

商品レビュー文における オノマトペと商品カテゴリーの関係についての予備調査

内田 ゆず^{†1} 長谷川 大[†] 荒木 健治[‡] 米山 淳[†]

[†] 青山学院大学 [‡] 北海道大学大学院情報科学研究科

1 はじめに

近年、商取引の電子化が進展しており、平成23年の日本国内の消費者向け電子商取引市場規模は8.5兆円まで拡大している¹⁾。それに伴い、Web上には一般消費者による商品レビューが大量に存在するようになった。それらのレビュー記事を分析して商品やサービスの評価を抽出し、消費者の情報収集・企業の市場調査に役立てようという動きが高まっている。

商品レビューには消費者の商品に対する主観的な情報が多く含まれており、商品の特徴や消費者の感情を表すためにオノマトペが用いられることがある。したがって、レビュー記事の分析を行う際に、オノマトペが有効な手がかりになると考えられる。

オノマトペは細かいニュアンスを伝えることに優れている。たとえば、やわらかいことを表す“ぐちゃぐちゃ”、“しんなり”、“ふっくら”では伝わるイメージが全く異なる。日本語オノマトペ辞典²⁾には、やわらかい様子を表すためのオノマトペだけでも71語が掲載されている。一方で、種類が多い、意味や用法にはっきりとした定義がない、といった点でオノマトペは機械的な処理には不向きである。

我々は、商品カテゴリーにおけるオノマトペの出現傾向を分析することで、オノマトペと商品カテゴリーの関係性を明らかにしようとしている。本稿では、Yahoo! ショッピング商品レビュー729,865件を対象として、オノマトペの出現頻度を調査する。さらに、“やわらかい”、“かわいた”を表現するオノマトペの出現頻度を変数としたクラスター分析によって商品カテゴリーを類型化する。

2章では、対象となる商品レビューデータの詳細について述べる。3章では、商品カテゴリーにおけるオノマトペ出現頻度について述べる。4章では、クラスター分析によって商品カテゴリーを類型化する試みの結果について述べる。最後に、5章で本稿の結論と今後の課題を述べる。

2 商品レビューデータ

本研究で用いる商品レビューは、Yahoo! Japan 商品レビュー検索APIを用いて収集されたYahoo! ショッピング²⁾の商品レビューである。また、APIで取得できる商品レビューの要素のうち、ReviewTitle (レビューのタイトル文) と Description (レビューの本文) の2つを用いる。

Yahoo! ショッピングでは、商品が階層化されたカテゴリーによって管理されている。本研究では最下層の商品カテゴリーを扱うものとする。レビューは9,170の商品カテゴリーに対するもので、総件数は729,865件、総単語数 (ReviewTitle と Description に含まれる単語数) は35,720,231語となっている。

表1に犬用シャンプーのカテゴリーに分類された実際のレビューの例を示す。

3 オノマトペ出現頻度の調査

3.1 手順

日本語オノマトペ辞典²⁾に事物の動き・変化、事物の形・状態、事物の音・道具・金銭、事物の程度を表すオノマトペとして掲載されている1,049語 (重複している語は除外) を対象として、9,170の商品カテゴリーのレビューに出現する頻度を調査する。表2に調査対象のオノマトペの一部を示す。

具体的な手順としては、2章で述べたレビューデータ (ReviewTitle と Description) に形態素解析を施し、上述のオノマトペと字面が完全に一致する形態素をカウントする。このとき、オノマトペはひらがな表記に加えてカタカナ表記も考慮する。形態素解析器には日本語形態素解析システムJUMAN 7.0³⁾を用いる。

一般に、オノマトペは形態素解析誤りを引き起こしやすい⁴⁾。JUMAN 7.0のオノマトペに対する解析精度の目安を調査するために、本研究で使用する1,049語のオノマトペをそれぞれ単語の形で入力したところ、510語が1形態素として出力された。解析精度は高くないが、ここでは形態素解析誤りについては考慮せずに出現頻度の調査を行う。

¹ yuzu@ee.aoyama.ac.jp

² <http://shopping.yahoo.co.jp/>

表 1: 商品レビューデータの例 (犬用シャンプー)

ReviewTitle	Description
つやつや	静電気もおきませんし、毛がつやつやになりとてもいいです。 うちのバグも気持ちよさそうにブラッシングされています。
高級感のある香り	リンスと一緒に購入しました (^ ^) 一番のお気に入りです。香りもさらさら感も一番です! うちでは 4 匹の犬がいますので業務用はとってもお得です。 このシャンプーは犬臭さが無くなりますのですぐに匂う子には とてもおすすめ!!

