知能と情報（日本知能情報ファジィ学会誌）Vol.28, No.4, pp.693 - 699 (2016)

ショートノート

係り先動詞に着目したオノマトペの語義分類に関する検討†

高丸　圭一*1・内田　ゆず*2・乙武　北斗*3・木村　泰知*4

本研究では複数の語義をもつオノマトペの語義分類に寄与する係り先動詞について検討した．語義分類済みコーパス「会議録 2010」において単一語義で一定頻度出現した「オノマトペ－係り先動詞」の組について，同一ジャンルのコーパス（会議録 2008－09）および，多様なジャンルの文を含むコーパス（BCCWJ）における用例の抽出と語義の分析を行った．「会議録 2008－09」の正例率(「会議録 2010」の語義と一致する用例の割合）は 99.4% と高く，同一ジャンルの文書では，係り先動詞はオノマトペの語義決定に寄与するといえる．「BCCWJ」の正例率は 78.0% であった．負例の主な要因は係り受け解析誤り，表記の揺れ，複数語義の包含であり，文書のジャンルが語義に与える影響は小さいことが明らかになった．ただし，各オノマトペに対する係り先動詞の出現傾向には文書のジャンルに起因した偏りがあることが示唆された．

キーワード：オノマトペ，コーパス，語義，係り先動詞

1. はじめに

オノマトペ（擬音語および擬態語）は音や状態，様子を伝える言語表現である．日本語のオノマトペには複数の語義を持つものが多い．例えば，『日本語オノマトペ辞典』[1] には「ごろごろ」の語義として“雷のとどろき響く音”(擬音の語義）や“仕事をしないでむだに暮らしているさま”(擬態の語義）など 6 種類が掲載されている．日本語文中で用いられるオノマトペの語義は表 1 に示す例のように，オノマトペの係り先動詞や文の表層格によって決まる．①「ごろごろ」が「鳴る」に係るとき“雷のとどろきひびく音”を，②「ごろごろ」が「鳴らす」に係るとき“猫などがよろこんで喉を鳴らす音”を表していると考えることができる．このような場合，文の他の部分（表層格）を見なくてもオノマトペの語義は決まる．一方，係り先が「する」のときは，係り先動詞のみでは語義が決まらず，③ガ格が「目」のとき“小さいものが間に入って違和感のあるさま”を表し，④デ格が「家」のとき“何もせずに時間を過ごすさま”を表すと考えることができる．「オノマトペとよく共起する語（コロケーション)」と「オノマトペの語義」の関係を整理することで，自然言語処理等においてオノマトペを適切に利活用することが可能となる．しかし，オノマトペは一般的な単語と比べて，語義が派生したり，新しい用法が生じやすいという性質をもつ．このため，コロケーションと語義の関係を作例によって網羅することは困難である．そこで，コーパスから実際の用例を収集し，オノマトペの語義とコロケーションの関係を明らかにする必要がある．

オノマトペと共起する語に関する先行研究［2］では，新聞コーパスを用いて「しんみり」「しみじみ」の 2 語と共起する動詞を調査し，人間が想起する動詞と比較している．また，［3］では，「がたん」「がたがた」等の 9 語を対象として，オノマトペと係り受け関係のある動詞および名詞を手がかりに擬態の語義と擬音の語義を判別する工学的手法が提案されている．

本論文では，複数の語義を持つオノマトペと係り先動詞のコロケーションについて検討する．まず，オノマトペの語義がアノテーションされた日本語コーパス

表 1　表層格・係り先動詞と語義の関係の例

	表層格	オノマトペ	係り先動詞	語義
①	—	ごろごろ	鳴る	雷のとどろきひびく音
②	—		鳴らす	猫などがよろこんで喉を鳴らす音
③	「目」ガ		する	小さいものが間に入って違和感のあるさま
④	「家」デ			何もせずに時間を過ごすさま

† An Investigation of Onomatopoeias' Senses Classification Based on Modified Verbs
Keiichi TAKAMARU, Yuzu UCHIDA, Hokuto OTOTAKE and Yasutomo KIMURA
*1 宇都宮共和大学
Utsunomiya Kyowa University
*2 北海学園大学
Hokkai-Gakuen University
*3 福岡大学
Fukuoka University
*4 小樽商科大学
Otaru University of Commerce

