

キャッチコピーの自動生成に向けた分析

丹羽 彩奈¹ 岡崎 直観¹

西口 佳佑² 亀山 千尋² 毛利 真崇²

¹ 東京工業大学 ² 株式会社サイバーエージェント

{ayana.niwa, naoaki.okazaki} at nlp.c.titech.ac.jp

1 はじめに

キャッチコピーは、商品や作品を端的に表現した文章である。消費者の心を動かし、興味を惹きつけることで、商品や作品の売れ行きを大きく左右する。キャッチコピーはマーケティングにおいて重要であるが、その制作には大胆な発想やユーモア、的確に魅力を伝える表現力など、高度な作文能力を要するため、キャッチコピーのテキスト分析や自動生成が試みられてきた。

山根ら [6] は、キャッチコピーの品詞、単語特徴、係り受け関係などを分析し、使われやすい品詞の割合や、頻出 n -gram が他のコーパスと異なることを報告した。キャッチコピーの制作方法により踏み込んだ分析として、宣言文・命令形・疑問文などの文のタイプ [2]、比喩表現や韻 [1]、受賞アピール型や内容紹介型などのコピータイプ [9] による分析が試みられた。しかし、日本語のキャッチコピーの制作過程を調査したり、修辞技法を系統的に分析した研究は、これまでに存在しない。

一方、キャッチコピーの生成を目指した研究も行われている。山根ら [5] は、消費者の嗜好を考慮したキャッチコピーの生成に取り組んだ。幅野ら [7] は、映画のあらすじとレビューに基づいてキャッチコピーを制作する支援システムを発表した。既存研究は、生成の種となるキャッチコピーを選択し、その中の単語を置換することでコピーを生成するものが多い。

本研究では、キャッチコピーの自動生成・制作支援への足掛かりとして、コーパスの収集、およびその分析と実験結果を報告する。まず、コピーライターによるキャッチコピーの制作過程を概観する。次に、本研究で収集したキャッチコピーのコーパスを紹介し、用いられる修辞技法を分析する。最後に、キャッチコピーの自動生成に向けて、文長や文字種などの基礎的な統計情報、商品カテゴリ毎の頻出単語などを分析したのち、単語穴埋め問題によるキャッチコピー制作支援について報告する。

2 プロによるキャッチコピーの制作

2.1 キャッチコピーの制作過程

キャッチコピーの自動生成・制作支援に取り掛かる前に、プロのコピーライターがどのような手順でコピーを作っているのか調査した。ここでは、著名なコピーライターである梅田氏の書籍 [8] に記載されている制作手順をまとめた。

1. 情報収集をして、訴求ポイント（伝えたいこと・競合他社にはない強みなど）を洗い出す
2. 手順1で得た訴求ポイントに対して、様々な視点に立ち、思いついた言葉を書き出す
3. 各言葉に対して、抽象度を上げたり、それによって起こる結果などを連想することで、言葉に広がりを持たせる
4. 挙げられた言葉を体系化・グルーピングする
5. 手順4の結果を踏まえ、足りない視点を補強する
6. 客観性を確保する
7. 真逆・逆転の発想を考える
8. 単語の言い換えなどにより表現を整える

このように、キャッチコピーの制作は、対象に関して多角的な視点から単語やフレーズを列挙し、文章としてまとめ上げる作業である。よって、自動生成においても対象とする商品名だけでなく、その商品の特徴や広告の目的、伝えたいメッセージを考慮することが有効であろう。

また、何を伝えるかを明確にした後は、日本語としての文章の正確さだけでなく、メッセージの伝え方を工夫する必要がある。例えば、「断トツの最下位から、つながりやすさ No.1 へ。」のような未来志向の表現方法を採用し、コピーを整えていく。

代表的な修辞技法	使用率	具体例
比喩	17.9%	FBI が撃つように、僕は撮った。 (秋山晶)
反復	4.5%	あなたはあなたでいて。それが、あなたの美しさだから。 (仲畑貴志)
対句	10.6%	変わるもの、変わらないもの。 (秋山晶)
断定	5.2%	宣誓。キューピーマヨネーズは安全です。 (秋山晶)
誇張	8.3%	安全が商品になる時代が終わりますように。 (秋山晶)
擬態・擬音	1.3%	乾杯の前にダラダラ話すな。 (仲畑貴志)
呼びかけ	7.8%	起きなさい、体内時計 (秋山晶)

