

# Web コーパスを用いたオノマトペ用例辞典の開発

浅賀 千里<sup>†</sup> 渡辺知恵美<sup>††</sup>

<sup>†</sup> 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: <sup>†</sup>asaga@db.is.ocha.ac.jp, <sup>††</sup>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いといわれている。オノマトペの意味・用法を理解するには、複数の用例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変わっていくことから、最新の用例を知ることが重要となる。そこで、今回は Web コーパスから数多くの新しい文章を、また、青空文庫から用例として適切な文章を自動抽出し、提示できるようなオノマトペ用例辞典の開発を提案する。また、学習者が用例をより理解しやすいよう、オノマトペの係り受け関係で用例をグループ化することも考えている。

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

## Developing an Online Onomatopoeia Example-based Dictionary using Web corpus

Chisato ASAGA<sup>†</sup> and Chiemi WATANABE<sup>††</sup>

<sup>†</sup> Department of Information Sciences, Faculty of Science, Ocyanomizu University

E-mail: <sup>†</sup>asaga@db.is.ocha.ac.jp, <sup>††</sup>chiemi@is.ocha.ac.jp

**Abstract** Onomatopoeia, which is imitative word such as "doki-doki", "shikkari", is express concrete phenomenon and it plays a crucial part in communication. It is difficult for learners of Japanese who study Japanese onomatopoeia to understand their means and usages even they have a good skill of Japanese. It is because onomatopia is sensuous and the words are different in different language. It is said that reading lot of sentences with onomatopoeia can help learners understanding how to use onomatopoeia effectively. In addition, onomatopoeia's mean change frequently with times, learners may want to know "lively" usage of onomatopoeia. Then we propose and develop an online onomatopoeia example-based dictionary which collects a lot of sentences with onomatopoeia from Web corpus. We collect sentences by two corpus, they are Yahoo! search results and "aozora bunko" which is digital library for non-copyrighted novels. We propose some heuristics for selecting good sentences for onomatopoeia's examples for learners, and show our proposed methods can select good sentence effectively by its experimentation.

**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、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。オノマトペは感覚的なものであるため、音に対する感性や、文化的、歴史的背景が異なっていると習得は難しい。また、オノマトペはひらがな、あるいは力

タカナで表記されるため、漢字を習得している中上級の日本語学習者であっても、その意味を理解するのは困難である。

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、新しいオノマトペが頻りに作られていることから、常に新しい用例を知ることが重要となる。

先行研究 [5] にて香林らは、オノマトペの用例を 3ヶ国語以上で同時に一画面に表示し、オノマトペの意味を用例を用いて表

のようなオンライン多言語辞書を構築した。

このオンライン多言語辞書は以下の特徴を持つ。

- 平易な日本語が使われている「窓際のトットちゃん」から手で文章を抽出し、それを意味がわかるようにまとめているので、用例として適切な文章のみ取得することができる。
- 日本語、英語、中国語、韓国語に訳された用例が同時に表示できる。

この辞書では、用例として適切な文章を得ることができるが、最新のオノマトペの用例まで網羅することはできない。更に、用例の抽出を手動で行っているため開発コストがかかる。

そこで、本研究では Web コーパスから数多くのオノマトペの最新の用例を自動抽出し、日本語学習者のためのオノマトペ用例辞典を開発する。コーパスとは言語研究のために大量に収集された書き言葉および読み言葉のテキストであり、Web 空間上の膨大な Web ページから収集されたコーパスを「Web コーパス」という。本辞典は Web コーパスを利用することで最新のオノマトペの用例を数多く提示することができる。Web サイトは個人の判断で作成されているため、適切ではない文章が含まれている可能性もある。そこで我々は、オノマトペの用例の適正度を高く保つための手法をいくつか提案する。なお、先行研究の辞書では多言語に渡って対訳を表示することができ学習者の理解をより深めることができるが、本研究では日本語のみの用例を抽出する。

