

機械処理のための日本語オノマトペ表現レキシコン： JMWEL_onomatopoeic

首藤公昭^{†1} 田辺利文^{†2} 高橋雅仁^{†3}

概要：オノマトペ（擬音語、擬態語）は日本語の重要な語彙資源であり、その多彩さは日本語の特徴とされているが、従来、機械処理において十分な対応がなされてこなかった。本論文では、筆者らが開発した日常の日本語処理を目的としたオノマトペ表現レキシコン JMWEL_onomatopoeic の概要を報告する。

キーワード：オノマトペ、複単語表現、機械辞書、言語資源、日本語処理、機械翻訳

A Lexicon of Onomatopoeic Expressions for Japanese Language Processing: JMWEL_onomatopoeic

KOSHO SHUDO^{†1} TOSHIFUMI TANABE^{†2}
MASAHITO TAKAHASHI^{†3}

Abstract: One of the remarkable features of Japanese Language is the richness of the onomatopoeias. In spite of this, almost no Japanese language processing system which is able to well process onomatopoeias has been developed yet. In this paper, an overview of a lexicon of Japanese onomatopoeic expressions developed by the authors for Japanese language processing is introduced.

Keywords: Onomatopoeia, Mimesis, Sound-Symbolic Word, Ideophone, Multiword Expression, Lexicon for NLP, Linguistic Resource for NLP

1. はじめに

オノマトペ（擬音語、擬態語）の豊富さは日本語の特徴の一つであり、日本語を母国語とする者にとって有用で不可欠な語彙資源であるが、従来、自然言語処理（NLP）、特に、機械翻訳などでは、十分な対応ができていない。また、近年、微妙、繊細な人の感覚を端的に表現するオノマトペの重要性が注目されており、特定領域文書に出現するオノマトペを調査・抽出する研究[1][2]をはじめ、オノマトペの意味に関する研究[3][4]、コーパスからオノマトペと他語の共起を調べる研究[5][6]、出現オノマトペを手掛かりに文書の種別判定をする研究[7]等々、多彩な研究が報告されているが、機械処理用のオノマトペ辞書を開発した報告は筆者らの知る限り見られない。本稿では、筆者らが開発した機械処理用の日本語オノマトペ表現レキシコン JMWEL_onomatopoeic の概要を紹介する。

自然言語文に使われている語の意味は、文中（あるいは文脈中）の他語との共起によって相補的に定まるため、語の共起をどのように捉えるかが NLP の基本課題であるが、共起の枠組みを論理式や格フレームなどで規定し、シソーラ

スと併用する旧来の方式に代わり、近年、コーパスから語の表層レベルの共起をベクトル化して捉える分散表現方式が広く試みられている。いっぽう、今世紀に入り、日常の言語ではコロケーション、決まり文句、慣用表現等の定型表現が予想外に多種、多量に使われていることが注目されるようになり、NLP 分野で複単語表現 Multiword Expression; MWE[8]、言語学では定型言語 Formulaic Language[9]、単語連鎖 Lexical Bundles[10]などの名称で種々の研究が行われるに至っている。筆者らは、語の共起を表層レベルで捉えることの重要性を古くから認識し、総括的な日本語複単語表現レキシコン JMWEL の開発を進めてきた[11][12]。本稿で述べる日本語オノマトペ表現レキシコン JMWEL_onomatopoeic（以後、本レキシコンと呼ぶ）は、JMWEL の一部をなすものである。

2. 採録表現

本レキシコンの見出しは、

(1) オノマトペ（単体）約 2,600 種

(2) オノマトペと他語が共起した日常よく現れる句

（以後、オノマトペ共起表現と呼ぶ）約 22,000 種

であり、方言、古語、ほとんど使われなくなったもの、特定年齢層で使われるものなどを除く。「悠々」、「懇々」、「生き生き」のような漢字表記可能なもの、感動詞に分類し得

^{†1} 福岡大学名誉教授
Emeritus, Fukuoka University

^{†2} 福岡大学
Fukuoka University

^{†3} 久留米工業大学
Kurume Institute of Technology

るものも採録している。

採録は、新聞、雑誌等の記事、小説、テレビ、ラジオの放送文から内省によって抽出したものをベースに、既存の不特定の辞典[13][14]等を使って補強したものである。

3. 記載情報

本レキシコンは、Microsoft Excel で作成した xls ファイルとして纏められ、管理されており、1 行に割り当てた 1 個の見出しに対して、A～I 欄に以下の情報を記載している。

3.1 オノマトペ (A 欄)

見出し表現中使用されているオノマトペを片仮名表記で与える。オノマトペからその共起表現を検索する際の指標となる。

3.2 見出し (B 欄)

