(特集論文) 「オノマトペの利活用」

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

Onomatopoeias in the Corpus of Japanese Regional Assembly Minutes

Analysis of the Appearance Tendency and the Word Sense

高丸 圭一
Keiichi Takamaru宇都宮共和大学
Utsunomiya Kyowa University
takamaru@kyowa-u.ac.jp内田 ゆず
Yuzu Uchida北海学園大学
Hokkai-gakuen University
yuzu@eli.hokkai-s-u.ac.jp乙武 北斗
Hokuto Ototake福岡大学
Fukuoka University
ototake@fukuoka-u.ac.jp木村 泰知
Yasutomo Kimura小樽商科大学
Otaru University of Commerce
kimura@res.otaru-uc.ac.jp**keywords:** onomatopoeia, spoken language, large scale corpus, regional assembly minutes, word-sense analysis

Summary

An onomatopoeia is a useful linguistic expression to describe sounds, conditions, degrees and so on. It is said Japanese is rich in onomatopoeic expressions. They are frequently used in daily conversations. The meaning and surface structure of an onomatopoeia varies diachronically. There seem to be regional variations in usage of onomatopoeias. It is necessary to investigate the actual condition of onomatopoeia quantitatively in order to apply onomatopoeias into artificial intelligence. This paper studies practical usages of onomatopoeias in spoken modern Japanese language. To explore Japanese onomatopoeias nowadays, we investigate regional assembly minutes collected from all areas in Japan. The corpus of regional assembly minutes, which has about 300 million words, is the target of the investigation of this study. The minutes of Japanese regional assemblies contain all transcriptions of the utterances in the assemblies. This corpus is suitable for our research since attributes of the speakers are clear and speakers are distributed nation-wide. The first research is about total frequency and regional distribution of onomatopoeias. The onomatopoeias, which represent a request for a promotion of policy, *e.g.*, “shikkari”, “dondon”, are used at high frequency in regional assemblies. There are no remarkable regional differences in frequencies of these onomatopoeias though western Japan has slight higher frequency. The second research is about the meaning of the onomatopoeias. Most of onomatopoeias are polysemous. The meaning of the onomatopoeia differs by context. The authors have manually checked through 10,827 sentences, which contain 153 kinds of onomatopoeia, and then classified the meaning of each onomatopoeic expression. We analyzed for the following subjects: i) ambiguity of onomatopoeic expression, ii) regional differences in meaning, iii) new meanings in modern spoken language, iv) special usage in assemblies, and v) onomatopoeias in the named entities. The third research is about false extraction of onomatopoeias in the morphological analysis. The extraction errors are analyzed from the viewpoint of surface structure and appearance position. In terms of surface structure, it is clear that the word length of an onomatopoeic expression, which has highly false extraction, is shorter. The onomatopoeic expressions, which end with special morae, namely moraic obstruent, moraic nasal and long vowel, have a higher rate of false extraction. In terms of appearance position, dialectal grammar is the main factor causing false extraction. About 25% of false extraction is found in the sentence-closing particles in dialectal grammar. The result of quantitative analysis of the onomatopoeia in modern spoken Japanese language serves as the basic data which contributes to engineering. The results of the analysis in our research are exhibited through the WWW. It is hoped that results will contribute broadly to the practical use of onomatopoeia in the engineering field.

1. は じ め に

オノマトペ(擬音語および擬態語)は音, 雰囲気, 程度, 様子を効果的に伝える手段であり, 日本語の話しことばでは多用されることが知られている。近年, オノマトペ

の工学的な利活用を目指した取り組みが盛んである。日本語学や音声学の分野では, 通時的な変遷, 語義, 表層的な形態, 音響的特徴との関係などが研究されている(例えば[斎藤 52, 宮地 78, 那須 07])。また, 方言オノマトペ

の収集も行われており、医療現場等におけるオノマトペの地域差について研究されている[方言 92, 竹田 12]。これらの言語学的に積み重ねられた知見は重要なものであるが、定性的な指摘に留まっているものがほとんどであり、直接工学的に活用することは困難である。

本研究は現代の話しことばにおけるオノマトペの諸相を定量的に明らかにすることを目的とする。このために、大規模コーパスからオノマトペの用例を抽出し、出現頻度や語義の分類について具体的な分析を行う。分析の対象は地方議会議録コーパスである。地方議会議録は都道府県議会または市区町村議会における議員や首長、行政職員などの発言を書き記したものである。これは日本全国に居住する者の公的な場におけるやや改まった話しことばの記録といえる。発言者の属性(年齢・性別・肩書きなど)が明らかで、かつ、特定の自治体に居住する者の発言が、地域別・年度別に記録されている。[竹安 04]における地方議会議員に対するアンケート(回答総数 17,062)によると、地方議会議員の 75.5 % の出生地は現在居住する自治体であり、12.1 % は同一都道府県内の他の市町村である。このことから、地方議会における発言者の大多数はその土地の方言話者であり、自治体別に公開された会議録はそれぞれの地域の方言を含んだ資料として取り扱うことができると考える。地方議会議録は自然言語処理、言語学、政治学等の様々な分野で利用すべき研究資源である。例えば、政治学[増田 12]、環境学[上田 12]、図書館情報学[松本 08]などの分野で、会議録を利用して地方議会の動向を探るいくつかの研究が行われている。また、言語学的観点から地方議会議録における地域差[高丸 13]や方言語彙[井上 13]についても分析が進められている。ただし会議録は自治体ごとに個別に提供されているため、横断的な研究は容易ではない。そこで近年、地方自治体がウェブに公開している地方議会議録を収集・整形し、関係データベースに登録することにより、コーパスとして学際的に利用することを目指した研究が進められている[木村 12]。現代の全国各地の話しことばを文字化した資料で、このように大規模なものは他に存在しないため、地方議会議録コーパスは本研究の目的に最も適したものであると考えられる。

本研究では、全国の地方自治体から収集した地方議会議録コーパスのうち 1 年度分(約 3 億語)を形態素解析し、オノマトペの抽出を試みる。地方議会議録にどのようなオノマトペが出現するのかということ自体が新たな知見となる。そこでまず、抽出したオノマトペについて、出現頻度と出現地域の傾向について分析することで、地方議会議録におけるオノマトペについて概観する。

次に、話しことばにおけるオノマトペの諸相を明らかにするために、抽出された会議録中のオノマトペを人手によって具体的に分析する。オノマトペの持つ意味や印象を定量的に表す研究[清水 14, 土屋 12]や、多義的なオノマトペを機械学習によって分類する研究[福島 14]が

進められている。オノマトペの利活用のためには、現代のオノマトペがどのような語義で使用されているかを明らかにする必要がある。本研究では、多義性の解消に寄与するために、会議録コーパスの実例におけるオノマトペの語義を分類し、語義の多様性や地域差の実態について分析する。

また、くだけたスタイルの文に対する形態素解析を高い精度で行うことは困難であることが知られている。その上、オノマトペは文字長の短いひらがな列、またはカタカナ列であり、品詞が一定ではない。このことから、単に形態素解析だけを用いて話しことばからオノマトペを正確に抽出することは難しい。しかしながら、オノマトペを工学的に利活用するためには、くだけたスタイルの文からオノマトペを適切に抜き出す必要がある。例えば、[内田 12]では、ブログ記事を対象としたオノマトペ抽出手法が提案されている。本研究では、オノマトペの自動抽出への寄与を指向して、形態素解析によって抽出されたオノマトペのうち、実際にはオノマトペではなかったもの(これを「非オノマトペ」と呼ぶ)の出現傾向について分析する。

本論文は以下のように構成される。2 章では、地方議会議録コーパスの概要、および、分析対象のオノマトペについて述べる。3 章では、地方議会議録に対する形態素解析によるオノマトペの抽出方法と、抽出結果の概要について述べる。4 章では、人手による分析作業について述べる。5 章では、会議録中のオノマトペを含む発言における用法や語義について分析する。6 章では、非オノマトペの表層構造と出現位置の特徴について分析する。

