

評価語ネットワークに基づく分野別のオノマトペ辞書拡張支援

千本 達也[†] 竹内 和広[‡]

[†] 大阪電気通信大学 工学研究科

[‡] 大阪電気通信大学 情報通信工学部

tatsuya17@takelab.jp

takeuchi@isc.osakac.ac.jp

1 はじめに

語義曖昧性解消タスクでは、辞書に記述されている複数の意味のうち、特定の文脈における意味を特定する。しかし、既存辞書で日本語のオノマトペの語義記述を調べると、オノマトペを用いる分野によっては、記載されている記述では特定分野の意味を十全に説明できない場合や、そもそもその意味が辞書に記載されていない場合がある。本稿では、特定分野の文章を使って当該分野の評価語の相関関係を表現した語彙ネットワークを用いて、当該分野のオノマトペの特徴的な意味の特定や、辞書に記述されていない意味の発見に寄与する手法を提案する。

2 オノマトペの特徴と意味表現

2.1 既存辞書の語義記述

日本語のオノマトペは、それが使用される分野や文脈に依存するといった特徴をもつ [1]。本稿では、オノマトペの特徴について、既存辞書の語義記述（語釈文）を用いて検討するため『暮らしのことば 擬音・擬態語辞典』¹ に掲載されていた 1,975 語を調査対象とする。また、分野に対する依存性を検討するために、服飾²・化粧品³・飲食店⁴の3分野について収集したレビュー文章集合（25,000 件ずつ）において、分野毎に出現頻度が高かった 20 語のオノマトペについて調査した。

左記に述べた辞書の語釈文では、オノマトペを用いる際の具体的な場面や評価対象における意味を逐一記述するのではなく、1つのオノマトペがとりうる広い文脈に対応できるような工夫がなされている。例えば左記の辞書において「どんどん」については「重い物が続けて衝突したり破裂したりして出す、低く大きな

音。たとえば、太鼓やドアを叩く音、花火が爆発する音、床を強く踏みならす音など。」という語釈文が、「重い・低い・大きな・強く」といった一般的な評価に関わる語を使うと同時に、音の性質の違いについて、複数の場合を列挙するように記述されている。つまり、一般化された大きな意味の枠組をいくつか示し、実際の文脈における細かな場面の違いに応じて複数の状況や場面を示しつつ、人間であれば文脈に応じて柔軟に意味が取れるような工夫がなされているが、計算機でオノマトペの意味を扱う上では必ずしも適さない。

2.2 意味表現に関する試み

オノマトペの意味的特徴をふまえて、その意味表現を扱った先行研究がある。古武ら [2] は、オノマトペと共通する用法で使用される言い換え表現候補から、オノマトペの言い換え表現を自動収集する手法を提案した。しかし、1語で言い換えることが困難であるオノマトペの存在を挙げ、オノマトペが持つ多様な意味をどのように解消するかが課題であるとしている。

一方、Semantic Differential Method [3]（以降、SD法）という感性調査アンケートにおける評価語（形容詞・副詞・形容動詞）の対（以降、感性軸）と Princeton WordNet⁵（以降、PWN）を用いて、オノマトペの意味表現の揭示を試みた研究に中部ら [4] による研究がある。

感性軸は、印象調査の対象毎に適切な評価語の選択と反意関係の設計がなされた、SD法のアンケートにおける質問項目に相当する。中部らは一つのオノマトペの意味表現として、予め用意した 53 対の感性軸上すべてに対して、計算機によるスコアの算出を行っている。これはオノマトペの特徴を表現する上で適切な意味表現であると考えられるが、オノマトペの特徴を意識した、オノマトペと感性軸間のスコアの算出や評価は行われていない。

¹ 山口仲美. 暮らしのことば 擬音・擬態語辞典. 講談社, 2003.

² 楽天市場 (<http://www.rakuten.co.jp/>)

³ @cosme (<http://www.cosme.net/>)

⁴ 食べログ (<http://tabelog.com/>)

