原著論文

複数コーパスを利用したキャッチフレーズの特徴分析

山根 宏彰, 萩原 将文

慶應義塾大学大学院理工学研究科

Feature Analyses for Catchphrases Using Multiple Corpora

Hiroaki YAMANE and Masafumi HAGIWARA

Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama-shi, Kanagawa 223-8522, Japan

Abstract : Catchphrases attract attention in a form of concise expressions; however, there are few researches which investigate them using the knowledge of natural language processing. First, for the analysis based on language engineering, we employed a catchphrase corpus containing 6,466 catchphrases, an encyclopedia corpus, a blog corpus and a dialog corpus. Moreover, we conducted an analysis on each theme for revealing the difference between themes on the catchphrase corpus. The result from the language engineering shows that forward parts of speeches tagging in the catchphrase corpus are similar to those of both the encyclopedia and blog corpora. In terms of backward parts of speeches tagging, a lot of catchphrases end with nouns, which differentiates the other corpora. In addition, for each theme in catchphrase corpus, we found that specific words such as “taste” and “skin” frequently appear.

Keywords : *Catchphrases, Multiple Corpora, Statistical Analysis*

1. は じ め に

言葉を用いたコミュニケーション、特に書き言葉は人間社会において重要な役割を果たしている。例えば、広告活動においてキャッチフレーズは大きな影響を私達に与える。一般の書き言葉と異なりキャッチフレーズは短い文字数で注目を引き、端的かつ効果的に対象の特徴を述べるものである。その目的は、受け手の注意を引いて、その関心と興味をかきたて、さらに送り手の期待するような欲望を起こさせて、最後にはその欲望にそって、受け手にある種の行動を行わせることである[1]。営利目的のキャッチフレーズの場合は、「利益を極大化するために、購買者を説得する意図を持った情報のマス・コミュニケーション」[2]という定義がある。

時代によってもキャッチフレーズの中身は変化してきており、単に商品の登場を告げることを第一にした「知らせばいい」時代から、マーケティング理論に基づいた生活提案型の広告が展開された時代、感性や気分を重視したいいわゆる「ムード広告」「フィーリング広告」の時代を経て、現在、広告のエンタテインメント化は一層進みつつあると言われている[3]。現代では生き残りをかけて、企業が価格競争よりもブランド構築を重要視するようになってきているとも言われている[4]。

また、広告文の側面も持つキャッチフレーズは、経済学や言語学からも盛んに研究が行われている。Kohliら[5]は、効果的なキャッチフレーズを作成するために必要な要素の考察を行っている。Lowreyら[6]は、ブランド名における語音象徴に関する研究をしている。新井[7]は、「言葉はなぜ通じるのか」という「関係性の理論」からキャッチフレーズ

の特性分析を行っている。日本社会における独特の文化を背景とする、「場」に焦点を当てた考察[8]も存在する。

キャッチフレーズの応用先には、雑誌・カタログ・CMなどに代表される広告利用、一般の人々でも関わる機会が多い出し物やイベント等の呼び込み広告がある。前者はプロのコピーライターがキャッチフレーズを作成するが、後者のような例も多く存在し、プロでない人がキャッチフレーズを作成するケースが少なくない。従って、人間の心を捉えるキャッチフレーズを自動生成するシステムへの需要もある。

松平ら[9,10]は、質より量、批判厳禁を旨とする発散的発想支援的な立場から、遺伝的プログラミングを用いたキャッチフレーズ生成システムを提案した。西原ら[11,12]は、研究発表に限定し、タイトル毎に興味を引きやすいかどうかをランク付けするアルゴリズムを提案した。中野ら[13,14]はブログ記事を用いて目を引くキャッチフレーズを作るための基礎的研究を行った。越川[15]はブランドネームについて語感に関する考察を行っている。また、脳科学の見地からのブランド考察[16]も存在する。

しかしながら筆者らの調査する限り、効果的なキャッチフレーズが具備する特性について、言語工学的観点からの深い議論はなされてこなかった。

キャッチフレーズは比較的短い文で、受け手に強いインパクトを与える。従って、通常の文とは異なる様々な特徴を有することが予想される。本論文は、他の種類の文章との比較分析を詳細に行うことにより、キャッチフレーズ独特の特徴を明らかにする。具体的には、文法的に正しく適切な説明文が用いられる百科事典、様々な言い回しや話題を扱っているブログ、更に文法的には厳密でなく自由度の高い会話文の3種類を比較対象とし、それぞれ膨大な情報量を有するコーパ

スを用いた分析を行う。

以降、本論文では第2章でキャッチフレーズコーパスと比較対象のコーパスの詳細について述べる。第3章ではキャッチフレーズの言語工学的観点からの分析結果を述べ、第4章をまとめとする。

2. キャッチフレーズと比較対象とするコーパス

キャッチフレーズには営利・非営利のものがある。第1章で述べたように、キャッチフレーズは受け手の注意を引き、関心と興味をかきたて、受け手に行動を行わせることを目的とする。営利目的のキャッチフレーズの場合は、「利益を極大化するために、購買者を説得する意図を持った情報のマス・コミュニケーション」[2]という定義がある。本論文では、比較的最近の営利目的のキャッチフレーズが大量に収録されている文献[17]（以下キャッチフレーズコーパス）を分析の対象とする。

