

国立国語研究所学術情報リポジトリ

日本語オノマトペ共起表現レキシコンJMWEL_onomatopoeic

メタデータ	言語: Japanese 出版者: 公開日: 2019-02-14 キーワード (Ja): キーワード (En): JMWEL_onomatopoeic 作成者: 首藤, 公昭, 田辺, 利文, 高橋, 雅仁 メールアドレス: 所属:
URL	https://doi.org/10.15084/00001685

日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic

首藤 公昭 (福岡大学名誉教授)

田辺 利文 (福岡大学)

高橋 雅仁 (久留米工業大学)

A Lexicon of Japanese Onomatopoeic Collocations: JMWEL_onomatopoeic

Kosho Shudo (Fukuoka University, professor emeritus)

Toshifumi Tanabe (Fukuoka University)

Masahito Takahashi (Kurume Institute of Technology)

要旨

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とされるが, 機械翻訳などの言語処理では十分な対応がなされていない. 筆者らが開発した, 平仮名ベた書き見出し, 形態素分かち書き, 構文機能, 形態・構文構造, 内部修飾可否情報, 文脈条件などを与えた見出し総数約 34,500 の日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic の概要を紹介する. 本レキシコンは, オノマトペと他語のコロケーション集であり, 大規模日本語複単語表現レキシコン JMWEL の部分レキシコンである.

1. はじめに

オノマトペ (擬音語, 擬態語) は日本語を母語とする人にとって有用・不可欠な語彙資源であるが, 従来の機械翻訳等の自然言語処理 Natural Language Processing, NLP では十分な対応がなされていない. 近年, 繊細・微妙な人の感覚を的確に表現するオノマトペが注目されており, 特定領域の文書に出現するオノマトペを調査する研究 (高丸ほか 2014, 井上ほか 2017) を始め, オノマトペの意味に関する研究 (小宮ほか 2016, 福島ほか 2014), オノマトペと他語の共起を調べる研究 (玉岡ほか 2011, 乙武ほか 2016), オノマトペを手掛かりに文書の種別を判定する研究 (渡辺ほか 2015) 等々, 種々の研究が行われている.

いっぽう, 今世紀に入り, 日常の言語にはコロケーション, 決まり文句, 慣用表現等の特異表現が予想外に多種, 多量に使われていることが重視されるようになり, NLP分野では複単語表現 Multiword Expression, MWE (Sag et al. 2002, Tanabe et al. 2014), 言語学では定型言語 Formulaic Language (Corrigan et al. 2009, Jiang et al. 2007), 単語連鎖 Lexical Bundles (Biber et al. 1999), 構文文法 Construction Grammar (Fillmore et al. 1988) といった枠組みで種々の研究が進められている. 筆者の一人は, 1960年代からこの種の日本語複単語表現の総括的なレキシコン Japanese MWE Lexicon, JMWELの開発を進めてきた. (Tanabe et al. 2014)

本稿では, JMWEL の一部をなす日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic (以下, 本レキシコンと記す) の概要を紹介する.

本レキシコンの主な特徴は,

- i オノマトペと他語の共起表現中にギャップ (内部修飾句) が介在する可能性を記載している
- ii オノマトペ (単体), オノマトペ共起表現中の語彙に漢字・片仮名異表記を与えている

iii オノマトペの連体，連用，動詞化用法を体系化して記載している点である。

2. 収録表現

本レキシコンの見出しは，

- (1) オノマトペ (単体) 約 3,000 種¹
- (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記す) 約31,500種

であり，「悠々」，「懇々」，「生き生き」のような漢語由来，漢字表記可能なもの，「オヤッ」のような感動詞に分類し得るものを含め，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基本として既存の不特定の辞典類 (小野 2007，阿刀田ほか 2004) を使って補強したものである。方言，古語や近年殆ど使われなくなったもの，特定年齢層でのみ使われるものは除外されている²。

3. 記載情報

本レキシコンは，Microsoft Excel で作成したxlsx ファイルに纏められており，1 行に割り当てた 1 個の見出しに対して，A～I 欄に以下の情報を与えている。例えば，「クンクンと犬が鳴く」という表現に対して与えた情報を A～I 欄の順に列挙すれば以下ようになる。(欄の区切りを・・・で，空データをφで示す。)

クンクン・・・くんくんといぬがなく・・・くんくんと-いぬ-がなく・・・クンクン-と-犬-が-鳴く・・・VP・・・[[Oto]*[[*Nga]*V30]]・・・φ・・・φ・・・音