（地方議会会議録コーパス［4］に収録された 2010 年度の発言文，以下「会議録 2010」という）から，オノマトペの係り先動詞を抽出する．ここで，一定の頻度で共起する「オノマトペ－係り先動詞」の組におけるオノマトペの語義が一意であるとき，この係り先動詞はオノマトペの語義決定に寄与する語であるといえる．そこで，「会議録 2010」において単一語義で一定の頻度出現する「オノマトペ－係り先動詞」の組を同一ジャンルのコーパス，および，多様な日本語文を含むコーパスから抽出し，オノマトペの語義が「会議録 2010」のものと同一であるかを調査する．これによりコロケーション（共起する係り先動詞）とオノマトペの語義との関係について検討する．同一ジャンルのコーパスとして，地方議会会議録コーパスに収録された 2008～2009 年度の発言文（以下「会議録 2008－09」という）を用いる．多様なジャンルの日本語文を含むコーパスとして「現代日本語書き言葉均衡コーパス」［5］（以下「BCCWJ」という）を用いる．

2. 語義分類済みコーパスの分析

2.1. オノマトペの係り先動詞抽出

オノマトペを含む文から係り先動詞を抽出するため，まず，CaboCha［5］によって係り受け解析を行う．解析の結果オノマトペを含む文節の係り先が動詞であるとき，これをオノマトペの係り先動詞として抽出する．例えば，

- 「石が**ごろごろ**出てきているんです。」
- 「…行政は**ごろごろ**っと方針が変わって本当に申しわけない…」

という文からは，それぞれ「ごろごろ－出る」「ごろごろ－変わる」が抽出される．ただし，

- 「…こんな大きな石ころが**ごろごろ**していて…」

のようにオノマトペがサ変名詞として使われ，オノマトペと動詞が同一の文節に含まれる場合は，例外として「ごろごろ－する」を抽出する．なお，オノマトペを含む文節の係り先に動詞が存在しない場合，分析対象から除外する．

2.2. 係り先動詞と語義数

高丸ら［7］は「会議録 2010」に含まれる 153 語のオノマトペ（約 1 万文）語義分類した．本論文では，このオノマトペのうち，複数の語義を持つ 4 拍のオノマトペ 14 語を分析対象とする．表 2 に「会議録 2010」においてオノマトペの係り先動詞が抽出された文数，係り先動詞の異なり数，このうちオノマトペが複数語義であった係り先動詞数，および，単一語義であった係り先動詞数をそれぞれ示す．複数の語義がある係り先動詞は延べ 62 語存在した．これらは表 1 の③④の例のように，係り先動詞のみからはオノマトペの語義が決まらないため，今後表層格を含めた分析を行う必要がある．

表 2 係り先動詞の抽出文数と係り先動詞の異なり数

オノマトペ	文数	異なり数	複数語義	単一語義
がたがた	123	25	6	19
がっちり	104	39	3	36
ぐずぐず	40	7	1	6
ごろごろ	58	19	5	14
とんとん	61	24	2	22
のんびり	209	38	6	32
ばたばた	196	42	4	38
ばっさり	102	28	2	26
ぱらぱら	65	27	2	25
ふらふら	84	15	4	11
ぼこぼこ	33	18	3	15
ぼつぼつ	74	37	10	27
ぽつぽつ	58	30	3	27
ゆったり	221	48	11	37
合計	1,428	397	62	335

表 3 単一語義で 10 回以上出現したコロケーション

オノマトペ―係り先動詞	オノマトペの語義
がたがた―言う	騒々しく聞き苦しい不平を言ったり、騒ぎまわったりするさま.
ばっさり―削る	思い切って捨てたり除いたりするさま.
がっちり―やる	抜け目がなく物事をするさま.
とんとん―なる	二つのもの差がほとんどなく，ちょうど同じくらいであるさま.
がっちり―組む	体格やものの構造，体系，組織などが，強く引き締まってゆるぎないさま.
ばっさり―切り捨てる	思い切って捨てたり除いたりするさま.
のんびり―やる	進みが遅いさま.
ぐずぐず―言う	不平などを不明瞭にあれこれとつぶやくさま.
のんびり―暮らす	何にとらわれるでもなく、心身ともにくつろいでいるさま.
ばたばた―倒れる	ものが続けざまに倒れたり衰えたりするさま.
ぱらぱら―見る	本などのページが軽く次々にめくれる音．また，そのさま.

