

日本語の呼びかけ・応答・挨拶・独言・間投表現機械辞書とその応用*

田 辺 利 文 **
 高 橋 雅 仁 ***
 首 藤 公 昭 ****

A Japanese MWE Lexicon of Call, Response, Greeting, Monologue, Interjection, and Its Applications

Toshifumi TANABE**, Masahito TAKAHASHI*** and Kosho SHUDO****

Abstract:

Since Sag et al.(2002), researchers of natural language processing(NLP) have become aware that proper treatment of multiword expressions(MWEs) is one of the most central and intriguing problems in the field. Recognizing the crucial importance of MWEs, Shudo began in the late 1960s to construct a Japanese MWE lexicon for general use.

In this paper, we begin with an overview of the lexicon of Japanese multiword expressions (Japanese MWE Lexicon : JMWEL). Next, we introduce a Japanese MWE lexicon of call, response, greeting, monologue, interjection, and its applications. The feature of this lexicon is that it contains about 1,100 header entries, and each entry has an emotion vector. It is a vector of ten levels of 27-dimensional emotional categories.

Key Words : Natural Language Processing, Multiword Expression (MWE), Collocation, Lexicon, Emotional Categories, Sentiment Analysis

1. はじめに

『油を売る』『赤の他人』『金魚のフン』などの、近年、構成性に基づいて意味を扱うことが難しいイディオムや語の強い結合によって成り立つ決まり文句など特異性のある複数の単語からなる表現を複単語表現 (Multi-Word Expression:MWE) と名付け、機械処理の立場からその全体像を俯瞰、考察した Sag らの論文 [3] がきっかけとなり、自然言語処理における MWE の重要性が広く認識されるようになった。著者の 1 人である福岡大学名誉教授の首藤は、長年にわたり、新聞記事、雑誌記事等から日本語の複単語表現の収集を行い、日本語複単語表現機械辞書（以下、JMWEL: Japanese MWE Lexicon と呼ぶ）と

してまとめている。

本論文では、JMWEL の概要を述べるとともに、『本当?』、『参ったなあ』、『今に見てろ』、『おととい来やがれ』等の表現に、発話者の感情情報を付与した新たな枠組みの、日本語の呼びかけ・応答・挨拶・独言・間投表現機械辞書について紹介する。本論文は [13] の具体的な応用を中心に拡充したものである。本論文の構成は以下の通りである。第 2 章で JMWEL の概要および特徴を紹介し、続く第 3 章では日本語の呼びかけ・応答・挨拶・独言・間投表現機械辞書の概要、特徴、性質および応用例について述べる。第 4 章では感情分類に関する関連研究について言及し、第 5 章でまとめ、今後の課題について論ずる。

2. 日本語複単語表現辞書 JMWEL

2.1 開発の歴史と見出しの採録基準

著者の 1 人である福岡大学名誉教授の首藤は日本語複単語表現辞書 (JMWEL: Japanese MWE Lexicon) を 1960

* 平成 28 年 12 月 2 日受付

** 電子情報工学科

*** 久留米工業大学

**** 福岡大学名誉教授

年代から開発してきており、現時点では、見出し数は基本形 14 万件を超え、一般の日本語処理用辞書としても有効なレベルに達している [8][9][10][12][14]。

JMWEL の見出しの採録基準としては基本的に次の 2 種の特異性に着目している。一つは、例えば、『油を売る』の意味が「油」、「を」、「売る」という要素語の通常の意味から導くことが難しいという性質 (イディオム性)、他の一つは、例えば、「こまねく」という動詞は『手をこまねく』以外にはほとんど使われないというような性質 (決まり文句性) を言う。これらの性質のうち少なくとも 1 つを有するものを JMWEL の見出しとして採録している。採録に関しては、新聞記事、雑誌記事等をもとに内省に基づいているものの、首藤らや田辺らのこれまでの研究で、(1) 収録表現は高頻度で文に現れるものが多く辞書は相対的にコンパクトにまとまっていること、(2) ビッグデータにも現れない表現をも収録していること、などから収録方法が妥当であることが示されている。