2. 研究の対象

2.1 地方議会議録

[齋藤 11]において、全国約 600 自治体の地方議会議録が機械的に収集され、コーパスの構築が進められている。収集した会議録の発言文字列は[菅原 12]の自動整形手法により、文単位に分割され、自治体名、開催年月日、発言者名などの情報を付与した上で、リレーショナルデータベースに登録されている。このコーパスの中で整形済みの文書数がもっとも多い、2010 年度の会議録を研究対象とする。2010 年度のコーパスには、すべての都道府県を網羅した 403 自治体(19 道県, 322 市, 13 特別区, 41 町, 8 村)の地方議会議録が収録されており、データ数は 10,848,883 文(約 3 億語)である。収集自治体数は都道府県ごとに開きがあり、最大で 31 自治体(埼玉県)905,744 文、最小で 1 自治体(青森県・奈良県)それぞれ 27,142 文、25,194 文である。

議会議録は議会における発言をすべて記録することを目的に作成される。しかし、読みやすさの観点から言い直しや単純な言い間違い、助詞の誤用、冗長な言い回しなど、そのまま文字化すると読みづらい表現は修正(整

表 1 オノマトベ辞典における意味分類の概要と語数

	大分類			総数
	自然	人間	事物	
中分類	天気 温度 水・液体 火・土	動作・状態 感情・感覚 性格・性質 体格・姿	音・道具・金銭 形・状態 程度 動き・変化	12
延べ語数	261	939	1,266	2,466
異なり語数	228	801	1,050	1,751

文)される。[高丸 11]によると、議会における実際の発言と会議録とのレーベンシュタイン距離(整文によって削除、追加、置換される文字数)は、会議録 1,000 文字あたり約 60 文字である。議会会議録から話しことばの特徴を分析する際には整文の過程での修正に注意が必要となる。しかし、整文の指針[野村 96]では、オノマトベは整文によって削除や修正される対象ではないため、本研究に対する整文の影響は少ないと考える。

2.2 オノマトベ

『日本語オノマトベ辞典』[小野 07]に収録された語の中から、意味分類別索引に掲載されている 2,466 語を対象とする。意味分類別索引ではオノマトベを 3 種類の大分類、12 種類の中分類、70 種類の小分類の 3 層に分けている。例えば、目的もなく歩き回るさまを表す「うろうろ」は大分類「人間」 中分類「動作・状態」に関するオノマトベ 小分類「歩く・走る」に分類される。複数の意味を持つオノマトベは重複して掲載されているため、異なり語数は 1,751 語である。上位 2 層の分類名と収録されたオノマトベの語数を表 1 に示す。

3. オノマトベの抽出

3.1 方法

オノマトベの抽出のために、2.1 節で述べた地方議会会議録に対して形態素解析を行った。形態素解析には日本語形態素解析システム JUMAN^{*1}(Ver.7.0)を用いた。オノマトベ用辞書として、JUMAN のユーザ形態素辞書に 2.2 節で述べた 1,751 語のひらがな表記とカタカナ表記をすべて登録したものを用いた^{*2}。形態素解析結果から、オノマトベ用辞書中の語が含まれている文をすべて抜き出した。

都道府県ごとにコーパスに収録された発言数が異なるため、抽出したオノマトベの出現頻度そのものを都道府県間で直接比較することは適当ではない。そこで、オノマトベの出現状況を定量的に比較する際には、出現確率を使用する。オノマトベ w の都道府県 g における出現確率 $P_{g,w}$ は、発言の総単語数 N_g 、オノマトベの出現頻度

表 2 全会議録における総出現数上位 15 語

オノマトベ	大分類	中分類	出現頻度
しっかり	事物	程度	77,464
どんだん	事物	動き・変化	20,680
はっきり	人間	動作・状態	19,382
だんだん	事物	音・道具・金銭	5,679
かん	事物	音・道具・金銭	5,171
びっくり	人間	感情・感覚	2,910
つい	事物	形・状態	2,728
もろもろ	事物	程度	2,372
そろそろ	事物	程度	2,021
だっ	事物	音・道具・金銭	1,984
たん	事物	音・道具・金銭	1,769
ゆっくり	事物	程度	1,610
じっくり	事物	程度	1,541
どん	事物	音・道具・金銭	1,393
かつ	自然 人間 人間 人間	天気 温度 動作・状態 感情・感覚	1,130

$C_{g,w}$ を用いて次式によって求める。

$$P_{g,w} = \frac{C_{g,w}}{N_g}$$

3.2 全体の傾向

形態素解析の結果、10,848,883 文の地方議会会議録コーパスから 186,416 例(約 1.7%)のオノマトベを含む文が抽出された。出現するオノマトベの異なり語数は 982 語(分析対象語の 56.1%)であった。出現頻度の高いオノマトベ上位 15 語を表 2 に示す。「しっかり」の出現頻度がもっとも高く、抽出したオノマトベ全体の約 42%を占める。地方議会会議録には「しっかり」「どんだん」などの政策の推進を求める際に使用されると考えられる表現や「はっきり」といった厳格な判断や言及を求める表現「そろそろ」「ゆっくり」「じっくり」といった議論や政策の時期や進行速度に関連するとみられるオノマトベなどが高頻度で出現することがわかる。[内田 13]では、Yahoo! ショッピングの商品レビューに出現する大分類「事物」のオノマトベ上位 10 語は「ぴったり」「ふわふわ」「さらさら」「きらきら」「さっぱり」「さらっ」「ぎりぎり」「ふんわり」「つい」「ふかふか」とであると述べられている。地方議会会議録における「事物」のオノマトベ上位 10 語と共通する語は 1 語のみ(「つい」)である。文書によって修飾する対象が異なるため、高頻度で使われるオノマトベには違いが見られる。しかし、地方議会という公的な場における話しことばにおいても、オノマトベは豊富に出現することが確認された。

[平田 13]によると、国会会議録における分析では「A っ B り型」のオノマトベが関西エリアに高頻度で出現すると述べている。地方議会会議録における出現頻度上位 3 語「しっかり」「どんだん」「はっきり」について、地方別の出現確率を図 1 に示す。各地方分類に含まれる自治

*1 <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

*2 ユーザ形態素辞書登録において、品詞は「副詞」としたが、名詞的用法等も抽出することができる。

図1 「しっかり」「どんどん」「はっきり」の出現確率（地方別）

体数を考慮して、「北海道・東北」「関東」「中部」「近畿」「中国・四国」「九州・沖縄」の6地方分類を用いる。

それぞれの語について、R^{*3}の prop.test 関数を用いて、地方間の比率の差の検定を総当たりで行った。この結果、「しっかり」では「中部」と「近畿」の間で $p = 0.243$ であり、有意差が認められなかった。また、「どんどん」では「北海道・東北」と「関東」の間で $p = 0.1384$ 、「北海道・東北」と「九州・沖縄」の間で $p = 0.07812$ 、「関東」と「九州・沖縄」間で $p = 0.4779$ であり、これらの組み合わせには有意差が認められなかった。「はっきり」では、「中国・四国」と「九州・沖縄」の間で $p = 0.7873$ となり、有意差が認められなかった。これ以外のすべての組み合わせでは $p < 0.05$ であり、それぞれのオノマトペの出現確率に有意差が認められた。東日本に比べて西日本においてオノマトペがやや多く出現する傾向を読み取ることができる。しかし、出現確率の順位は語ごとに異なり、これらのオノマトペが特定の地方で顕著に多く出現するという特徴は見られない。

4. オノマトペを含む文の分析

4.1 方法

次に、会議録に見られるオノマトペの語義の多様性や地域差について述べる。出現確率がきわめて高いオノマトペは議会発言中の多様な文脈で使用することができるため、どの地域においても高い頻度で一様に出現すると考えられる。一方、出現確率がきわめて低いオノマトペは事例が少なく、個人性に大きく影響されるため出現分布を分析することに適していない。これらのことから形態素解析によって抽出されたオノマトペのうち出現確率が中程度の語を対象とする。具体的には、都道府県別の出現確率の和が 50×10^{-7} 以上 500×10^{-7} 未満の 177 語である（付録 A 参照）。