単一語義の係り先動詞は延べ 335 語存在した．出現頻度の低い係り先動詞が多く，約 70% は出現頻度が 1 であった．出現頻度の低い係り先動詞は，現段階ではオノマトペとよく共起するコロケーションであるとはいえない．本論文では単一語義の「オノマトペ－係り先動詞」の 335 組のうち出現頻度が 10 以上の 11 組を検討の対象とする．11 組の「オノマトペ－係り先動詞」とオノマトペの語義を表 3 に示す．表中のオノ

マトペの語義は原則として『日本語オノマトペ辞典』のものを用いた．ただし，「のんびり－やる」の語義“進みが遅いさま”は「のんびり」の語義として［7］において追加された語義であり，辞典に記載された語義“何にとらわれるでもなく，心身ともにくつろいでいるさま”から派生したものである．

3. コロケーションとオノマトペの語義の評価

3.1. 評価データにおける係り先動詞の抽出と語義分類

「会議録 2008－09」(総文数約 2200 万文）および「BCCWJ」(総文数約 100 万文)を用いて，コロケーション(「オノマトペ－係り先動詞」の組）とオノマトペの語義との関係について評価する．2.1 節と同様の方法で 2 つのコーパスからオノマトペの係り先動詞を抽出した．表 3 に示した 11 組の「オノマトペ－係り先動詞」を含む文は，「会議録 2008－09」に計 486 文，「BCCWJ」に計 168 文存在した．これらの文におけるオノマトペの語義を［7］と同様の方法で，筆者らが人手で確認し，「会議録 2010」における語義と照らし合わせた．語義分類が困難な例では，用例文から読み取ることができる範囲でもっとも妥当性の高いと考えられる語義を筆者らの合意の元で選択した．各コーパスから得られた用例数および，正例率(「会議録 2010」の語義と一致する用例の割合）を表 4 に示す．

3.2. 同一ジャンルの文書における結果の考察

「会議録 2010」と同一ジャンルの文書である「会議録 2008－09」では，11 組のうち 9 組が正例率 100% であった．負例は「がたがた－言う」の 2 例，および，「がっちり－組む」の 1 例のみであった．この 2 組における正例および負例を以下に示す．

表 4 各コーパスにおける用例数（括弧内は正例率）

オノマトペ―係り先動詞	会議録 2010	会議録 2008-09	BCCWJ
がたがた―言う	19	53（96.2%）	12（75.0%）
ばっさり―削る	18	59（100.0%）	0（――）
がっちり―やる	17	24（100.0%）	4（100.0%）
とんとん―なる	15	162（100.0%）	15（60.0%）
がっちり―組む	14	24（95.8%）	10（100.0%）
ばっさり―切り捨てる	13	38（100.0%）	6（100.0%）
のんびり―やる	12	29（100.0%）	35（40.0%）
ぐずぐず―言う	11	38（100.0%）	28（85.7%）
のんびり―暮らす	11	10（100.0%）	29（100.0%）
ばたばた―倒れる	11	26（100.0%）	17（100.0%）
ばらばら―見る	11	23（100.0%）	12（75.0%）
合計	152	486（99.4%）	168（78.0%）

「がたがた－言う」の正例

- 「…これほど**がたがた**言うようになったのは，あの財政健全化法という悪法ができてからなんですね。」

「がたがた－言う」の負例

- 「…そういうことを一々言い合いしよると臨時会も**がたがた**なりますので言いませんが…」

「がっちり－組む」の正例

- 「…市がおのおのの特徴を生かしながら，**がっちり**パートナーシップを組んで進めていくことが重要です。」

「がっちり－組む」の負例

- 「…そのためにも本来職員がそういう質疑応答を**がっちり**組む，評価者とのやりとりが重要になってくるんですけれども，残念ながら中にはちょっと十分対応できてない場合もあったりします。」

