

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

加藤祥[†] 浅原正幸^{††}
[†]目白大学 ^{††}国立国語研究所
 skato@mejiro.ac.jp

1 はじめに

一般的に読み手が語の意味を判定するとき、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、読み手の判定根拠は不明瞭である。そこで、文の自然さや意味の取りやすさ、用法の古さや新しさ、修辞性などの印象評定を収集することで、語義の判定根拠を取得し、語義認識や語義間関係認識を調査することにした。すなわち、語義判定の揺れが生じやすい用言を対象とし、読み手が意味を判定する際の印象情報を収集し、多義語の語義との関係について調査する。具体的には、計算機用日本語基本辞書 IPAL [1] に含まれる用言の全文型の例文について、一般的な読み手の語義感覚を収集するため、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) [2] の代表義を付与した。本稿は、作成したデータの形容詞・形容動詞について、代表義と読み手の語義判定との関係を調査し、報告する。一般的な読み手の語義感覚が語義の判定にどのように関わるかを確かめ、中心的な意味の認定や語義間の派生関係認定への活用可能性を考える。

今後、動詞類の付与情報整備を進め、用法や文型を含めた読み手の印象の観点から、用言全体について語彙間の関係を整理する予定である。

2 先行研究

2.1 多義語の中心的意味

多義語の語義は当該言語において定着した意味 [3] であり、一般に辞書へ記載されるものと考えられる。しかし、いわゆる多義語の語義数は辞書においてもまちまちであり、それらの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。いずれの語義が中心的であるのか、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[4] 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、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[5])。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) [6] 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[7], 2002[8], 2003[9] など) ともされる。なお、松本 (2009) [10] 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。中心的な意味は一概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。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 [2] が、『分類語彙表』増補改訂版 (国立国語研究所, 2004[11]) 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 [2]) 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、一部に頻度調査（加藤ら、2018[12]）もある。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。

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を進めてきた（BCCWJ-WLSP, 加藤・浅原・山崎, 2019[13]）。BCCWJ-WLSPでは、掛詞等の特例のほか一つの語に対し一つの語義のみを付与するため、多義語の意味を付与するにあたっては、作業者が当該文脈によって意味の判定を行った。しかし、語義情報の付与作業においては、作業者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, 2019[14]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（靱山, 2003[9]など多数）ものの、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も、何に基づいて当該語の意味を決定しているのか明確ではない。

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設定することにした。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。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。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/>）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）。

以下の表現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: まったく違う	<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: まったく違う	<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: まったく違う	<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: まったく違う	<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. 何かを他の物事でたとえ（比喩）していますか。

<input type="radio"/> 0: まったく違う	<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 判定画面例

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。とって、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[1]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用例が記されている。調査にあたっては、IPALに含まれた用言ⁱ（動詞 861 語とサ変動詞 50 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。動詞類の 5,385 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。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[11]において該当せず、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た場合、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い場合、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。以降、本稿では形容詞類のデータを取り上げ、付与状況とデータに見られる傾向を報告する。

4 データの分析と考察

4.1 印象評定調査の基礎統計

表 1 に本調査で収集した形容詞類の評定の平均値を示す。形容詞・形容動詞の種類に関わらず、類似した評定値が得られている。「自然さ」と「わかりやすさ」は平均 4 以上の高い評定値が得られた。なお、名詞的な形容動詞では「比喩性」が若干下がる傾向が見られる。

ⁱ 「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

表 1 印象評定調査結果（評定は平均）

品詞	語数	例数	自	わ	古	新	比
形容	135	2122	4.19	4.21	1.99	1.37	1.17
形動	30	197	4.29	4.31	1.97	1.37	1.04
計	165	2319	4.20	4.22	1.98	1.37	1.16

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。また、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[2]を付与した（付与情報例は付録を参照）。分析は代表義情報を 5~1 の代表義度として数値化したうえで、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。

代表義度~自然さ+わかりやすさ+古さ+新しさ+比喩性+(1)用例。

表 2 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）

	自	わ	古	新	比
推定値	-0.002	0.025 (p<.01)	-0.008 (p<.05)	0.018 (p<.01)	-0.040 (p<.01)

表 2 に分析結果について示す。「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。一般にいわれ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

290 語を記述としている。なお、動詞辞書では漢語サ変動詞を扱わず、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るだけであったが、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。」