この 177 語のオノマトペが会議録にどのように出現しているのかを具体的に確認する。このため、形態素解析によって抽出されたオノマトペを含む文（計 18,545 文）を

表3 分類の基準と該当する用例数

オノマトペか？	分類基準	出現頻度
オノマトペである	① オノマトペとして発言	10,827
	② 固有表現に利用	1,380
オノマトペではない	③ 方言	1,522
	④ 固有表現の一部	1,066
	⑤ 同音異義語	523
	⑥ オノマトペの一部	238
	⑦ 言い間違い・入力ミス	69
	⑧ その他	2,920

すべて人手で読み、共著者相互の合意により表3に示す8種類に分類した。最初の分類基準は形態素解析結果が実際にオノマトペであったか否かである。「オノマトペである」と判断したものは、「①オノマトペとして発言」されているものと、「②固有表現に利用」されているもの^{*4}に細分類した。さらにオノマトペとして発言されているものを対象に語義の分類を行った。

「オノマトペではない」（非オノマトペ）と判断したものについてはどのような文脈に表れたかを確認した。出現位置に基づいて「③方言」「④固有表現の一部」「⑤同音異義語」「⑥他のオノマトペの一部」「⑦言い間違い・入力ミス」「⑧その他」に細分類した。

4.2 オノマトペか否か

人手による分析の結果、形態素解析によって抽出したオノマトペのうち、実際にオノマトペであったものは 12,207 例 (65.8%)、オノマトペでなかったものは 6,338 例 (34.2%) であった。

5章に、オノマトペとして発言された（表3①）10,827 例（153 語）の語義の分析、および、固有表現としてオノマトペが使用された（表3②）1,380 例（34 語）の具体例について述べる。オノマトペではないと分類された用例については6章に述べる。

5. 会議録にみられるオノマトペ

人手によってオノマトペであると確認された 10,827 例（153 語）のオノマトペについて、語義の分類を行った。語義分類は原則的に『オノマトペ辞典』に従う。『オノマトペ辞典』の語義にあてはまらない用例のうち、語義が解釈可能なものは当該オノマトペの語義として新たな項目を立てて分類した。また、オノマトペが表す語義が解釈ができない場合には「語義不明」と分類した。5.1 節では、オノマトペの多義性について述べる。さらに、語義の地域差（5.2 節）、辞典に記載されていない語義（5.3 節）、議会における特徴的な語義（5.4 節）、固有表現中のオノマトペ（5.5 節）についてそれぞれ具体的に述べる。

*3 www.r-project.org/

図2 オノマトペの語義分類

表4 会議録における「ごろごろ」の語義

語義分類	用例数
①雷の響く音	1
②猫がのどを鳴らす声	0
③丸いものなどが転がる音	5
④ものが無造作に転がっているさま	24
⑤仕事をせず無駄に暮らしているさま	3
⑥間に入って違和感のあるさま	0
たくさんある(いる)さま	17
変わりゆくさま	13
グラウンドをならす音	1
語義不明	2

5.1 多義性

オノマトペは多義的に用いられることが知られている。本分析の範囲では、153語のオノマトペが348の語義に分類された。1つのオノマトペの持つ語義分類数の分布を図2に示す。語義を「オノマトペ辞典内の語義」「辞典に記載された語義から派生したと考えられる語義」「辞典に記載されていないの語義(または語義不明)」の3種類に分けた。図2には、それらの割合を合わせて示す。

153語のうち32.0%にあたる49語は語義が単一のオノマトペであった(例えば「ぼん」(271例)、「すんなり」(246例)、「ぼっ」(234例)、「ずらっ」(209例))。一方、68.0%にあたる104語は複数の語義を持つものであった。このうち、5種類以上の語義に分類されたものが9語見られた。

もっとも語義の分類数が最も多かった「ごろごろ」の分類結果を表4に示す。「ごろごろ」は、会議録中の64の用例が7種類の語義に分類された。さらに「語義不明」と分類した用例が2例あった。表4のうち①から⑥は『オノマトペ辞典』に掲載されているものである。語義②と語義⑥に該当する用例は出現しなかった。一方、辞典にない語義が3項目追加された。このうち17例出現する「たくさんある(いる)さま」は、語義④から派生した語義であると考えられるが、例えば「一千万円超える職員がごろごろしている。(徳島県)」などの用例を語義④に分類することは適当ではないため新たな語義の追加して分類した「変わりゆくさま」(13例)は語義③から派生して、担当者や制度等が変わるさまを表すものである。例

えば「…担当者がごろごろかわるとその3%減額の制度があるよということが引き継がれない…」(奈良県奈良市)^{*5}のように用いられる。この語義は8都府県に出現する。関東においても使用されるが西日本において出現例がやや多い^{*6}。

語義「グラウンドをならす音」に分類されたものは以下の用例である。本研究の分類作業では、辞典にない語義として項目を追加したが、この音を「ごろごろ」と表すことの一般性、共通性は本研究の範囲からは明らかではない^{*7}。

- 「吉田中学校におかれましては、河川に近い状態のところをグラウンドにしておりますために、例えばトンボでごろごろなると石が出てきて、非常に危ない環境が見受けられました。」(埼玉県秩父市)

また「語義不明」に分類したものは以下の2例である。

- 「それに引きかえて、私のは原稿もございませんし、またしゃべりがごつごつ、ごろごろしてまして、質問の大部分がお金の要る話ばかりでありまして…」(和歌山県和歌山市)
- 「ただ、その中に、じゃみんなも言ったことだけ、ごろごろまとめたのかと、そういうことでもなく、私は私でこういうふうになりたいというものを持ってお示しをさせていただいたものについて…」(福岡県苅田町)

1例目では「ごろごろ」はしゃべる様子を表し、2例目ではまとめ方々を表していると考えられる。しかし、これらの「ごろごろ」がどのような様子表現しているかを解釈できず、語義が記述できないと判断し「語義不明」と分類した。

話しことばにおけるオノマトペの利活用において、多義性に対処することは重要である。オノマトペは、辞典に記載された語義にかかわらず、音象徴を利用して雰囲気伝えるという使い方が可能である。このため、上述の「ごろごろ」の語義分類において「語義不明」に分類したような、個性が高く共通認識を持ちづらい用例も存在する。しかし、図2から読み取ることができるように、実際のオノマトペの用例においては、辞典に記載された語義が多く使用される(73.6%(256語義))。辞典の語義からは派生した語義を含めると86.5%(301語義)であ

*5 用例の引用における「…」の表記は、紙幅の都合から文の前方または後方を省略したものである。また、オノマトペはゴシック体で表記した。これを除く用例内の表記はすべて原文のままである。

*6 東京都(1例)、神奈川県(1例)、長野県(1例)、京都府(1例)、大阪府(2例)、兵庫県(3例)、奈良県(1例)、福岡県(3例)

*7 この文におけるグラウンドをならす作業が、仮にローラー等を用いたものであれば語義③の「丸いものなどが転がる音」と解釈することができる。また、話しことばにおける語順の逆転が起きており、この「ごろごろ」は次の句「石が出てきて」にかかっていると解釈することもできる。その場合、語義④にあてはまる。しかし、もっとも単純に読むと、グラウンドをならす作業の擬音語として「ごろごろ」を用いていると解釈できる。

*4 「組織の名称」「イベントの名称」「施策の名称」等に取り入れられたオノマトペを指す

表5 出現頻度に地域差が見られる語義 (丸付き数字はオノマトペ辞典に対応)

オノマトペ	語義	出現 頻度	変動 係数	北海道 東北	関東	中部	近畿	中国 四国	九州 沖縄
びしゃっ	きちんと、完全に、明確に	99	2.0	0	0	3	1	5	90
ばくっ	漠然とした、大雑把なさま	68	1.6	1	3	9	52	2	1
かちっ	②ちょうどよい具合にかみ合うさま	80	1.3	4	10	2	50	13	1
めちゃくちゃ	②どうにもならないほどひどい状態になるさま	64	1.2	1	7	9	38	6	3
のんびり	進みが遅いさま	137	1.1	8	80	16	20	8	5
すかっ	①遮るものがなく快いさま	37	1.1	1	2	3	21	2	8
ころっ	⑥急に、まったく変わってしまうさま	94	1.1	8	9	4	54	11	8
がっちり	②抜け目がないさま	64	1.1	27	27	3	5	1	1
ぴちっ	きちんと、明確に	53	1.0	0	9	1	11	27	5

