

オノマトペ文例の理解を支援する画像データの収集と提示

Kanwipa Lertsumruaypun (指導教員：渡辺知恵美)

1 はじめに

オノマトペとは「がりがり」「ふわふわ」のような擬音語・擬態語である。オノマトペは感覚的な言葉であるため、外国人日本語学習者にとって理解することが難しい。我々はこれまで日本語学習者の中上級者を対象に、オノマトペのオンライン用例辞典オノマトペディアを開発してきた [1]。オノマトペディアは、Web 上から文章を自動収集し、オノマトペの文例として適切に表示する。しかし、オノマトペディアでは文例のみ表示させるため、オノマトペを理解しにくい場合もある。そこで、本研究の目的は画像検索エンジンを用いてオノマトペディアの文例に合った画像を収集し、文例とともに表示することを目的とする。しかし、オノマトペで画像検索をすると文例に関連のない画像が多数含まれる。そこで、本研究は対象文例と画像との意味の合致度に着目した下記の 3 パターン (A, B, C) の画像収集法を提案する。

- 収集法 A : 文例からオノマトペとオノマトペに係る語を抽出し、それらをキーワードとして画像を検索
- 収集法 B : A のキーワードを助詞や付属語を繋げて文章の一部としたキーワードで画像を検索
- 収集法 C : 文例からオノマトペ、係る語、主語、目的語を抽出し、それらをキーワードとして画像を検索

これらの手法は画像の合致度と収集画像数がトレードオフの関係にある。そこで、どの方法で取得した画像がオノマトペを理解するうえで最も有用かを日本語学習者 (中上級) によるアンケートで選ぶ。また、検索結果画像の付随文章に着目した 3 つの画像表示方法を提案し、検索キーワード同様アンケートによって最も理解しやすい表示方法を選ぶ。最終的に文例と画像を表示するインタフェースを提案する。

2 画像収集法の検討

2.1 単純なキーワード検索による収集

収集法 A は文例からのオノマトペ及びそれに係る語を検索キーワードにして画像収集する。例えば、図 1 で示す文例から「がりがり」及び「噛む」を抽出する。続いて、係る語である「噛む」の全活用形を「がりがり」また「がりがり」と組み合わせた全パターンで画像検索を行う。図 2 は全検索キーワードを示す。

図 1 のように、収集法 A による検索結果画像数 (計 290 件) は多いが、7 人の日本人を対象とした文例と検索結果画像 (ランダムに 30 件を選択) との合致度を評価した結果、合致する画像の割合の平均は 28.1 % となった。よって、検索結果の中には合致しない画像が多数含まれることがわかった。また、収集法 A は同じオノマトペと係る語を持つ異なる文例の場合は同じ結果が得られてしまう。



図 1: 収集法 A における画像検索結果



図 2: 収集法 A における画像検索キーワード

つまり、収集法 A の結果は多数であり、また合致する結果画像の割合が低いため、オノマトペを学習する際どの画像がオノマトペの意味を表すかを判断することが困難であると考えられる。そこで、より文例に合致した画像を収集する方法を提案する。

2.2 より文例に合致した画像収集法

収集法 B はより文例に合致した画像を収集するには助詞や付属語を検索キーワードに付ける。図 1 の文例からの助詞 (は, が, を), 付属語 (に, と, -(付属語なし)) を収集法 A のキーワードに加えると「をがりがり」と「噛む」のキーワードとなる。しかし、これによる検索結果は 0 件であった。検索結果画像数を増加させるために、動詞の全活用形を求め、図 3 に示すようにこれらの全組合せをキーワードにして画像を検索する。

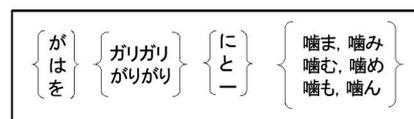


図 3: 収集法 B における図 1 の文例の画像検索結果キーワード

収集法 B による検索結果画像数は 44 件となった。また、助詞や付属語をキーワードに含めることによって、文法的に正確な画像の付随文章が得られる。しかし、収集法 B の検索キーワードには主語や目的語のような名詞を含んでいないため、検索結果画像のシチュエーションが様々であり、学習の際に混乱が生じると考えられる。そこで、より文例のシチュエーションに合った画像を取得するために収集法 C を提案する。

