社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－DBS－143 (65)
2007／7／3

オノマトペ用例辞典における用例を意味により分類するためのクラスタリング手法の諸検討

浅賀 千里† YusufMukarramah†† 渡辺知恵美†

† お茶の水女子大学大学院人間文化創成科学研究科 〒112-8610 東京都文京区大塚2-1-1
†† お茶の水女子大学理学部情報科学科 〒112-8610 東京都文京区大塚2-1-1
E-mail: †{asaga,mukarramah}@db.is.ocha.ac.jp, ††chiemi@is.ocha.ac.jp

あらまし オノマトペとはいわゆる擬態語・擬音語のことである．事象を的確に表現でき，コミュニケーションを図る上で重要なものである．ところが，オノマトペは感覚的なものであるので外国人の日本語学習者がオノマトペの用例を習得するのは難しく，その学習に有効なのはオノマトペを含む文章を知ることだと言われている．そこで，我々はWebから数多くの新しい文章を抽出し，学習者に提示できるようなオノマトペ用例辞典の開発を進めている．本辞典では現在，オノマトペの品詞的役割によって用例を分類し画面上に提示しているが，それでは様々な用例が混在し，オノマトペの意味が理解しにくい．そこで，学習者がより理解を深められるよう，本稿ではオノマトペの用例をオノマトペの意味ごとに分類し提示する手法を検討している．また，オノマトペがどのような語に係るのか，その語にどのようなオノマトペが係るのかを知ることができるようその関係を可視化する．

キーワード e-learning，Webコーパス，WebDB，データマイニング

Examinations of clustering technique for classifying sentences by meaning of onomatopoeia on Online Onomatopoeia Example-based Dictionary

Chisato ASAGA†, Yusuf MUKARRAMAH††, and Chiemi WATANABE†

† Ochanomizu University Graduate School of Humanities and Sciences
Otsuka 2–1–1, Bunkyou-ku, Tokyo, 112–8610 Japan
†† Department of Information Sciences, Faculty of Science, Ochanomizu University
Otsuka 2–1–1, Bunkyou-ku, Tokyo, 112–8610 Japan
E-mail: †{asaga,mukarramah}@db.is.ocha.ac.jp, ††chiemi@is.ocha.ac.jp

Abstract Onomatopoeia which is imitative word is exepress concrete phenomenon and it plays a crucial part in communication. But it is difficult for learners of Japanese who study Japanese onomatopoeia to understand their means and usages because onomatopia is sensuous. An effective method for mastering onomatopoeia is to read a lot of sentences with onomatopoeia. Then we are developing an online onomatopoeia example-based dictionary which collects a lot of sentences with onomatopoeia from the Web. Now we presents sentences which is classified in a part of speech role of onomatopoeia the screen. However each onomatopoeia has multiple meaning, in the current version, sentences which have different meaning of onomatopoeia are mixed in the example list. Then, we attempt classifying these examples of onomatopoeia by onomatopoeia meaning to presrent them to Japanese-language learners in more understandable way. Moreover, to study co-occurrence relationships between onomatopoeias and verb (or noun), we also attempt to visualize relationships of onomatopoeias and the other words.

Key words e-learning, Web corpus, WebDB, data mining

1. はじめに

オノマトペとは,「どきどき」や「しっかり」などのいわゆる擬態語・擬音語のことである．具体的な事象を的確に表現できる語彙であり，コミュニケーションを図る上で重要なものである．ところが，オノマトペが感覚的な語であることや，外国語

にオノマトペの対応語がないこと，1つのオノマトペが複数の意味を持つことなどから，日本語学習者にとってオノマトペの用例を習得するのは難しいといわれている．

日本語学習者がオノマトペの意味・用法を理解するには，複数の用例として適切な文章からオノマトペが文中でどのように使われているのかを知ることが有効である．また，オノマトペは時代と共に意味が変わっていくことから常に新しい用例を知ることが重要となる．そこで，我々はWebから数多くのオノマトペの最新の用例を自動抽出し，日本語学習者に提示するようなオノマトペ用例辞典の開発を進めている．

本辞典はWebから用例を抽出するため，あまり適切ではない文章も多く抽出されてしまう．そこで，我々はどのように適切な文章を効率よく抽出できるのかについて検討し，その結果，オノマトペの文法的性質を利用して用例を収集することにした．その文法的性質とは，オノマトペの語尾に付属語をつけるとそのオノマトペが特定の品詞の役割を果たすというもので，例えば，「くるくる」に付属語の「と」を付けて「くるくると」にするとその「くるくると」は「回る」や「転がる」などの動詞にかかる副詞としての役割を持つ，というものである．オノマトペに付属語をつけたものを見出し語として検索し用例を抽出する実験を行ったところ，つけないで抽出するよりも良い結果が得られたので，この方法を採用し，実装している[1] [2].