⁵ <http://wordnet.princeton.edu/>

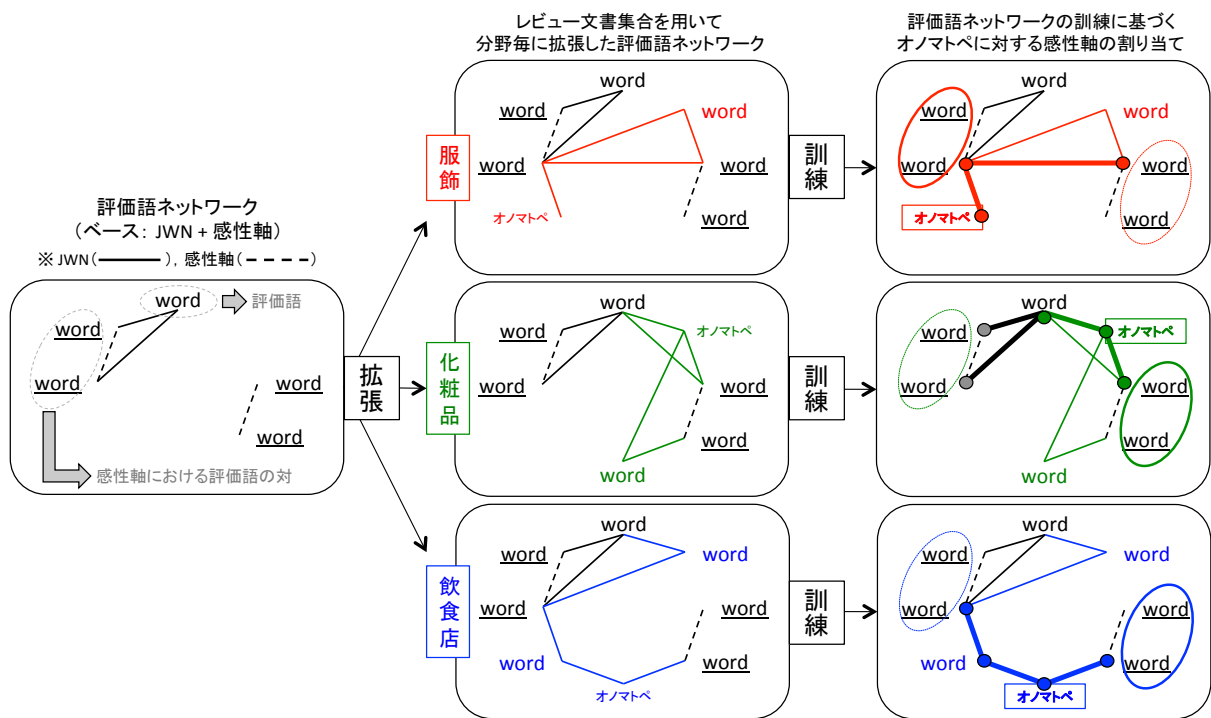


図 1: 評価語ネットワークの構築とオノマトペに対する感性軸の割り当て

中部らの研究に対して我々は、オノマトペの特徴を考慮して、多くのオノマトペと整合性を持つような意味表現を扱うために、SD 法を扱った書籍⁶と複数の論文で用いられていた、オノマトペを含まない258対の感性軸を収集した。感性軸は調査対象の評価に特定の評価語が関係する可能性があるという知識であり、我々はこの258対の感性軸上の評価語の対を、日本語において代表的な語彙概念知識である日本語 WordNet⁷ (以降、JWN) に反意語知識として導入する。

そして、特定の語に依存するオノマトペと感性軸を結びつけるのではなく、JWN をベースに、オノマトペを含めた評価語の相関関係を表現した語彙ネットワーク (以降、評価語ネットワーク) を用いて、オノマトペの意味表現を行う。さらに、左記のネットワークに対して分野毎に、レビュー文章集合におけるオノマトペと評価語の使用文脈を考慮した拡張を行う。

ここで、評価語ネットワークとは JWN 上の評価語と感性軸から再構築したネットワークに対して、3 分野 (服飾・化粧品・飲食店) のレビュー文章集合から獲得した知識を用いて拡張を行った語彙ネットワークである。

以上に基づき、分野毎の評価語ネットワーク上の各語に対し、258 対の感性軸を用いたスピンモデルに基づくスコア付けを実施することで、分野毎のオノマト

ペの意味を複数の感性軸により表現する。そして、オノマトペが特定分野で担う評価的な意味における特徴を考慮し、既存辞書の語釈文に対して、分野間で異なるオノマトペの意味を拡張するための計算機による支援手法を提案する。

3 提案手法

前節で紹介した中部らの研究では、オノマトペの持つ複数の意味を、複数の感性軸上にプロットすることにより表現している。我々もオノマトペの特徴を鑑み、感性軸の数を増やして、オノマトペの意味表現として取り入れたい。

また、中部らは PWN と英和辞書を用いてオノマトペの意味表現を計算したことに対し、本稿では、オノマトペの意味を表現する感性軸の特定に、JWN と 258 対の感性軸をベースとした評価語ネットワークを用いる。その動機付けは、この評価語ネットワークをオノマトペの特徴に対応できるように、分野毎の文章データを用いて語彙間の関係付けを変更したいからである。

このようにして構築した評価語ネットワークに対して、分野毎の特性を反映するために収集した各分野のレビュー文章集合を用いた調整 (以降、分野調整) を行う。そして、分野調整済みの評価語ネットワークに基づいて、オノマトペの意味表現として感性軸を割り

⁶岩下豊彦. SD 法によるイメージ測定. 川島書店, 1983.