## 2. オノマトペ用例辞典

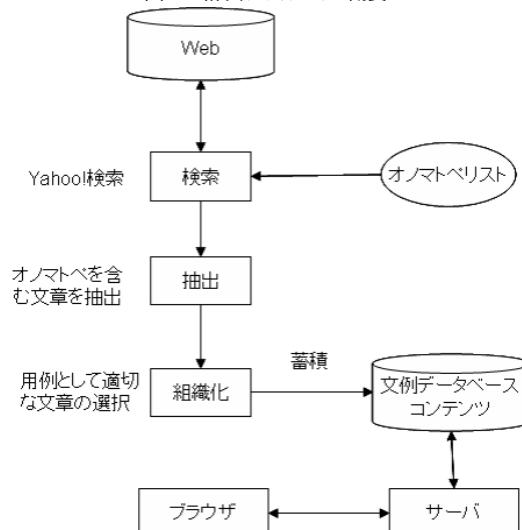
本研究で開発するシステムはユーザがオンライン上で検索したオノマトペを含んでいる文章を Web から抽出し画面上に表示することで、日本語学習者にオノマトペの用例を提示するシステムである。

本辞典の特徴は、Web サイトと青空文庫 [11] の両方から用例を取得することにある。Web サイトから抽出する用例は新しく、また、数も多いが、Web サイトは個人が経営していることが多いため、用例として適切ではない文章も多く抽出されてしまう。それに対して、青空文庫から抽出した文章は著作権の消滅した作品から取得してきているため用例として適切である文章のみ抽出できる。だが、青空文庫から抽出した用例は新しさにかける。このように性質の異なる 2 つから用例を抽出する。

### 2.1 システムの流れ

オノマトペ用例辞典のシステムの流れを図 1 に示す。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、あるいは青空文庫のページを Yahoo!API [10] を用いて検索し自動取得する。それらのページの中からそのオノマトペを含んでいる文章を抽出する。また、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、オノマトペに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 [1]。用例の抽出方法については 2.2 節、用例の適正度については 3 章、用例を格納するテーブルについては 2.3 節、検索インタフェースについては 4 章にて詳しい内容を述べる。

図 1 辞典システムの概要



### 2.2 用例の抽出

自動取得したページからオノマトペの含まれている文章を抽出する。まず、取得ページの HTML 文から指定したオノマトペよりも後ろにある文から終了タグを抽出する。終了タグが抽出できたら、その終了タグの開始タグを指定したオノマトペより前にある文から抽出する。そして、抽出された開始タグと終了タグに挟まれている部分を抽出する。その抽出された部分のうちオノマトペより前にある文と後ろにある文から文の終了を表すのに一般的に用いられる「。」「…」「・」「!」「?」「:」などの記号で挟まれている部分を抽出し、それを用例とする。

### 2.3 スキーマ定義

自動抽出してきた用例を格納し、検索インタフェースを作成するために以下のようなテーブルを定義した。

```
onomatopes
(id int ,
 onomatope varchar)
```

```
sentences
(id int ,
 onomatope_id int ,
 url varchar ,
 cache varchar ,
 sentence text ,
 modword varchar ,
 modpart varchar ,
 attach varchar ,
 reftitle varchar)
```

onomatopes は我々がオノマトペ (onomatope) を手動で入力したオノマトペリストである。それぞれのオノマトペには ID(id) を付与する。sentences は自動抽出してきた用例 (sentence)、オノマトペが用例内で係っている語 (modword) とその品詞 (modpart)、付属語 (attach)、抽出元の URL(url) とキャッシュページ (cache)、そのページのタイトル (reftitle) を格納

するためのテーブルである．それぞれの用例には ID(id) を与える．

オノマトペの係り受け関係，付属語については 3 章において詳しい内容を述べる．

### 3. 用例として適正な文章の抽出

本辞典は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，また「ほかほか」などオノマトペ単体で使われているものなどが多い．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，オノマトペの文法的性質を辞典に取り入れることを考えている．

#### 3.1 オノマトペの文法的性質を利用した用例の抽出

オノマトペは，基本的に副詞で，用言（動詞，形容詞，形容動詞）を修飾するものが主であるが，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果たす．オノマトペの品詞と付属語の関係を表 1 に示す．また，オノマトペは「ほかほか陽気」などのように複合名詞として用いられる場合も多い．

我々は，Web から抽出された文章の用例としての適正度を向上させるために以下の 2 点の手法を提案する．

1) Yahoo!検索時に表 1 にある付属語を付けての検索と付属語を付けない検索の両方を行い文章を収集する．付属語をつけることにより，効率的に用例として適切な文章を集めることができる．