3.2 結果

調査の結果, 1,049 語のオノマトベのうち, 613 語が合計で 66,952 回レビューに出現していることが明らかになった. 頻度が高い 10 語を順に挙げると, “しっかり” (23,025 回), “ぴったり” (5,444 回), “ぼたん³” (2,840 回), “ちゃん⁴” (1,876 回), “さらさら” (1,325 回), “ぱん” (1,263 回), “ふわふわ” (1,230 回), “きらきら” (1,067 回), “たっぷり” (1,063 回), “きー” (997 回) となっている. 100 回以上出現したオノマトベは 78 語である.

3.3 考察

“しっかり” の出現頻度が突出して多く, 全体の約 34% を占めている. これは, どのような商品カテゴリにも使用される汎用的な語であるためだと考えられる. 一方, 出現が確認された 613 語のうち 535 語は出現頻度が 1 ~ 99 回となっており, オノマトベの多様性とオノマトベ使用の個人差がうかがえる.

頻度上位 10 語のうち, “ぼたん”, “ぱん”, “きー” に関しては, カタカナ表記にした場合に別の語と一致する. したがって, 以下のようなレビューによって出現頻度が増加している.

- イヤホンもついてて機能充実です でも個人的に ボタン が少し押しにくいです。(電子辞書)
- 3 日に一度は パン を焼くので, まとめ買い出来て助かります (牛乳)
- 見た目も綺麗で, キー も打ちやすくてとても気に入ってます. 薄くて軽くて扱いやすいです。(日本語キーボード)

また, “ちゃん” は “ワンちゃん”, “キティちゃん” などの語で出現頻度が増加している. 今後, このような同音異義語の問題を解決する必要がある.

³ 水滴などが高い所から落ちる音

⁴ 鐘や陶器などがぶつかる音

表 2: 調査対象としたオノマトベ

大カテゴリ	小カテゴリ	オノマトベの例
動き・変化	進む・運ぶ	ぐいぐい
	落ちる	ばさっ
	折れる	ぼきん
	ゆれる	ぐらぐら
	切る・切れる	すばっ
	破る・破れる	びりびり
	貼る・くっつく	べたっ
	刺す	ちくり
	曲がる・しわがよる	うねうね
	張る・膨らむ	むくむく
	光る・輝く	ぴかぴか
	焼く・焼ける	こんがり
	煮る	ぐつぐつ
	浮く・浮かぶ	ぶかっ
	粘る・ぬめる	ぎとぎと
形・状態	かわいた	ばさばさ
	のびる・縮む	すらり
	乱雑	ぐちゃぐちゃ
	あいまい	もやもや
	こすれる・きしむ	きーきー
	打つ・たたく	びしびし
	ぶつかる	こちっ
	楽器	びーひゃら
	武器	どんばち
	金銭	ぼっきり
	やわらかい	しんなり
	かたい	かちかち
	多い	どっさり
	少ない	ちらほら
	はやい・すばやい	さっ
音・道具・金銭	遅い	もたもた
程度		

4 商品カテゴリの類型化

オノマトベの出現頻度によって商品カテゴリを類型化するために, クラスタ分析を行う.

4.1 頻度データ

4.1.1 “やわらかい” オノマトベ

3.2 節で得た頻度情報から, “やわらかい” に属するオノマトベ 71 語のうち, レビューに 5 回以上出現した 12 語 (ふわふわ, ふかふか, ふんわり, ふっくら, ぷるぷる, ふわっ, とろとろ, ふにゃふ