現在，本辞典は，抽出した用例をそのオノマトペの品詞的役割を元に分類し，画面に提示しているが，学習者の意味・用法の理解をより深めることができるよう，用例の組織化の手法や提示方法を考えている．本稿では，主に，複数の意味を持つオノマトペの用例をオノマトペの意味ごとに分類するために現在行っている諸検討の項目について報告する．「がりがり」を例にあげると，「氷をがりがり食べる.」や「がりがり勉強する.」，「がりがりの体.」など同じ「がりがり」でも全く違う意味を持つ．よって，このような異なる意味ごとに用例を分類し，提示する．具体的には，情報検索の手法を適用し，用例を文書ベクトルで表してクラスタリングする．また，適切な重み付けやオノマトペ辞書の効果的な利用，URLによる用例の一般性の判定について検討する．また，オノマトペとそれが係る用言に着目し，あるオノマトペが係る用言や，その用言がかかるオノマトペの関係を可視化することで，連鎖的に語彙の習得を支援するシステムについても検討する．

2. オノマトペ用例辞典

本研究で開発しているシステムはユーザがオンライン上で検索したオノマトペを含んでいる文章をWebから抽出し画面上に表示することで，日本語学習者にオノマトペの用例を提示するシステムである．用例の表示画面のイメージを図1に示す．

学習者が用例を知りたいオノマトペを選択すると，そのオノマトペの用例が画面に一括表示される．現在は，「がりがりと氷を食べる.」のようにオノマトペが副詞的に使われている(用言に係る)もの，「がりがりの体.」のように連体詞的に使われている(体言に係る)もの，「がりがり君.」のように複合名詞として使われているものに用例を分類し表示するようにしている．また，それぞれの用例に対して出典元のページのリンクがはってある．更に，本辞典では，用例の編集機能と削除機能を追加することで，用例として不適切である文章が抽出された場合には編集，もしくは削除をすることができるようにしている．今後は，利用者参加型の手法をシステムに追加することでユーザにコメントを残してもらえるようにすることを考えている．

図1 用例辞典のイメージ
Fig. 1 image of dictionary

2.1 システムの流れ

オノマトペリストからユーザが選択したオノマトペを含むWebページをYahoo!API[3]を用いて検索し自動取得し，その中からそのオノマトペを含んでいる文章を抽出する．検索を行う際，オノマトペに付属語を付けたものを見出し語として検索したページからも文章を抽出する．付属語については，2.2節で述べる．抽出した文章を日本語係り受け解析器CaboCha[4]を用いてその文章の係り受け解析を行い，オノマトペが係る語句とその品詞を求める．抽出した文章と情報から用例としての適正を判定し，適正と判断したものをテーブルに格納する．用例提示用のWebサーバを設置し，データベースへの検索インタフェースを提供する[5].

2.2 用例として適正な文章の抽出

本辞典はWebから文章を抽出しているため，用例として適正ではないものがいくつも検出される．その中には，オノマトペが複合名詞として使われているものやオノマトペ単体で使われているものが多い．そこで，しっかりとした文章になっている用例を抽出するために以下の手法を用例抽出手法に取り入れている．

1) Webから抽出した文章の中から，日本語係り受け解析器CaboChaを用いた時にオノマトペが何らかの語に係っていると判断された文章を抽出する．

2) オノマトペは語尾に付属語をつけることによって様々な品詞の役割を持つという文法的性質があることから，オノマトペの語尾に付属語をつけたものを見出し語としてYahoo!検索をし，文章を抽出する．オノマトペの品詞と付属語の関係は表1に示す．

