 2.1533 (漏れ・吸入): 「差す」、分類番号 2.3042 (欲望・期待・失望): 「指す」、分類番号 2.3092 (見せる): 「指す」、分類番号 2.3370 (遊楽): 「挿す」、分類番号 2.5010 (光): 「差す」の掲載がある。今後、表記との関わりについても調査を進める必要が残る。

さて、「さす」の例では、まず IPAL の最重要動詞 2 において、最も代表義度の高い用例は「部屋に西日が差している(分類番号: 2.5010, 光)」であった。代表義ではない「水を注す(分類番号: 2.1533)」は「自然さ」「わかりやすさ」の評定値が低いことから代表義度が低くなっており、分類語彙表に「差す」の語義として掲載のない「赤みが差す(分類番号: 2.1210)」のような派生義は「比喩性」の評定値が高いことから代表義度が低くなっており、代表義とされる光(分類番号: 2.5010)の用例が、それぞれ代表義度の高い評定傾向にある。

しかし、「部屋に西日が差している」と「窓から朝日が差している」において「比喩性」の評定値に差が生じているという現象が見られる。「部屋に」と「窓から」という助詞の違いにより、想起された読み手の位置が異なることや、「窓から」と客観的な位置を想起した場合に「日」を擬人化と認識した影響が考えられる。また、「西日が差す」は一般的な現象として見聞きしやすい状況であるのに対し、「朝日が差す」は「朝日さす(富裕にかかる)」枕詞や「朝日」から連想される明るいイメージなどにより「比喩性」や「新しさ」の評定

度を高めた可能性があるほか、あえて「朝日が差す」状態であることを言及しているという特殊な状況⁷として「比喩性」があると読まれた可能性もある。同語義とされた用例であっても、読み手の印象評定が異なることから、典型的用例の認定については、文型によって想起される文脈の差異に留意する必要がある。

表 9: 「さす」

用例	自然さ	わかりやすさ	古さ	新しさ	比喩性	代表義	推定代表義度	分類番号	分類項目
部屋に西日が差している	4.80	4.85	1.30	0.70	0.45	●	2.16	2.5010	光
窓から朝日が差している	4.40	4.45	1.85	1.55	1.80	●	2.12	2.5010	光
彼女の頬に赤みが差した	4.05	4.15	2.65	1.30	2.35	なし	2.08	2.1210	出没
彼女は花器に水を注した	3.85	3.45	2.80	1.70	1.05	×	2.09	2.1533	漏れ・吸入など
母は鍋の湯に水を注した	3.45	3.60	1.70	1.10	0.90	×	2.10	2.1533	漏れ・吸入など
時計の針が正午を指した	4.70	4.80	3.20	0.90	1.10	●	2.12	2.3092	見せる
証人は3人目の男を指して「彼が犯人だ」と言った	4.25	4.25	1.40	0.65	0.80	●	2.13	2.3092	見せる
先生が太郎を指してイオウの周期番号を尋ねた	4.05	3.80	1.80	1.35	1.10	なし	2.11	2.3630	人事
あの言葉は暗に君を指している	3.00	2.95	2.10	1.40	1.70	なし	2.05	2.3070	意味・問題・趣旨など
傍線部の指示語は何を指しているのか	3.40	3.55	1.75	1.40	1.05	●	2.10	2.3092	見せる
一行は極点を指して進んで行った	3.45	2.90	2.70	1.30	1.55	0	2.05	2.3042	欲望・期待・失望

また、IPAL の最重要動詞 229 では、最も代表義度の高い用例が「3 人目の男を指して (分類番号: 2.3092)」であった。代表義とされる「見せる」(分類番号: 2.3092) の用例の代表義度がいずれも高めとなっており、代表義度の下がった例では「時計が」という主語により「比喩性」の上がることや「傍線部の指示語」のような一般性が下がる用例では「自然さ」「わかりやすさ」が下がることの影響が考えられる。

なお、IPAL の最重要動詞 2 と 229 を、UniDic では別語と認めていないが、「さす」用例群中で代表義度の高い例を見ると、IPAL の最重要動詞 2 における「部屋に西日が差している (分類番号: 2.5010, 光)」であり、最重要動詞 229 の「3 人目の男を指して (分類番号: 2.3092)」が次ぐ。代表義度の高い用例が別義であり、語義間に隣接性や因果関係等が認められないことから、同音異義語の認定が可能であるともいえよう。今後、用例の類似度の調査を進めることにより、代表義度を用いた派生関係の調査を進めたい。

5. まとめと展望

本稿は、多義語の用例に対する印象評定を収集し、代表義との対照を行うことで、一般的な読み手が多義語の代表義度をどのように認識しているのか、語義を認識するに際し何を判定根拠としているのかという観点で分析を試みた。

IPAL 辞書に掲載された最重要動詞、形容詞類の例文全てに対し、印象評定と分類語彙

⁷ 加藤 (2018) は、高頻度であることが期待される要素の言及が少ない場合のあること、その場合には比喩表現に特頻度の低い例と比喩表現の関わりのあることを指摘している。