表1: キャッチコピーで使用される修辞技法の割合と具体例

2.2 キャッチコピーにおける修辞技法

文献 [8] には、キャッチコピーによく用いられる修辞技法として、物事を共通点を持った他のもので例える「比喩 (直喩, 隠喩, 擬人など)」、同じ言葉を繰り返す「反復」、対照的な意味を持つ物事を並べる「対句」、言い切る「断定」、実際よりも大きさに表現する「誇張」、物事の様子や音声の描写に用いる「擬態・擬音」、そして「呼びかけ」などが挙げられている。そこで、著名なコピーライターによる作品集 [3, 10, 13] に掲載されているキャッチコピーを分析し、用いられる修辞技法を調査した。今回は、作品集に掲載されているコピーの一部である 1,240 件について、用いられている修辞技法を手作業で認識することで調査を行った。表1に、各技法が使用された割合 (1,240 件に占める割合) と具体例を示した。

最も頻繁に用いられる「比喩」では、「FBI が撃つように」といった直喩を用い、被写体をしっかりと見据えて撮影している様子が伝わる。このように比喩表現を用いることで、内容をより感覚的に捉えやすくなる。「断定」では、「安全です」と言い切ることで、コピーに強い意志を込めている。

コピーに用いられる修辞技法は他にもあり、それらを使うことでキャッチコピーの表現力を高め、伝えたいメッセージをより明確にできる。コピーライターの創造力は用いる単語やフレーズを選択に加えて、修辞技法の部分にも強く現れるため、これらを自動生成においてどう扱うかが課題となる。

3 キャッチコピーの統計的分析

第2節で用いたキャッチコピーは著名なライターの作品であるが、統計的な分析を行うには数が足りない。そこで、キャッチコピーの言語的な特徴や自動生成の可能性を調査するため、本節では日本最大の公募広告賞であ

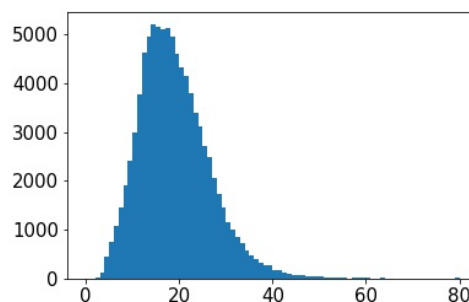


図1: キャッチコピーの長さの分布

る宣伝会議賞*1の一次審査の通過作品がすべて収録されている SKAT [4] を利用する。具体的には、SKAT.2 ~ SKAT.17 の 16 冊のコピーを書き起こし、91,682 件のコピーを収集した。採用に至らなかった作品もあるが、プロのコピーライターや企業が選定した、食品メーカーや不動産会社など幅広い職種のコピーが収められている。

3.1 キャッチコピーの長さと言読点の使い方

キャッチコピーの自動生成に向けた基本情報として、キャッチコピーの長さや句読点の使用回数について調べた。図1にキャッチコピーの長さの分布を示した。平均文字数は 18.91 文字となっており、この短い文字数に全ての情報がコンパクトに分かりやすく詰め込まれる。句点の一つ含むキャッチコピーは全体の 77.9%、句点なしも 14.0% を占める。つまり約 9 割のキャッチコピーは一文から構成されていることになる。句点なしのキャッチコピーは、言葉のみで表現されているものや、「リゲイン前 (>_<) リゲイン後 (~o~)」のように顔文字で終わっているものなどがある。

一文の長さ (句点間の文字数) は、JAMUL コーパス [12] の新聞見出しの 25.0 文字に対し、キャッチコピ

*1 <https://senden.co/>

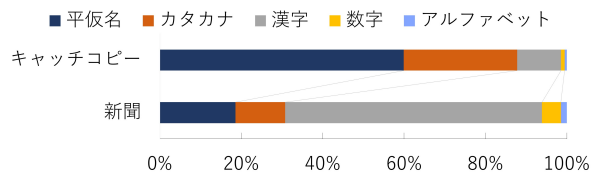


図2: キャッチコピーと新聞見出しで用いられる文字種

ーは16.2文字であり、JAMUL コーパスの短文要約(10文字や13文字)に近い。また、54.5%のキャッチコピーは読点を1つ含み、38.5%は読点を含まない。読点間の平均文字数は、キャッチコピーと新聞見出し(JAMUL)ではそれぞれ10.7文字および13.8文字となっている。

3.2 文字の表記方法

次に、キャッチコピーの重要な言語的特徴として平仮名、カタカナ、漢字、数字、アルファベットが用いられる割合を調べた。キャッチコピーと新聞の見出し(JAMUL)の比較を図2に示す。