1)，2)の手法を用いることで，不適正な文章を用例から除去し，用例として適切な文章が抽出できる割合を高めることができる．

表 1　オノマトペと付属語の関連

付属語	オノマトペの品詞	例
と	副詞	髪がさらさらと揺れる
な，の	連体詞	さらさらな髪
だ	形容動詞	髪がさらさらだ
する	動詞	髪がさらさらする

3. オノマトペの意味による用例の分類

本辞典のインタフェースは，用例中のオノマトペが副詞的に使われているもの，連体詞的に使われているもの，複合名詞として使われているものというようにオノマトペの用法別に用例を分類して表示しているが，これでは，同じような意味で使われているオノマトペの用例が別用法の用例として表示されてしまう．実際に抽出した用例を副詞的用法，名詞的用法に分類したものを表 2，表 3 に示す．表 2，表 3 から 1 つのグループに，「ひどく痩せている」や「ものを引っ掻いたり，削ったりする時の音」などのいろいろな意味のオノマトペの用例が混在していて意味が統一されていないことがわかる．例えば，副詞的用法の「がりがりのやせた雌犬で，そこそこ年はいってます.」と名詞的用法の「かなりがりがりな体.」と「肋骨の浮き出たがりがりの体.」という用例の「がりがり」は全て「ひどく痩せている」という意味であるので，これらの用例は同じグループとして表示した方が，より学習者は見やすくなるはずである．

そこで，用例中のオノマトペの意味ごとに用例を分類して表示する手法を検討している．

前述した「がりがり」の場合，既存の擬態語・擬音語の辞書 [6] [7] にある「がりがり」の意味は大きくわけて 4 つある．

1) かたい物を繰り返し引っ掻いたり，削ったり，噛み砕いたりした時などに発する音．

2) 引っ掻いたり，削ったり，噛み砕いたりした場合に，1) の音が発するようなかたさの様子．

3) ひどく痩せている様子．

4) 自分の欲望のために一途に物事に打ち込む様子

表 2，表 3 の用例を上記の意味的に分類したものを表 4，表 5，表 6，表 7 に示す．

明らかに表 2，表 3 より用例の意味がわかりやすくなった．このように意味によってオノマトペの用例を分類して提示することによって，学習者が理解しやすくなるようにする．

3.1 分類の基本手法

本研究では，用例分類の基本手法として情報検索手法を使用する．用例に出現する単語に重みをつけ，それを元に用例をそれぞれベクトル化し，ベクトルの距離から用例のクラスタリングを行う．

表 2　副詞的用法の用例

がりがりと頭をかきながら，ベッドから起き上がる
がりがりのやせた雌犬で，そこそこ年はいっています
がりがりと格闘派小説を書きまくる俺ですが
がりがりと氷を噛み砕く音がする
パイナップルの葉っぱ (がりがりする)
これから数ヶ月は特にリリースがりがりするわけでもない…

手順は以下のようになっている．

あるオノマトペの用例が N 個あったと仮定する．

1) それぞれの用例から主要な単語 (用言，名詞等) の語幹を抽出する．それ自体があまり意味を持たない単語 (助詞，接続詞等) の語幹は抽出しない．

2) それぞれから抽出した単語の IDF を計算する．

単語 j の IDF=log(N/n_j)

N:全用例数　n_j:単語 j が含まれている用例数

3) それぞれの用例におけるその単語の重みを計算する．

用例 i における単語 j の重み w_{ij}=t_{ij} ×単語 j の IDF

t_{ij}:用例 i に含まれている単語 j の数

4) それぞれの用例について，単語 j の方向に w_{ij} の大きさの用例ベクトルを作成する．

5) それぞれの用例ベクトルからクラスタリングを行う．

3.2 文書ベクトルの重みの調整

3.1 節の基本手法を用いた場合，用例中に含まれている単語はどれも同様に重要だとされて重みが計算されている．「私は昨日，がりがりと氷を削った.」という用例だと，基本手法のままでは「昨日」というあまり「がりがり」と関係のない単語と，「氷」や「削る」という「がりがり」に非常に関連ある単語の重要性が等しくなってしまう．オノマトペの用例から意味・用法を理解する際，オノマトペに関係する語が他の語に比べて特に重要であるので，オノマトペが係っている語（「削る」）・その語に係っている他の語 (「氷」) の重みを強くする．