表番号を付与し、代表義情報と対照した結果、「自然さ」や「わかりやすさ」が高く、「比喩性」が低いと判断されれば代表義度が高く、一般的に認識される中心的な意味に近いといえそうであった。そこで、代表義度を推定することにより、各語の用例群から最も代表義度の高い用例を抽出し、一般的な代表義と典型的な用例を得た。一般的な読み手の印象評定を用い、多義語における代表義と当該語の典型的な用例を抽出することができる。但し、一般的な読み手の認識における代表義の特徴として、見聞きしやすい用例であることが「自然さ」「わかりやすさ」に影響する可能性、助詞の違いをはじめとする文型によって想起される文脈が「比喩性」に影響する可能性が考えられた。同語義と判定される用例であっても、一般的な読み手によって代表義度が高く典型的と考えられる例と代表義度が低い例が見られる。今後、評定値と用法・文型の関係性の整理を進めるとともに、用例の類似度の調査を進めることで、典型的な用例と派生関係についても明らかにできる可能性がある。また、今回取得できた「比喩性」の高い文型と派生義との関係を調査すること、「古さ」と「比喩性」の評定値を「慣用的」比喩表現の判定に利用することなど、本データの更なる利活用を図りたい。

謝辞

本研究は国立国語研究所コーパス開発センター共同研究プロジェクトの成果物です。また、科研費 17H00917, 18K18519, 19K00591, 19K00655 の支援を受けました。

文献

- Cruse, Alan. (2011) *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press.
- Gries, Stefan Th. (2006) Corpus-based Methods and Cognitive Semantics: the Many Meanings of to Run. In Stefan Th. Gries and Anatol Stefanowitsch (eds.) *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. 57-99. Berlin: Mouton de Gruyter.
- Tyler, Andrea, and Vyvyan Evans. (2001) Reconsidering Prepositional Polysemy Networks: The case of over. *Language* 77(4): 724-765.
- 加藤祥 (2018) 「特徴的な要素と用例頻度の関係：角を例とした一考察」『国立国語研究所論集』14: 55-72.
- 加藤祥, 浅原正幸, 山崎誠 (2019) 「分類語彙表番号を付与した『現代日本語書き言葉均衡コーパス』の書籍・新聞・雑誌データ」『日本語の研究』15(2): 134-141.
- 加藤祥, 浅原正幸 (2021) 「多義語語義調査を目指した IPAL 形容詞例文への印象評定情報付与」『言語処理学会第 27 回年次大会発表論文集』
- 国広哲弥 (1982) 『意味論の方法』大修館書店.
- 国立国語研究所 (2004) 国立国語研究所資料集 14 『分類語彙表 -増補改訂版-』大日本図書.
- 国立国語研究所 (2013) 国立国語研究所共同研究報告 12-07 『基本動詞ハンドブック』 (<https://verbhandbook.ninjal.ac.jp/>) (2021 年 3 月 15 日確認)
- 情報処理振興事業協会 (IPAL). 計算機用日本語基本辞書 IPAL: 動詞・形容詞・名詞.

- 瀬戸賢一 (2007) 「メタファーと多義語の記述」 楠見孝 (編) 『メタファー研究の最前線』 31-61. ひつじ書房.
- 瀬戸賢一 (2019) 「多義記述の問題点とその解法—日本における正しい多義記述の出発点—」 『日本認知言語学会論文集』 507-517.
- 西内沙恵, 加藤祥, 浅原正幸 (2020) 「語義間類似度の双方向評定に基づくプロトタイプの意味の解明—クラウドソーシングを用いた量的調査による多義的形容詞分析—」 『日本認知言語学会論文集』 256-268.
- 松本曜 (2009) 「多義語における中心的意味とその典型性: 概念的 center 性と機能的 center 性」 *Sophia linguistica : working papers in linguistics* 57: 89-99.
- 松本曜 (2010) 「多義性とカテゴリー構造」 澤田治美 (編) 『ひつじ意味論講座』 1:23-43. ひつじ書房.
- 榎山洋介 (2001) 「多義語の複数の意味を統括するモデルと比喻」 『認知言語学論考』 1: 29-58. ひつじ書房.
- 榎山洋介 (2002) 『認知意味論のしくみ』 研究社出版.
- 榎山洋介 (2003) 「多義性」 松本曜 (編) 『認知意味論』 135-186. 大修館書店.
- 山崎誠, 柏野和佳子 (2017) 『『分類語彙表』の多義語に対する代表義情報のアノテーション』 『言語処理学会第 23 回年次大会発表論文集』 302-305.

(2021 年 3 月 15 日受付)

Invited Paper (B) to the Special Issue 2021

Impression Rating Evaluation of Example Sentences for Verbs and Adjectives in the IPAL Dictionary: Their Representative Senses and Typical Usages

KATO Sachi (Mejiro University)

ASAHARA Masayuki (National Institute for Japanese Language and Linguistics)

Abstract:

We performed large-scale impression rating experiments to investigate the evaluation of the senses of polysemous words in example sentences. For the impression rating, we used 5,125 example sentences for 530 words in the IPA Lexicon of Basic Japanese Verbs, Adjectives, and Nouns (IPAL). In addition, we annotated the sense tags of Word List by Semantic Principles and their representative senses given by Yamazaki and Kashino (2017).

A contrastive survey between the representative senses and impression ratings was performed. Further, linear regression analysis was performed to estimate the degree of representativeness and extract their typical usages. We found that differences in contextual words, including case particles, affect the recognition of word senses.

Keywords: IPAL, polysemous, representative sense, impression rating, questionnaire study