2.2 記載情報

JMWEL には、原則として以下の a. ~ f. の情報が記載されている。

a. **区切り, 表記情報**: 辞書見出しは『ばかをみる』のように平仮名ベタ表記で与え、「ばか/を/みる」と要素単語に / で区切られること、そのうち、「ばか」は「バカ」「馬鹿」「莫迦」と 4 種、「みる」は「見る」「観る」と 3 種に表記可能であることが記載されている。従って、この場合、 $12(=4 \times 3)$ 種の異表記が与えられていることになる。

b. **文法機能情報**: 例えば、『墓穴を掘る』は全体として動詞句 (VP)、『命の洗濯』は名詞句 (NP)、『年がら年中』は副詞句 (AdvP) の働きをする、など、表現全体の相当文法カテゴリーが記載されている。

c. **文法構造情報**: 例えば、『目が点になる』は、目 → が → なる、点 → に → なる という依存構造を持つことを自立語は品詞記号、付属語はローマ字つづりを使い、カッコ [] で [[Nga][[Nni]V30]] と 2 項の句表示をしている¹。また、並列構造はカッコ <>、あるいは《 》で、並列要素はカッコ () で表示している。例えば、『泣く子と地頭』の構造は <[[V40N]]to(N)> である²。

d. **内部修飾可否情報**: 例えば、慣用句『油を売る』は「油

をいつもの店で売る」のように内部修飾句をとり、ギャップが生じることがある。このことを c. の文法構造情報中にアスタリスクで [[Nwo]*V30] のように記載する。この情報は、慣用句などをいつも単語化して扱うことのデメリットを避け、より柔軟なギャップ付きフレーズ (不連続フレーズ) として扱うための枠組みである³。

e. **文脈情報**: 例えば、軽動詞構文『顔をする』は『困った顔をする』の様に文頭側に連体修飾句を要求する。また、副詞的表現『一つたりとも』は後方に「与えない」の様な否定句を要求する。この種の必須(あるいは選好)文脈が記載されている。

f. **連体, 連用, 動詞化情報**: 本辞書は『独りよがり』、『針で突いた程』、『子供だまし』のような、物事の様態を表わす形容動詞的と言える表現を含んでおり、これらが連体、連用修飾句として用いられ、動詞化して用いられる際の後続要素を記載している。例えば、擬態語「フラフラ」は「フラフラの」、「フラフラした」、「フラフラとした」で連体修飾、「フラフラ」、「フラフラと」、「フラフラして」、「フラフラとして」で連用修飾、「フラフラする」、「フラフラとする」と動詞化すること、これに対して「グングン」は、「グングン」、「グングンと」の連用修飾形のみ存在することなどが記載されている。

2.3 部分辞書の構成

JMWEL を、見出しの文法機能 (相当品詞) で分割した 11 種の部分辞書の構成を以下に記す。その他、トピック別の部分辞書が 8 種ある⁴。

- (1) 日本語名詞性複単語表現機械辞書 (23,500 件)
(例) 『無二の親友』、『あれやこれや』、『愚の骨頂』
- (2) 日本語動詞性複単語表現 (1 類) 機械辞書 (35,800 件)
(例) 『手を結ぶ』、『意味がある』、『沽券に関わる』
- (3) 日本語動詞性複単語表現 (2 類) 機械辞書 (13,800 件)
(例) 『骨の髄までしゃぶる』、『目から鱗が落ちる』
- (4) 日本語動詞性複単語表現 (3 類) 機械辞書 (3,700 件)
(例) 『放り出す』、『飲んだくれる』、『秋めく』
- (5) 日本語形容詞性複単語表現機械辞書 (4,800 件)
(例) 『頭が痛い』、『性格がきつい』、『途方も無い』
- (6) 日本語形容動詞性 (様態) 複単語表現機械辞書 (2,500 件) (例) 『願ったり叶ったり』、『足手纏い』

1 ここで、ga は格助詞「が」、ni は格助詞「に」、V30 は動詞終止形の品詞記号である。JMWEL での依存構造は、(文節間ではなく) 単語間の係り受け構造を採用している。カッコを用いていることから文脈自由文法の枠組みで記述していることが自明である。

2 V40 は動詞連体形の品詞記号である。

3 この内部修飾可否情報を持つことが JMWEL の大きな特徴の 1 つである。JMWEL では依存構造が文脈自由文法で記述されていることから、アスタリスクの位置に単語が存在する場合でも、文脈自由文法の性質の 1 つである「非交差性」を満たすように部分木を作成する必要があるものとしており、構文解析時に、表現に与えられた依存構造とマッチしないような部分木が生成される場合にはその候補を却下することで構文解析の曖昧さを軽減することが可能となる。