2) 1) で収集した文章の中で，オノマトペが用言または名詞を修飾するもののみを抽出する．ただし「～だ」で終わるものと「～する」動詞はその限りでない．

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

#### 3.2 文法的性質を用いた用例抽出の精度を調べる実験

提案手法の有効性を検証するにあたって以下の実験を行った．

##### 3.2.1 実験の内容

本辞典システムを使用し，3.1 節の手法を適用した場合，抽出した文章がどのくらいの割合で用例として適切であるかを調べる目的で行った．

「ほかほか」「しとしと」「しっかり」「のんびり」「ぴちぴち」「すっぱり」「くるくる」「ちらほら」「しんみり」「まったり」という 10 個のオノマトペに対して，以下の手法で用例を抽出した．

- 1) 付属語をつけて Yahoo!検索した結果から用例を抽出
- 2) 付属語をつけずに Yahoo!検索した結果から用例を抽出
- 3) 付属語をつけて Yahoo!検索した結果から取得した文章を手動でオノマトペの係っている語の品詞ごとに用例を分類し，それぞれ抽出
- 4) 付属語をつけずに Yahoo!検索した結果から取得した文章を手動でオノマトペの係っている語の品詞ごとに用例を分類し，

それぞれ抽出

5) 付属語をつけて Yahoo!検索した結果から取得した文章を cabocha にかける，オノマトペの係っている語の品詞ごとに用例を分類し，それぞれ抽出

6) 付属語をつけずに Yahoo!検索した結果から取得した文章を cabocha にかける，オノマトペの係っている語の品詞ごとに用例を分類し，それぞれ抽出

用例を抽出する対象は Yahoo!検索で上位 50 ページにあげられたページである．

以上の手法を表 2 にまとめる．

表 2 実験内容

	付属語あり	付属語なし
品詞限定なし	1) (図 2)	2) (図 3)
手動で品詞限定	3) (図 4)	4) (図 5)
cabocha で品詞限定	5) (図 6)	6) (図 7)

##### 3.2.2 実験の結果

1) の結果を図 2，2) の結果を図 3，3) の結果を図 4，4) の結果を図 5，5) の結果を図 6，6) の結果を図 7 に示す．図 2，図 4，図 6 では抽出された文章のうち，用例として適正である文章の数と不適正である文章の数，そしてそれらの割合をそれぞれ示す．図 3，図 5，図 7 ではそれに加え，オノマトペが複合名詞として使われている文章の数と割合を示す．また，上記の手法で抽出された文章の例を図 8 に示す．

用例として不適正なものは「ほかほか」などオノマトペ単体が 1 文となっているもの「ほかほかの夢」（『ほかほか』はページの筆者のニックネームとして用いられている）などオノマトペが固有名詞として使われているもの「楽しかったんでよしとしとするしかない。」など文章に誤字や脱字，明らかな造語が含まれているもの「『しとしと』という言葉は...」のようにオノマトペが名詞として使われているものなどである．

複合名詞とは，「ほかほか陽気」など他の名詞と結合して 1 語の名詞となっているものである．

### 3.3 考察

#### 3.3.1 提案手法の有効性に関する考察

図 2，図 3 より，オノマトペ単体を見出し語として検索をかけるより，オノマトペの語尾に付属語をつけて特定の品詞の役割を持たせたものを見出し語として検索をかけた方が適正な用例を高い割合で抽出できることがわかった．図 3 では「ほかほか」などオノマトペ単体を 1 文としている文章が不適正な文章の高い割合を占めているが，図 2 で付属語をつけることでこれらの語の多くを除去することができた．

また，図 3 では複合名詞が多く抽出された．複合名詞は「ほかほか陽気」や「しっかり者」などオノマトペの意味・用法を知るのに役立つものもあるが，抽出された複合名詞のほとんどは「ほかほか日記」や「ぴちぴちピッチ（アニメ名）」など用例として適正だと判断し兼ねるものばかりであった．よって，付属語をつけて検索することで複合名詞も用例から除去できることは有意義であると考えられる．