る。このことから、例えば、オノマトペの語義曖昧性解消手法を検討する際には、まず、辞典に記載された語義への分類を検討することが重要であることが示唆される。

本章で行った人手によるオノマトペの語義分類をさらに進め、より多くの具体例を蓄積することで、一般性・共通性を持つ語義への対処が可能になると期待される。辞典に記載されていない語義については、5.3 節で述べる。

5.2 語義の地域差

本節では、6 地方分類によって語義出現の地域差を分析する。まず、語義ごとに地方別の出現頻度の平均と標準偏差を求めた。とりうる値に大きな幅のある語義の出現頻度の標準偏差を直接比較することは適切ではないため、次に変動係数(=標準偏差/平均)を求めた。会議録コーパスに複数の語義が出現するオノマトペのうち、一定の出現頻度があり、かつ、出現地域に偏りがある語義として、全国出現頻度が30回以上で変動係数が1.0以上の語義を抜き出した。これを表5に示す。

変動係数の高い語義は西日本に多い傾向が観察される。「びしゃっ」は九州・沖縄地方において、「ばくっ」「かちっ」「めちゃくちゃ」「すかっ」「ころっ」は近畿地方において、「ぴちっ」は中国・四国地方において、それぞれ表に示した語義が高頻度で出現した。一方、「のんびり」は関東地方で、「がっちり」は、北海道・東北地方および関東地方において、それぞれ表に示した語義が高頻度で出現した。

これらの語義は「びしゃっ」と「ばくっ」を除くと、辞典に記載された語義、または、そこから派生した語義であり、全国的に使用可能な語義であるものの、出現地域には偏りが見られた。このような語義の出現頻度の偏りにおいては、同様の語義を持つ他のオノマトペと相補分布をなす類似した語義を持つ複数のオノマトペの中で、地域によって優先的に使用されるオノマトペが異なる可能性が考えられる。この点に付いては、個別の語義について、それぞれさらに詳しく検討する必要がある。

辞典に記載されていない語義で出現に地域差がみられた「びしゃっ」について、以下に具体的に述べる。

「びしゃっ」は17県に112例出現し、3種類の語義に分類された。戸口などを閉めるさまから派生した「完全に閉める(やめる、止める)さま」を表す以下のような用例は12例出現する。関東地方から中国四国地方までの8

都県に分布する。

- 「… 市長はお金がないというたった一言でびしゃっ」と切ってしまったんです。」(和歌山県和歌山市)
- 「逆にきょう呼んでしっかり話をして、それでびしゃっ」とやめるとなれば…」(長野県松本市)
- 「あの地域内に水が入るのをびしゃっ」と入り口でとめていただいたということで、大変対応がよかった」(山口県山陽小野田市)

一方、「きちんと、完全に、明確に」に相当する語義で用いられるものが99例みられた。このうちの90例が九州地方の7県に出現しており、九州地方以外で観察されたものは4県9例のみであった。この語義は『オノマトペ辞典』への掲載はなく、出現地域に顕著な偏りがみられることから、方言オノマトペであると考えられる。

- 「… そういうものについてびしゃっ」と明らかにして示す必要があるんじゃないかというふうに思うんですが…」(福岡県嘉麻市)
- 「そりゃ、職員は黒字が出ようが赤字が出ようがびしゃっ」とボーナスも出ておる。」(長崎県雲仙市)
- 「こんな問題は、びしゃっとした方がおれはいいんじゃないねえかと思うんだけど…」(大分県)

このほか「平らにつぶれるさま」(「うちがびしゃっ」となくなってしまうみたい、あれは建物が欠陥だと思うのです。))が東京都豊島区に1例みられた。

5.3 辞典に記載されていない語義

「辞典に記載されていない語義」と分類された語義は42語に47種類存在した(「語義不明」としたものの26種類を含む)。このうち用例数が1例の語義が21種類、2例の語義が16種類、3例の語義が3種類存在した。辞典に記載されていない語義は出現頻度が低いものが多く、これらは個性的な表現に基づく用法である可能性が高い。辞典外の語義のうち頻度が5以上のものを表6に示す。

辞典外の語義には、前節で述べた共通語にない語義(方言)のほかに、言語使用の変化や生活様式の発展に伴う新たな語義が存在した。以下に言語使用の変化による新たな語義の例である「ばくっ」および、生活様式の発展に伴う語義の例である「ぴっ」についてそれぞれ具体的

表 6 出現頻度が 5 以上の辞典外の語義

オノマトペ	語義	頻度
びしゃっ	きちんと、完全に、明確に	99
ばくっ	漠然、大雑把	68
ふわっ	曖昧、大ざっぱ、簡単	26
すとな	ちょうど、ぴったり、順調に、すんなり	25
ぼっ	突然現れるさま	17
ふわふわ	曖昧なさま	9
びっ	ボタン等を押すとなる電子音	8

に述べる。

§1 漠然とした「ばくっ」

会議録に 69 例出現する「ばくっ」は「裂け目が突然大きく開くさま」で用いられた岩手県紫波町の 1 例（「…何か重たいものが通ると、すぐそれがばくっと折れてしまう性格のものです…」）のほかは、すべて以下のような「漠然とした、大ざっぱな」の意味で用いられた。全国 12 都府県で出現する語義であるが、兵庫県 26 例、大阪府 17 例と近畿地方に多い。

- 「… 議会としてのばくっとした考え方なりを聞きたいというそういった思いで…」(愛知県尾張旭市)
- 「… この請願の項目については、割にばくっ書いてあって…」(兵庫県宝塚市)
- 「… ちょっとばくっとした聞き方になりますが、改めて代表委員としてのご見解を…」(熊本県合志市)

会議録では「漠とした」という表現も使用される。「ばくっ」は「漠」が促音を伴ってオノマトペ化したものであると考えられる。また、この語には「大きな口を開けて勢いよく食いつくさま」という語義もあり、「大枠で概略を捉える」という印象を付与しやすいものと考えられる。

§2 電子音を表す「びっ」

「びっ」は 35 例がオノマトペとして会議録に出現した。内訳は「確実に行うさま」が 16 例、「瞬時に行われるすばやいさま」が 9 例、「電子音」が 8 例、「布などを裂く音」が 1 例、「笛の音」が 1 例であった。このうち「電子音」は電子音自体がなかった時代には存在しない用例であり、生活様態の変化とともに生じた新しい語義といえる。以下に用例を示す。

- 「… たまにバスに乗ると、PASMOとか持っても、最初にどこにびっとやったらいいのかわからなかったり…」(東京都町田市)
- 「今、携帯なんかも普及してますので、QRコードでびっとやったら…」(大阪府河内長野市)

5.4 議会発言における語義の偏り

複数の語義を持つオノマトペにおいて、議会に特有であると考えられる語義の偏りが確認された。

比較のために現代日本語書き言葉均衡コーパス「少納言」*8 における出現頻度を示す。現代日本語書き言葉均

表 7 会議録における語義の偏りの例

	語義	会議録	少納言
ふ わ っ	曖昧・大ざっぱ・簡単	26	1
	軽く動く・浮く	5	94
	優しい・柔らかい	3	84
	その他	2	4
	総数	36	183
と ん と ん	順調にはかどる	73	10
	2 者が同じくらい	28	9
	軽くたく	6	41
	総数	107	60
が た が た	固いものが触れ合う音	24	41
	勢いよく行われるさま	34	7
	具合が悪い状態に陥る	39	37
	騒々しく不平を言う	35	9
	激しく震えるさま	0	39
	その他	0	2
	総数	132	95
が く が く	侃々(喧々) 諤々	135	1
	震え続けるさま	1	74
	その他	1	4
	総数	137	79