4 ここでのトピックの具体例は、<http://jefi.info/> の「辞書構成」を参照されたい。本論文で紹介する「日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書」もトピック別の部分辞書の 1 つである。

- (7) 日本語連用修飾複単語表現機械辞書 (16,100 件)
(例)『思いもよらず』、『気を引き締めて』
- (8) 日本語連体修飾複単語表現機械辞書 (16,100 件)
(例)『世に言う』、『筋の通った』、『得も言われぬ』
- (9) 日本語 MWE 辞書__談話指標表現編 (1,200 件)
(例)『そうは言っても』、『とはいえ』、『驚くべき事に』
- (10) 日本語文末表現 (終助詞、助動詞性表現) 機械辞書 (4,400 件) (例)『～かもしれない』、『～てもよるしい』
- (11) 日本語関係表現 (格助詞、副助詞、接続助詞性表現) 機械辞書 (2,600 件)
(例)『～における』、『～のいかにかわらず』、『～の甲斐あって』、『～ところの』、『～を励みに』

3. 日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書

日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書は、基本的に JMWEL に収録された見出しのうち、呼びかけ、応答、挨拶、独言、間投表現としての機能をもつ口語表現をまとめて機械辞書としたもので、例えば、『あーおもしろい』、『今に見てる』、『おととい来やがれ』、『冗談じゃない』、『すぐこれだ』、『そうなんだよね』、『そんなアホな』、『なんとということでしょう』、『ほんとですかあ』、『もう一度言ってみろ』、『いやだ』などの多彩な表現が数多く収録されている⁵。現時点では約 1,100 件の見出しが収録されており、日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書の記載情報は、JMWEL の記載情報を全て含むものとなっている。

3.1 特徴

日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書の記載情報としては、一般的な JMWEL の記載情報には含まれていない、種別情報、意味用法が新たに与えられている⁶。種別情報は、Call (呼びかけ)、Res (応答)、Self (独言)、Lgh (笑い)、Grt (挨拶)、Fil (間投表現)、Sht (叫び) の 7 種が与えられている⁷。また、意味用法としては、27 種の感情情報 (感情タグと呼ぶ) に対応する項目欄を設け、感情情報ごとにその強さが 0 (無い)

～9 (強い) の 10 レベルで付与されている⁸。27 種類の感情タグは、< 疑問 >、< 驚き >、< 理解 >、< 感謝 >、< 肯定 >、< 否定・拒否 >、< 納得・了解 >、< 同情 >、< 詫び・謝罪 >、< 合槌・単純応答 >、< 困惑 >、< 不満 >、< 喜び・満足 >、< 共感 >、< 寂しさ >、< つらさ・痛さ・苦しさ >、< 感動 >、< 鬱陶しさ >、< 悲しさ >、< 羨ましさ >、< 面白さ >、< しくじり >、< 怒り・不快 >、< 恨み >、< 怖さ >、< 依頼 >、< 応援 > と極めて詳細に設定されている。また、感情タグやレベルの設定は人手で設計していること、なども辞書の特徴である。例えば、『本当かなあ』の意味用法は、< 疑問 > が 4、< 否定・拒否 > が 1、のように設定されている。

意味用法は、複雑な感情情報が記載された行ベクトルとみなすことができ、意味用法を感情ベクトルとして用いることで、対話処理などにおいて重要な役割を持つものと考えられる⁹。この辞書を自然言語処理システムに組み込むことで、発話の強弱の調整や、状況に応じた言い換えも可能になるなど、これまでに類をみない、汎用性のある繊細な感情情報処理システムが実現できると考えられる。

3.2 辞書の性質

見出しに与えられた感情ベクトルを使うことで、辞書の性質をうかがうことができる。感情ベクトルにおけるそれぞれの意味タグには 0～9 と 10 レベルの数値が与えられているため、感情ベクトルの大きさ (絶対値) を計算して、絶対値が大きい表現は「感情的な度合いが強い」という解釈が可能であると考えられる。表現に対する感情ベクトルから絶対値を算出すると、例えば、『なにつ』が 14.3、『だいじょうぶ』が 13.3、『おもてにでろ』が 12.7、『おぼえていろ』が 10.7、『おとといきやがれ』が 9.9、また、『こまっちゃう』は 14.0 となっており、おおむね、強い感情をもつ表現にはベクトルの絶対値が高くなるように値が設定されていると考えることができる。

