

駄洒落に含まれるオノマトペの特徴分析

内田ゆず[†] 荒木健治[‡]
 yuzu@hgu.jp araki@ist.hokudai.ac.jp

[†]北海学園大学 [‡]北海道大学大学院情報科学研究科

1 はじめに

昨今の人工知能技術の発展により、接客ロボットやスマートスピーカーなどが身近になりつつある。それに伴い、対話システムへの需要・期待が増しているが、十分な性能が得られているとは言いがたい。特に非タスク指向型対話システムには、対話が破綻する、システムの発話が不自然である、ユーザが飽きてしまうといった問題がある。

ユーザを退屈させず、自然な対話を続けるためにユーモアを導入することは有効な手段である。しかし、ユーモアの認識や生成を工学的なアプローチで研究しようとする、評価の難しさという壁にぶつかる。ユーモアの面白さを評価する手法が確立されていない上、評価用の標準データセットが存在しないため相対的な評価も難しいのが現状である。

そこで、我々はユーモア研究の基盤整備の第一歩として、駄洒落のデータベースを構築した[1]。現在は、コンピュータによる“面白い駄洒落”の認識や自動生成を目指し、データベース中の駄洒落について様々な観点からの分析に着手している。

佐山らは人間が駄洒落を理解する手続きについて、認知科学的な考察を行った[2]。そして、駄洒落の理解のしやすさと駄洒落の面白さの関係を示した。このように、人間が面白いと感じる駄洒落の特徴を明らかにすることで、コンピュータによる駄洒落の面白さの評価が可能になると考えられる。

駄洒落は同じ音、あるいは非常に近い音をもつフレーズをかけて楽しむ言葉遊びの一種である。駄洒落を成立させるためには、2つのフレーズが全く同じ音である必要はなく、撥音、拗音、促音、長音の追加・削除や、濁音、半濁音への変化にも寛容である。

- 3073 薬を飲んだら、ぐっすり

上記の例では、「薬」と「ぐっすり」がかかっており、濁音への変化と促音の追加がみられる。

なお、この例は我々が構築した駄洒落データベースに収録されており、駄洒落の前の数字「3073」はデータベース中の通し番号である。以降の例でも通し番号を付記する。

例 3073 では、「ぐっすり」というオノマトペ（擬音語・擬態語の総称）が使われている。他にも駄洒落データベースにはオノマトペを使用した駄洒落が収録されている。以下に、オノマトペを含む駄洒落の例を挙げる。

- 3174 ポッキーが折れちゃったポッキーン
- 36899 きちきちのキッチン
- 50645 キッチンはきちんと片付けよう
- 32637 沼の水はヌマヌマしている

オノマトペは撥音、拗音、促音、長音の挿入が比較的自由にできる上、「はらはら」「ばらばら」「ばらばら」など、濁音、半濁音への変化にも柔軟に対応できる。さらに、新語の作成も容易であり、駄洒落との相性が良い。例 3174 では、オノマトペ「ポキン」に促音、長音を追加して、「ポッキー」と近い音をもたせている。例 36899 および 50645 は、「キッチン」に対応するフレーズをそれぞれオノマトペ「きちきち」「きちんと」としている。これらの例から、促音を無視しても駄洒落は成立することがわかる。例 32637 の「ヌマヌマ」は、オノマトペの辞典[3]には掲載されていない新語である。

我々は、駄洒落におけるオノマトペの特徴を捉えることで、駄洒落の自動生成に役立つ知見が得られると考えている。そこで、本稿では駄洒落に含まれるオノマトペについての分析を行う。

2 使用するデータ

2.1 駄洒落データベース[1]

本データベースには 9 つの Web サイトから収集した 51,000 件の駄洒落が収録されている。駄

洒落には、音韻的に似た2つの区間が存在する。駄洒落の基となる区間を種表現、種表現から連想される音韻的に類似した区間を変形表現と呼ぶ。

データベース中の駄洒落にはタグが付与されている。種表現は(), 変形表現は[]で囲まれている。複数の種表現、変形表現が含まれる駄洒落も存在する。たとえば、「猛吹雪の車で、毛布にくるまる」(14408)の種表現は「猛吹雪」と「車」であり、それぞれに対応する変形表現は「毛布」と「くるまる」となる。このような場合、対応する種表現-変形表現のペアに対して各括弧の後ろに同じ番号が付与されている。さらに、形態素解析ツール MeCab[4]を用いて、形態素解析を行い、形態素の区切りに半角スペースが挿入されている。形態素解析誤りは修正されていない。