オノマトペ、オノマトペ共起表現ともに平仮名ベタ書きで見出しを与える。同音異義、同音異機能のオノマトペに対しては別見出しを立てる。例えば、「ばらばら」は擬音と擬態で別見出し、「こんこん」では、擬音とは別に擬態の多義でも「懇々」、「昏昏」、「滾々」を別見出しとする。

3.3 分かち書き (C 欄)

オノマトペ共起表現に対する分かち書きを平仮名表記上にハイフン「-」で与える。分かち書きの単位は、単語、接頭語、接尾語、接頭造語要素、接尾造語要素とし、活用語尾は切り離さない。ただし、形容動詞語尾の「な」、「に」、「たる」、「と」は切り離す。複合語は基本的にアンダースコア「_」で要素語に分ける。

3.4 異表記 (D 欄)

漢字、カタカナなど、異表記可能な表現には、C 欄の分かち書きの上で、一種の正規表現によってオプションを与える。例えば、「ポッチャリ-と-し-た-(身)体_付き」は「ポッチャリ-と-し-た-身体_付き」、「ポッチャリ-と-し-た-体_付き」の 2 様の可能性、「満(満/々)-たる-自身」は、「満満-たる-自身」、「満々-たる-自身」の 2 様の可能性を示す。

3.5 構文機能 (E 欄)

採録しているオノマトペには、構文・形態的機能により

- 1 単純オノマトペ
- 2 連用オノマトペ (副詞的オノマトペ)
- 3 接頭オノマトペ
- 4 接尾オノマトペ
- 5 名詞性オノマトペ

がある。1 は格助詞「と」を後接して連用修飾機能を持つもの、2 はそのままでも連用修飾機能を持つもの、3, 4 は他語に接続して造語する機能を持つ、あるいは造語されているものである。以上を区別して表 1 の記号を記載する。H 欄の情報で 1, 2 はさらに詳細化される。

以上の機能を異にする同音レキシコンは別見出しとしている。以上を区別して表 2 の記号を記載する。表 1 に本レキシコンにおける 1～5 の分布と例を示す。

いっぽう、採録しているオノマトペ共起表現には、

- i 名詞句
- ii 動詞句
- iii 形容詞句
- iv 形容動詞 (語幹) 句
- v 連用修飾句
- vi 連体修飾句

が有る。表 2 に本レキシコンにおける i～vi の分布と表現例を示す。

3.6 構文構造 (F 欄)

C 欄のハイフンによる分かち書きに基づき、オノマトペ共起表現に対して係り受け構造を、修飾子、被修飾子の対をカッコ[]で括って記載する。すなわち、句 α の主辞が句 β の主辞を修飾して出来た句 $\alpha \beta$ の構造記述を α 、 β の構造記述 a, b を使って[ab]とする。

表 1 採録したオノマトペの分布と例

Table 1 Morpho-syntactic categories, their sizes and examples of onomatopoeias in the lexicon.

機能	記号	採録数	例
単純オノマトペ	O	1,439	ツルリ, ホワッ, ドッカーン, グネツ, ピューン, ゴロリ, ヒヤッ, ドブン
連用オノマトペ	AdvO	1,055	ドッカー, フラフラ, ミッチリ, フワフワ, チャリチャリ, ゴリゴリ
接頭オノマトペ	Op	106	ドタ, ジリ, グラ, ゴタ, ソヨ, ビリ
接尾オノマトペ	Os	19	タツブリ, タラタラ, モリモリ, ピカ
名詞性オノマトペ	NPO	13	ブツブツ, コリコリ, フリフリ, デコボコ
計		2,632	

表 2 採録したオノマトペ共起表現の分布と例

Table 2 Syntactic categories, their sizes and examples of onomatopoeic expressions in the lexicon.

構文機能	記号	採録表現数	表現例
名詞句	NP	1,817	サッパリと-し-た-性格, カリッと-し-た-口 当り, サラサラ-し-た-肌 触り, ギリギリの-妥協
動詞句	VP	15,081	ドタッと-音-が-する, 鼻-先-に-人參-を-ブラ-(下/提)げる, 肌-が-バサバサ-に-乾く, 馬-が-ヒューン-と-嘶く, ニャーン-と-(ネコ/猫)-が-(鳴/啼)く, フツと-胸-に-浮かぶ, カツカツと-靴音-が-する, クラクラと-(眩暈/目眩)-が-する, (深(深/々)/シンシン)-と-夜-が-(更/深)ける
形容詞句	AP	239	モチモチと-柔らかい, ポンポンと-威勢-が-良い
形容動詞(語幹)句	AVP	117	愛嬌-タツプリ, フンワリと-柔らかか, ツンツンと-無愛想
連用修飾句	AdvP	2,969	キョトンと-し-て, ガラガラ-音-を-立て-て, 熟(熟/々)-思う-に, ワイノワイノと
連体修飾句	AdnP	1,844	シドロモドロの, グチョグチョ-し-た, キラキラ-し-た-瞳-の
計		22,067	