衡コーパス」は偏りのないサンプルを提供することを目標として書籍、新聞、ウェブなどから抽出したコーパスである。このため、語義の分布の比較に適していると考えられる。少納言は「現代日本語書き言葉均衡コーパス」を全文検索できるウェブサービスである。本節では、複数の語義を持つオノマトペを少納言で検索し、この検索結果に対して会議録の分析と同様の方法で人手によって語義分類した。

例えば、会議録に 149 例出現する「さらっ」は以下の例のような「人の態度・性格ややり方に気負いやくどさがないさま」として 148 例出現するが、「ものに湿りや粘りけがなくて、かわいた手触りがするさま」が 1 例のみである。

- 「それで、ちょっとその辺について率直にどのように認識しているのか、再度さらっとでよいですので、お願いします。」(岩手県滝沢村)
- 「前半は大変長かったということで、2 問目はさらっと流していきたいと思います。」(和歌山県和歌山市)
- 「登壇時は主題だけをさらっと申し上げて、詳細は以降発言席で行いたいと思います。」(熊本県八代市)

少納言には「さらっ」が 106 例出現し、前者が 62 例、後者が 44 例である。

このほか、表 7 に示すように、議会の場で使用されるオノマトペと書きことばには、語義の偏りが見られる。会議録には音や物質の具体的な状態を表す語義が少なく、程度や抽象的な状態を表す語義が多い傾向が見られる。

5.5 固有表現中のオノマトペ

分析対象のうち、1,380 例(異なり語数 34 語)のオノマトペは固有表現の中で使用されているものであった。これらは、それぞれの地域で独自に命名した取り組み、イベント、施設等の名称であるため、出現地域は限定され

*8 <http://www.kotonoha.gr.jp/shonagon/>

表 8 固有表現中のオノマトペ

	用例種類	用例数
にこにこ	46	105
わいわい	44	170
のびのび	36	246
きらきら	28	145
きらり	28	168
ぶらり	28	38
きらら	26	109
ほのぼの	24	47
さんさん	15	97
ぐんぐん	8	28

る。固有表現のバリエーションの多い10語を表8に示す。

「うれしそうに笑みを浮かべるさま」を表す「にこにこ」はバリエーションが最も多く、「にこにこすくーる」(東京都荒川区・18例)、「にこにこ保育支援」(大分県日田市・13例)、「にこにこ市場」(大阪府富田林市・6例)といった使用例がみられた。「大勢の人が一緒になって賑やかに騒ぐさま」を表す「わいわい」は「わいわい市」(神奈川県藤沢市・40例)、「わいわいスクール」(神奈川県横須賀市・17例)、「わいわいルーム入間」(埼玉県狭山市・11例)、「わいわい談義」(岐阜県関市・11例)などの使用例がみられた。このほか「子育ての駅ぐんぐん」(施設名称/新潟県長岡市)、「ぶんぶんテレビ」(第三セクターのケーブルテレビ局名/佐賀県佐賀市)、「ぶらぶら歩きがここちよいまち」(まちづくりの理念/東京都昭島市)、「ふわふわドーム」(公園遊具の名称/埼玉県北本市)というようにオノマトペの持つ意味を利用して、直感的に理解しやすい名称を付与している例が観察される。また、「きらり」「きらら」「きらきら」「さんさん」などは「子ども総合支援センターきらり」(北海道北見市・38例)、「海きらら」(長崎県佐世保市・17例)、「次世代育成きらきらプラン」(神奈川県鎌倉市・40例)、「塚口さんさんタウン」(兵庫県尼崎市・43例)のように用いられる。これらは語義を直接利用しているのではなく、そのオノマトペが表す明るい印象、輝く印象を利用している。

6. 「オノマトペではない」と分類されたもの

分析対象の177語の範囲では、6,338例(34.2%)が形態素解析によって抽出されたものの、オノマトペではない文脈で出現した。オノマトペは短いひらがなまたはカタカナの文字列であるため、オノマトペではない部分文字列とマッチしやすい。オノマトペの工学的活用を目指すためには、オノマトペとマッチする部分文字列のうち、実際にはオノマトペではないもの(非オノマトペ)がどのように出現するかを検討する必要がある。本章では、形態素解析によって抽出されたもののうち人手によって「オノマトペではない」と分類した用例を対象に、表層的な構造(6.1節)と出現位置(6.2節)の観点から分析を行う。

6.1 表層構造

形態素解析によるオノマトペの抽出結果において、すべてが非オノマトペだった語(表9「オノ0」)が22語、一部が非オノマトペであった語(表9「中間」)が56語、すべてがオノマトペであった語(表9「オノ100」)が99語存在した。この3分類について、拍数、最終拍の文字、表層パタン(繰り返し等)の観点から比較する。これを表9に示す。

すべてがオノマトペのグループの平均拍数は3.78で、すべてが非オノマトペのグループの平均拍数は2.65であった。すべてが非オノマトペのグループの拍数は約1拍短い。短い文字列の方がさまざまな文脈で部分文字列としてマッチしやすいため、これは妥当な結果といえる。4拍のオノマトペにおける代表的な表層構造についてみると、ABAB型(「がんがん」など)は、オノ100の割合が高い一方、オノマトペではない副詞の畳語と一致する(表3の⑤同音異義語に分類される)語が見られた。このうち、「なんなん」「おいおい」の2語は、オノマトペとしては会議録に出現せず、すべて非オノマトペ(それぞれ「垂んとす」「追々」のひらがな表記)であった。ABCB型(「どたばた」など)、「A→B」型(「しっかり」など)、「A→B」型(「すんなり」など)ではオノ0が存在しなかった。これらの表層構造の語は他の部分文字列にマッチしづらく、オノマトペ特有のものであると考えることができる。また、最終拍の種類による分類では、最終拍が特殊拍(撥音、促音、長音)のオノマトペは最終拍が非特殊拍のものとは比べて、非オノマトペの割合が高い。本分析の範囲では「非特殊拍」<「撥音」<「促音」<「長音」の順に非オノマトペの割合が増加している。

6.2 出現位置

非オノマトペの出現位置について、語義分類の過程で表3の③から⑦に示す5つに分類した。方言語彙または方言文法の一部をオノマトペとして誤抽出したものが1,522例と最も多かった。例えば、「咳払いの声」を表すオノマトペである「えへん」は京都府、大阪府、滋賀県、三重県等に計527例出現したが、これらはすべて以下に示すような関西方言における動詞の否定形であった。

- 「ほんで、残業せえへんかったらもう行くところないわと。」(京都府宇治市)
- 「やっぱり農協だけではね、そら金のやっぱり入ってけえへんところは、なかなか手出しにくいんですわ。」(大阪府河内長野市)
- 「評価がこれでは見えへんねんね。」(兵庫県尼崎市)

また、「しとり」(岡山県、島根県等13府県、146例)、「はった」(大阪府、京都府等8府県、127例)、「かったん」(大阪府、兵庫県等18府県、112例)、「かんから」(兵庫県、大阪府等16府県、102例)の4語が方言に起因する非オノマトペとして100例以上出現した。

表 9 オノマトペ / 非オノマトペと表層構造の関係

	語数	平均拍数	最終拍				4 拍語の表層構造			
			促音	撥音	長音	非特殊拍	ABAB 型	ABCB 型	A っ B リ	A ん B リ
オノ 100	99	3.78	26.2%	35.5%	0.0%	74.8%	75.7%	100.0%	91.7%	100.0%
中間	56	2.96	54.8%	51.6%	0.0%	16.5%	21.6%	0.0%	8.3%	0.0%
オノ 0	22	2.65	19.0%	12.9%	100.0%	8.7%	2.7%	0.0%	0.0%	0.0%
合計 (語数)	177	3.39	42	31	1	103	74	7	12	4