新聞の見出しと比較すると、キャッチコピーは圧倒的に平仮名・カタカナの使用率が高く、逆に漢字の使用率は低い。また、新聞では数字の使用率が高い。これは、新聞では漢字で書くべき単語を平仮名やカタカナで表記することはないが、キャッチコピーでは易しい単語が用いられる傾向にあり、漢字で書ける単語も平仮名やカタカナで表記することがあるためと考えられる。

平仮名によりキャッチコピーの表現力が高まる事例もある。例えば、「あなた、「しょうゆ」って言えなかったの覚えてる?」では、「醤油」を「しょうゆ」と表記することで漢字が読めない子供時代にうまく発音ができなかった微笑ましい情景を映し出している。

3.3 職種ごとの頻出単語

キャッチコピーの自動生成や制作支援の第一歩は、商品の特徴を表すキーワードを選択することである。本研究では、キャッチコピーに関連付けられている企業の業種ごとに、頻出単語のランキングを作成した。企業名から業種への対応付けは、企業名の Wikipedia 記事のカテゴリ名を活用した。本稿では、例として日本の食品メーカーの頻出単語とその出現頻度を表2に示した。

「飲む」「食べる」「美味しい」という単語が上位に現れることは想像に容易いが、「私」や「あなた」といった生活の身近な場面を表す単語や、「母」「帰る」といった家庭の情景を表す単語もよく用いられている。紙面の都合で今回は省略したが、化学メーカーであれば社会背景を

名詞	動詞	形容詞
味 (193)	飲む (336)	おいしい (151)
人 (177)	食べる (223)	うまい (53)
しょうゆ (153)	思う (138)	美味しい (44)
醤油 (115)	言う (126)	多い (39)
料理 (114)	知る (51)	よい (21)
もの (102)	作る (49)	早い (20)
日本 (102)	買う (48)	高い (17)
母 (100)	見る (48)	まずい (16)
私 (92)	変える (45)	悪い (16)
魚 (91)	帰る (41)	長い (16)

表2: 食品メーカーに関するコピーの頻出単語とその頻度

オリジナルコピー: 今日の失敗に、乾杯しよう。

単語穴埋め問題

正解ラベル

<noun> の失敗に、乾杯しよう。	今日
今日の <noun> に、乾杯しよう。	失敗
今日の失敗に、<noun> しよう。	乾杯
今日の失敗に、乾杯 <verb> う。	しよ

図3: 単語穴埋め問題によるキャッチコピー制作支援

反映してか、「空気」「変える」「変わる」のような環境問題やその改善に関する単語が頻出する。

3.4 単語穴埋めによる制作支援

第2.2節で述べたキャッチコピーの制作過程の一部は、未完成のキャッチコピーに適切な単語を当てはめるという作業である。そこで、本研究ではキャッチコピーの単語穴埋め問題に取り組む。具体的には、キャッチコピー中の1つの単語(名詞、動詞)を隠した文を入力、その隠された単語を出力としたエンコーダ・デコーダモデルを学習した。学習モデルには、2層の双方向 LSTM エンコーダー、入力側へのアテンション付きデコーダを採用した*2。今回の実験では、まず SKAT のキャッチコピーのデータを83,000件の学習用データ、682件の開発用データ、8,000件の評価用データに分割した。次に各データセットのキャッチコピーに対し、名詞および動詞を1つずつ隠し、それを予測するタスクを作成することで、3種類のデータからそれぞれ417,868件、3,194件、38,956件の単語穴埋め事例を得た(図3)。モデルの単語ベクトルは、日本語 Wikipedia で事前学習済みの単語

*2実装には OpenNMT (<http://opennmt.net/>) を用いた。

番号	コピー例
(1)	正解: 女将の名前は忘れたが、あの酒の名前は忘れない。 生成例: 父の名前は忘れたが、あの酒の名前は忘れない。
(2)	正解: 酔った記憶は <u>忘れて</u> も、太った記憶は忘れない。 生成例: 酔った記憶は <u>残って</u> も、太った記憶は忘れない。

表3: 単語穴埋め問題によって生成されたコピー例

ベクトル^{*3} [11] で初期化した。

評価セットにおける単語穴埋め問題の正解率は、0.274 という結果になった。なお、予測に用いられた単語候補数は 26,798 である。

表3に、学習モデルによる生成例を示した。モデルが単語穴埋めを行う際の正解率は高くないものの、(1) のように意味の伝わる生成結果が多い。人間が作ったコピーでは、女将がいるようなお店で飲んだお酒がとても美味しく、印象に残った様子が浮かぶ。一方、生成例では名詞を一つ変えただけで「父の名前は忘れたが」という現実では考えづらい誇張表現が現れ、その相乗効果によりさらに対句の印象が強まったコピーになっている。