オノマトペの用例部分だけでは，オノマトペの意味がよく表せていないものもしばしば出てくるので，用例部分だけでなく，オノマトペの用例の周辺にある文章の全ての語に重みをつける．例えば，「猫が家の壁をかりかりと引っ掻いていた．犬もがりがりしていた．壁に引っ掻き傷がたくさんできた.」の場合，用例部分は「犬もがりがりしていた.」であるが，これだけだと何を表しているのかが明確に判断できない．そこで，周辺文章である「猫が家の壁をかりかりと引っ掻いていた.」,「壁に引っ掻き傷がたくさんできた.」に含まれるものに重みをつける．その場合，そのオノマトペの用例に位置的に近い文章の単語ほど重みは強く，用例から離れている文の単語の重みは弱くする．

3.3 クラスタリングの方法の検討

オノマトペの用例を分類する際に，どの手法がどのような利用に適切であるかを考察する．

クラスタリング手法は，非階層的か階層的かに分かれ，更にそれぞれに対して，教師なし学習と教師あり学習に分かれる．それらの代表的な手法を表 8 に示す．

表 3　連体詞的用法の用例

かなりがりがりな身体
がりがりの塊が口の中に残るような食べ物はどうにも苦手だ
肋骨の浮き出たがりがりの身体
天気は晴天，雪質は朝のうちはがりがりのアイスバーン
わんこのいたずらで壁をがりがりの被害もけっこう多いです

表 4 「かたい物を繰り返し引っ掻いたり，削ったり，噛み砕いたりした時などに発する音」の用例

がりがりと頭をかきながら，ベッドから起き上がる
がりがりと氷を噛み砕く音がする
わんこのいたずらで壁をがりがりの被害もけっこう多いです

表 5 「引っ掻いたり，削ったり，噛み砕いたりした場合に，1) の音が発するようなかたさの様子」の用例

パイナップルの葉っぱ (がりがりする)
がりがりの塊が口の中に残るような食べ物はどうにも苦手だ
天気は晴天，雪質は朝のうちはがりがりのアイスバーン

表 6 「ひどく痩せている様子」の用例

がりがりのやせた雌犬で，そこそこ年はいっています
かなりがりがりな身体
肋骨の浮き出たがりがりの身体

表 7 「自分の欲望のために一途に物事に打ち込む様子」の用例

がりがりと格闘系小説を書きまくる俺ですが
これから数ヶ月は特にリリースがりがりするわけでもない…

表 8 代表的なクラスタリング手法

	教師なし学習	教師あり学習
階層的	群平均法，ウォード法など	
非階層的	k-means 法，SOM など	LVQ，SVM など

まず，3 節冒頭に述べたように，辞書に定義されている意味ごとに分類したい場合には非階層的クラスタリングが適切であると考えられる．非階層的クラスタリングの代表的なものに k-means 法と SOM がある．

k-means 法とは，クラスタの平均を用い，与えられたものを k 個のクラスタに分類する手法である．最初にシードとなるものを k 個用意し，それぞれのものを一番距離の近いシードにまず分類する．そして，次にそのそれぞれのクラスタの中心をシードとし，同様の分類を変化がなくなるまで繰り返す．

SOM とは，m 個のサンプルについて，n 個の特性が観測できる場合，n 次元の座標空間に m 個のデータポイントが存在していると考える．データ空間中で，距離の近いサンプルは類似していると考えられ，その集合が意味を形成する．観測された現象から，空間に内在する構造をコピーして，概念のテンプレートを生成する．

k-means 法では辞書の定義の数を k に当てはめて分類するなどが考えられる．ただし，辞書では想定されないような使い方を感覚的に行えるのがオノマトペの特徴である．例えば，「がりがりの予算で働く.」という用例は辞書のどの定義にも当てはまらない．そのため，辞書定義以外のクラスタも扱えるよう考慮すべきである．

階層型クラスタリングは，グラフ間リンクを 1 つずつ含む初期クラスタ群から始めて，クラスタ間距離が最短のペアを 1 つの新しいクラスタに併合することを繰り返していくクラスタリングである．階層的なクラスタリングでは，非階層的なクラスタリングよりも細かい分類ができるようになる．例えば，「がりがり」では，3 節冒頭で述べたように大きくわけて 4 つの意味があるが，それぞれの意味に分類された用例の中で，更に細かく用例を分けることができる．1 つ目の「かたい物を繰り返し引っ掻いたり，削ったり，噛み砕いたりした時などに発する音」という意味で「がりがり」が使われるている用例には，「犬が壁をがりがり引っ掻いている.」のようにペット等が家や壁，床を引っ掻いている様子を表す用例や，「がりがりと氷を食べる.」など，噛んだ時にがりがりと音がするような物を食べている様子を表す用例などがある．そこで，このような更に細かい意味や，使う場面などにより，階層的に分類することによって，きめ細やかな日本語の学習支援ができる．