3.1 オノマトペ：A 欄

見出し表現中に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に利用できる。

3.2 見出し：B 欄

オノマトペ (単体)，オノマトペ共起表現ともに平仮名べた書きで見出しを与える。同音異義，同音異機能オノマトペは原則として別見出しとする。例えば，「ぱらぱら」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し，その分かち書きを平仮名表記上にハイフン「-」で区切って与える。分かち書き単位は，単語，接頭語，接尾語，接頭造語要素，接尾造語要素とし，活用語尾は形容動詞語尾「な」，「に」，「たる」，「と」以外は切り離していない。造語要素とは造語機能を持つ拘束形態素であり，多くの場合，「緊張-感」の「感」のように音読みの一漢字である。

複合語は基本的にアンダースコア「_」で要素語に区切っている。

3.4 異表記：D 欄

¹ オノマトペ単体は複単語表現ではないが，便宜上，本レキシコンに含めている。

² 「クルリ」に対する「クルリッ」のような末尾が促音化されたものも原則として独立のオノマトペとみなしている。

オノマトペ (単体), オノマトペ共起表現に対して, 漢字, カタカナなど, 異表記可能な部分には, C 欄の分ち書きの上で, 正規表現類似の形式で選択肢を与える. 例えば, 「ポツチャリ-と-した-(身)体_付き」という D 欄の記載は「ポツチャリ-と-した-身体_付き」, 「ポツチャリ-と-した-体_付き」の可能性, 「満(満/々)-たる-自信」は, 「満満-たる-自信」, 「満々-たる-自信」の可能性を表す. ハイフンやアンダースコアで区切られた C, D 欄の記載から種々の表記形を簡単に生成できる, 例えば, C 欄の「からだ_つき」と D 欄から得られた「身体_付き」, 「体_付き」から「からだつき」, 「身体つき」, 「身体付き」, 「からだ付き」, 「体付き」, 「体つき」の 6 通りの表記形が得られる.

3.5 構文的機能 : E 欄

収録したオノマトペ (単体) は, 形態・構文上の機能により,

- 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

に分類される.

1 は, 格助詞「と」を後接して連用修飾機能を持つものである. 本レキシコンでは, 「ころっと」などは, オノマトペ「ころっ」と格助詞「と」の共起表現とみなしている. 末尾促音型オノマトペの殆どは 1 に分類される.

2 は, そのままでも「と」を後接して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ものである.

表1に本レキシコンにおける1~5 の分布と例を示す.

E欄には表1の記号が記載されている. 1, 2 の機能は, 後述するH欄の情報でより詳細化される.

表1 収録したオノマトペの分布と例

種別, 記号	見出し数	例
単純オノマトペ, O	1,588	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	1,155	ドツカリ, 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	148	ドタ, ジリ, グラ, ゴタ, ソヨ, ビリ
接尾オノマトペ, Os	32	タツプリ, タラタラ, モリモリ, ピカ
名詞性オノマトペ, NPO	81	ブツブツ, コリコリ, フリフリ, デコボコ
計	3,004	

いっぽう, 本レキシコンに収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
- vii 名詞文形式

が有る。

表2に本レキシコンにおける i ~vii の分布と表現例を示す。

オノマトペ共起表現の E 欄には表2の記号が記載されている。

表2 収録オノマトペ共起表現の分布と例

種別, 記号	見出し数	例
名詞句, NP	3,688	サッパリ-と-した-性格, カリッ-と-した-口_当り, ホクホク-した-食感, サラサラ-した-肌_触り, ギリギリ-の-妥協
動詞句, VP	21,248	ドタッ-と-音-が-する, 鼻-先-に-人参-を-ブラ-(下/提)げる, 肌-が-パサパサ-に-乾く, 馬-が-ヒューン-と-嘶く, ニャーン-と-(ネコ/猫)-が-(鳴/啼)く, フッ-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, キリキリ-痛む, (深(深/々)/シンシン)-と-夜-が-(更/深)ける
形容詞句, AP	485	モチモチ-と-柔らかい, ポンポン-と-威勢-が-良い, ガンガン-と-痛い
形容動詞(語幹)句, AVP	193	愛嬌-タップリ, フンワリ-と-柔らかか, ツンツン-と-無_愛想
連用修飾句, AdvP	3,721	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	2,187	シドロモドロ-の, グチョグチョ-した, ガチガチ-な
名詞文形式など, NPS, INC, OPS	15	英語-が-ペラペラ, 予定-が-ビッシリ, 収支-が-トントン
計	31,535	

3.6 構文構造と内部修飾可能性表示 : F 欄

オノマトペ共起表現に対して C 欄のハイフンによる分かち書きに基づき, 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]と記載する.

ここで, 要素単語の構造記述は, 以下の英記号とする.