先述の通り、キャッチコピーは数多くの単語の組合せから表現を選ぶため、その選び方によって与える印象が大きく異なる。したがって、今回のように様々な関連語を埋め込んだコピーを出力することで、ライターがよりよいコピーを推敲するプロセスを支援できる。

しかし、修辞技法を捉えていない生成結果も多く見受けられる。その典型例を(2)に示した。正解コピーでは、「忘れるもの、忘れないもの」というギャップを表現する対句と、「忘れる」という同じ単語を繰り返す反復の修辞技法が使われている。しかし生成例ではそのような関係性が失われてしまい、伝えたいメッセージが不明瞭になっている。修辞技法は広告の効果を高めるために有用であるが、修辞技法などの大域的な構造をエンコーダ・デコーダのような文生成モデルにどのように取り込むか、今後の課題となるであろう。

4 おわりに

本稿では、キャッチコピーの自動生成・制作支援を目的として、実際の制作過程や言語的特徴の分析、そして単語穴埋めによる実験結果を報告した。キャッチコピー

^{*3}http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector

の制作は、商品の特性や伝えたいメッセージに関するキーワードを挙げ、修辞技法などを活用しながら文章にまとめ上げる作業である。これらの作業を支援するような手法を今後検討していきたい。

また、キャッチコピーをテキスト生成の問題と捉えた場合、約9割は単文で短い(平均18.91文字)こと、ひらがなやカタカナが8割以上を占めることから、平易な単語を用いながら(文脈を考慮することなく)単文を生成するタスクと考えることができる。また、対句や反復などの修辞技法を用いることも重要なポイントである。例えば見出し生成と比較すると、用いる単語や文字種、修辞技法がもたらす意味的な関係など、様々な制約を考慮しながら文を生成する必要がある。今後は、キャッチコピーの自動生成をきっかけとして、新しい文生成タスクの提案やその生成手法を考えていきたい。

参考文献

- [1] Tatjana Dubovičienė and Pavel Skorupa. “The Analysis of some Stylistic Features of English Advertising Slogans”. In: *Žmogus ir žodis* 16 (2014), pp. 61–75.
- [2] Narumon Huadhom and Wannapa Trakulkasemsuk. “Syntactic Analysis of Online Tourism Slogans: Frequency, Forms and Functions.” In: *PASAA: Journal of Language Teaching and Learning in Thailand* 53 (2017), pp. 182–213.
- [3] 仲畑貴志. ホントのことを言うと、よく、しかられる。勝つコピーのぜんぶ。株式会社宣伝会議, 2018.
- [4] 宣伝会議賞実行委員会. *SKAT.2-SKAT.17*. 株式会社宣伝会議, 2003-2018.
- [5] 山根 宏彰, 萩原 将文. “ウェブにおけるユーザの嗜好を反映させたキャッチコピー自動生成システム”. In: *日本感性工学会論文誌* 13.4 (2014), pp. 493–500.
- [6] 山根 宏彰, 萩原 将文. “複数コーパスを利用したキャッチフレーズの特徴分析”. In: *日本感性工学会論文誌* 11.2 (2012), pp. 233–239.
- [7] 幅野 裕貴, 浦谷 則好. “あらすじとレビューを用いた映画のキャッチコピー作成支援手法の提案”. In: *言語処理学会第20回年次大会*. 2014, pp. 864–867.
- [8] 梅田悟司. 「言葉にできる」は武器になる。日本経済新聞出版社, 2016.
- [9] 橋本 浩典, 安藤 一秋. “D-027 書籍のキャッチコピー自動生成に向けたキャッチコピーの収集と分析 (D 分野: データベース, 一般論文)”. In: *情報科学技術フォーラム講演論文集*. Vol. 14. 2. Aug. 2015, pp. 129–130.
- [10] 谷山雅計. 広告コピーってこう書くんだ! 読本. 株式会社宣伝会議, 2007.
- [11] 鈴木 正敏, 松田 耕史, 関根 聡, 岡崎 直観, 乾 健太郎. “Wikipedia 記事に対する拡張固有表現ラベルの多重付与”. In: *言語処理学会第22回年次大会* (2016), pp. 797–800.
- [12] 雄太 人見, 雄哉 田口, 秀明 田森, 二郎 西鳥羽, 洸 菊田, 直観 岡崎, 健太郎 乾, 学 奥村. “出力長制御を考慮した見出し生成モデルのための大規模コーパス”. In: *言語処理学会年次大会*. 2019, (to appear).
- [13] 青田 光章, 秋山 晶, 東 秀紀, ほか. 最新約コピーバイブル. 株式会社宣伝会議, 2007.