図 4 と図 5 はそれぞれ図 2 と図 3 で抽出された文章の内，オ

図 2 付属語をつけて検索し品詞を限定しなかった場合の結果

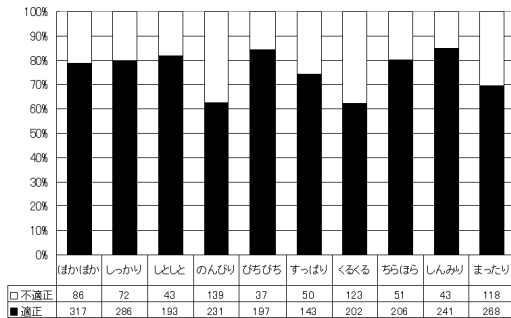


図 4 付属語をつけて検索し品詞を手動で限定した場合の結果

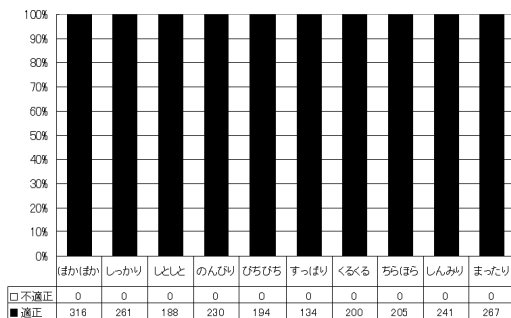


図 6 付属語をつけて検索し品詞を cabocha で限定した場合の結果

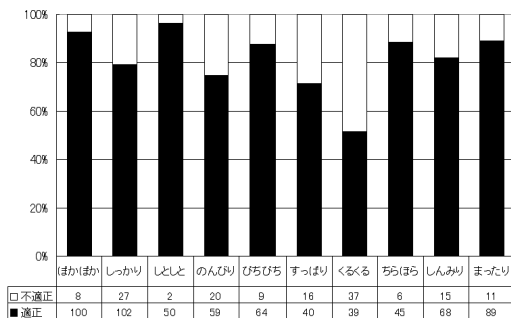


図 3 付属語をつけずに検索し品詞を限定しなかった場合の結果

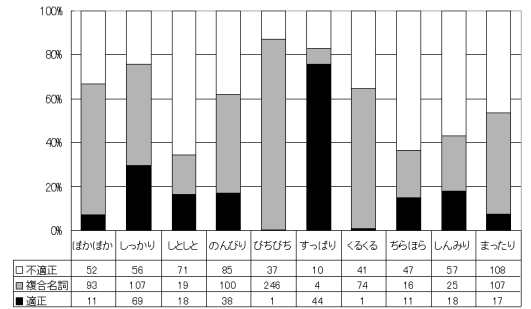


図 5 付属語をつけずに検索し品詞を手動で限定した場合の結果

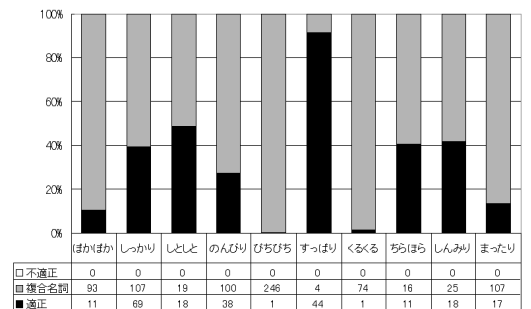


図 7 付属語をつけずに検索し品詞を cabocha で限定した場合の結果

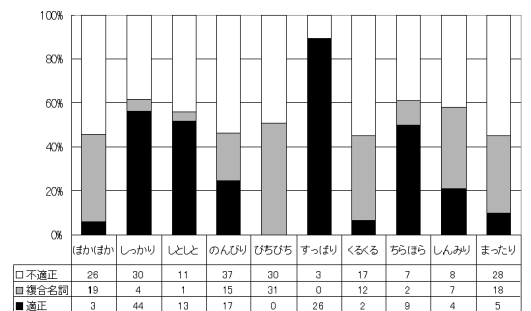


図 8 抽出された文章

付属語なし	適正	「雨がしとと降り出した。」「しっかり温める。」
	不適正	「しととさんのプロフィール」「しとと」
	複合名詞	「のんびり温泉」「しとと横浜」
「と」	適正	「プランをしっかり立てよう。」「ぼかぼかと午後の陽気が背中にあたる。」
	不適正	「しっかりと涙」「しっかりと。」
「な」	適正	「ぼかぼかな陽気」「歯ごたえしっかりな鶏砂肝」
	不適正	「ぼかぼかなの」
「の」	適正	「しととの雨の中を歩いた。」「ぼかぼかのマットで寒い日も春の暖かさ」
	不適正	「ぼかぼかの...」「のんびりのほほ～ん生活」
「だ」	適正	「魚介の風味はしっかりだが強すぎない。」「心も体もぼかぼかだよ。」
	不適正	「のんびりだら～り」「ぼかぼかだ。」
「する」	適正	「笑う君を見てると心がぼかぼかするんだ。」「ぼかぼかする陽気になってきました。」
	不適正	「楽しかったんでよしととするしかない。」「ぼかぼかする。」

ノマトペの係っている語の品詞で用例を限定したものであるが、不適正な文章は全く抽出されなかった。図 2 では、タイトルなどで「のんびりと。」など用例として適正でないものも多ク抽

出されたが、これらの文章に含まれているオノマトペは係っている語がないため、係っている語句の品詞を限定することで除去される。また、「1日1日をしっかりと。」などオノマトペの

係っている語が省略されているものも除去できる。これらのことから、抽出された文章をそのまま用例とするよりも、その中からオノマトペの係っている語の品詞で用例を限定する方が、適正な用例を高い割合で得る上で有効であると考えられる。