構成単語の構造記述は、以下の通りとする。

- ・単純オノマトペ, 連用オノマトペ: O
- ・接頭オノマトペ: Op
- ・接尾オノマトペ: Os
- ・接頭語: P
- ・接尾語: S
- ・接頭造語要素: Q
- ・接尾造語要素: R
- ・名詞: N
- ・動詞: V (未然形 V11,V12, 連用形 V22,V23, 終止形 V30, 連体形 V40, 仮定形 V50, 命令形 V60)
- ・形容詞: A (未然形 A13, 連用形 A22, A23, 終止形 A30, 連体形 A40, 仮定形 A50, 命令形 A60)
- ・形容動詞(語幹): K00
- ・副詞: D
- ・連体詞: T
- ・接続詞: C
- ・機能語及び機能性自立語: 活用形も含め英小文字綴り

文節内の語の接続も、便宜上、左 2 分岐句構造とみなして同様の記述を行う。

例えば、「クンクン-と-犬-が-鳴く」の構造記述は、

「クンクン」: オノマトペ O

「と」: 格助詞 to

「犬」: 名詞 N

「が」: 格助詞 ga

「鳴く」: 動詞終止形 V30

から[[Oto]*[[*Nga]*V30]]と記載する。図 1 にこれの意味する構文木と係り受け構造を示す。

3.7 内部修飾可能性 (ギャップ付き構造記述)

F 欄の構造記述内には前節の例の如く、必要に応じてアスタリスク「*」が含まれている。アスタリスクは、直後の句の主辞に対する修飾句がこの位置に入り得ることを意味する。従って、図 1 の構造記述[[Oto]*[[*Nga]*V30]]は、「クンクン-と-朝-から-隣-の-犬-が-寂し-そう-に-鳴く」のような拡大表現の可能性を示している。図 2 にこの模様を示す。JMWEL のこのようなギャップ付き構造記述は定型表現の柔軟性を留保するのに有効である。

3.8 後方文脈条件 (G 欄)

オノマトペに対し、文末側に呼応する語句がある場合にその情報を与える。例えば、「おちおち」には文末側に否定句が要求されることを<negation>と記載する。

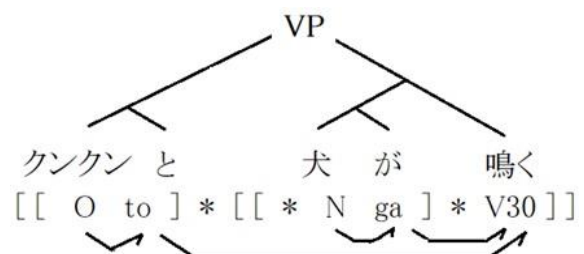


図 1 「クンクン-と-犬-が-鳴く」の構造記述
Figure 1 Syntactic structure denoted by [[Oto]*[[*Nga]*V30]] given to the verbal expression クンクン-と-犬-が-鳴く kunkun-to-inu-ga-naku “a dog whines”.

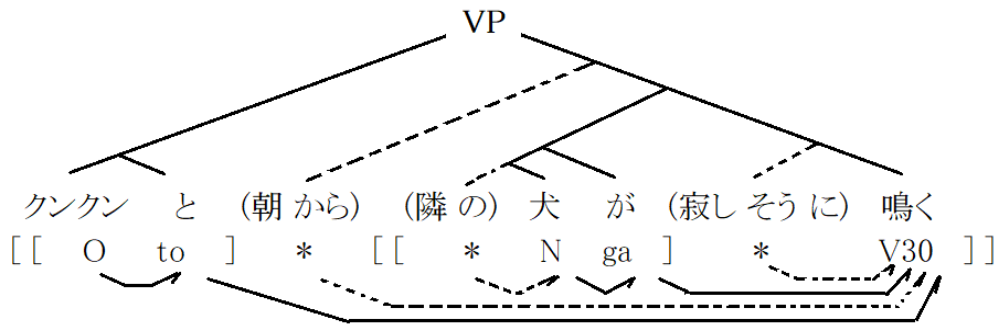


図2 「クンクン-と-朝-から-隣-の-犬-が-寂し-そう-に-鳴く」の構造記述

Figure 2 Syntactic structure of expression クンクン-と-朝-から-隣-の-犬-が-寂し-そう-に-鳴く *kunkun-to-asu-kara-tonari-no-inu-ga-sabisi-sou-ni-naku* “a dog of the neighboring house whines sadly since the morning” derivable from the description $[[Oto]*[[*Nga]*V30]]$ given to the expression クンクン-と-犬-が-鳴く *kunkun-to-inu-ga-naku* “a dog whines”.

