

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 Collection and Analysis of Book Advertising Copies toward Construction of an Automatic Copy Generation System

橋本 浩典[†] 安藤 一秋[‡]
Hironori Hashimoto Kazuaki Ando

1. はじめに

キャッチコピーとは、短い文字数で対象の魅力伝えるものである。大辞林[1]によると「消費者の心を強くとらえる効果を狙った印象的な宣伝文句」と定義されている。キャッチコピーは、様々な商品やサービス、イベントなどの宣伝文句として利用されているが、人々の心を動かすものを生み出すことは難しい。このことは、コピーライターというキャッチコピーを考える専門職の存在からもわかる。

本研究では、書籍に注目し、各書籍にふさわしいキャッチコピーを自動生成するシステムの構築を目的とする。対象として書籍を選択した理由は、以下の2点である。

- ・書籍のキャッチコピーは、本の帯に利用されるだけでなく、書店の店員が独自に作成する場合や自費出版で執筆者が作成する場合などがある。

- ・書籍には、あらすじやレビューなど、自動生成に活用できるテキスト情報が存在し、Web上から収集できる。

本研究では、書籍のキャッチコピーを Web から自動収集し、キャッチコピーDBとして構築・利用する。常に新しいキャッチコピーを収集することで、時代や流行に合わせたキャッチコピーの自動生成を目指す。なお、キャッチコピーの自動生成に関する研究[2-3]はいくつか行われているが、キャッチコピー事例集という静的情報源から得られた知識を利用するものが多く、本研究の目的とは異なる。

本稿では、キャッチコピーの自動生成手法を検討するための前段階として、書籍のキャッチコピーを Web 上から収集し、分析した結果について述べる。

2. 書籍のキャッチコピーの収集

分析に利用するキャッチコピーデータは、インターネット書店の楽天ブックスから API を利用して自動収集する。本稿での調査では、楽天ブックスの 24 種のジャンルの内、キャッチコピー数が多いと考えられる小説ジャンルに注目する。小説ジャンルには、8 種のサブジャンルがあるが、比較的登録書籍が多いミステリ・サスペンスと SF・ホラーを収集対象とする。

収集手順としては、まず、楽天ブックス書籍検索 API [4]を利用して、各商品ページの書籍情報を獲得する。書籍情報には、書誌情報の他に様々な情報が含まれている。そこで、書籍情報を確認した結果、内容説明情報にキャッチコピーが含まれていることを確認した。また、キャッチコピーは、内容説明情報の内、最後の 1 文に多く出現する傾向も確認できた。しかし、最後の 1 文はキャッチコピー文を含まない文（例“宮田の説に耳を傾けながら、歴史に興味を持ち始めた若い頃のようにワクワクするジョゼフであっ

た”）もあるため、一部を機械的に排除した後、人手で取り除くこととした。なお、最後の一文以外にもキャッチコピーが記述される場合や、複数のキャッチコピーが記述される場合などもあるが、人手での判断が必要となるため、本調査からは除外した。

最終的に、楽天ブックスのミステリ・サスペンスと SF・ホラージャンルからは、それぞれ 694 文と 432 文のキャッチコピーを収集できた。次章では、このキャッチコピーを分析対象とする。

3. 収集したキャッチコピーの分析

本章では、楽天ブックスのミステリ・サスペンスと SF・ホラージャンルから収集した、694 文と 432 文のキャッチコピーを、品詞と構造パターン、タイプで分析する。

3.1 品詞に基づく分析

McCab を用いてキャッチコピーを形態素単位に分割し、品詞別の特徴を分析する。

3.1.1 名詞

ジャンル毎に、キャッチコピーに頻出する名詞の特徴を調査する。名詞は複合名詞化して集約し、その頻度をカウントして調査する。

表 1 に、各ジャンルに出現した複合名詞の上位 8 件を示す。表 1 に示すように、各ジャンルを表わす「ミステリ」「ホラー」といった単語が上位に出現した。また、両ジャンル共に「著者」が上位に出現していることから、書籍のキャッチコピーは、本の内容以外にも、著者の情報を用いて生成されることがわかる。なお、名詞の全体での出現回数は、ミステリ・サスペンスでは 2,505 回、SF・ホラーでは 1,600 回であった。

表 1 各ジャンルにおける複合名詞の出現頻度 (上位 8 件)

ミステリ・サスペンス		SF・ホラー	
傑作	45	恐怖	19
謎	30	テーマ	14
著者	23	ホラー	14
ミステリ	18	謎	13
本格ミステリ	15	開幕	12
物語	14	待望	12
待望	11	世界	10
女王	11	著者	9

3.1.2 動詞

次に、キャッチコピーに頻出する動詞の特徴を調査する。なお、動詞の頻度は、原形化して集約し、その頻度をカウントして調査する。なお、「する」や「なる」は抽象的な動詞であるため、調査から除外する。

[†] 香川大学大学院工学研究科

Graduate School of Engineering, Kagawa University

[‡] 香川大学工学部

Faculty of Engineering, Kagawa University

表2に、各ジャンルに出現した動詞の上位8件を示す。表2に示すように、両ジャンルにおいて「描く」「贈る」「放つ」といった動詞が上位に共通して出現している。この結果から、ジャンル固有の動詞は、あまり多くない可能性がある。なお、動詞の全体での出現回数は、ミステリ・サスペンスでは706回、SF・ホラーでは401回であった。