全体を通して、文法的には適正であるがあまり一般的には使われないような文章がいくつか抽出された。例えば「ぼかぼかな親子」や「ぼかぼかの願い」などである。オノマトペは感覚的な語彙であるため用法も個人で異なる。今回はこれらの文章も適正な文章の内にいれたが、「ぼかぼかな陽気」など万人が意味・用法認識できるようなものとは別格であるため、本辞典システムでは抽出された文章のオノマトペ、付加した付属語、オノマトペの係っている語を1つの文章として再度検索し用例を集め、その用例の件数でランキングすることを考えている。「今日もぼかぼかと暖かい天気で気分がいい。」という文章が抽出されたら「ぼかぼかと暖かい」で再度文章を Web から集め直すということである。これにより、一般的な用例数が増える一方で「ぼかぼかな親子」のような一般的でない用例との差が大きくなり、上位に現れなくなると思われる。このようにすることで、学習者にオノマトペのより適当な意味・用法を理解してもらえないか考えている。

### 3.3.2 品詞限定の自動化による影響

本辞典では cabocha を用いてオノマトペの係っている語とその品詞を抽出する。そこで、実際に cabocha を使用した結果と手動で係っている語と品詞を判断した結果(理想的な結果)の違いを考察する。

図6と図7は、図4と図5でオノマトペの係っている語の品詞の分類を手動で行っていたところを実際に cabocha を使った行ったものの結果である。図4と図5では用例として不適正な文章が抽出されなかったのに対し、図6と図7では不適正な文章も抽出されてしまっている。その原因は cabocha の係り受けの誤判断であることが多い。誤判断が原因で抽出された不適正な用例としては、「ぼかぼか ~やよいの育児日記」、「ぼかぼかは天然鉱石を精製した製品で...」などがあった。この例では、前者は「ぼかぼか」は「育児」に係り、後者は「製品」に係ると誤判断された。cabocha が係り受けを誤判断する原因としては、オノマトペを副詞と判断するように作られていることが挙げられる。cabocha ではオノマトペは用言に係るというルールがある他、いくつかのルールをベースに係り受けを判断しており、誤認識の例にあるようなサブタイトルや固有名詞としてのオノマトペの利用は想定されていない。このような cabocha の性質により、用例として適正な文章の割合が低くなってしまっている。また、Web ページから文章を抽出しているため、完璧に文章を抽出できているわけではない。そうなると文法が正しくないものが含まれているということになり、cabocha は対処できない。そこで、これらの対策として、用例を手動で修正することを考えている。