教師あり学習では，本辞典で適用する場合，辞書や辞典的 Web サイトの用例が教師データとなるが，それぞれのオノマトペに対する用例が少ないため，効果的な学習が行えない．また，本の辞書から用例を抽出する場合，オノマトペの数自体が多いので用例数が膨大になり取得が困難である．このような理由から我々は教師なし学習を利用することを検討している．

これらのクラスタリング手法について，本辞典での使用方法や予想される結果等をより明確に考え，検討していく．

3.4 辞書を利用した分類手法

本研究では，既存の辞書や Web 上の辞典的なページを用いて用例の分類する．その手法を以下に示す．

1) k-means 法を使う場合，辞書にあるオノマトペのある意味の説明文をベクトル化し，それをシードとし，用例のクラスタリングを行う．

2) 辞書や辞典的なページにある，オノマトペのある意味に関する説明文の中からいくつかキーとなる単語を抽出し，その単語の重みを強くし，より用例が明確に分別できるようにする．例えば，3.1 節の「がりがり」の 1) の説明文の場合，「かたい」，「物」，「引っ掻く」，「削る」，「噛み砕く」，「発する」，「音」が重要な単語となるのでその語の重みが強くなるようにする．

このように辞書を用いてクラスタリングを行うことを検討している．

3.5 クラスタ内の用例の一般性の判定

クラスタの中に同じ用例がいくつも含まれている場合がある．そのオノマトペの用例が一般的に広く使われているという可能性もあるが，同じページで何度も同じ表現が使われていう可能性もある．実際，本システムで収集している用例の中には，「がりがりの予算で働く.」(「ぎりぎり」を感覚的に書き換えている) や，「がりがりの穴熊党.」(「ばりばり」と近い意味で使っている) など，個人の感覚で独特の利用法をしている例もみられる．また，あるオンラインゲーム上の表現で「がりがり狩る.」という表現が多く抽出されたが，これも特定のコミュニティでのみ用いられる特殊な用例である．学習者に用例を提示する際，このような用例の一般性を提示することはとても重要である．このような用例の一般性を判定するための手法にクラスタ内の用例の抽出元のドメイン数を利用することを考えている．抽出元のドメイン数が多ければ，幅広く使われていることになり，ドメイン数が少ない場合は個人，もしくは特定の組織でのみ多く使われていることになると考えられる．このように，同じ数

の用例を含むクラスタでもその内容が異なることがあるのでドメイン数からクラスタの精度の判断を行う．

3.6 予備実験

用例のオノマトペの意味による非階層的なクラスタリングを行った際，それがどのくらい有効であるのか調べる実験を行った．

3.6.1 実験内容

オノマトペの意味ごとに用例を分類することを目的として，以下の手法を適用した場合，それぞれクラスタ内の用例がどのように分類されるのかを調べる目的で実験を行った．使用したのは「がりがり」というオノマトペである．分類する用例はWebから取得した230件である．

実験で利用する「がりがり」の意味は3節冒頭で紹介した「かたい物を繰り返し引っ掻いたり，削ったり，噛み砕いたりした時などに発する音」(意味(1))，「引っ掻いたり，削ったり，噛み砕いたりした場合に，(1)の音が発するようなかたさの様子」(意味(2))，「ひどく痩せている様子」(意味(3))，「自分の欲望のために一途に物事に打ち込む様子」(意味(4))である．

今回は以下の3手法の比較を行った．

1) ランダムに取得した4つのシードを元にk-means法でクラスタリングをする．どのシードにもあてはまらなかった用例はクラスタにランダムに挿入される．用例内の単語の重みはすべて同等である．

2)「がりがり」の4つの意味の説明文をベクトル化したものをシードにし，それを元にk-means法でクラスタリングをする．どのシードにもあてはまらなかった用例はクラスタにランダムに挿入される．用例内で，オノマトペが係っている語の重みを通常の単語の重みの4倍にし，その語の係っているオノマトペ以外の語の重みを2倍にする．