表2 各ジャンルにおける動詞の出現頻度 (上位8件)

ミステリ・サスペンス		SF・ホラー	
描く	60	描く	32
贈る	26	贈る	10
放つ	15	書き下ろす	7
満ちる	10	放つ	7
続く	10	挑む	6
挑む	9	受賞する	6
受賞する	8	続く	6
輝く	7	迫る	6

3.1.3 その他の品詞

その他の品詞として、形容詞、形容動詞、副詞についても調査を行った。しかし、それぞれの品詞の出現回数は、ミステリ・サスペンスでは67, 234, 46回であり、SF・ホラーでは51, 122, 41回であった。名詞や動詞の出現回数と比較して、出現回数が少なく、書籍のキャッチコピーには、あまり利用されない可能性がある。

3.2 構造パターンに基づく分析

キャッチコピーの構造を確認するため、キャッチコピー中の複合名詞を「N」に置換してパターンを抽出する。

出現回数が多いパターン上位8件を表3に示す。表3に示すように、名詞の組み合わせのみで構成されているキャッチコピーが上位に出現している。このようなキャッチコピーを確認した所、受賞について言及したものや、シリーズ番号を示す「第N弾」のような言い回しものが多く、汎用性が高いキャッチコピーであった。他にも「NのN」のような主に名詞で構成されるものを含めると上位すべてが名詞で構成されており、名詞の重要性が高いことがわかる。

表3 各ジャンルで出現回数が多いパターン (上位8件)

ミステリ・サスペンス		SF・ホラー	
N第N弾	31	NのNはN	9
NのN	12	N	9
N賞N作	9	NのN	7
N	9	N賞N	5
第N回N賞N作	8	N、第N弾	5
N賞N	7	N第N弾	4
NのN第N弾	6	N第N弾!!	4
NのN、N第N弾	5	N賞N作	4

また、3.1.2の調査で出現回数の多かった動詞について、出現したパターンの調査を行った。その結果、「描く」や「贈る」のように著者が主語になるもの(例“英国のベテ

ランジャーナリストが描く、清新な警察ミステリ”)と「挑む」のように主役が主語になるもの(例“主婦探偵が映画界に挑む第五弾”)に大別できることがわかった。さらに、主語(著者)が省略され、動詞とその目的語でキャッチコピーを構成しているもの(例“受け継がれる“絆”と“謎”の行方を描く、静謐な恋愛譚”)が存在しており、数としては、著者を明記するものとほぼ同数であった。

3.3 キャッチコピーのタイプ分析

収集したキャッチコピーはいくつかのパターンに分類することができた。本稿では、比較的出現数が多い、2パターンについて説明する。

3.3.1 受賞アピール型

このタイプは、受賞した賞がそのままキャッチコピーに利用されているものである。このタイプは、著名な賞であるほど有効に働くと考えられる。さらに、このタイプは、受賞した書籍をアピールするもの(例“全米図書館協会アレックス賞受賞作”)と賞を受賞した著者をアピールするもの(例“CWA賞受賞作家が描きあげた、現代英国本格ミステリの傑作”)の2つに大別できることがわかった。このタイプのキャッチコピーを自動生成するには、書籍の情報に加えて、著者の情報を収集する必要がある。

3.3.2 書籍の内容紹介型

このタイプは、対象書籍の内容を紹介するキャッチコピーである。このタイプには、書籍のあらましを短くまとめたと考えられるもの(例えば“児童文学の気鋭が、頼りない人形作家の父と、日々成長する娘の姿を優しく綴った快作”)、クライマックスについて言及したもの(例えば、“切ないラストが胸を打つ、きらめく青春群像劇”)が存在した。このタイプのキャッチコピーを自動生成するためには、書籍の本文やレビュー情報が必要になる。すべての書籍本文の取得は困難であるが、著作権の切れたものは利用できる。また、レビューは、掲示板やブログなどから収集できる可能性が高い。

4. おわりに

本稿では、キャッチコピーの自動生成手法を検討するための前段階として、Webから収集した書籍のキャッチコピーを分析した結果について述べた。分析結果として、書籍のキャッチコピーには、名詞と動詞がよく利用されることがわかった。また、キャッチコピーはいくつかの構造パターンやタイプがあり、生成する対象ごとに必要となる情報が異なることを確認した。

今後の課題としては、売上情報とキャッチコピーを絡めることで、より特徴的なキャッチコピーの特性をつかみ、自動生成手法の考案に利用する。

参考文献

- [1]三省堂, “大辞林第3版”, 三省堂 (2006).
- [2] 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”, 言語処理学会第20回年次大会発表論文集, pp.864-867 (2014).
- [3] 山根 宏彰, 萩原将文, “Web情報をを用いたキャッチコピーの自動生成システム”, ファジィシステムシンポジウム講演論文集 28, pp.1008-1013 (2012).
- [4] 楽天ブックス書籍検索 API, “<https://webservice.rakuten.co.jp/api/booksbooksearch/>”