1M1-OS-8a-5

# 共起情報抽出による日韓オノマトペの意味差の分析に向けて

# Towards A Contrastive Analysis of Japanese and Korean Onomatopoeic Expressions with Co-occurrence Extraction

橋本 喜代太[*1] Kiyota Hashimoto
竹内 和広[*2] Kazuhiro Takeuchi

[*1] 大阪府立大学現代システム科学域 College of Sustainable System Sciences, Osaka Prefecture University
[*2] 大阪電気通信大学情報通信工学部 Faculty of Information and Communication Engineering Osaka Electro-Communication University

岡田 真[*3] Makoto Okada
廣川 佐千男[*4] Sachio Hirokawa

[*3] 大阪府立大学大学院理学研究科 Graduate School of Science, Osaka Prefecture University
[*4] 九州大学情報基盤研究開発センター Research Institute for Information Technology, Kyushu University

Onomatopoeic expressions are unique as part of sentiment expressions in that their descriptions/meanings are largely shared by a particular linguistic community, and this characteristics opens a possibility to specify a finer-detailed meaning extraction with computer. On the other hand, onomatopoeic expressions are bound by language variations and some difference is to be found between very similar onomatopoeic expressions in different languages. In this study, as a starting point of a project of multi-language analysis of onomatopoeic expressions, we picked up Japanese and Korean and analyzed a small set of similar onomatopoeic expressions in the two languages by observing co-occurring words. We revealed that their different behaviors are successfully captured and discussed their implications.

## 1. はじめに

オノマトペ(擬音語・擬声語・擬態語・擬容語・擬情語などを総称して本稿ではオノマトペと呼称する)は現実世界の音を模倣したり，行為・状態の様態を音表象したりする表現である．どの言語にもオノマトペは見られるが，その模倣・表象は各言語の音韻体系等に従うため，各言語によって大きな異なりを見せると同時に，その数・利用頻度も異なる．よく知られるように英語をはじめとする欧米の言語ではオノマトペは幼児的表現とみなされていることが多く，特に擬態語の数や頻度は少ない．一方，日本語や韓国語などは口語はもちろん文語においてもきわめて広範囲にオノマトペを使用し，その数は少なく見積もっても数千，変異形や地域的・世代的・個人的な造語も含めれば万の位にも及ぶと考えられる．これはオノマトペの中でも擬態語ないし擬態語用法に分類される使用が多いということでもある．

日本語では学術的にも一般的にもオノマトペは古くから関心を持たれ，[小野 2007]を代表としていくつものオノマトペ辞典が公刊され，[田守 1999]のような言語学的研究，[奥村 2003]のような計算機による抽出や辞書構築の研究などが多数行なわれてきている．一般にオノマトペは行為・状態の様態を音表象するから一般的な形容詞や副詞などと同じく，感性評価表現の一種と考えることもできる．この点に着目して[橋本 2011]のようなオノマトペの感性極性値の推定のような研究も出てきている．

オノマトペを感性評価表現と考える場合，重要な点が 2 つある．一つは当然ながら，その描写対象によって同じオノマトペが異なった意味を持ち，感性極性値もそれに応じて異なる，ということである．このため，粒度のよい意味抽出や感性極性値抽出を行なうのであれば，文脈情報，もっとも典型的な共起情報，の分析が欠かせない．いま一つは一般に感性表現はある対象の描写において感性極性値は共有しても，その具体的な描写・意味そのものは話者の主観に依存するため，大幅に違うことがあり得るが，オノマトペについては具体的な描写・意味が話者間で驚くほど共有されていることが多い，ということである．例えば，「このケーキは固い」といっても具体的にどのような固さなのかは分からない．しかし，「このケーキはかちかちだ」，「このケーキはガチガチだ」，「このケーキはかちんかちんだ」といったオノマトペ表現では大半の日本語話者はその差異を感覚的に共有する．こうした 2 つの特徴を考えれば，より微細な違いを捉えた感性評価分析についてオノマトペの活用はきわめて有望であると同時に，その基盤となるオノマトペの詳細な分析も今後に待つところが大きく，特に複数言語間での翻訳のための意味差の把握は従来以上の慎重さと丁寧さが求められることになる．