- ・単純, 連用, 名詞性オノマトペ : 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ローマ字綴り

文節内の語接続も便宜上、左2分岐句構造とみなして上記と同様の記述を行っている。

例えば、オノマトペ共起表現「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であることから、[[Oto]*[[*Nga]*V30]] と記載する。図1にその意味する構文木と係り受け構造を示す³。

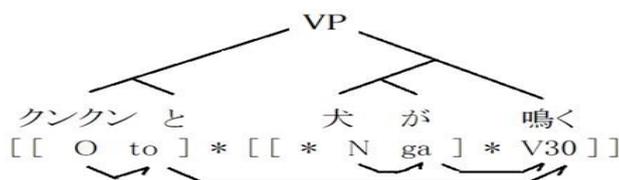


図1 オノマトペ共起表現「クンクン-と-犬-が-鳴く」の構造記述

この例の如く F 欄の構造記述内には適所にアスタリスク「*」が含まれており、その位置に、直後の句の主辞に対する修飾句が入り得ることを意味している。従って、図1の構造記述 [[Oto]*[[*Nga]*V30]] は、例えば、「クンクン-と-朝-から-隣-の-犬-が-寂し-そう-に-鳴く」のような拡張表現の可能性を示している。図2にこの模様を示す⁴。

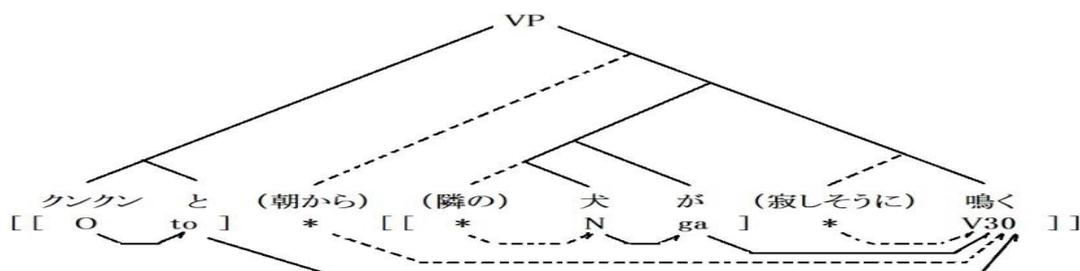


図2 「クンクン-と-朝-から-隣-の-犬-が-寂し-そう-に-鳴く」の構造記述

JMWELのこのようなギャップ付き構造記述は、定形表現が持ち得る部分的な柔軟性を構文・

³ この表現に対する2018年6月時におけるGoogleオンライン英訳は“Cunk and dogs cry”であった。

⁴ この表現に対する2018年6月時におけるGoogleオンライン英訳は“Cunngun and morning next dog crying lonesome”であった。

意味解析機に反映させるための重要な仕組みである。

3.7 後方文脈条件：G 欄

オノマトペ (単体) , オノマトペ共起表現に対し, 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, 「オチオチと」に対しては, 文末側に「休んではいけない」のような否定句が要求されることを<negation>と記す。

3.8 連体化, 連用化, 動詞化情報：H 欄

オノマトペ (単体) に対し, E欄の構文機能情報を詳細化して与える。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通常使われる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」
- ・動詞化：「する」, 「になる」, 「とする」

ここで, εは空列を表し, オノマトペが後接語句なしで連用修飾できる場合を表す。

例えば, オノマトペ「フラフラ」は, 「フラフラの (...状態)」で連体修飾, 「フラフラと (...歩く)」, 「フラフラ (...歩く)」で連用修飾, 「フラフラする」, 「フラフラになる」, 「フラフラとする」と動詞化すること, 「フツ」の場合は, 「フツと (...気が付く)」で連用修飾する以外には考えにくいことを, それぞれ, 後接する語句集合の三つ組によって{no}-{to, ε}-{suru, ninaru, tosuru}, Φ-{to}-Φ と記載する。Φは空集合を表わす。三つ組のパターンは100種程度である。

3.9 擬音, 擬態の別：I 欄

オノマトペ (単体) に対し, 擬音, 擬態の別を「音」, 「態」と記載する。

4. 応用例

機械翻訳等の自然言語処理においては, オノマトペをその共起表現やコロケーションとして一括して単位的に捉えることが必要である。例えば, 日英機械翻訳において, 前出の MWE 「クンクンと犬が鳴く」全体を単位的に捉えなければ, 的確な英語動詞句 “a dog whines”などを与えることは難しい。これを与えたのち, 本レキシコンに記載された構造記述 [[Oto]*[[*Nga]*V30]] を用いれば, 拡張表現「クンクンと朝から隣の犬が寂しそうに鳴く」に対しても適切な訳文, 例えば, “The neighbor’s dog sadly whines from the morning”などを生成出来る可能性が生じる。図3にこの翻訳過程のイメージを与える。