方言に起因する解析誤りは文法に関わるものが大半であるが、「きっぱし」(干し芋)、「こじゃんと」(ひどく / 徹底的に)などの例も見られた。

このほか「④固有表現の一部」(1,066 例)、「⑥オノマトペの一部」(238 例)はそれぞれ固有表現や文字長が長く辞書に登録されていないオノマトペの部分文字列と一致する例である。「あびっ子クラブ」「ふっつ」「あさびー」「あそびっ子」「ごんごバス」「ぱりっ子」(以上④)、「くちゅっ」「ちんたら」「ぴちん」(以上⑥)の例が見られた。「⑤同音異義語」(437 例)はオノマトペと表記が同一の語が存在する例である。「しゅん(旬)」(241 例)、「ぼたん(牡丹)」(13 例)、「ずきん(頭巾)」(12 例)など、漢字表記が可能な名詞のひらがな表記がオノマトペとマッチした。

このように文中の様々な位置で、ひらがなの文字列が表層的にオノマトペと一致する。ユーザ辞書を用いた形態素解析だけでは、オノマトペ以外も広く抽出してしまう。例えば、本分析の範囲では「えへん」「おいおい」など「オノ 0」の語は、辞書登録によって誤抽出だけを生じさせる結果となる。本分析の結果をオノマトペ抽出の正解データとして利用することで、今後、オノマトペ自動抽出手法の高度化の研究を進めていくことが重要である。

7. ま と め

本研究では、全国 403 自治体から収集した 2010 年度の地方議会会議録コーパス約 1000 万文 (3 億語) を対象として、現代の話しことばにおけるオノマトペを分析した。形態素解析を用いて、186,416 例 (異なり語数 982 語) のオノマトペ候補を抽出した。地方議会会議録コーパスを分析対象とした本研究の第一の知見は、やや改まった話し方をするであろう公的な場での話しことばにおいてもオノマトペが豊富に使用されていることである。特に、「しっかり」「どンドン」など、政策の推進を求める際に使用されると考えられる表現や「はっきり」といった厳格な判断や言及を求める表現が高い頻度で出現した。さらに、「そろそろ」「ゆっくり」「じっくり」といった政策実行の時期や議論の進行に関連するとみられるオノマトペなどが出現することを確認した。また、オノマトペの出現確率は西日本においてやや高い傾向があり、統計的有意差が見られるものの、特定の地方で顕著に多く出現するという特徴は見られないことを明らかにした。

次に、出現頻度が中程度のオノマトペ 177 語 (18,545 例) を人手により分析した。人手によりオノマトペである

と判断したものについては、オノマトペ辞典に基づいて語義を分類した。半数以上の語は複数の語義で使用されており、分類した語義の 86.5% が「辞典に記載された語義」または「辞典の語義から派生した語義」であった。このため、複数の語義を持つオノマトペにおける語義の曖昧性解消は、オノマトペ辞典に基づく分類によって概ね可能となることを指摘した。さらに、語義の地域差、辞典に記載されていない語義、固有表現に利用されている語義の特徴を具体例を挙げて述べた。

また、人手によりオノマトペではないと判断したものを対象として、表層構造・出現位置の観点から分析を行った。表層構造では、拍数が短く、特殊拍で終わるオノマトペは非オノマトペとして出現しやすく、繰り返し型 (ABAB 型、ABCB 型) や A っ B リ型、A ん B リ型のような典型的な構造は非オノマトペとして出現しづらいことが明らかになった。出現位置では、方言文法や方言語彙において非オノマトペが多くみられた (非オノマトペ全体の約 25%)。

本研究では、地方議会会議録コーパスを用いて現代の話しことばにおけるオノマトペの用例を分析した。本研究における人手による分析結果は重要な知見であるが、分析結果をすべて述べることは紙幅の都合上困難である。本研究の成果である事例に基づく定量的なデータはオノマトペの工学的利活用を指向した種々の研究において活用が可能である。人手によるオノマトペ、非オノマトペの判別結果はオノマトペ自動抽出の学習データや正解データとして利用することができる。また、語義分類の結果は語義の曖昧性解消 (語義の自動推定) の研究に利用することができる。例えば「つるつる」は「表面が滑らかなさま」を表すオノマトペであり、北海道から沖縄県までの 14 都道府県に 37 例出現する。このうち、北海道に出現する「つるつる」(11 例) はすべて道路の凍結を表すものであった。語義分類済みのオノマトペを含む文を係り受け解析することで、主題や係り先を決定づける要因を見つけ出すことにも役立てることができる。

人手による分析結果は今後の研究の進展に寄与するものであると考えられるため、分析データを提供可能な形に集計・整理し、研究プロジェクトのウェブサイト^{*9}を通じて提供する。地方議会会議録に対する研究においては、発言の賛否への積極性の判定システム [筒井 13] や、議会情報提供システム [木村 11] へのオノマトペの利用を検討したい。

*9 <http://www.local-politics.jp/>

謝 辞

本研究の一部は科学研究費 No.22300086 および、No.23700256、No.25370524 による。

◇ 参 考 文 献 ◇

- [平田 13] 平田佐智子, 中村聡史, 小松孝徳, 秋田喜美: 国会会議録コーパスを用いたオノマトペ使用の地域比較, 第 27 回人工知能学会全国大会, 3N4-OS-01c-2 (2013)
- [方言 92] 方言研究ゼミナール編: 身体感覚を表すオノマトペ, 方言資料叢刊, 第 2 巻 (1992)
- [福島 14] 福島弘識, 内田ゆず, 荒木健治: 2 つの意味をも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, P3-10 (2014)
- [井上 13] 井上史雄: 去った 日, ことばの散歩道, 明治書院, pp.154-155 (2013)
- [木村 11] 木村泰知, 渋谷英潔, 高丸圭一, 乙武北斗, 小林哲郎, 森辰則: 地方議員マッチングシステムにおける能動的質問のための質問生成手法, 人工知能学会論文誌, Vol.26, No.5, pp.580-593 (2011)
- [木村 12] 木村泰知, 渋谷英潔, 高丸圭一, 乙武北斗, 森辰則: 地方議会会議録コーパスの構築とその利用, 第 26 回人工知能学会全国大会, 3B3-NFC-4-3 (2012)
- [増田 12] 増田正: 地方議会の会議録に関するテキストマイニング分析: 高崎市議会を事例として, 地域政策研究 15(1), pp.17-31 (2012)
- [松本 08] 松本直樹: 地方議員の図書館への関心に関する予備的考察: 埼玉県市議会の議会会議録分析をもとに, 日本図書館情報学会誌 54(1), pp.39-56 (2008)
- [宮地 78] 宮地裕: 擬音語・擬態語の形態論小考, 国語学, 115 号, pp.33-39 (1978)
- [那須 07] 那須昭夫: オノマトペの語末促音, 音声研究, 第 11 巻, 第 1 号, pp.47-57 (2007)
- [野村 96] 野村稔・鶴沼信二: 地方議会実務講座, 第 3 巻, ギョウゼイ (1996)
- [小野 07] 小野正弘編: 日本語オノマトペ辞典, 小学館 (2007)
- [齋藤 11] 齋藤誠, 大城卓, 菅原晃平, 永井隆広, 渋谷英潔, 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, P2-21 (2011)
- [齋藤 52] 齋藤義七郎: 擬声語擬態語の音韻論的考察, 音声学会会報, 第 79 号, pp.11-17 (1952)
- [清水 14] 清水祐一郎, 土斐崎龍一, 坂本真樹: オノマトペごとの微細な印象を推定するシステム, 人工知能学会論文誌 29(1), 41-52 (2014)
- [菅原 12] 菅原晃平, 大城卓, 齋藤誠, 永井隆広, 渋谷英潔, 木村泰知, 森辰則: 地方議会会議録コーパスの拡充における問題点の分析と対処, 言語処理学会第 18 回年次大会論文集, P1-15 (2012)
- [高丸 11] 高丸圭一: 規模の異なる自治体における地方議会会議録の整文の比較, 社会言語科学会第 27 回研究大会発表論文集, pp.256-259 (2011)
- [高丸 13] 高丸圭一, 乙武北斗, 渋谷英潔, 木村泰知, 森辰則: 形態素 N-gram を用いた地方議会会議録コーパスの地域変異検出の試み 文末表現を例に, 言語処理学会第 19 回年次大会発表論文集, pp.737-740 (2013)
- [竹田 12] 竹田晃子: 円滑な医療コミュニケーションのための方言集 『東北方言オノマトペ用例集』の取り組み, 日本方言研究会研究発表会発表原稿集, 94, pp.55-58 (2012)
- [竹安 04] 竹安栄子: 地方議員のジェンダー差異 「2002 年全国地方議員調査」結果の分析より, 京都女子大学現代社会研究, 7, pp.99-118 (2004)
- [土屋 12] 土屋誠司, 鈴木基之, 任福継, 渡部広一: モーラ系列と音象徴ベクトルによるオノマトペの印象推定法, 自然言語処理, 19(5), pp.367-379 (2012)
- [筒井 13] 筒井貴士, 我満拓弥, 渋谷英潔, 木村泰知, 森辰則: 地方政治情報コーパスへの賛否の積極性に関する注釈作業, 言語処理学会第 19 回年次大会発表論文集, pp.874-877 (2013)
- [内田 12] 内田ゆず, 荒木健治, 米山淳: ブログ記事からのオノマトペ用例文の自動抽出手法, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics 24(3), pp.811-820 (2012)
- [内田 13] 内田ゆず, 荒木健治, 米山淳: 商品レビュー文に含まれ