筆者らは現在、日本語、中国語、韓国語を含むアジア諸言語のオノマトペ表現の計算機的活用並びにその前提としての言語学的記述の高度化を狙う包括的な研究プロジェクトを始めており、その出発点として言語構造の類似点が多く、かつ、オノマトペの使用頻度がきわめて高い日本語と韓国語にまず着目することが有用であると考えている。そこで、以上を前提にして，本研究では日本語と韓国語の類似したオノマトペについてどのような意味差が見られるかをオンラインのテキストデータを収集したものから共起情報を抽出することによってどの程度明らかにできるかを考察する．ただし，本研究は端緒についたばかりであり，抽出・分析の自動化や精度評価ではなく，それに向けての事前調査的考察であることを了解されたい。

連絡先：橋本喜代太，大阪府立大学学術研究院電気情報系，
599-8531 堺市中区学園町 1-1, hash@kis.osakafu-u.ac.jp

## 2. 日本語と韓国語のオノマトペの形態特徴

まず，今後のオノマトペ表現の自動抽出を見据え，日本語と韓国語のオノマトペについてその形態的な特徴を見ておこう．いずれの言語でも重畳(反復)があるかないか，語幹に促音，鼻撥音，流音などが加わるか否かで変異形が存在する．こうした変異形は日本語について見ると，概略，表 1 に示すような変異形がほとんどのオノマトペに存在する．これに対して[チェ 2005]などで示されるように韓国語の場合は，語幹に対して，それが単独で使われるもの，それが重畳されるもの，流音をはさんで重畳されるものの大きく 3 種の変異があり，それらについてさらに母音や子音の交替が加わるため，その変異形の数は数え方によっては日本語以上となる．なお，言うまでもなく重畳されるかどうかは描写される様態が繰り返されるものであるかどうかを判断基準とする点は日韓とも同じである．

表 1 日本語のオノマトペの形態類型

| 語幹 | 変異 | 例 |
|---|---|---|
| (C)V<br>eg. さ | (C)V + っと | さっと |
| | (C)V[long] + っと | さあっと |
| | (C)V(C)V + っと | ささっと |
| | (C)V(C)V[long] + っと | ささあっと |
| $(C_1)V_1(C_2)V_2$<br>eg. きらきら | $(C_1)V_1(C_2)V_2$N | きらん |
| | $(C_1)V_1(C_2)V_2$ +っと | きらっと |
| | 重畳形<br>( X = $(C_1)V_1(C_2)V_2$, N = 撥音, G = 促音) | |
| | X X | きらきら |
| | X X N | きらきらん |
| | X N X N | きらんきらん |
| | X G X | きらっきら(と) |
| | $(C_1)V_1G(C_2)V_2$ X | きっらきら |
| | X G X G | きらっきらっ(と) |
| | $(C_1)V_1$[long]$(C_2)V_2$N | きーらきら |
| | 後続する助詞 | |
| | と | きらきらと |
| | に | きらきらに |
| | で | きらきらで |
| | 動詞化 | |
| | する | きらきらする |

いずれの言語においてもオノマトペは他の表現と形態的に際立った違いを見せており，これにより形態的パターンを指定することによるテキストからの自動獲得がある程度可能である．参考ながらこの点は中国語やインドネシア語のようにやはりオノマトペで重畳を多用する言語にもある程度成り立つ．

## 3. 類似した日韓のオノマトペの意味差

周知のように日本語と韓国語はそれぞれの言語的来歴が比較言語学的には未だに確定されていないにも関わらず，その語形成や統語法は驚くほど類似している．このため，一般に互いの言語を学習しようとした場合，単語的な対応を学習することを中心としてもある程度の習得が可能であり，機械翻訳的にも単語置き換えがある程度成功する．しかしながら，それはあくまでも「ある程度」であって，さまざまな微細な違いがあることは言うまでもなく，これはオノマトペについても言えると考えられる．

そこで本研究では，少数の類似した意味を持つ日韓のオノマトペを取り上げ，それがどのような共起関係を持つかを分析することで意味の違いがどの程度あるかを明らかにすることを事前調査研究として試みた．