3)「がりがり」の4つの意味の説明文をベクトル化したものをシードにして，まず，それぞれの用例を一番距離が近いシードに分類する．そして，どのシードのクラスタにもあてはまらなかったものがあった場合，クラスタを増やす．そして，クラスタ内の中心を決め，それぞれの用例を一番距離が近いシードに分類する．以上のことを繰り返すことで用例すべてをクラスタにわける．用例内で，オノマトペが係っている語の重みを通常の単語の重みの4倍にし，その語の係っているオノマトペ以外の語の重みを2倍にする．

以上の手法を表9にまとめる．

表9 実験内容

	重み調整なし	重み調整あり
k-means法(ランダム)	1)	
k-means法(辞書)		2)
クラスタを追加していく方法(辞書)		3)

3.6.2 実験結果

予備実験により，それぞれのクラスタ内の，意味(1)〜(4)の意味でオノマトペが使われている用例の数をそれぞれ示す．2)，3)の手法では，はじめに意味(1)がシードとなっていたのはクラスタ1，意味(2)がシードとなっていたのはクラスタ2，意味(3)がシードとなっていたのはクラスタ3，意味(4)がシードとなっていたのはクラスタ4である．よって，クラスタ1では意味(1)，クラスタ2では意味(2)，クラスタ3では意味(3)，クラスタ4では意味(4)の用例を中心に構成されているのが理想的である．

1)の結果を表10，2)の結果を表11，3)の結果を表13に示す．また，2)の結果の表11を元に再現率と適合率を計算したものを表12に示す．再現率は意味にあった用例の中でそのクラスタがどれだけの用例を含んでいるかという網羅性の指標であり，適合率はクラスタ内に得られた用例中にどれだけ意味に合った用例を含んでいるかという正確性の指標である．3)では，クラスタを増やしていった結果，最終的に15個のクラスタに分類された．

表10 1) k-means法(シードはランダム)，重み調整なし

クラスタ	意味(1)	意味(2)	意味(3)	意味(4)
1	18	4	9	1
2	16	3	6	1
3	9	5	10	0
4	31	11	14	2

表11 2) k-means法(シードは辞書)，重み調整あり

クラスタ	意味(1)	意味(2)	意味(3)	意味(4)
1	18	8	9	2
2	11	8	8	1
3	9	3	11	0
4	32	10	10	1

表12 2)の再現率，適合率

クラスタ，意味	再現率	適合率
クラスタ1，意味(1)	18/70	18/37
クラスタ2，意味(2)	8/29	8/28
クラスタ3，意味(3)	11/38	11/23
クラスタ4，意味(4)	2/4	1/53

3.6.3 考察

表10からわかるように，1)の手法では，1つのクラスタにいろいろな意味の用例が混在してしまっている．よって，1)の手法をそのまま本辞典の用例の分類に適用することはできない．

2)の手法では，辞書の説明文をシードとし，オノマトペが係っている語と，その語に係る語の重みを強くしたので，1)の手法よりも良い結果が得られるのではないかと期待していたが，表12の再現率，適合率の低さからもわかるように，それぞれのクラスタは理想的な意味番号の用例で構成されていない．1)の結果と同様にいろいろな意味の用例が混在してしまった．その理由として，各ベクトルをクラスタに分類する際，どのクラスタにもあてはまらなかったベクトルを，ランダムに選んだクラスタに入れていることが挙げられる．実際，第1回目のクラスタリング処理で，どのクラスタにもあてはまらなかったものは8割以上であった．そのため，重みの調整や初期シードに関

表 13　3) クラスタを追加していく方法 (シードは辞書), 重み調整あり

クラスタ	意味 (1)	意味 (2)	意味 (3)	意味 (4)
1	21	5	2	2
2	6	2	4	0
3	3	0	9	0
4	0	6	3	0
5	3	0	1	0
6	6	1	1	0
7	2	5	0	0
8	9	2	4	0
9	3	0	0	0
10	3	0	2	1
11	1	0	3	0
12	0	0	6	0
13	3	0	0	0
14	4	0	1	0
15	3	0	1	1

わらず，複数の意味の用例が混在してしまっている．

3) の手法に関しては，2) と同じく，クラスタ 1〜4 に対して，意味 1)〜4) の用例数は少なくなっているが，クラスタ 6 内に意味 (1) の用例のみが多く集まっているように，特定の意味を多く持つクラスタがたくさんある．他にはクラスタ 7 の意味 (2),8 の (1)，9 の (1)，12 の (3)，13 の (1)，14 の (1) がある．ただし，3) の手法でも依然として意味の異なる用例が混在している他，辞書上で同じ定義に入るものが複数のクラスタに分割されてしまっている．