以上より、例 14408 はデータベースに以下の形式で格納されている¹。

- ・ 14408 (猛吹雪)1 の (車)2 で、[毛布]1 に [くるまる]2

2.2 オノマトペリスト

ある単語がオノマトペであるかを判断する際に、日本語オノマトペ辞典[3]を基準として用いる。この辞典の「五十音順さくいん」には、辞典の本編ならびにコラム、付録(漢語オノマトペ、鳴き声オノマトペ)に収録されている全 4,506 語が掲載されている。このうち、漢語オノマトペを含む表現(例: 焔焔に滅せせんば炎炎を若何せん)や、オノマトペではないがコラムで言及されている語(例: あいまい)を除いた 4,314 語を分析対象のオノマトペとする。

3 オノマトペを含む駄洒落の抽出

駄洒落におけるオノマトペの出現傾向を定量的に評価するため、オノマトペを含む駄洒落を抽出する。

著者らによるこれまでの研究で、3文字以下の短いオノマトペは形態素解析誤りが多く発生することが報告されている[5]。さらに、駄洒落は口語表現や新語が多く含まれることに加え、文法上の不自然さが許容されるため、形態素解析の精度は低くなることが予想される。したがって、オノマトペの抽出には形態素解析を用いないこととした。

¹ 実際には、駄洒落の種類(併置型、重畳型)の情報も付与されているが、紙面の都合上説明を割愛する。また、これ以降の記述では形態素区切りのスペースは省略する。

具体的な抽出手法は、以下の3ステップである。

- ① 種表現および変形表現中に、2.2 で述べたオノマトペと字面が完全一致する文字列を含む駄洒落をすべて抽出する
※種表現、変形表現以外の部分にあるオノマトペは分析対象外
- ② ①で一致した文字列がオノマトペかどうかを第一著者が人手で判断する
- ③ 駄洒落を成立させるために表層が変化しているオノマトペや新語オノマトペを第一著者が人手で抽出する

③は、リスト中のオノマトペと字面が一致しないオノマトペ(つまり、①で機械的に抽出できないオノマトペ)を発見することを目的としている。③で抽出された例 30774 では、オノマトペ「どっさり」を「どっしゃり」へと変形させ、種表現「土砂」の音に近づけている。

- ・ 30774 (土砂)が[どっしゃり]

①のステップで、23,267 件のオノマトペ候補を含む 14,689 件の駄洒落が抽出された。②、③のステップで 553 種のオノマトペが 1,715 件の駄洒落に合計 1,737 回出現していることが明らかになった。

4 オノマトペの出現頻度の特徴

553 語のオノマトペのうち、10回以上出現したオノマトペは表 1 に示す 33 語であった。ただし、駄洒落データベースには類似した駄洒落も多く存在するため、同じ種表現-変形表現が複数回出現することもある。

たとえば、「ネバネバ」は 14 回出現しているが、すべて「ネバーギブアップ」「動詞+ねば」「ネバダ州」のいずれかとの対応である。したがって、種表現-変形表現のペアとしては実質 3 パターンである。今後、オノマトペとそれに対応する表現のペアの単位で分析する必要があるだろう。

- ・ 3814 納豆の(ネバネバ)が手についた。[ネバー・ギブアップ]
- ・ 23604 (ネバネバ)したものを[食べねば]
- ・ 27547 (ネバダ)州の、[ネバネバだ]

表 1 駄洒落中に 10 回以上出現したオノマトペ

順位	オノマトペ	出現回数
1	チュー	37
2	フッ	27
3	ボーッ	26
4	ハッ	23
4	キッ	23
4	チャン	23
7	エーン	22
7	ビックリ	22
9	キヤー	21
9	タップリ	21
11	ヒー	19
12	ホッ	18
13	チョコッ	17
14	サッパリ	16
14	グッ	16
14	スカッ	16
17	ギリギリ	15
17	ツイ	15
19	アッサリ	14
19	ジッ	14
19	ネバネバ	14
22	ジックリ	13
22	ポテッ	13
24	ガッカリ	12
24	サバサバ	12
24	ムッ	12
27	シッカリ	11
27	フト	11
29	ギョッ	10
29	コケッココー	10
29	ツクヅク	10
29	ボイン	10
29	ワッ	10

出現頻度上位に 2~3 文字の短いオノマトペが多い。以下に短いオノマトペを含む駄洒落を挙げる。

- ・ 2585 (中国人)がジュースを飲んだ。[チュー、ゴクッ]

- ・ 17909 (中央政界)では、[チュー旺盛かい]?
- ・ 41597 (フットサル)の途中で、[フッと去る]
- ・ 43759 (富士山)でよく見かけるのは、[ふっ、爺さん]
- ・ 72 (ハゲタカ)だと思ったら、[はっ、ゲタか]
- ・ 17920 (キャベツ)に虫がついてる。[キヤー、別]のにして