### 3.1 データ

分析のため，まず日本語，韓国語それぞれについていくつかの商品口コミサイト(Trip Advisor (http://www.tripadvisor.com/)など)から，5 万文字ずつを対象とした．今回は事前調査という点からデータの母集団に対する網羅性や代表性は問題としておらず，一方で明らかな理由から話し言葉に類似した特徴を持つブログ，掲示板，口コミサイトの文章の方が新聞などよりオノマトペを多く含む．特に複数言語による投稿がなされる Trip Advisor などは複数言語間である程度類似した表現が登場することが期待できる．このため，関連研究で収集済みのデータの中から選択した．

次に，調査対象としてまた，[HANA 韓国語教育研究会 2009]などを参考にして，日韓で類似したオノマトペとして比較的頻度が高いと思われる 30 対を選んだ．その例は「ふんわり」，「かちかち」，「きらきら」などである(詳細は口頭発表時に示す)．

### 3.2 分析

取り上げた 30 対のオノマトペが収集したデータ内に 10 回以上現れるかどうかを確認し，それのみを取り出し，共起する描写対象の異同について観察を行なった．

この結果，まず一方の言語には多く見られるがもう一方の言語にはあまり出てこないオノマトペ対が約半数あった．この理由についてはいくつか考えられるが，後の課題とする．

類似した日韓オノマトペ対で顕著な異同が見られたものの例としては，次の例があった．

「ふんわり」

나긋나긋

いずれも基本的には「柔らかさ」を示すが，日本語では事物を描写するものがほとんどすべてであったのに対して(今回のデータ中では一例のみ，「ふんわり落ちた」があった)，韓国語では「ふんわり笑う」といった日本語では通常出てこない共起が観察された．一方、上記対は[HANA 韓国語教育研究会 2009]で紹介されているものであるが，「ふんわり」の訳語として말랑말랑などを挙げることもある．こうしたこうした対象範囲のずれはオノマトペに限らずどんな言語表現でも複数言語間類義表現に見られるものであるが、オノマトペは他の語集団に比べ、はっきりした意味記述が難しい分、共起関係からこうしたずれを明らかにし，そのずれがどのような意味差を表わしているのかを大規模に調査していく必要があることが分かる．実際、こうした差異についてはいくつかのオノマトペについては既にいくつかの文献や学習書などでも指摘されているが，大規模な調査は行われていない．

## 4. 終わりに

本研究は端緒についたばかりであり，たぶんに印象的な事例のみを報告したきらいはあるが，複数言語においてオノマトペの描写内容，意味を考える際，いかに類似しているように思えても多くのオノマトペ対は共起する語のグループにずれが見られ，それが言語間の類似オノマトペの意味の違いを示唆しているこ

とを示した. この示唆を受けて, 今後, データ数を増やすと同時に, 共起の重み付け等を含めて分析を本格化する予定である.

大規模データを用いて複数言語間のオノマトペの比較記述を行なうことは言語学的関心でもあるわけだが, 機械翻訳などの直接的な計算機的利用はもちろん, 複数言語の記述を含むレビューサイトなどでの活用などへの応用が考えられる. そうした新たな活用についても今後提案していきたい.

## 参考文献

[小野 2007] 小野正弘(編):『日本語オノマトペ辞典』, 小学館, 2007.

[田守 1999] 田守育啓, ローレンス・スコウラップ:『オノマトペ -形式と意味-』くろしお出版, 東京, 1999.

[橋本 2011] 橋本喜代太, 竹内和弘:“ジャンル, 文脈に応じたオノマトペ表現の感性極性値の推定手法の提案, ”人工知能学会第 25 回年次大会, 1C2-OS4b-6, 2011.

[奥村 2003] 奥村敦史, 齋藤豪, 奥村学:“Web 上のテキストコーパスを利用したオノマトペ概念辞書の自動構築, ” 情報処理学会 自然言語処理研究会 2003-NL-154-10, pp.63-70, 2003.

[チェ 2005] チェ・エリカ・ユンジョン:“日韓両語の対照研究ー両語のオノマトペの聴覚的印象の異同の検討ー,”『日本語・日本語文化研修プログラム研修レポート集』第 21 期, 広島大学, 2005, http://www.iie.hiroshima-u.ac.jp/center/activities/japanese/pdf/2006/choi.pdf

[HANA 韓国語教育研究会 2009] HANA 韓国語教育研究会:『音で覚える韓国語の擬声語・擬態語』, アルク, 2009.