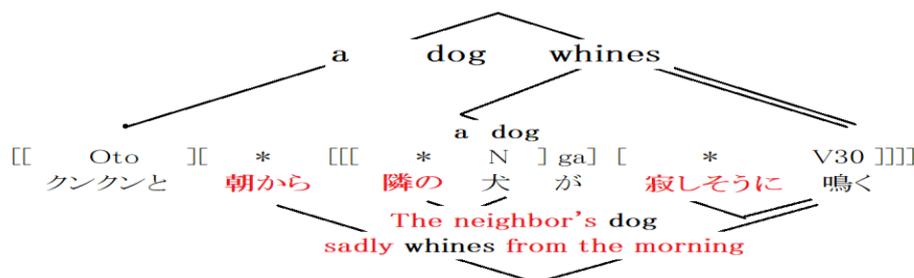


図3 「クンクンと朝から隣の犬が寂しそうに鳴く」に対する英訳
“The neighbor’s dog sadly whines from the morning”を生成する過程のイメージ

5. おわりに

「コロコロと音がする」, 「コロコロ転がる」, 「話がコロコロ変わる」, 「目がコロコロと痛い」, 「コロコロと笑う」など, 他語との共起によって, 「コロコロ」の多義性が低減されるように, 語の共起を如何に捉えるかが NLP における基本的な課題である. 筆者の一人は 1960 年代に始めた機械翻訳の研究を通して, 意味素性や語類による語の共起ルールとは別に表層レベルの相当大規模な語の共起辞書が不可欠であるという認識を得, 大規模日本語フレーズレキシコン JMWEL の開発を始めた. JMWEL の基本的な特徴の一つは必ずしも隣接しない語の共起をデータ化している点にある. 現在, JMWEL は見出し数 14 万件 (異なり) に達し, 未だ十分網羅的とは言い切れないが, プロトタイプとしてのレベルには達したのではないかと考えている.

本稿で紹介したオノマトペ共起表現レキシコンは JMWEL のサブレキシコンの一つである. 本レキシコン及び JMWEL が今後の日本語処理, 機械翻訳, 日本語教育・研究に役立てば幸いである⁵.

文 献

- 阿刀田稔子・星野和子 (2004). 「擬音語擬態語使い方辞典第2版」 創拓社出版.
- Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (eds.) (1999). “Longman Grammar of Spoken and Written English”, Harlow: Pearson Education Limited.
- Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen Wheatley (eds.) (2009). “Formulaic Language, vol.1, Distribution and historical change”, John Benjamins Publishing Company.
- Charles J. Fillmore, Paul Kay and Mary Catherine O’Connor (1988). “Regularity and Idiomaticity Grammatical Construction: The Case of Let Alone” *Language* 64, pp.501-538.
- 福島弘識・内田ゆず・荒木健治 (2014). 「2つの意味を持つオノマトペの意味判別における素性の検討」 言語処理学会第20回年次大会発表論文集, pp.181-184.
- 井上音々・望月源 (2017). 「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回年次大会発表論文集, pp.963-966.
- Nan Jiang and Tatiana M. Nekrasova (2007). “The Processing of Formulaic Sequences by Second Language Speakers”, *The Modern Language Journal*, 91:3, pp.433-445.
- 小宮嘉那子・佐々木稔・新納浩幸 (2016). 「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回年次大会発表論文集, pp.473-476.
- 小野正弘 (2007). 「日本語オノマトペ辞典」 小学館.
- 乙武北斗・内田ゆず・高丸圭一・木村奏知 (2016). 「表層格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回年次大会発表論文集.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger (2002). “Multiword Expressions: A Pain in the Neck for NLP” *Proc. of the 3rd CICLING*, pp.1-15.
- 高丸圭一・内田ゆず・乙武北斗・木村奏知 (2014). 「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回年次大会発表論文集, pp.566-569.
- 玉岡賀津雄・木山幸子・宮岡弥生 (2011). 「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, pp.57-84.

⁵ 本レキシコンの利用については関連サイト <http://jefi.info> を参照されたい.

Toshifumi Tanabe, Masahito Takahashi and Kosho Shudo (2014). “A lexicon of multiword expressions for linguistically precise, wide-coverage natural language processing”, *Computer Speech and Language*, 28:6, Elsevier, pp.1317-1339.

渡辺知恵美・中村聡 (2015). 「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, pp.340-352.

関連 URL

日本語処理研究工房ことばの森 <https://jefi.info>