5 『ヤッター』のような、MWE でないものも含まれている。

6 JMWEL の記載情報には 2.2 で述べたものの他に、前接語情報、係り先情報、意味タグがある。これらの情報をもつ部分辞書としては、日本語文末表現機械辞書、日本語関係表現機械辞書があり、これらを用いた日本語文末の推論など、意味に基づく応用処理が可能であることを示している [1][11]。

7 同一の見出しで複数の種別情報をもつものも存在する。例えば、『こまったな』は Res (応答) と Self (独言) の 2 つの用法がある。また本論文では種別情報を、以降、カテゴリとも呼ぶことにする。

8 実際の辞書記述では、意味の強さが 0 の場合には空欄となっており、空欄でない場合には 1～9 までの整数の数値が記述されている。また、見出しと種別情報が同一でも、複数の意味用法をもつものも存在する。例えば『そうですか』の種別情報が Res (応答) になるものは意味用法が 2 つ存在し、1 つは、< 疑問 > が 7、< 否定・拒否 > が 2 であり、もう 1 つは、< 理解 > が 2、< 納得・了解 > が 2、< 合槌・単純 > が 5 である。これは意味の曖昧さがあることを示している。

9 日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書は、それぞれの行が、見出しとその記載情報の形式となっている。意味用法は 27 個の列をもって記載されているため、感情ベクトルは行ベクトルの形となる。

表1 カテゴリごとのデータ

カテゴリ	表現数	強さの平均	強さの分散	最小値	最大値	最大値を取る表現の例
Call(呼びかけ)	108	4.28	20.01	0	13.30	だいじょうぶ
Res(応答)	505	7.24	5.87	0	14.35	なにっ
Self(独言)	224	5.76	10.58	0	14.04	こまっちゃう
Lgh(笑い)	19	3.30	5.76	0	6.40	うふふ
Grt(挨拶)	136	1.69	7.39	0	8	ごめんなさい
Fil(間投表現)	11	1.45	5.70	0	6	なんといいますかー
Sht(叫び)	42	5.73	13.26	0	12.25	やったー

3.2.1 カテゴリごとのデータ

同様の方法で、カテゴリごとについて感情ベクトルの大きさを計算する。表1に、それぞれのカテゴリに対して、そのカテゴリに属する表現数、感情的な強さの平均、感情的な強さの分散¹⁰、最小値、最大値、最大値を取る表現の例を示す。

表1から、感情的な強さの平均は、Res(応答)に属するものが高く、それにSelf(独言)とSht(叫び)が続き、Fil(間投表現)やGrt(挨拶)が低いという結果となった。Grtの強さが低いのは、一部を除いて、感情があまりこもっておらず、感情伝達の上では挨拶という行為が形骸化していると考えられる。また、Self(独言)は、例えば『こまっちゃう』はSelfの強さの最大値である14.04を取っているなど、強い感情をもつ表現が見られた。この表現は感情は強いものの、機能的には独言であるため、その感情を自分自身で消化しようとしていると捉えることもできる¹¹。Call(呼びかけ)は、強さの平均はあまり高くはないものの、強さの分散が大きく、表現ごとの強さの差が大きいことが分かった。Callで強さの高いものとして『だいじょうぶ』が13.3、『おもてにでる』が12.7、『おぼえている』が10.7、『おとといきやがれ』が9.9であったが、その反面、感情の強さが0であるもの(後述する)も少なくないことが分かった。また、すべてのカテゴリにおいて、感情の強さの最小値はいずれも0であることが分かった。

3.2.2 言い換えへの応用

日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書を用いた応用の1つとして、言い換えが挙げられる。辞書に記載されている表現には感情ベクトルが与えられているため、感情ベクトルを用いて2つの表現間のコサイン類似度を計算することが可能となり、コサイン類似

度が1に近い場合には、それらの表現同士は意味的に似ているものと解釈できる。今回、コサイン類似度の算出結果が1になる表現のペアが複数存在することが確認された。このような表現のペアがある場合、カテゴリが異なっている場合には工学的に応用する際に都合がよいことがある。例えば、Self(独言)である『うれしいなー』とRes(応答)である『よかったです』は、類似度が1になったペアの1つであるが、このようなポジティブな意味を含む表現の場合には、独言ではなく、相手への応答という形式に言い換えて、直接ポジティブな気持ちを相手に伝えることが考えられる。また、ネガティブな意味を含む表現の場合でも類似度が1になるペアがあり、具体的にはRes(応答)である『じょうだんも休み休み言え』とSelf(独言)『いやだなあ』であったが、このようなネガティブケースである場合には、応答という形式を用いずに、独言の形式として、間接的に、それとなく気持ちを伝えるような発話をするということが考えられる¹²。