3.9 連体化, 連用化, 動詞化情報 (H 欄)

オノマトペに対し, E 欄の構文機能情報を詳細化して記載する. オノマトペを連体修飾, 連用修飾に使用する場合と動詞化して使用する場合に必要な後接語句を以下のように整理した.

- 連体修飾: 「な」, 「の」, 「たる」
- 連用修飾: 「に」, 「と」, 「ε」
- 動詞: 「する」, 「になる」, 「とする」

ε はオノマトペがそのまま連用修飾可能な場合を表す. 例えば, 「ふらふら」の場合は, 「ふらふらの (姿)」で連体修飾, 「ふらふらと (歩く)」, 「ふらふら (歩く)」で連用修飾, 「ふらふらする」, 「ふらふらとする」と動詞化し, それ以外の可能性はない. また, 「ふっ」は「ふっと (気が付く)」で連用修飾する以外には考えにくい. これらの特徴をそれぞれ, {no}- {to, ε} - {suru, tosuru}, Φ - {to} - Φ と集合のトリプルで記載する. Φ は空集合を表す. 実際のトリプルは 110 種程度となっている.

3.10 擬音, 擬態の別 (I 欄)

オノマトペに対し, 擬音, 擬態の別を「音」, 「態」と記載する.

4. むすび

本論文では, 筆者らが開発した日本語複単語表現レキシコン JMWEL (Japanese Multiword Expression Lexicon)の一部をオノマトペ表現レキシコン JMWEL_onomatopoeic の現在における概要を述べた. 自然言語処理をフレーズベースで行うには, フレーズレキシコンの存在が望まれ, ACL 等でも積極的に開発の試みが報告されているが[15], これまで, 十分な規模, 内容の成果は得られていない様である. JMWEL は, 現在, 11 万件を超える見出し数を数え, 最初のプロトタイプ日本語フレーズレキシコンとして機能するのではないかと考えている. JMWEL の設計思想, 構成, 特徴などの細部については, [11][12]等を参照されたい.

参考文献

- [1] 高丸圭一, 内田ゆず, 乙武北斗, 木村奏知. 地方議会会議録におけるオノマトペの出現傾向に関する基礎的検討. 言語処理学会第 20 回大会発表論文集, 2014.
- [2] 井上音々, 望月源. 日本語歌謡曲のオノマトペに関する調査. 言語処理学会第 23 回大会発表論文集, 2017.
- [3] 小宮嘉那子, 佐々木稔, 新納浩幸. 分散表現と文脈ベクトルによるオノマトペの分類の比較. 言語処理学会第 22 回大会発表論文集, 2016.
- [4] 福島弘識, 内田ゆず, 荒木健治. 2 つの意味を持つオノマトペの意味判別における組成の検討. 言語処理学会第 20 回大会発表論文集, 2014.
- [5] 玉岡賀津雄, 木山幸子, 宮岡弥生. 新聞と小説のコーパスにおけるオノマトペと動詞の共起パターン. 言語研究 139, 2011.
- [6] 乙武北斗, 内田ゆず, 高丸圭一, 木村奏知. 表層核に着目したオノマトペ共起語の抽出と分析. 言語処理学会第 22 回大会発表論文集, 2016.
- [7] 渡辺知恵美, 中村聡. オノマトペロリ: 味覚や食感を表すオノマトペによる料理レシピのランキング. 人工知能学会論文誌, 30-1, 2015.
- [8] I.A.Sag, T. Baldwin, F. Bond, A. Copestake, D. Flickinger. A Pain in the Neck for NLP. Proc. of the 3rd CICLING, 2002.
- [9] R. Corrigan, E. A. Moravcsik, H. Ouali, K. Wheatley (eds.). Formulaic Language Vol.1, Distribution and historical change. John Benjamins Publishing Company, 2009.
- [10] D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan (eds.). Longman Grammar of Spoken and written English. Harlow: Pearson Education Limited, 1999.
- [11] 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, 2014.
- [12] 首藤公昭. “日本語処理研究工房 ことばの森”. <http://jefi.info>, 2011 開設.
- [13] 小野正弘. 日本語オノマトペ辞典. 小学館, 2007.
- [14] 阿刀田稔子, 星野和子. 擬音語擬態語使い方辞典 第 2 版. 創拓社出版, 2004.
- [15] "Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)". http://multiword.sourceforge.net/PHITE.php?sitesig=CONF&page=CONF_04_LAW-MWE-CxG_2018, COLING 2018.