「がたがた－言う」の負例をみると，「がたがた」の本来の係り先動詞は「なる」であることが分かる．すなわち係り受け解析における解析誤りにより，適切な係り先動詞が抽出されないことが負例の原因となっている．一方，「がっちり－組む」の負例における「がっちり」は“抜け目がなく物事をするさま”と語義分類される．これは今後，表層格（例えばヲ格の「質疑応答」)を手がかりに分析することで解決される可能性がある．ただし，コーパスの実例に基づく分析では，用例の例外性や臨時性の観点からの検討も必要である．これは今後の検討課題である．

3.3. 多様なジャンルを含むコーパスにおける結果の考察

多様なジャンルの日本語文が収録されたコーパスである「BCCWJ」から，11 組の「オノマトペ－係り先動詞」を含む用例が合計 168 個得られた．このうち正例は 78.0%，負例は 22.0% であった．「BCCWJ」から得られた用例にみられる特徴を以下に述べる．

(1) 出現頻度の偏り

BCCWJ における「オノマトペ－係り先動詞」の出現頻度の地方議会会議録コーパス(「会議録 2010」と「会議録 2008－2009」の和）に対する出現比率を図 1 に示す．これは，会議録コーパスにおける出現頻度を 100 としたときの BCCWJ における出現頻度を表す値である．「オノマトペ－係り先動詞」の組ごとに出現比率が異なり，使用傾向に偏りがあることがわかる．具体的には，「がたがた－言う」「ばっさり－削る」「がっちり－やる」「とんとん－なる」「がっちり－組む」「ばっさり－切り捨てる」等は BCCWJ における出現比率が

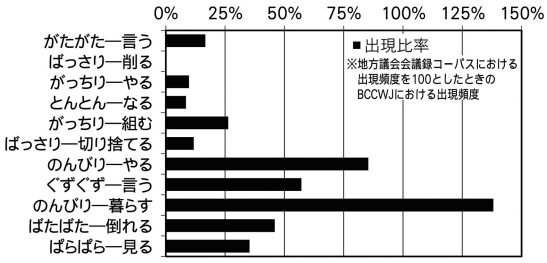

図1　BCCWJにおける出現比率

低い．BCCWJは多様なジャンルの日本語文を含むコーパスであることから，これらの組は会議録のジャンルにおいて，相対的に高頻度で使用されるものであると考えることができる．「会議録2008－09」における「ばっさり－削る」「がっちり－やる」「ばっさり－切り捨てる」の例を以下に示す．用例の観察から，これらの組が予算や施策の審議において特に使用されやすいものであることが示唆される．このように，文書のジャンルや話題に応じて，特徴的な共起語が存在することが予想されるため，さらなる分析が必要である．

「ばっさり－削る」の例（会議録2008－09）

- 「…暮らしや福祉にかかわる事業に対する補助金が**ばっさり削ら**れようとしています。」
- 「…こうしたいわば市民の命綱になっている政策を**ばっさり削る**と…」

「がっちり－やる」の例（会議録2008－09）

- 「…それと営業所の実態など，事前調査を私は本当に**がっちりやら**れたのかどうか，疑問が出てきておりますので…」
- 「…総合型の子育て施策を**がっちりやっ**ていくということが必要だと思うんです。」

「ばっさり－切り捨てる」の例（会議録2008－09）

- 「これは，今まであった手当をいきなり**ばっさり**とすべて**切り捨てる**ものです。」
- 「…区民の意見聴取を形式だけにし，出された意見を**ばっさり切り捨てる**不当性は際立っています。」