これらの例では、種表現をいくつかの文字列に分割して、複数単語から成る変形表現を作っている。たとえば、例 17909 は種表現「中央政界」を「チュウ/オウセイ/カイ」と分割している。このような手法で駄洒落を作成する場合に短いオノマトペは都合が良く、出現頻度が高くなったと考えられる。

なお、553 語のオノマトペのうち、245 語は 1 回しか出現しなかった。分析対象の駄洒落を増やして、大規模な分析を行う必要がある。

5 駄洒落中のオノマトペの特徴

5.1 出現位置

オノマトペの出現位置を種表現・変形表現で分類したところ、種表現中のオノマトペは 215 件、変形表現中のオノマトペは 1,522 件であった。変形表現中の出現が種表現中の約 7 倍となっている。

オノマトペは他の語群にはない独特な音韻的特徴をもつ。駄洒落を作るために種表現の音に近い言葉を探するときオノマトペが多用され、変形表現に出現すると予想される。また、種表現は文頭に位置して主語になることが多いため、必然的に名詞(名詞句の一部)が多くなる傾向がある。オノマトペが種表現中に出現する場合も、以下の例のように連体詞的用法や名詞的用法が多く見られる。

- ・ 8582 (ぼーっと)した奴が[ボート]に乗る
- ・ 46315 (へナチョコ)がもらった[変なチョコ]

5.2 音の変化

先述のとおり、オノマトペは比較的自由に音を

変化させることができる上、新語も容易に作成できる。駄洒落に出現したオノマトペのうち、オノマトペリストと字面が完全一致しなかったものを詳細に分析し、駄洒落におけるオノマトペの音の変化を考察する。分析対象となる駄洒落は211例である。

オノマトペの変化の度合いを次の4パターンに分類した。(a)は駄洒落とは関係なく音の変化を起こしているもの、(b)～(d)は駄洒落を成立させるために強制的に音を変化させたものである。(b)→(d)の順に変化の度合いが強くなる。

(a) 促音や長音を追加したり、繰り返し回数を変更しているが、種表現-変形表現の対応が目的ではない(46例)

- 21652 (母艦)が爆発した！[ボッカーン]
→「ボカン」に促音・長音追加
- 39332 (閣下)が笑った「[かっかっか]！」
→「かっかっ」の繰り返し回数変更

(b) 種表現-変形表現の対応をとるために促音や長音を追加したり、繰り返し回数を変更している(83例)

- 9039 (日光)で[ニッコーリ]
→「ニッコリ」に長音追加
- 26879 (本)が沈んだ。「[ブックブック]」
→「ブクブク」に促音追加

(c) 種表現-変形表現の対応をとるために音を変更したり追加したりしている(62例)

- 25820 (秘書)が[ひしょひしょ話]をしていたよ
→「ひそひそ」からの子音変更
- 50420 (外野)が[ガイヤガイヤうるさい]
→「ガヤガヤ」への母音追加

(d) 種表現-変形表現の対応をとるために新しいオノマトペを作っている(20例)

- 1276 (日本)が沈んだ![ジャッパーン]
→新語「ジャッパーン」
- 35994 空から(馬糞)が降ってきた。[バッフン]

→新語「バッフン」

- 44725 (部長)の鼻から鼻血が[ブッチョ～～ウ]
→新語「ブッチョ～～ウ」

以上より、駄洒落を成立させるためにオノマトペの音を様々な形式で変化させていることが明らかになった。今後は、オノマトペ以外の表現ではどの程度の変形が許容されるのかを検証して比較する必要がある。

6 おわりに

本稿では、駄洒落に出現するオノマトペの出現傾向を定量的に示し、オノマトペの出現位置や音の変化について考察を行った。

今後は、駄洒落データベースを拡張し、大規模な分析を行いたい。また、分析から得られた知見を駄洒落の自動生成手法に活用することも視野に入れている。

謝辞

本研究は科研費(基盤研究(C)17K00294 および若手(B)17K1279102)の助成を受けたものである。

参考文献

- [1] 荒木健治, 内田ゆず, 佐山公一, 谷津元樹: 駄洒落データベースの構築及び分析, 人工知能学会第2種研究会 ことば工学会資料, SIG-LSE-B702-3, pp.13-24, 2017.
- [2] 佐山公一, 荒木健治: コンピュータが駄洒落で笑わせる?—駄洒落の面白さを認識する理解の手続き—, 人工知能学会第2種研究会 ことば工学会資料, SIG-LSE-B702-4, pp.25-32, 2017.
- [3] 小野正弘編: 日本語オノマトペ辞典, 小学館, 2007.
- [4] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.
- [5] 内田ゆず, 高丸圭一, 乙武北斗, 木村泰知: BCCWJ コアデータにおけるオノマトペ出現実態の分析—現代オノマトペ実例辞書アプリ構築に向けて, 2015年度人工知能学会全国大会, 3G4-OS-05b-5, 2015.