今後，これらの手法の改善や，他の手法についての検証を行い，より効率的に用例を分類できる手法を検討していく．

4. オノマトペとオノマトペの関連性の可視化

オノマトペの意味・用法を知る際，オノマトペと，そのオノマトペが用例中で係っている語の関係が重要となる．学習者が，オノマトペ同士をオノマトペが係っている語を介して関連付けて考え，より理解を深められるようにするため，我々は，オノマトペがどのような語に係っているのか，またその語にどのようなオノマトペが係るのかを学習者が視覚的に判断できるようにオノマトペと係っている語を Java インタフェースを用いて可視化することにした．可視化の結果例を図 2 に示す．例えば，「雨がしとしと降る.」の場合,「しとしと」と「降る」が結ばれる．あるオノマトペが特定の単語に係っている用例数によって線の長さが変わる．例えば,「しとしと」というオノマトペが「降る」という単語に係っている用例が多く抽出された場合は,「しとしと」と「降る」を結ぶ線の長さは短くなる．このように，オノマトペと用言の関係を視覚化することにより，以下の 2 点が利用者に分かりやすく提示される．

1) あるオノマトペに係る用言のリストとその数.

図 2 では「しとしと」は,「降る」,「降り注ぐ」,「うつ」などの用言に係ることが一目で見てとれる．

2) ある用言に対して使うことのできるオノマトペリスト.

例えば,「降る」は,「ちらほら降る.」や「しとしと降る.」というように使うことができることが分かる．

このような可視化結果を元に，連鎖的にオノマトペの語彙を増やすことができるため，学習者にとって効果的なツールとなり得ることが分かった．

また, ノードの分布や偏りからもいろいろなことが見てとれる．例えば，複数のオノマトペが共有の用言に係ることが多い場合，そのオノマトペのノード同士は近くに表示されることになる．結果として，係る用言が「行く」,「する」,「見る」など汎用的なものが多い場合，そのオノマトペは中心に集中するようになる．例えば，図 1 では,「しっかり」,「のんびり」などがあたる．また逆に,「しとしと」,「すっぱり」は特徴的な用言が使われることが分かる．

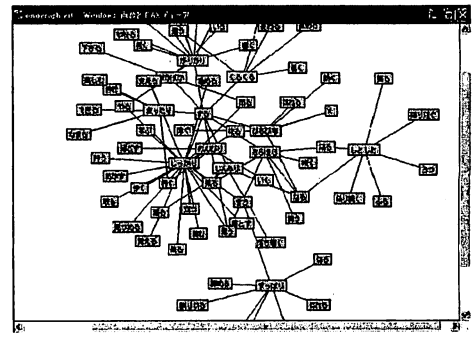

図 2　可視化のイメージ
Fig. 2　image of view

5. まとめと今後の課題

本稿では，オノマトペの用例を自動抽出し，それらを日本語学習者に提示するというオノマトペ用例辞典における用例の分類手法への取り組みについて述べた．

今後は，今回提案した手法を実際に実装し，検証していき，また検索インタフェースに機能を追加したりレイアウトを考えていくことでオノマトペ用例辞典が学習者にとってより使いやすいものになるようにしていきたい．

文　献

[1] 浅賀　千里，渡辺　知恵美.: “Web コーパスを用いたオノマトペ用例辞典の開発,” 電子情報通信学会 第 18 回データ工学ワークショップ，B9-2 2007.

[2] 浅賀　千里，渡辺　知恵美.: “オノマトペのオンライン文例辞典の構築に向けて ,” 第 25 回ことば工学研究会.

[3] “YahooAPI,” http://developer.yahoo.co.jp/category/.

[4] 工藤 拓.: “CaboCha/南瓜,” http://chasen.org/ taku/software/cabocha/.

[5] George Chang , Marcus J.Healey , James A.M.McHugh and Jason T.L. Wang. : “ Web マイニング,” 共立出版, 197p.

[6] “英語表現辞典　擬態語・擬音語集 スペースアルク,” http://home.alc.co.jp/.

[7] “擬音語・擬態語 - 日本語を楽しもう! - ,” http://jweb.kokken.go.jp/gitaigo/index.html.