(2)　係り受け解析誤りによる係り先動詞の誤抽出

係り受け解析の誤りに起因する係り先動詞の誤抽出が「がたがた－言う」に1例，「ぐずぐず－言う」に2例，「ぱらぱら－見る」に3例みられた．

誤抽出の例（がたがた－言う）

- 「基本的に信頼しており，結局大丈夫だから，そのつどの細かい現象に「**ガタガタ**しなさんな」と言うのです。」

誤抽出の例（ぱらぱら－見る）

- 「…まるで窓の外で**パラパラ**と降っている時雨を部屋の中から静かに見ているような，隠やかな表情だった。」

「がたがた－言う」の例では，「ガタガタ」の直後に方言形があるため解析誤りが生じたものと考えられる．また，「ぱらぱら－見る」の例は，「パラパラ」の係り先は「降る」であるが，これがさらに「見る」にかかるため，抽出処理に誤りが生じたと考えられる．「会議録2008－09」に見られた解析誤りと併せて，係り先動詞の抽出手法の精度向上は今後の課題である．

(3)　正例率100％の「オノマトペ－係り先動詞」

前述の係り受け解析誤りを集計から除外すると，11組中6組において正例率が100％であった（「がっちり－やる」，「がっちり－組む」，「ばっさり－切り捨てる」，「のんびり－暮らす」，「ばたばた－倒れる」，「ぱらぱら－見る）．このうち「がっちり－組む」を除く5組は「会議録2008－09」における正例率も100％であった．正例率の高い係り先動詞は，オノマトペの語義決定に寄与すると考えられる．ただし，「会議録2008－2009」における「がっちり－組む」に負例が存在するように，これらの係り先動詞を用いてオノマトペの語義が異なる文例を作ることは可能である．このため，語義を確定するためには，さらに文中の他の部分を手がかりにする必要がある．

(4)　表記の揺れ

「がたがた－言う」「ぐずぐず－言う」において，「言う」の表記に起因するとみられる負例が存在した．

「がたがた－言う」の負例

- 「サーブのタイヤは，これまで出したこともないスピードに悲鳴を上げ，車体を**ガタガタ言わ**せている。」
- 「今朝は強風であおられて何かが**ガタガタ**って言う音で目が醒めました。」

「ぐずぐず－言う」の負例

- 「元同居人はまたちょっと咳をし，鼻を**ぐずぐず言わ**せた。」
- 「乾杯のシャンペンでさえ　鼻が**ぐずぐず言う**のでこれも苦手です。」

「がたがた－言う」の負例はいずれも，“かたいものがふれ合って発する，重く騒々しい音”を表している．また，「ぐずぐず－言う」の負例はいずれも“風邪などで洟（はな）をすする音”（辞典には記載のない語義）を表している．これらの用例の係り先動詞「言う」は人間による発話行為を指していないため，ひら

がなで「いう」と表記することが一般的である．しかしながら，漢字表記の「言う」が誤謬であるともいえないため，実例の観察に基づく分析においては，表記の揺れに対する処理を検討する必要がある．

(5) 語義分類の判断が難しい例（複数語義の包含）

2.2 節で述べたように，「会議録 2010」における「のんびり－やる」の用例はすべて，辞典の語義“くつろいでいるさま”から派生した“進みが遅いさま”に分類された．「会議録 2008－09」でも同様の語義で使用されていると分類することを確認した．しかしながら，「BCCWJ」では，「…**のんびり**ペースでやっていきたい…」というように進みの遅さを表す用例（14 例）よりも，「**のんびり**パソコンやったり…」というようにくつろいで物事を行う様子を表す用例（21 例）が多く見られた．後者の用例の中には，“物事をくつろいで行うこと”の当然の帰結として“進みが遅くなること”を含意しているものも観察される．そのような文脈の「のんびり」を 2 つの語義のどちらか一方に決定することは困難であり，人間による語義分類においても判断が分かれる可能性がある．「会議録 2008－09」の用例では，行政の施策（他者の行動）について批判的な文脈で「のんびり－やる」が用いられている．このため，くつろぐという意味より遅いという意味が強調されているものと解釈できる．

雰囲気や状況を表現するオノマトペの語義は「語義 A である」と決定できる文脈もあれば，「語義 A，B どちらも含意している」ように解釈できる文脈（複数語義の包含），さらには「包含する複数の語義のうち，語義 A が比較的強調されている」という文脈もあり得る．本論文の語義分類では，包含される語義のうちもっとも妥当性の高いものを 1 つ選択したが，語義を一意に決定することが困難な用例の取り扱い方について今後検討が必要であると考えられる．その上で，包含する複数の語義の傾向について分析を行う必要がある．