⁷<http://nlpwww.nict.go.jp/wn-ja/>

表 1: 分野毎の出現頻度上位 20 語のオノマトペに対して計算機から割り当てられた感性軸

服飾分野			化粧品分野			飲食店分野					
オノマトペ	感性軸 1	感性軸 2	感性軸 3	オノマトペ	感性軸 1	感性軸 2	感性軸 3	オノマトペ	感性軸 1	感性軸 2	感性軸 3
しつかり	薄い-諷い	可愛い-憎い	嫌い-好き	しつとり	嫌い-好き	男らしい-女らしい	少ない-豊か	しつかり	嫌い-好き	美味しい-不味い	特殊-普通
びつたり	薄い-諷い	可愛い-憎い	嫌い-好き	しつかり	嫌い-好き	眩しい-優しい	優しい-優しい	たつぷり	嫌い-好き	大変-楽	甘い-渋い
ゆつたり	薄い-諷い	可愛い-憎い	嫌い-好き	さらさら	嫌い-好き	特別-平凡	眩しい-優しい	あっさり	困難-容易	華麗-質素	簡単-精密
がっかり	可愛い-憎い	嫌い-好き	黒っぽい-白っぽい	さっぱり	四角い-丸い	鋭い-丸い	完全-不完全	さっぱり	四角い-丸い	鋭い-丸い	完全-不完全
べらべら	薄い-諷い	可愛い-憎い	丁寧-乱暴	すっきり	嫌い-好き	新鮮-古い	眩しい-優しい	ゆつくり	嫌い-好き	騒がしい-静か	温かい-冷たい
すっきり	薄い-諷い	可愛い-憎い	特殊-普通	たつぷり	疎い-密	心地良い-悪い	嫌い-好き	しつとり	穏やか-激しい	穏やか-激しい	嫌い-好き
ちくちく	鈍感-敏感	四角い-丸い	完全-不完全	べたべた	嫌い-好き	少ない-豊か	薄い-諷い	こつてり	嫌い-好き	特殊-普通	大人しい-激しい
びちびち	陰気-陽気	寂しい-賑やか	静か-賑やか	つるつる	薄い-諷い	嫌い-好き	粗野-優雅	がっかり	粗野-優雅	乏しい-豊か	自然-不自然
ぎりぎり	特殊-普通	普通-珍しい	薄い-諷い	びつたり	青白い-黒い	嫌い-好き	質素-華やか	さくさく	安全-危険	乏しい-豊か	地味-華やか
はつきり	不明-明瞭	薄い-諷い	青白い-黒い	びりびり	心地良い-悪い	少ない-豊か	華奢-丈夫	ふわふわ	眩しい-優しい	眩しい-優しい	残酷-優しい
ふわふわ	眩しい-優しい	眩しい-優しい	残酷-優しい	くるくる	丈夫-脆い	嫌い-好き	眩しい-優しい	どんどん	新しい-古臭い	悲しい-幸せ	美味しい-不味い
ぱつぱつ	淡い-濃い	鮮やか-淡い	寂しい-楽しい	はつきり	不明-明瞭	嫌い-好き	簡単-難解	びつたり	公平-不公平	質素-贅沢	賢沢-貧しい
さっぱり	完全-不完全	四角い-丸い	鋭い-丸い	ふんわり	眩しい-優しい	眩しい-優しい	嫌い-好き	はつきり	クラシック-モダン	不明-明瞭	繊細-力強い
ふんわり	眩しい-優しい	眩しい-優しい	残酷-優しい	ひりひり	嫌い-好き	眩しい-優しい	嫌い-好き	ゆつたり	快適-深い	心地良い-耳障り	快い-不快
ぶかぶか	可愛い-憎い	恐ろしい-可愛い	多い-少ない	ほんのり	嫌い-好き	独特-平凡	眩しい-優しい	とろとろ	黒い-白い	嫌い-好き	温かい-冷たい
ゆるゆる	薄い-諷い	嫌い-好き	可愛い-憎い	かさかさ	辛い-甘い	安価-高価	甘い-苦い	ほんのり	甘い-渋い	特殊-普通	甘い-酸っぱい
さらさら	固い-柔らかい	有難い-迷惑	豪華-安っぽい	ふくら	嫌い-好き	眩しい-優しい	眩しい-優しい	すっきり	美味しい-不味い	清潔-不潔	黒い-白い
たつぷり	重い-軽い	大変-楽	薄い-諷い	するする	素早い-鈍い	速い-遅い	薄い-諷い	すんなり	困難-容易	簡単-精密	簡単-困難
めちやくちや	賢い-馬鹿	臭い-水っぽい	心地良い-悪い	ぬるぬる	心地良い-耳障り	心地良い-不安	心地良い-優しい	ふんわり	眩しい-優しい	眩しい-優しい	残酷-優しい
どんどん	遅い-速い	可愛い-憎い	素早い-鈍い	がっかり	貧弱-立派	急-緩やか	青白い-黒い	しゃきしゃき	簡潔-複雑	単純-複雑	シンプル-複雑