### 3.4 青空文庫からの用例抽出

本辞典では、用例として適正な文章を抽出するために青空文庫から文章を抽出する。青空文庫には、旧仮名遣いが使われているページが存在し、それでは外国人学習者はオノマトペの用法を理解しにくくなってしまおうと考えられるため、現代仮名遣

いで書かれているページからのみ用例を抽出する。

### 3.5 青空文庫からの用例の精度を調べる実験

青空文庫から用例を抽出した内、用例として適正な文章の割合を調べるため以下の実験を行った。

#### 3.5.1 実験の内容

青空文庫でどのくらいの割合で用例として適正な文章を得られるのか、また cabocha にかけた際に適切な係り受けがどのくらいの割合で抽出できるのかを調べるため、3.2.1の1)~6)の手法で青空文庫から用例を抽出した。対象は上位50ページである。見出し語としたオノマトペは3.2.1と同じものである。

#### 3.5.2 実験の結果

1)~6)の結果を表3に示す。結果はオノマトペそれぞれについて用例として適正、不適正、また複合名詞を数を出しまとめたものである。

表3 青空文庫での結果

	付属語	品詞限定	適正	複合名詞	不適正
1)	あり	なし	41	0	0
2)	なし	なし	16	0	3
3)	あり	手動	39	0	0
4)	なし	手動	16	0	0
5)	あり	cabocha	12	0	4
6)	なし	cabocha	9	5	0

#### 3.5.3 実験の考察

表3より、かなり高い割合で適正な文章を抽出できていることがわかる。だが、用例数がとても少ない。それは現代仮名遣いのページに限定しているために50ページ全てを抽出できていないためである。また、現代仮名遣いのページの限定しても日本人でも理解しにくい用例も抽出される。例えば「割れたる面はぴちぴちと氷を砕くが如く粉微塵になって室の中に飛ぶ」(「薙露行」夏目漱石)は躍動感の伝わる文章であるが、日本語を勉強する外国人には多少難しいかもしれない。また、青空文庫からは最新の用例が抽出されない。「さくさく」というオノマトペは昔から擬音語として使われてきたが、現代では「さくさく片付けて遊ぼう。」など「手際よく、迅速に」という意味で使われることもある。このように現代になって出てきた用例は青空文庫からは抽出されなかった。よって、一般のWebページからの抽出も必要となる。

## 4. 検索インターフェース

あらかじめユーザにはWebページと青空文庫のどちらから用例を抽出するのを選択してもらう。Webページから抽出した用例の表示画面のイメージを図9に示す。

ユーザが用例を知りたいオノマトペを選択すると、その用例と、用例の中でオノマトペが係っている語句が一括表示される。用例は、上からオノマトペに「と」、「の」、「な」、「だ」、「する」をつけて抽出した用例、付属語を付けずに抽出した用例の順で表示される。また、そのそれぞれについて、オノマトペの係っている語の数でソートされている。1つの用例でしか係られていない語を含む文章は用例から削除している。用例にはそれぞれ「URL」がついておりクリックするとその用例の抽出元

のキャッシュページが参照できるようになっている。また、用例として不適切だと思われる文章はユーザが編集できるようになっている。

また、学習者の理解をより深めるためにそれぞれの用例にユーザがコメントをつけられるようにする。

図 9 用例表示画面のイメージ

オノマトペを選択:  ぽかぽか  暖かい  暖かい

「ぽかぽか」の用例(07件)

用例	係る語	出典	編集
ぽかぽかと暖かい日差しが辺りを照らす	暖かい	URL	Edit
ぽかぽかと暖かい日差しが降り注ぐ縁側で、雅は自分の引っこかっていた桜の木を見上げながらぼんやりとしていた	暖かい	URL	Edit
ぽかぽかと暖かくなり、風に透明感を感じるようになったら、桜の花も春の訪れを告いだ	暖かい	URL	Edit
ぽかぽかと頬から暖かく、道すがら紅梅の匂いが鼻についてくる	暖かい	URL	Edit
今日はぽかぽかと暖かい日になりました	暖かい	URL	Edit
今日はぽかぽかと暖かい横浜です	暖かい	URL	Edit