「のんびりーやる」の正例（進みが遅いさま）

- 「どちらも少し**のんびり**ペースでやっていきたいです。」
- 「しかし，事件によっては，そんな**ノンビリ**したことをやってはおれず，もっと迅速で簡易な処理がどうしても必要だというものも多い。」
- 「僕としてはもっと**のんびり**やりたかったんだけれどな。」

「のんびりーやる」の負例（くつろいでいるさま）

- 「14 時頃までは，**のんびり**パソコンやったり，ご飯食べたり，テレビみたりそう，植え替えなんて全然してませんでした w」
- 「仕事が終わったあとのお楽しみ，ビールを飲みながらの野球観戦くらい，世界の流行なんてほっといて，**ノンビリ**やりたいものだとは思いませんか？」
- 「**のんびり**と民宿やランチをやりながら，毎日を楽しんでいきたいと思っています」

「会議録 2008－09」における「のんびりーやる」の例

- 「…やはりあんまり**のんびり**やっているうちに子どもは大きくなってしまいますから…」
- 「…これを**のんびり**やっていると万が一のことがあったときに間に合わないことになってしまいますので…」

4. まとめ

本論文では，オノマトペの語義分類に寄与する係り先動詞について検討した．語義分類済みコーパス「会議録 2010」において単一語義で一定頻度出現した「オノマトペ－係り先動詞」の組について，同一ジャンルのコーパスおよび，多様なジャンルの文を含むコーパスにおける用例の抽出と語義の分析を行った．

「会議録 2008－09」の正例率は高く（99.4%），同一ジャンルの文書では，係り先動詞はオノマトペの語義決定に寄与するといえる．「BCCWJ」の正例率は 78.0% であった．分析対象の「オノマトペ－係り先動詞」の組のうち約半数はオノマトペの語義決定に文書のジャンルを問わず寄与することが確認された．用例分析の結果，係り受け解析誤り，表記の揺れ，複数語義の包含が負例の主な要因であることが明らかになった．コーパスの実例の分析に基づくアプローチでは，係り受け解析誤りや表記の揺れがある程度存在することは避けられない．一定以上の出現頻度がある共起語を対象として分析することで，例外的な現象をある程度無視することができると考えられる．ただし，頻度の低い共起語がすべて例外であるとはいえないため，更なる分析，検討を進めていく．また，複数語義の包含については，含意する語義の取り扱い方や語義分類の方法ついて今後検討を行う．

本論文において，多様なジャンルの文を含むコーパスを対象に分析を行った結果，係り先動詞の種類自体に文書のジャンルに起因した偏りがあることが示唆された．このため，今後は分析対象を拡大し，種々の大規模コーパスを対象に，オノマトペと共起する語についての検討を進める．また，次の段階として本論文で取り扱わなかった，文の表層格について検討を進める．

謝辞

本研究の一部は科研費 No. 26370498 による．

参 考 文 献

［1］ 小野正弘編：日本語オノマトペ辞典，小学館，2007.
［2］ 郜楓：コーパスを利用した類義語のコロケーション分析－擬態語「しんみり，しみじみ」と動詞の共起から－，ことばの科学, 19, pp. 129－140, 2006.
［3］ Hironori Fukushima, Kenji Araki, and Yuzu Uchida : Disambiguation of Japanese Onomatopoeias Using Nouns and Verbs, TSD 2014, LNAI 8655, pp. 141－149, 2014.
［4］ 齋藤誠，大城卓，菅原晃平，永井隆広，渋木英潔，木村泰知，森辰則：地方議会会議録の収集とコーパスの構築，言語処理学会第 17 回年次大会論文集，P 2－21, 2011.
［5］ Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den : Balanced corpus of contemporary written Japanese, Language Resources and Evaluation 48(2), pp. 345－371, 2014.
［6］ 工藤拓，松本裕治：チャンキングの段階適用による日本語係り受け解析，情報処理学会論文誌, 43(6), pp 1834－1842, 2002.
［7］ 高丸圭一，内田ゆず，乙武北斗，木村泰知：地方議会会議録コーパスにおけるオノマトペ－出現傾向と語義の分析－，人工知能学会論文誌，30(1)，pp. 306－318, 2015.