るオノマトペに基づく商品カテゴリーの分類, 第 29 回ファジィシステムシンポジウム (FSS2013) 講演論文集, pp.762-777 (2013)

[上田 12] 上田翔, 八木田浩史: 地方議会会議録における環境用語の出現頻度に基づく自治体の環境問題対応の解析, 環境情報科学学術研究論文集 (26), pp.283-288 (2012)

〔担当委員: 小松 孝徳〕

2014 年 04 月 28 日 受理

—— 著 者 紹 介 ——

高丸 圭一(正会員)

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 年宇都宮共和国シティアライフ学部准教授 (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, 社会言語科学会, 日本音声学会, 日本語学会各会員。

内田 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 年北海道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。日本知能情報ファジィ学会, 電子情報通信学会, 言語処理学会, 各会員。

乙武 北斗(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兼任, 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, 日本知能情報ファジィ学会, 電子情報通信学会各会員。

木村 泰知(正会員)

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兼任。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。言語処理学会, 情報処理学会, 電子情報通信学会各会員。

◇ 付 録 ◇

A. 分析対象の 177 語

4~6 章における分析対象 (全 177 語) の形態素解析結果 (都道府県出現確率の全国計, 総出現頻度, 出現都道府県数, オノマトベの正抽出率), 語義分類における語義数 (合計, 辞典内語義, 辞典派生語義, 辞典外語義), オノマトベ分類におけるオノマトベ, 固有表現の頻度を下表に示す.

オノマトベ	出現確率 ($\times 10^{-7}$)	抽出 頻度	都道 府県数	正抽出率	語 義				出現頻度	
					合計	辞典内	派生	辞典外	オノマトベ	固有表現
あたふた	57.48	34	20	100.0%	1	1	0	0	34	0
いらいら	235.45	156	35	100.0%	1	1	0	0	156	0
うろうろ	204.75	137	30	100.0%	2	2	0	0	137	0
うんざり	73.35	37	21	100.0%	1	1	0	0	37	0
えへん	455.54	527	13	0.0%						
おいおい	196.06	139	35	0.0%						
がくがく	251.41	137	36	100.0%	3	2	1	0	137	0
がくっ	83.61	53	23	96.2%	1	1	0	0	51	0
がくん	74.77	59	21	100.0%	2	2	0	0	59	0
がしっ	58.45	39	20	2.6%	1	1	0	0	1	0
がたがた	177.85	132	32	100.0%	4	4	0	0	132	0
がたっ	77.57	38	21	86.8%	1	1	0	0	33	0
がちがち	61.85	41	20	100.0%	2	1	1	0	41	0
かちっ	94.02	81	18	100.0%	2	1	0	1	81	0
がっくり	79.74	49	23	100.0%	3	2	0	1	49	0
かったん	170.16	120	18	0.0%						
がちり	144.18	120	30	100.0%	3	2	1	0	119	1
がばっ	167.47	123	32	35.0%	2	1	1	0	43	0
からっ	80.29	37	19	27.0%	4	3	0	1	10	0
がらり	62.65	48	20	100.0%	1	1	0	0	48	0
かんから	126.53	104	17	0.0%						
かんかん	137.59	72	27	100.0%	4	4	0	0	72	0
がんがん	190.69	130	31	100.0%	2	2	0	0	130	0
ぎくしゃく	237.91	117	35	100.0%	2	1	1	0	117	0
きちっ	57.85	37	19	97.6%	1	1	0	0	41	0
きちん	73.6	47	23	100.0%	1	1	0	0	47	0
きっ	59.65	35	19	0.0%						
きらきら	327.35	197	30	100.0%	1	1	0	0	52	145
きらら	216.09	110	19	99.1%	0	0	0	0	0	109
きらり	390.13	273	32	100.0%	1	1	0	0	99	174
ぐい	342.4	221	38	1.8%	1	1	0	0	2	2
ぐいぐい	56.67	22	13	100.0%	2	2	0	0	22	0
ぐずぐず	56.59	47	20	100.0%	3	3	0	0	47	0
くすり	105.04	74	6	0.0%						
くっ	237.19	130	27	3.8%	1	1	0	0	5	0
くどくど	157.18	134	29	98.5%	1	1	0	0	132	0
ぐんぐん	60.75	52	17	100.0%	2	1	1	0	24	28
ごくごく	248.84	206	34	1.0%	1	1	0	0	2	0
こそこそ	56.75	48	22	97.9%	3	2	0	1	47	0
ごたごた	170.41	119	31	100.0%	4	3	0	1	119	0
ごちゃ	106.43	68	24	100.0%	2	1	0	1	67	0
ごちゃごちゃ	325.95	240	35	100.0%	2	2	0	0	240	0
こつこつ	104.74	79	22	100.0%	3	2	1	0	78	1
ごっちゃ	173.75	133	31	100.0%	1	1	0	0	133	0
ことっ	206.74	128	21	0.0%						
ごろごろ	126.63	66	27	100.0%	8	4	3	1	66	0
ごろっ	215.46	117	30	95.7%	7	6	0	1	112	0
ころり	57.81	39	17	100.0%	2	2	0	0	38	1
ごん	110.37	66	12	10.8%	1	1	0	0	7	0
さっ	113.99	78	13	0.0%						
さらさら	171.13	118	27	100.0%	5	3	1	1	32	0
さらっ	236.57	150	34	99.3%	2	2	0	0	149	0
ざわざわ	60.16	30	16	100.0%	3	2	1	0	30	0
さんさん	424.91	133	18	74.4%	1	1	0	0	2	97
しく	153.04	110	30	0.0%						
じっ	50.45	12	9	16.7%	1	1	0	0	2	0
しとっ	174.24	85	18	0.0%						
しとり	347.33	147	14	0.0%						
じゃん	126.55	106	16	0.0%						