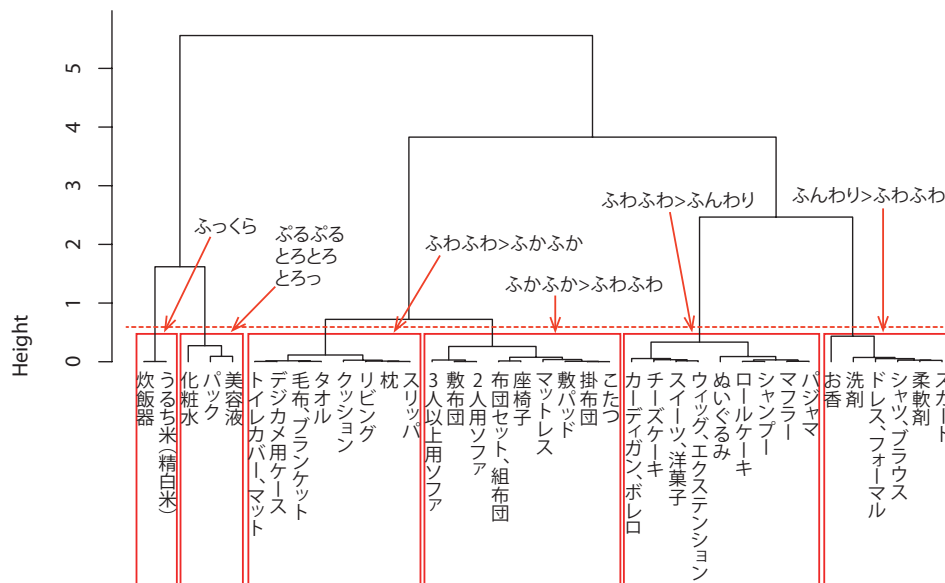


図 1: デンドログラム (やわらかい)

にや, ゆるゆる, とろっ, どろどろ, ふわり) を変数, それらのオノマトペを含むレビューが 20 件以上存在する 37 の商品カテゴリをケースとした頻度表を作成する。

4.1.2 “かわいた” オノマトペ

3.2 節で得た頻度情報から, “かわいた” に属するオノマトペ 24 語のうち, レビューに 5 回以上出現した 12 語 (さらさら, さっぱり, さらっ, ぱさぱさ, かさかさ, かりかり, ぼろぼろ, ぱさぱさ, ぼろぼろ, がさがさ, がびがび, かすかす) を変数, それらのオノマトペを含むレビューが 20 件以上存在する 31 の商品カテゴリをケースとした頻度表を作成する。

4.2 手順

ケースクラスター分析により, オノマトペ (変数) を手がかりに商品カテゴリ (ケース) を分類する。4.1 節で作成した頻度表のデータは粗頻度となっているため, 事前に標準化を行う。個体間非類似度計算には平方ユークリッド距離を用い, クラスター結合手法にはウォード法を用いる⁵⁾。

4.3 結果

“やわらかい” に関するオノマトペを変数にしてクラスター分析を行うことによって図 1 のデンドログラムが得られた。デンドログラムを目視で確認した上で, 分析のためのクラスター数は 6 としてカットポイントを定めた。

“かわいた” に関するオノマトペを変数にしてク

ラスター分析を行うことによって図 2 のデンドログラムが得られた。デンドログラムを目視で確認した上で, 分析のためのクラスター数は 6 としてカットポイントを定めた。

4.4 考察

頻度表と比較したところ, 図 1 の第 1 クラスター (炊飯器, うるち米) は, “ふっくら” の出現頻度が特徴的なカテゴリであった。また, 第 2 クラスターは “ふっくら” に加えて “ぷるぷる”, “とろとろ”, “とろっ” の影響が大きい。第 3, 4 クラスターは距離が近いので “ふわふわ”, “ふかふか” が高頻度であるという類似した特徴をもっているが, 第 3 クラスターは “ふわふわ” の, 第 4 クラスターは “ふかふか” の頻度がより高いという差異がある。第 5 クラスターは “ふかふか” がほとんど用いられず, “ふわふわ” が突出して高頻度となり, さらに “ふんわり” が続くグループである。最後の第 6 クラスターも “ふわふわ” の頻度が比較的高いが, “ふんわり” がより高頻度で出現するグループとなっている。

次に, 図 2 から “かわいた” に関するオノマトペを変数にしたときのクラスター分析の結果について考察する。この図で最も特徴的なのは, 第 4 クラスター (ズワイガニ) である。ズワイガニに対するレビューには, “食べ物などに水分がなく, うまみのないさま” を表すオノマトペ²⁾である “かすかす” が出現しており, これが第 5, 6 グループとの分離の要因だと考えられる。第 3 クラスターに含