収集法 C は文例から抽出した主語や目的語を検索法 B のキーワードに加え、画像を検索する。例えば、図 1 の文例についての検索キーワードは図 4 に示す。

図 4 で示しているキーワードの検索結果画像数は 2 件となった。収集法 C を用いることによって文例と同じ主語や目的語を持ち、文例との合致性がある画像が得られるが、検索キーワードの限定度が高いため、



図 4: 収集法 C における画像検索結果

文例によって画像が収集できない場合が多い。どの方法で取得した画像がオノマトペを理解するうえで最も有用かを日本語学習者（中上級）によるアンケートを行った。図 5 はアンケートの結果を示す。

収集法A	収集法B	収集法C	計
2	12	2	16人

図 5: 収集法についてのアンケート結果

アンケートの結果、収集法 B が最も理解しやすいことがわかった。さらに、収集法 C は画像の数が少な過ぎる。また、収集法 A は画像が多いが付随文章が文法的に正しくないためオノマトペの使い方がわかりにくいという意見が多かった。

2.3 合致度を高めるための文例と画像の再収集法

画像を収集した後、さらに文例と画像の合致度を高めるために図 6 に示すように文例と画像の再収集を行う。

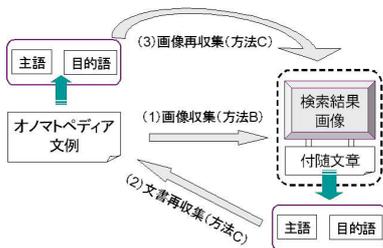


図 6: 文例及び画像の収集手順

図 6 においての手順を下記で説明する。

1. まず、アンケートで最も選ばれた収集法 B で画像及び付随文章を収集する。
2. 付随文章から主語や目的語を抽出し、収集法 C を用いてオノマトペディアの文例を再収集する。
3. 再収集した文例から主語や目的語を抽出し、収集法 C を用いて画像及び付随文章を再収集する。

画像の付随文章からの主語や目的語を利用し、文例を再収集する。また再収集した文例からの主語や目的語を含むキーワードで画像再収集することによって文例と画像の合致度が高められる。

3 画像表示法

オノマトペの学習を支援するためにオノマトペディアにおいて収集した画像及び付随文章の表示方法を以下の 3 通り提案する。

- 表示法 1 : 1 つの文例に画像及び付随文章が割り当てられる。

- 表示法 2 : 文例及び付随文章付きの検索結果画像を別々に表示し、一つの「オノマトペ オノマトペの係る語」のグループ毎に画像が割り当てられる。

- 表示法 3 : 検索結果画像及び付随文章のみを表示する。

どの表示方法が最もオノマトペを理解しやすいかを選ぶために、収集法と同様にアンケートを行った。

表示 I	表示 II	表示 III	計
7	5	4	16人

図 7: 表示法についてのアンケート結果

アンケート結果より、表示法 1 が最も選ばれた。しかし、表示法 1 のように例文と関連がある画像があったら学習しやすいが、文例に対応する画像がない場合は表示法 2 で画像を提示して欲しいという意見が多かった。

4 文例と画像を表示するインタフェース

ユーザによるアンケート結果及びコメントより、図 8 のようなインタフェースを提案する。



図 8: 提案するインタフェース

インタフェースの概要については、収集法 C を利用し、オノマトペディアの文例に一致する画像を得られる場合は、表示法 1 を適用する。そして、各「オノマトペ 係る語」グループ毎に表示法 2 で収集法 B の検索結果画像を表示する。

5 まとめと今後の課題

本稿は検索エンジンを用いて画像をインターネット上から自動的に収集し、文例とともに適切な画像を表示する方法を提案した。また、日本語学習者を対象とするアンケートを行った。今後、日本人による結果検証実験を行う。

参考文献

- [1] Chisato Asaga, Yusuf Mukarramah and Chiemi Watanabe.: ONOMATOPEDIA: Onomatopoeia Online Example Dictionary System Extracted from Data on the Web, Proceedings of Asia-Pacific Web Conference (APWeb'08), pp.601-612 (2008)
- [2] L.Kanwipa, 浅賀千里, 渡辺知恵美: オノマトペ学習支援システムにおける画像検索結果と文例の関連付け, データ工学と情報マネジメントに関するフォーラム (DEIM2009), D6-5 (2009) (発表予定)