感性軸 1,2,3 は、感性極性値の絶対値順に対応

当てる (図 1)。

具体的には、我々は高村ら [5] の手法を参考に、対象となる分野のレビュー文章集合を用いたスピンモデルに基づくネットワークの重み付け訓練 (以降、分野調整) を実施する。高村らの研究は、PWN をベースに当該分野の文章データを用いて拡張した語彙ネットワークに対して、少数の初期単語集合 (「良い-悪い」といった Positive/Negative の極性が自明である語の対) を用いて訓練してやることにより、語彙ネットワーク上の各語彙が、Positive/Negative のどちらの極性 (以降、P/N 極性値) であるかを決定している。このような語彙ネットワークの訓練は、評価語の分野調整に肯定的な結果を得ている [6]。

従って本稿では、左記の手法をより発展させ、258 対の感性軸それぞれについて分野調整を行うことにより、評価語ネットワーク上の各評価語 i に対して、感性軸 k に基づく感性極性値 ($-1 \leq m_{i,k} \leq 1$) を獲得する。そして、複数の感性極性値 $m_{i,k}$ からなるベクトルを感性極性ベクトル:

$$M_i = \{m_{i,1}, m_{i,2}, \dots, m_{i,k}, \dots, m_{i,258}\} \quad (1)$$

とする。

4 実験と結果の検討

本節では、分野調整済みの評価語ネットワークを用いて、分野毎にオノマトペの意味表現として感性軸の割り当てを行う。具体的には、評価語ネットワークの分野調整により獲得したオノマトペ i の感性極性ベクトル M_i 中の要素から、感性極性値の絶対値 $|m_{i,k}|$ の上位から 3 軸 (それぞれ上位順に感性軸 1,2,3 と呼ぶ) を抽出し、オノマトペ i に対する感性軸として割り当

てた。表 1 に、分野毎のレビュー文章集合における出現頻度上位 20 語のオノマトペと、それらのオノマトペに対して割り当てられた感性軸 1,2,3 を示す。

表 1 に示した各オノマトペの意味表現として、既存辞書の語釈文と分野調整済みの評価語ネットワークにより割り当てられた感性軸との意味的な対応関係について議論したい。表 1 のオノマトペについて、語釈文と感性軸との意味的な対応について人手で整理した一部を表 2 に示す。表 2 では、各オノマトペの語釈文に対して、分野毎に割り当てられた感性軸が、語釈文の表現する意味を含んでいると考えられるならば、語釈文と感性軸を対応付けた。また、語釈文と対応が見られない感性軸は【その他】としてまとめた。

例えば「はつきり」の「あいまいでなく、確かな様子。」において、服飾分野・化粧品分野では「不明-明瞭」という 1 対の感性軸が対応しているが、飲食店分野では「不明-明瞭・繊細-力強い」という 2 対の感性軸が対応している。このことから、計算機から割り当てられた感性軸が、分野毎に異なる意味を適切に表現していることを確認した。また、各オノマトペにおける【その他】に含まれる感性軸は、語釈文との対応は見られなかったが、これらの感性軸は語釈文には記述されていない発見的な意味である可能性が考えられる。

例えば「すっきり」の【その他】において、化粧品分野では「新鮮-古い」という化粧品の新しさに関する感性軸が対応しているが、飲食店分野では「美味しい-不味い・清潔-不潔」という料理の評価や店内の清潔さに関する感性軸が対応している。このことから、既存辞書の語釈文にはない分野毎に発見的な意味を、特定分野について調整を行った評価語ネットワークから獲得できた可能性がある。