（2015 年 11 月 19 日　受付）
（2015 年 12 月 25 日　採録）

［問い合わせ先］
〒320-0811　栃木県宇都宮市大通り 1-3-18
宇都宮共和大学
高丸　圭一
TEL：028-650-6611
E-mail：takamaru@kyowa-u.ac.jp

著 者 紹 介

高丸（たかまる）　圭一（けいいち）［非会員］

2004 年北海道大学大学院工学研究科電子情報工学専攻博士後期課程単位修得退学．同年那須大学都市経済学部専任講師，2013 年宇都宮共和大学シティライフ学部准教授（現在）．修士（工学），博士（応用言語学）．自然言語処理，音声学，社会言語学，コーパス言語学などの研究に従事．電子情報通信学会，人工知能学会，社会言語科学会，日本音声学会，日本語学会各会員．

内田（うちだ）　ゆず［正会員］

2005 年北海道大学工学部情報工学科卒業．2010 年同大大学院博士後期課程修了．博士（情報科学）．同年，青山学院大学理工学部助手．2012 年より同助教．2014 年北海学園大学工学部電子情報工学科准教授，現在に至る．自然言語処理，特に概念獲得システム，日本語オノマトペなどに関する研究に従事．人工知能学会，電子情報通信学会，言語処理学会，各会員．

乙武（おとたけ）　北斗（ほくと）［正会員］

2010 年北海道大学大学院情報科学研究科博士後期課程修了．博士（情報科学）．同年，福岡大学工学部助教着任，現在に至る．自然言語処理，教育工学などの研究に従事．言語処理学会，情報処理学会，人工知能学会，電子情報通信学会各会員．

木村（きむら）　泰知（やすとも）［非会員］

2004 年北海道大学大学院工学研究科電子情報工学専攻博士後期課程修了．博士(工学)．2005 年，小樽商科大学商学部助教授着任．2007 年，同准教授，現在に至る．この間，2010年10月より 2011 年 9 月まで New York 大学客員研究員．自然言語処理，情報抽出などの研究に従事．人工知能学会，言語処理学会，情報処理学会，電子情報通信学会各会員．

An Investigation of Onomatopoeias' Senses Classification Based on Modified Verbs

by

Keiichi TAKAMARU, Yuzu UCHIDA, Hokuto OTOTAKE and Yasutomo KIMURA

Abstract :

One onomatopoeia word has plural senses. A modified verb of an onomatopoeia, which is clue to decide a meaning of onomatopoeia, is investigated in this paper. Modified verbs are extracted from sentences in the "local assembly corpus 2010" (LAC 2010). The sets of "onomatopoeia and modified verb" (O－MV sets) used with high frequency with a single sense in the LAC 2010 are examined by using two other corpuses. One is a same genre corpus, the "local assembly corpus 2008－09" (LAC 2008－09). The other is the corpus which includes various genres of documents, the BCCWJ. 99.4% of O－MV sets in the LAC 2008－09 have a same sense with the LAC 2010. A modified verb contributes to decide the sense of onomatopoeia among certain genre of documents. 78.0% of O－MV sets in the BCCWJ have a same sense with the LAC 2010. Negative examples in the BCCWJ are mainly caused by (i) errors of dependency analysis, (ii) orthographical variants and (iii) implication of plural senses. Onomatopoeia's sense which is decided with modified verb is not depending on the genre of document. However, there are some biases in occurrence of modified verb among the genre of documents.

Keywords : onomatopoeia, corpus, word sense, modified verb

Contact Address : **Keiichi TAKAMARU**
Utsunomiya Kyowa University
1-3-18, Odori, Utsunomiya 3200811, Tochigi, Japan
TEL : +81-28-650-6611
E-mail : takamaru@kyowa-u.ac.jp