

日本語の呼びかけ・応答・挨拶・独言・間投表現機械辞書

高橋 雅仁* 田辺 利文** 首藤 公昭***

*久留米工業大学情報ネットワーク工学科 **福岡大学工学部電子情報工学科 ***福岡大学名誉教授

1 はじめに

近年、構成性に基づいて意味を扱うことが難しいイデ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[1]が端緒となり、自然言語処理におけるその重要性が広く認識されるようになった。筆者らの一人(首藤)は、長年に亘り、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[2][3]。一方、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、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ため、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。筆者らは、このような用途に対応した会話表現辞書を JMWEL に追加したいと考えている。本稿では、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・間投表現辞書に収集されている、「本当?」、「参ったなあ」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会話表現辞書について紹介を行う。

2 日本語複単語表現機械辞書 JMWEL

2.1 特徴

- (1)網羅性が比較的高い。見出し数約 145,000 件。
- (2)構文構造を与えている。構文解析処理に、本辞書を組み込むことで、構文解析精度の向上が見込まれる。

2.2 辞書構成

表現の文法機能(相当品詞)で分けた 11 種の部分辞書の構成を以下に記す。その他、「日本語呼びかけ・応答・挨拶・独言・間投表現機械辞書(見出し数約 1,100 件)」を含むトピック別の部分辞書が 8 種ある。

- (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」
- (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」、「～でもよろしい」
- (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」、「～のいかんにかかわらず」、「～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

3.1 特徴

- (1)網羅性が高い。約 1,100 件の口語表現を収録している。
- (2)本辞書の各表現に、新たに、<驚き>、<疑問>、<困惑>など 26 種の感情に対応する項目欄を設け、感情毎にその度合いを 0~10 点で付与した。複雑な感情を感情ベクトルとして表現可能である。

3.2 辞書の内容

辞書エントリーの内容項目を以下に記す。

- (1)辞書種別: Call-Res-Self-Lgh-Grt-Fil
- (2)見出し
- (3)分かち書き
- (4)表記形
- (5)種別: Call:呼びかけ、Res:応答、Self:独言、Lgh:笑い、Grt:挨拶、Fil:間投表現
- (6)構文構造
- (7)文脈条件:文頭および文末
- (8)意味用法(感情との対応をベクトルとして記す):
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 計 26 種

- (例)「本当かなあ」 見出し:ほんとうかなあ、分かち書き:ほんとう-か-なあ、表記形:本当-か-なあ、構文構造:[Nka]naa、意味用法:疑問:4 点、拒否・否定:1 点

4 まとめ

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・間投表現に感情情報を付与した会話表現辞書の試案を紹介した。

参考文献

- [1] I. A. Sag et al., “Multiword Expressions: A Pain in the Neck for NLP”, Computational Linguistics and Intelligent Text Processing Volume 2276 pp 1-15, 2002.
- [2] 首藤公昭, “日本語複単語表現機械辞書 1.37 版: JMWEL v1.37”, 2016. <http://jefi.info/>
- [3] T.Tanabe, M.Takahashi, K.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.