表 2 の【その他】に対応する感性軸が、語釈文には見

表 2: 既存辞書の語釈文と計算機から割り当てられた感性軸との人手による対応付け

オノマトベ	語釈文	服飾分野			化粧品分野			飲食店分野		
		感性軸 1	感性軸 2	感性軸 3	感性軸 1	感性軸 2	感性軸 3	感性軸 1	感性軸 2	感性軸 3
がっかり	「予想や期待が外れて落胆した様子。」		嫌い-好き							
	【その他】	可愛い-憎い	黒っぽい-白っぽい		貧弱-立派	急-緩やか	青白い-黒い	粗野-優雅	乏しい-豊か	自然-不自然
すっきり	「無駄がなく洗練されている様子。」									
	【その他】	薄い-深い	可愛い-憎い	特殊-普通		新鮮-古い	眩しい-優しい	美味しい-不味い	清潔-不潔	黒い-白い
たっぶり	「物の量が余るほど十分にある様子。液体にも固体にも用いる。」									
	【その他】	重い-軽い	大変-楽	薄い-深い		心地良い-悪い	嫌い-好き	嫌い-好き	大変-楽	甘い-渋い
どんだん	「物が続けて衝突したり破裂したりして出す、低く大きな音。たとえば、...」									
	【その他】	遅い-速い		素早い-鈍い		計算機による感性軸の割り当て未実行				
はつきり	「他と紛れることなく、明らかな様子。」		不明-明瞭							
	【その他】		可愛い-憎い					新しい-古臭い	悲しい-幸せ	美味しい-不味い

【その他】: 語釈文との対応が見られない感性軸

表 3: 分野毎に計算機から割り当てられた感性軸の分析

感性軸	分野	計算機から割り当てられた感性軸			合計
		語釈文中に意味が存在する感性軸	【その他】		
			発見的な感性軸	発見的ではない感性軸	
感性軸	服飾分野	10 (16.67%)	29 (48.33%)	21 (35.00%)	60 (100.00%)
	化粧品分野	16 (26.67%)	27 (45.00%)	17 (28.33%)	60 (100.00%)
	飲食店分野	16 (26.67%)	37 (61.67%)	7 (11.67%)	60 (100.00%)
	合計	42 (23.33%)	93 (51.67%)	45 (25.00%)	180 (100.00%)

発見的: 語釈文との対応は見られないが、対象分野の意味として適切だと考えられる感性軸

られない発見的な意味(発見的な感性軸)であるか否か(発見的ではない感性軸)を人手により分類し、整理した結果を表3に示す。結果として、計算機から割り当てられた180対の感性軸(3分野 × 20語のオノマトベ × 3対の感性軸)のうち75%(= (42+93)/180)の感性軸が、オノマトベの意味表現として適切であることを確認した。

以上の結果から、既存辞書の語釈文に対して、分野間で異なるオノマトベの意味を拡張する際に、特定分野について調整を行った評価語ネットワークから獲得した感性軸を、表2で示したように既存辞書の語釈文と対応・対照させることで、オノマトベが用いられる分野毎に、特徴的な意味表現の特定や語釈文中には見られない意味を発見することに寄与すると考えられる。

5 おわりに

本稿では、分野毎の文章データに基づいて構築した評価語ネットワークを用いて、当該分野のオノマトベの特徴的な意味の特定や、既存辞書に記述されていない意味の発見に寄与する手法を提案し、それに基づいて、オノマトベの意味の特徴を踏まえた分析が可能であることを示した。また、提案手法では、文章データに出現する評価語の相関関係を扱うため、辞書に登録

されていない新規語に対しても感性軸を割り当てることが可能である。

今後の課題としては、本稿の提案である複数の感性軸と評価語ネットワークをさらに精緻化することにより、オノマトベに対して割り当てた感性軸を自動的かつ高度に認定する処理を開発していきたいと考えている。その際、計算機が割り当てたオノマトベの意味表現としての感性軸をより客観的に評価する方法の開発が重要になる。

参考文献

- [1] 西村由美, 竹内和広. 言い換え調査に基づくオノマトベ表現の選択志向性-看護介護場面での例文調査を題材に-. 比較文化研究 No.112, pp.55-70, 2014.
- [2] 古武泰樹, 佐藤理史, 駒谷和範. オノマトベを言い換える表現の自動収集. 言語処理学会第17回年次大会, pp.904-907, 2011.
- [3] Charles E. Osgood. The Measurement of Meaning. University of Illinois Press, 1957.
- [4] 中部文字, 渡辺知恵美. 感性情報を利用したオノマトベ学習システムの開発. DEIM2009, E5-1, pp.1-7, 2009.
- [5] Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting Semantic Orientations of Words Using Spin Model. Proc.ACL 2005, pp.133-140. 2005.
- [6] 千本達也, 竹内和広. アンケート知識を導入した日本語 WordNet の分野調整に基づくテキスト評価分析. 知能と情報 (to appear 2015).