オノマトペ	出現確率 ($\times 10^{-7}$)	抽出 頻度	都道 府県数	正抽出率	語 義				出現頻度	
					合計	辞典内	派生	辞典外	オノマトペ	固有表現
しゃんしゃん	61.03	17	12	100.0%	2	2	0	0	17	0
じゃんじゃん	56.91	44	15	100.0%	1	0	1	0	44	0
しゅん	484.59	275	40	1.8%	2	1	0	1	5	0
じりじり	62.14	39	18	100.0%	4	3	0	1	39	0
じわじわ	205.8	121	35	100.0%	2	2	0	0	121	0
じん	496.99	303	37	2.3%	1	1	0	0	7	0
すい	54.99	29	12	0.0%						
すいすい	65.03	59	15	100.0%	2	2	0	0	22	37
すかつ	229.82	109	23	57.8%	2	2	0	0	38	25
ずきん	302.44	39	12	7.7%	1	1	0	0	3	0
ずっ	129.75	85	29	79.3%	1	1	0	0	134	0
ずとん	143.84	91	27	100.0%	4	1	1	2	91	0
ずばっ	65.3	39	22	100.0%	2	2	0	0	38	1
ずばり	378.29	298	35	99.7%	2	2	0	0	296	1
ずらずら	130.9	61	28	100.0%	2	2	0	0	61	0
ずらっ	300.34	209	34	100.0%	1	1	0	0	209	0
すんなり	372.61	246	40	100.0%	1	1	0	0	246	0
ぞろぞろ	77.98	50	15	100.0%	2	2	0	0	50	0
たっ	438.87	214	32	0.5%	1	1	0	0	1	0
たった	464.56	338	42	0.0%						
たっぶり	130.02	88	30	100.0%	3	3	0	0	83	5
だらだら	346.91	234	36	100.0%	2	2	0	0	234	0
ちゃっ	130.66	87	22	11.5%	1	1	0	0	10	0
ちゃん	382.23	221	27	11.3%	2	2	0	0	25	0
ちゅっ	52.99	30	8	0.0%						
ちょこちょこ	352.89	275	37	100.0%	2	2	0	0	275	0
ちょろちょろ	94.55	65	26	100.0%	5	3	1	1	65	0
ちょん	84.48	38	15	50.0%	5	3	2	0	19	0
ちらちら	128.46	101	29	100.0%	5	4	0	1	101	0
ちらほら	181.49	125	33	100.0%	2	2	0	0	125	0
ちん	66.97	51	16	3.9%	2	1	0	1	2	0
つつい	338.87	195	39	100.0%	1	1	0	0	195	0
つらつら	125.9	61	26	100.0%	2	1	1	0	61	0
つるつる	54.63	37	14	86.5%	5	1	3	1	32	0
てきばき	94.35	45	21	100.0%	2	2	0	0	45	0
どかん	59.65	46	19	97.8%	1	1	0	0	45	0
どきっ	98	44	25	100.0%	1	1	0	0	44	0
どきどき	113.21	87	23	100.0%	2	1	0	1	76	11
とくとく	176.3	40	11	17.5%	2	0	1	1	7	0
とことこ	64.74	90	15	44.4%	1	1	0	0	19	21
どさくさ	58.42	34	17	100.0%	2	1	1	0	34	0
どしどし	187.52	97	29	100.0%	1	1	0	0	97	0
どたばた	63.02	49	22	100.0%	1	1	0	0	49	0
どっさり	73.61	22	12	100.0%	1	1	0	0	21	1
とっと	211.27	89	22	5.6%	1	1	0	0	5	0
どっと	329.02	231	38	96.5%	0	0	0	0	0	4
どっぶり	92.91	50	20	100.0%	2	2	0	0	50	0
どろどろ	56.04	43	19	100.0%	4	3	1	0	43	0
とんとん	138.59	112	29	95.5%	3	3	0	0	107	0
なんなん	72.82	54	23	0.0%						
にこにこ	323.05	192	41	100.0%	1	1	0	0	87	105
にっ	98.56	70	18	0.0%						
にやっ	306.99	194	36	3.6%	1	1	0	0	7	0
にやり	227.61	132	33	0.0%						
のこのこ	183.85	111	33	10.8%	2	1	0	1	12	0
のびのび	328.71	286	28	99.3%	2	2	0	0	38	246
のんびり	286.68	244	36	100.0%	3	1	1	1	224	20
ばくっ	67.37	69	13	100.0%	2	1	0	1	69	0
ばさっ	98.14	81	25	100.0%	3	1	1	1	80	0

オノマトペ	出現確率 ($\times 10^{-7}$)	抽出 頻度	都道 府県数	正抽出率	語 義				出現頻度	
					合計	辞典内	派生	辞典外	オノマトペ	固有表現
ばたばた	380.62	230	40	100.0%	3	3	0	0	230	0
はっ	164.71	130	31	0.0%						
ばっさり	173.31	132	30	100.0%	3	2	0	1	132	0
はっし	52.13	33	17	100.0%	1	1	0	0	33	0
はった	148.84	129	10	0.0%						
ばっばっ	58.6	60	18	100.0%	2	0	2	0	61	0
はらはら	61.18	35	20	100.0%	3	2	0	1	35	0
ばらばら	298.09	218	35	99.5%	2	1	0	1	217	0
ばらばら	98.13	78	26	100.0%	4	3	1	0	78	0
ばりっ	59.08	32	2	3.1%	1	1	0	0	1	0
ばりばり	56.54	47	21	100.0%	6	5	0	1	48	0
ばん	94.66	72	20	63.9%	3	2	1	0	46	0
ばんばん	219.76	179	28	99.4%	2	2	0	0	178	0
びー	71.67	75	13	0.0%						
びかびか	69.53	53	22	100.0%	3	2	0	1	51	2
びしゃっ	387.74	112	17	100.0%	3	1	2	0	112	0
びたっ	152.31	96	29	100.0%	3	3	0	0	93	3
びちっ	104.96	66	15	100.0%	3	1	1	1	66	0
ひっ	165.01	106	29	0.0%						
びっ	129.05	134	17	2.2%	2	0	1	1	3	0
びっ	91.26	73	22	47.9%	5	2	0	3	35	0
びっしり	139	78	29	100.0%	1	1	0	0	78	0
びったり	328.25	230	38	100.0%	4	4	0	0	225	5
びんびん	73.05	51	22	100.0%	3	1	1	1	21	30
ぶくぶく	53.05	26	8	100.0%	3	2	1	0	25	1
ふっ	71.3	69	7	1.4%	1	1	0	0	1	0
ぶつぶつ	77.89	39	19	100.0%	3	1	2	0	39	0
ふらふら	147.49	94	26	100.0%	4	3	0	1	94	0
ぶらぶら	67.39	56	19	100.0%	2	1	1	0	43	13
ぶらり	104.52	48	22	100.0%	1	1	0	0	10	38
ふわっ	53.16	36	18	100.0%	4	0	2	2	36	0
ふわふわ	59.77	36	15	100.0%	3	2	0	1	30	6
ぶんぶん	64.41	30	9	100.0%	3	2	0	1	10	20
ぼーん	55	27	16	100.0%	1	0	1	0	27	0
ぼこぼこ	51.91	40	17	100.0%	3	2	0	1	40	0
ぼたん	81.66	89	14	1.1%	1	0	0	1	1	0
ぼっ	110.9	74	28	29.7%	3	1	0	2	22	0
ぼっ	444.44	384	37	60.9%	1	1	0	0	234	0
ぼっかり	54.13	27	17	100.0%	3	2	1	0	27	0
ぼっきり	52.31	32	18	100.0%	3	3	0	0	32	0
ぼつぼつ	152.35	98	26	100.0%	3	3	0	0	98	0
ぼつぼつ	124.44	72	28	100.0%	4	4	0	0	72	0
ほのほの	175.4	81	30	100.0%	2	2	0	0	34	47
ぼやっ	94.84	76	21	56.6%	2	2	0	0	43	0
ぼろぼろ	56.31	38	20	100.0%	3	2	1	0	38	0
ぼん	353.19	281	36	96.4%	1	1	0	0	271	0
ぼんぼん	141.54	113	29	93.8%	2	1	0	1	106	0
ぼんぼん	122.47	103	26	98.1%	4	3	0	1	101	0
ぼんやり	104.71	62	23	100.0%	4	4	0	0	62	0
まざまざ	125.44	60	27	100.0%	1	1	0	0	60	0
めちゃくちゃ	100.92	75	22	100.0%	2	2	0	0	74	1
めっきり	85	64	21	100.0%	2	1	0	1	64	0
もくもく	56.3	21	9	52.4%	2	2	0	0	11	0
もたもた	79.79	52	24	100.0%	2	1	1	0	52	0
もやっ	172.64	116	29	6.0%	1	1	0	0	7	0
もやもや	116.13	73	25	100.0%	2	2	0	0	73	0
やきもき	56.48	29	20	100.0%	1	1	0	0	29	0
ゆったり	480.61	269	41	99.3%	3	2	0	1	258	9
わいわい	251.9	240	32	100.0%	2	2	0	0	70	170