コサイン類似度は、感情ベクトルの方向がどれだけ一致しているかの指標であり、感情の強さ(感情ベクトルの大きさ)には無関係である。そこで、コサイン類似度が1である表現のペアに対して、それらの表現の感情の強さも考慮に入れることができれば、相手に対して、意味は変えないまま感情の強さを変化させるようなより高度な言い換えも可能となる。

3.2.3 感情タグの設定

日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書で定義された27種の感情タグには、粒度が細かいことから似た意味をもつ感情タグがあることも推測される。そこで本研究では、感情タグごとの相関係数を算出することを試みる。表現間の意味的類似度を算出する際

10 ここでの分散は標本分散を用いている。

11 この場合、形式は独言であるが、実際には、応答としての機能があるものとしても考えることができる。実際、本辞書では、『こまっちゃう』は、Res(応答)の種別情報をもつものも別見出しとして収録されている。

12 この例はあくまでも応用例の1つにすぎないが、よりスムーズな対話を実現するためには必要かもしれない。

には、行ベクトルとしてみた感情ベクトルを用いたが、感情タグごとの相関係数を算出するには、その感情タグの意味を持つ単語の感情の強さを要素とするベクトル、すなわち要素の値が0～9までの整数をもつ列ベクトルを用いる。この列ベクトルの要素数は辞書の見出し数となる¹³。相関係数を r とすると、 r は-1から1の間の値をとり、一般的には、 $0.7 \leq |r| \leq 1$ の場合にはかなり強い相関があり、 $0.4 \leq |r| \leq 0.7$ の場合にはやや相関があり、 $0.2 \leq |r| \leq 0.4$ の場合には弱い相関があり、 $0 \leq |r| \leq 0.2$ の場合にはほとんど相関はない、とされている。

感情タグ間の相関係数を求めた結果、相関係数の絶対値が高かったものから、<怒り・不快>と<不満>の間が0.54、<納得・了解>と<理解>の間が0.50、<否定・拒否>と<怒り・不快>の間が0.42、<合槌・単純>と<理解>の間が0.42と算出された。相関係数の値が負のものの絶対値の最大値をとったものは<疑問>と<共感>の間の-0.16であった。そこで具体的にしてみると、感情タグ<怒り・不快>と<不満>にかかわる表現として、『じょうだんじゃないよ』では、<怒り・不快>の値は6、<不満>は8であり、『じょうだんをいうな』では、<怒り・不快>の値は5、<不満>は0であった。このように、相関係数がやや高い値である場合でも感情タグの値は異なることから、若干の依存関係はあるものの感情タグはほぼ独立であるとみなすことが可能であると考えられる¹⁴。

4. 関連研究

感情分類の体系は研究者によって大きく異なる。日本語において、表現と感情との対応関係が示されているものとして感情表現辞典[6]がある。感情表現辞典は、<喜>,<怒>,<哀>,<怖>,<恥>,<好>,<厭>,<昂>,<安>,<驚>の10種類の感情を表した意味タグ(感情タグ)を定義し、感情を持つ表現(複単語表現を含む)がどの感情タグをもつかの対応関係が示されている。また、ラッセルの円環モデル[4]は、「すべての感情は、快(pleasure)－不快(unpleasure)、覚醒(arousing)－眠気(sleepy)を2軸とする平面に円環上に並んで存在する」というもので、すべての感情は2次元ベクトルとして表現できることを示しており、28種類の感情が円環状にどのように配置されるかが例として示されている。ラッ

セルの円環モデルを用いた研究として[5]がある。感情情報が2次元ベクトルで表現されることから計算機での取り扱いには適切であるが、表現と感情との対応関係は明確に示されておらずモデルの提案に留まっている。

本研究で用いる日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書は、電子化されており、見出し約1,100、感情タグが27種類、感情の強さを10レベルで設けているため、詳細でありながら計算機で利用しやすいという特徴をもつ。