やすさ」と相関係数が高く(0.8以上 ($p<.01$))多重線形性の問題があったと考える。

4.3 事例分析「うるさい」「くさい」

個別の用例を取り上げ、データの傾向を確認しておく。ここでは、分類語彙表でも IPAL でも多義語とされ、かつ各語義に複数例文の見られた「うるさい」「くさい」を例とし、付与情報に見られる傾向を示す(付録の「うるさい」「くさい」データを参照)。

「うるさい」の具体的な用例を見てみると、代表義とされる「音」に関する「プリンターの音」「(犬の)鳴き声」などの古さ・新しさ・比喩性が低い。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。文型と文脈的な情報量の影響を整理する必要がある。

「近所づきあいがうるさい」の語義「邪魔あるいは面倒で、できれば避けたいと思う。(IPAL)」(分類語彙表番号:3.3014)は、自然さとわかりやすさが低く、「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る。

また、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」(分類語彙表番号: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。

同様に、「くさい」の例では、代表義とされる「におい」に関する用例で比喩性が低く、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」のみ若干比喩性が高くなっていた。同文型の「足は臭い」におい

て類似した比喩性の値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、比喩性の印象が異なる傾向があり得る。

なお、なお(分類語彙表番号: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)」(分類語彙表番号:3.3068)として読まれた可能性がある。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、前節(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る。

5 おわりに

本稿は、一般的な読み手が多義語の語義を認識するに際し何を判定根拠としているのか、また、一般的にどのような語義感覚を有しているのか、多義語の用例に対する印象評定を収集することで分析を試みた。IPAL辞書[1]に掲載された用言の例文全てに対し、印象評定と分類語彙表番号を付与し、代表義情報[2]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中心的な意味に近いと言えそうである。

付与した情報を用いることで、今回取得できた比喩性の高い文型や「慣用的」な判断の活用についても検証を行いたい。古くて比喩性が感じられるとされる用例が一般的に「慣用的」な比喩表現であり、被喩辞の説明が少ない文型ほどわかりやすさが低く、比喩性が高く判断されると考えられるため、このような用法や文型が比喩表現の収集に有用な可能性がある。

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。

参考文献

1. 情報処理振興事業協会 (IPAL) . GSK 配布版「計算機用日本語基本辞書 IPAL 一動詞・形容詞・名詞一」, GSK2007-D, 2007.
2. 『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. 山崎誠・柏野和佳子. 言語処理学会第 23 回年次大会発表論文集, 2017.
3. Cruse, Alan. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press, 2011.
4. 瀬戸賢一, 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』, ひつじ書房, 2007.
5. Tyler, Andrea, and Vyvyan Evans. Reconsidering prepositional polysemy networks: The case of over. *Language* 77, 2001.
6. Gries, Stefan Th. “Corpus-based methods and cognitive semantics: the many meanings of to run.” In *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Ed. by Stefan Th. Gries and Anatol Stefanowitsch. Berlin: Mouton de Gruyter, 2006.
7. 榎山洋介. 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, ひつじ書房, 2001.
8. 榎山洋介. 『認知意味論のしくみ』研究社出版, 2002.
9. 榎山洋介. 多義性. 松本曜編『認知意味論』, 大修館書店, 2003.
10. 松本曜. 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), 2009.
11. 国立国語研究所資料集 14『分類語彙表 -増補改訂版-』, 2004.
12. 多義語の語義分布と語義間の派生関係調査の試み—一相の類を中心に, 加藤祥, 田邊絢, 浅原正幸, 古宮嘉那子, 新納浩幸, 言語処理学会第 25 回年次大会発表論文集, 2019.
13. 加藤祥, 浅原正幸, 山崎誠. 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), 2019.
14. 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—, 西内沙恵, 加藤祥, 浅原正幸, 日本認知言語学会第 20 回大会予稿集, 2019.

付録 (評定値は実験協力者 20 名の平均)