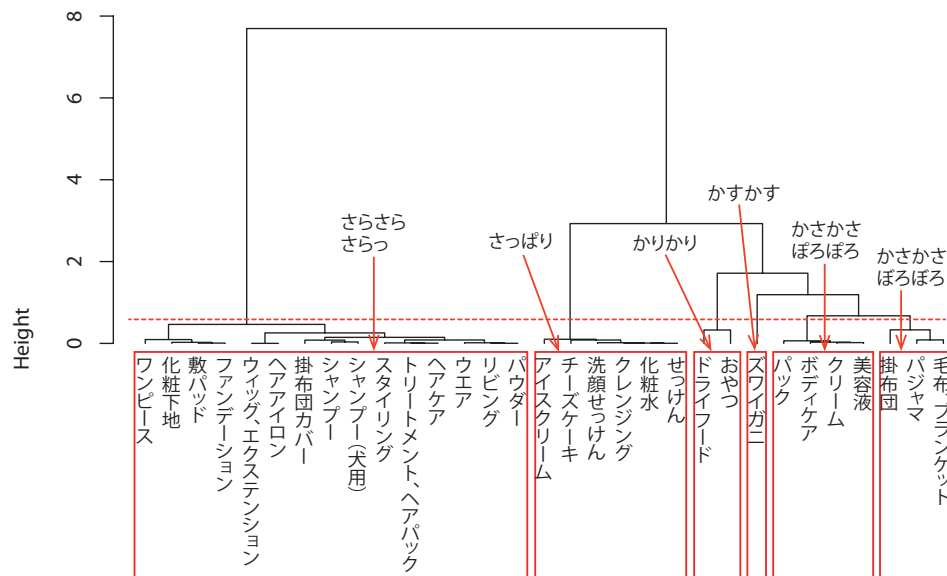


図 2: デンドログラム (かわいい)

まれる“おやつ”という商品カテゴリは親カテゴリが“ドライフード”であり，“ドライフード”とともに“かりかり”によって分類されている．第 1 クラスタは“さらさら”，“さらっ”によるグループであり，第 2 クラスタは“さっぱり”，第 5 クラスタは“かさかさ”と“ぼろぼろ”，第 6 クラスタは“かさかさ”と“ぼろぼろ”による分類だと考えられる．

このように，オノマトペが一定数含まれる商品カテゴリのレビューのクラスター分析を行うことによって，妥当性のある類型化が可能であることが示された．このデータは，レビュー分析のみならずオノマトペの学習支援にも有益であろう．

5 おわりに

本稿では，Yahoo! ショッピング商品レビュー 729,865 件を対象として，1,049 語のオノマトペの出現頻度を調査した結果について述べた．1,049 語のオノマトペのうち，613 語が合計で 66,952 回レビューに出現しており，100 回以上出現したオノマトペは 78 語であった．また，オノマトペの出現を判定する際に同音異義語が大きな問題になることが明らかになった．

さらに，“やわらかい”，“かわいい”を表現するオノマトペを変数としてクラスター分析を行い，商品カテゴリを分類した．その結果，日本語母語話者にとって，直感的に違和感のないデンドログラムが得られた．

今回の調査で，商品カテゴリによって使用され

るオノマトペに一定の傾向があることが明らかになった．一方，同じオノマトペでも，商品カテゴリによって使用される意図が異なる場合がある．今後は，レビュー中に含まれるオノマトペ，商品カテゴリ，レビュー評価点の関係を調査する予定である．

謝辞

本研究は科研費（課題番号：23700256）の助成を受けたものである．

参考文献

- 1) 経済産業省: 平成 23 年度我が国情報経済社会における基盤整備（電子商取引に関する市場調査），2012.
- 2) 小野 正弘: 擬音語・擬態語 4500 日本語オノマトベ辞典, 小学館, 2007.
- 3) 日本語形態素解析システム JUMAN 7.0, <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>.
- 4) 古武泰樹, 佐藤理史: 用例に基づくオノマトペ用法辞書の編纂, 言語処理学会第 16 回年次大会発表論文集, pp.994-997, 2010.
- 5) 石川慎一郎, 前田忠彦, 山崎誠 編: 言語研究のための統計入門, くろしお出版, 2010.