5. おわりに

本論文では、日本語の複単語表現を収録したJMWELの概要を述べるとともに、日本語の呼びかけ・応答・挨拶・独言・間投表現辞書を紹介した。日本語の呼びかけ・応答・挨拶・独言・間投表現辞書には、従来の感情分類体系には見られない、人手で設計した詳細な感情タグとレベル設定が大きな特徴である計算機用機械辞書である。本論文の後半ではこの辞書の性質と感情ベクトルを用いたいくつかの応用についての可能性について論じた。カテゴリと感情の強さに関しては言語学的な洞察がさらに必要となるだろう。

今後の課題としては、まずは、感情表現辞典やラッセルの円環モデルに挙げられる、既存の感情分類との対応付けが考えられる。これらの異なる感情分類間での感情タグの対応関係を整理し、この対応関係の妥当性を評価する必要がある。次に、感情の強さを10レベルで表現していることの評価が必要となる。それぞれの感情タグには0～9の10レベルの数値で感情の強さを表記している。そこで、例えば、10レベルの数値を、0～4、5～9の2レベルに分けたのち、10レベルと2レベルでの感情タグのレベルが異なることでの違いを評価することが考えられる。また、感情ベクトルのレベルだけが異なる表現同士を言い換えて、どちらが感情の程度が強いかの評価も行う。次はベクトル演算の可能性を示すことである。見出しに感情ベクトルが対応していることから、例えば、表現 e の感情ベクトルを $V(e)$ とすると、表現 α 、 β に対して、 $V(\alpha) + V(\beta) = V(\gamma)$ となる表現 γ が存在すれば、表現列 $\alpha\beta$ (または $\beta\alpha$)が γ と言い換えられるか等を調査する。

次のステップとしては、対話システムへの組み込みが考えられる。対話システムへの組み込みには、対人関係

13 言い換えると、意味タグ $m1$ と $m2$ の、表現 e における値、 $m1(e)$ と $m2(e)$ による $(m1(e), m2(e))$ を xy 座標系における座標と考える。 $m1(e)$ と $m2(e)$ の値はそれぞれ0～9までの整数値をとるため、座標は $10 \times 10 = 100$ 通りの点のどれかになる。直感的には、その点の分布がどれだけ $y=x$ のグラフに近いかと考えるとよい。

14 『じょうだんをいうな』と『じょうだんをいえ』は、共に<怒り・不快>と<不満>の値がそれぞれ5、0になっている。これらの表現は同じ意味として使われることから感情ベクトルも同一になっている。この場合、単語単位で意味を認識、合成しようとする否定表現の有無により、表現同士が同じ意味であることの認識がしにくくなってしまうため、表現の意味の認識には、単語単位ではなくMWEを単位とみなす考え方が重要であることを示唆している。

のモデルも組み込むことを考えており、ラッセルの円環モデルと類似したモデルとして対人円環モデルがある[2][7][15]。対人円環モデルの考え方は、温和-冷淡(親密性)、支配-服従(支配性)を2軸とする平面上に対人特性がマッピングされる、というものであり、ラッセルの円環モデルと同様に低次元のベクトルで表現可能である。対話の状態は固定的ではなく、対話が進むにつれて対人特性が円環上で変化することを仮定した上で、27次元の感情ベクトルが円環上のどの位置に対応するかが分かれば、発話をモデルに組み込むことも可能であると考えている。対話システムの組み込みの延長線上にあるのが、対話ロボットへの実装である。現在、1人暮らしの高齢者向けに、会話ができるペトロボットのニーズなどが高まってきているが、会話・口語表現を収集した言語資源が十分に整備されていないことから、現状では、ペトロボットなどで日常的な会話のやりとりを実現するのは難しい。本論文で紹介した日本語の呼びかけ・応答・挨拶・独言・間投表現辞書を用いることで、繊細な感情認識にも耐えうるような対話ロボットシステムの構築も考えている。

本論文で提案した27種の感情は、日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書の見出しのみ存在するものではなく、一般的な単語やMWEにも同様に含まれているものと推測できる。MWEに含まれる感情タグとしては、例えば、動詞性複単語表現『手に余る』には<鬱陶しさ>、『目に余る』には<怒り・不快>、『手を結ぶ』には<納得・了解>などが、また動詞性以外の複単語表現、例えば『火の車』には<つらさ・痛さ・苦しさ>、『荷が重い』には<困惑>、『願ったり叶ったり』には<喜び・満足>などが考えられる。このように感情情報を含むMWEは少なくとも、JMWELに登録されている見出しに対して27種類の感情タグによる感情ベクトルを与えることができれば、より柔軟な感情情報処理が可能になるだろう。