「うるさい」例

用例	自	わ	古	新	比	意味記述	分類番号	代表義	代表義	部門	中項目	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	(私は) 音や声などが大きく耳ざわりだと感じる。	3.5030	●	5	自然	自然	音
プリンターの音がうるさい	4.70	4.60	1.50	0.75	0.85	音や声などが大きく耳ざわりである。	3.5030	●	5	自然	自然	音
隣の犬がうるさい	4.05	4.10	2.00	2.45	2.15	音や声などが大きく耳ざわりである。	3.5030	●	5	自然	自然	音
室内がうるさい	4.10	4.05	1.50	1.05	1.25	音や声などが大きく耳ざわりである。	3.5030	●	5	自然	自然	音
この犬は鳴き声がうるさい	4.30	4.35	1.75	0.85	0.50	音や声などが大きく耳ざわりである。	3.5030	●	5	自然	自然	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	音や声などが大きく耳ざわりである。	3.5030	●	5	自然	自然	音
この通りは車がうるさい	4.10	3.80	1.95	1.60	1.20	音や声などが大きく耳ざわりである。	3.5030	●	5	自然	自然	音
僕は紅茶にうるさい	4.65	4.50	2.70	1.35	1.75	いろいろな細かい文句や注文をつけて厳しい。	3.3100		1	活動	言語	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	いろいろな細かい文句や注文をつけて厳しい。	3.3100		1	活動	言語	言語活動
その労働は労働条件にうるさい	3.95	4.05	1.35	1.05	1.30	いろいろな細かい文句や注文をつけて厳しい。	3.3100		1	活動	言語	言語活動
親の注文があれこれうるさいので、おもしろくない	4.10	4.15	2.25	1.80	1.55	注文や要求などが、いやになるほど細かくしつこい。	3.3100		1	活動	言語	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	注文や要求などが、いやになるほど細かくしつこい。	3.3100		1	活動	言語	言語活動
あの先生はうるさい	4.15	4.30	2.45	1.25	1.45	注文や要求などが、いやになるほど細かくしつこい。	3.3100		1	活動	言語	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	注文や要求などが、いやになるほど細かくしつこい。	3.3100		1	活動	言語	言語活動
あの会社はあれこれ注文がうるさい	4.75	4.70	1.80	1.15	1.30	注文や要求などが、いやになるほど細かくしつこい。	3.3100		1	活動	言語	言語活動
あの老人はここがうるさい	4.35	4.10	2.45	1.45	1.20	注文や要求などが、いやになるほど細かくしつこい。	3.3100		1	活動	言語	言語活動
私には蚊がうるさい	4.35	4.55	1.45	0.40	0.35	(私は) 邪魔あるいは面倒で、できれば避けたいと思う。	3.3014		1	活動	心	苦悶・悲哀
近所づきあいがうるさい	2.65	2.60	2.10	1.90	2.15	(私は) 邪魔あるいは面倒で、できれば避けたいと思う。	3.3014		1	活動	心	苦悶・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	議論が活発に広く行われていて、わずらわしい感じを与える。	3.3100		1	活動	言語	言語活動
世間では、原発反対の音がうるさい	4.30	4.20	3.00	1.55	1.20	議論が活発に広く行われていて、わずらわしい感じを与える。	3.3100		1	活動	言語	言語活動

「くさい」例

用例	自	わ	古	新	比	意味記述	分類番号	代表義	代表義	部門	中項目	分類項目
私には煙草のにおいが臭い	3.85	4.20	2.30	1.40	0.45	(私は)臭いがよくないと感じる。	3.5040	●	5	自然	自然	におい
あいつの足は臭いなあ	4.70	4.50	2.10	0.95	1.20	(私は)臭いがよくないと感じる。	3.5040	●	5	自然	自然	におい
あ、何か臭いな!	4.10	4.30	1.55	1.90	1.60	(私は)臭いがよくないと感じる。	3.5040	●	5	自然	自然	におい
にんにくのにおいは臭い	4.55	4.60	1.55	1.45	1.10	においがよくない。	3.5040	●	5	自然	自然	におい
たくあんは臭い	4.25	4.40	1.90	1.10	1.00	においがよくない。	3.5040	●	5	自然	自然	におい
この部屋は煙草で臭い	4.55	4.65	1.10	1.25	0.95	においがよくない。	3.5040	●	5	自然	自然	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	においがよくない。	3.5040	●	5	自然	自然	におい
あいつは足が臭い	4.60	4.70	2.05	1.20	0.75	においがよくない。	3.5040	●	5	自然	自然	におい
うちの子は汗で体が臭い	4.20	4.45	1.90	1.65	0.70	においがよくない。	3.5040	●	5	自然	自然	におい
酔っぱらいは息が臭い	4.30	4.40	1.60	1.40	0.55	においがよくない。	3.5040	●	5	自然	自然	におい
目撃者の証言が臭い	3.20	3.35	2.30	1.60	2.10	重要な情報が隠されていると感じさせる。	3.3068		1	活動	心	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	重要な情報が隠されていると感じさせる。	3.3068		1	活動	心	詳細・正確・不思議