## 5. 関連研究

近年膨大な Web 空間からの情報抽出が注目を集めており多くの研究が行われている。藤井らが開発している Web コーパスによる事典 cyclone [8] は Web 空間から用語説明として適切な文章を抽出することで膨大な内容量の事典を実現している。また、Web による人物情報の抽出 [3] や人物のニックネームの抽出 [7] などの研究もある。これらは目的に応じて適切な文章を抽出するためにヒューリスティックスを利用しており、我々もこれらの研究と同様、オノマトペを含んだ例文に関するヒューリスティックスを適用している。しかしながら、ヒューリスティックスを利用した情報抽出はある程度の精度の向上は期待できるが、確実性は保障できない。そのため抽出結果の人的修正が必要である。その点で我々は Wikipedia [9] のような利用者参加型の手法を組み合わせることも検討している。

外国人に対する日本語教育のための e-learning 教材に関しては、東工大留学生センターが提供している「あすなる」[6] がある。あすなるは科学技術用語を含んだ文書を対象に、数力国語の対訳と日本語文書の文法や語句の説明などを表示することできめ細かい学習コンテンツを提供している。オンラインのオノマトペ用例辞典に関しては国立国語研究所の提供するページ [13] やアルクが提供するページ [12] があるほか、音声や画像を用いたマルチメディア辞書 [2] などもある。しかしながらこれらのシステムが用いる用例はすべて提供者が集めたものであり、その数には限りがあるため、我々が Web 空間から膨大な用例を抽出することができれば、これらのシステム構築・運営の一助になると考えられる。

## 6. 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ついて記した。

今後は、オノマトペとオノマトペが係っている語を用いた用例パターンをだけでなく、オノマトペが係っている語以外の語との用例パターンを抽出することも考えている。例えば「陽射しがぽかぽかと暖かい」の「陽射し」と「ぽかぽか」のパターンなどである。また、それを使ってオノマトペ同士を関連付けることも考えている。

また、既存の国語辞典等からオノマトペの複数の意味を取得し、検索条件式や意味の分類に利用することも考えている。

そして、4章で述べた用例として適正でない文章を修正できるような利用者参加型の手法をシステムに組み込む。それに関連してユーザが自作の用例を追加できるようにするシステムを考えている。ただし、個人の判断で用例を追加してしまうと用例として不適正な文章が用例に追加されてしまう恐れがあるため削除フラグをそれぞれの用例に付加し、ユーザに削除した方が良いと思う文章にはフラグを立ててもらい、それが一定量を越えたら削除されるというようなシステムも考えている。

## 文献

- [1] 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.
- [2] 越智 洋司, 川崎 桂司, 矢野 米雄, 林 敏浩.: “外国人のための擬態語・擬音語辞書システム” JAMIOS”の構築,” 電子情報通信学会論文誌 D- Vol.J80-D- No.12 pp.3210-3219 1997 年 12 月.
- [3] 木村 隼, 小山 聡, 田中 克己.: “人物情報辞典の自動生成のための時系列情報自動収集” 情報処理学会データベースシステム研究会研究報告, No.2006-DBS-140(11), pp.51-58 (2006).
- [4] 工藤 拓.: “CaboCha/南瓜,” <http://chasen.org/taku/software/cabocho/>.
- [5] 香林 隆子.: “オノマトペのオンライン多言語辞書の構築,” DEWS2002 論文集, DEWS2002 A4-4, March 2002.
- [6] 仁科 喜久子, 奥村 学.: “Japanese Reading System of Multi-Lingual Environment ” ASUNARO”” 日本語教育方法研究会誌, Vol.7, No.1, pp.16-17.
- [7] 外間 智子, 北川 博之.: “Web データを用いた人物の呼称抽出” 日本データベース学会 Letters, Vol 5, No.2, pp49-52 (2006).
- [8] 藤井 敦, 石川 徹也.: “World Wide Web を用いた事典知識情報の抽出と組織化,” 電子情報通信学会論文誌, J85-D-(2):300-307, 2002.
- [9] “Wikipedia,” <http://ja.wikipedia.org/wiki/>.
- [10] “YahooAPI,” <http://developer.yahoo.co.jp/category/>.
- [11] “青空文庫,” <http://www.aozora.gr.jp/>.
- [12] “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/>.
- [13] “擬音語・擬態語 - 日本語を楽しもう! - ,” <http://jweb.kokken.go.jp/gitaigo/index.html>.