その一方で、本論文で紹介した日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書での感情タグは27種類に設定されているが、感情の粒度は27種類で良いのかという問題がある。辞書中の感情ベクトルの大きさが0である表現は少なくとも、そのような表現としては、Callの『ってきます』『いってらっしゃい』『こらーっ』『すみませんが』、Filの『うーん』、Grtの『おつかれさまです』『こんにちは』『じゃあね』、Lghの『おほほ』『がっはっは』、Resの『おそれいました』『そいつはありがたい』『まいったか』、Selfの『おー』『さーてね』『そうねえ』『しめたっ』、Shtである『おととと』『えいやっ』があるようである。このような表現は意味がないものと考えて、文の解析処理中に除去することも考えられるが、このような表現にも何らかの意味があると考え、既存のどれかの感情タグが示すわずかな意味が入っている

か、既存の感情タグには入っていないような感情タグが必要になる、などが考えられる。感情タグの種類は研究者ごとにまちまちであり、粒度の細かいものとしては、感情表現辞典のあとがきによれば、「表現を素材として見た感情の諸相は「慶・喜・謝・幸・誇・楽・明・爽・快・温・浮・恍・躍・弾・笑・怒・憤・屹・不愉快・煮・痛・不機嫌・悲・淋・怖・恥・辱・好・厭・嫌・憎・悔・妬・惑・鬱・惰・苦・焦・苛・緊・昂・動・安・驚・衝・呆」という四六種ほどの広がりがある」とある。このように、感情ベクトルの大きさが0になっている表現の意味を再考しつつ感情の粒度について再検討する必要もある。

参考文献

- [1] 雨崎翔悟, 田辺利文, 首藤公昭. 推論ルールを用いた日本語文末表現の言い換え. 情報処理学会研究報告. NL-205. pp.1-7. 2012.
- [2] 橋本泰央, 小塩真司. 対人円環モデルに基づいたIPIP-IPC-Jの作成. 心理学研究. 87, pp.395-404. 2016.
- [3] Ivan. A. Sag et al.. Multiword Expressions: A Pain in the Neck for NLP. Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing, Volume 2276, pp 1-15. 2002.
- [4] J.A.Russell. A circumplex model of affect. Journal of Personality and Social Psychology, Vol.36, pp.1161-1178, 1980.
- [5] 松浦有容, 渥美幸雄. 感情表現による書評情報の可視化手法の提案と実装. 専修大学情報科学研究所報 (78), pp.11-28, 専修大学情報科学研究所. 2012.
- [6] 中村明. 感情表現辞典. 東京堂出版. 1993.
- [7] 大淵憲一, 堀毛一也. パーソナリティと対人行動. 誠信書房. 1996.
- [8] 首藤公昭. 日本語複単語表現機械辞書 1. 37版: JMWEL v1.37. <http://jefi.info/>. 日本語処理研究工房・ことばの森. 2016.
- [9] Kosho Shudo, Akira Kurahone, and Toshifumi Tanabe. A Comprehensive Dictionary of Multiword Expressions, Proceedings of the 49th Annual Meeting of the ACL, pp.169-177, 2011.
- [10] 首藤公昭, 田辺利文. 日本語の複単語表現辞書: JDMWE, 自然言語処理, Vol.17, No. 5, pp.51-74. 2010.
- [11] Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi, and Kenji Yoshimura. MWEs as non-propositional content indicators. Proceedings of ACL2004 Workshop on Multiword Expressions: Integrating Processing, pp.31-39. 2004.
- [12] 首藤公昭, 田辺利文, 高橋雅仁. 文の部分構造を与

える日本語フレーズレキシコン. 言語処理学会第
20 回年次大会発表論文集, pp.1134-1137. 2014.

- [13] 高橋雅仁, 田辺利文, 首藤公昭. 日本語の呼びかけ・
応答・挨拶・独言・間投表現機械辞書. 電気関係
学会九州支部第 69 回連合大会論文集. 2016.
- [14] Toshifumi Tanabe, Masahito Takahashi, and Kosho
Shudo. A lexicon of multiword expressions for
linguistically precise, wide-coverage natural language
processing. *Computer Speech and Language*, 28-6,
pp.1317-1339. Elsevier. 2014.
- [15] Wiggins, J. S., Trapnell, P., and Phillips, N.
Psychometric and geometric characteristics of the
Revised Interpersonal Adjective Scales (IAS-R).
Multivariate Behavioral Research, 23, pp.517-530.
1988.