さらに、他の種類の文章が含まれているコーパスとの比較・分析を行うことによりキャッチフレーズの特徴を明らかにする。比較対象とするコーパスは、文法的に正しく適切な説明文が用いられているインターネット百科事典Wikipediaコーパス[18]、様々な言い回しや話題を扱っているブログコーパスであるKNBコーパス[19]、文法的には必ずしも厳密でないが自由度の高い会話文コーパス[20]である。表1にそれぞれの詳細を示す。表からわかるように、それぞれ数千の文からなるコーパスを用いて分析実験を行う。キャッチフレーズコーパス[17]は、食品・美容・生活用品を始めとする12の分野を包含しており、全体で6,466個のキャッチフレーズが掲載されている。百科事典コーパス[18]については、情報量が多く処理に時間がかかるため、全体から約1万文をランダムに抽出し用いる。

3. キャッチフレーズの言語工学的観点からの分析

ここでは、キャッチフレーズの言語工学的観点からの分析を行う。具体的にはまず3.1で、用いるコーパスの前処理と解析方法について述べた後、3.2で品詞、単語特徴、係り受け関係について他種類のコーパスとの比較を行い、キャッチフレーズ固有の特徴について明らかにする。さらに3.3で、

表1 コーパスの種類と含まれる文の数

コーパスの種類		文の数
キャッチ フレーズ コーパス [17]	内訳: 食品(495 個), 美容(451 個) 生活用品(1,133 個), ファッション(1,066 個) デジタル関連製品(671 個) 家庭用電化製品(649 個) 交通・レジャー(450 個), サービス(253 個) ハウジング(429 個), 自動車(385 個) メディア(341 個), その他(143 個)	6,466
百科事典コーパス[18]		11,406
ブログコーパス[19]		3,937
会話文コーパス[20]		5,384

図1 解析の流れ

キャッチフレーズコーパスそのものに含まれる単語、係り受け関係、分野ごとの名詞に着目した分析・考察を行う。

3.1 コーパスの前処理と解析方法

図1に、コーパスの前処理と解析の流れを示す。前処理では、各コーパスに含まれている分析に不必要な成分の削除を行う。例えば、括弧“()”内に入っている文や不必要な記号などがあげられる。さらに形態素解析器MeCab[21]、構文解析器CaboCha[22]を用いた解析を行う際に、エラーを発生させる成分の削除・変更も行う。例えばMeCabでは、半角記号に対しては全て名詞判定を行ってしまう。そこで、全てのテキストデータに関して半角記号を全角に変更する。ただし年号のように数字が連続する場合には半角に戻し、数字列を名詞として扱わせる、などの例外的な処理も行っている。前処理の後、一文ずつ形態素解析を行い、各コーパスにおける品詞を取得する。CaboChaの場合には、係り受け関係において係る側と係られる側の形態素列を取得する。

3.2 他コーパスとの比較による特徴分析

ここでは、キャッチフレーズコーパスと他コーパスについて、品詞の使用割合および品詞列に着目した分析を行う。

3.2.1 全体の品詞使用割合の分析

図2に、解析によって得られたキャッチフレーズコーパス全体における品詞の使用割合を示す。全体の品詞使用割合の傾向としては名詞の割合が一番多く、助詞・記号・動詞と続いていることがわかる。キャッチフレーズにおいて、名詞が重要であることがわかる。

図2 キャッチフレーズコーパスにおける品詞の使用割合

図3 抽出した百科事典コーパスにおける品詞の使用割合

次に、図3に百科事典（Wikipedia）コーパス [18] における品詞の使用割合を示す。図2のキャッチフレーズコーパスの場合と比較すると、割合が違うものの品詞割合の順位がかなりの部分で一致していることがわかる。一方、百科事典コーパスにおいては特に名詞割合が多い。理由として、Wikipediaは名詞として扱われる年号や専門用語が多く、名詞の連語が多かったからと推測できる。また、Wikipediaにはほとんど形容詞が見られなかった。

図4には、ブログコーパス [19] の品詞割合を示す。ブログコーパスは、図3の百科事典コーパスと比較すると、より図2のキャッチフレーズコーパスの割合に近いことがわかる。理由としては、ブログでは様々な言い回しが用いられているため、品詞の使用割合がよりキャッチフレーズに類似したためと考えられる。

図5に会話文コーパス [20] の品詞割合を示す。図から会話文コーパスは、他のコーパスの品詞使用割合と大きく異なっていることがわかる。これは会話文特有に、話の途中で終わってしまったり、主語となる名詞の省略などが起こったりするケースが頻繁に見られ、名詞の使用割合が下がったためと考えられる。また、感動詞の割合が多いことも特徴として挙げられる。

以上から、会話文コーパスを除く、3種類のコーパスにおいては名詞が占める割合が最も多く、キャッチフレーズコーパスに含まれる名詞の割合は百科事典コーパスとブログコーパスの中間程度であることがわかった。

図4 ブログコーパスにおける品詞の使用割合

図5 会話文コーパスにおける品詞の使用割合

3.2.2 品詞列に着目した分析

キャッチフレーズの文法構造に着目した分析を行うために単語の連なりを扱う N -gram モデル [23] を用いた。まず、形態素解析で得られた品詞列により、文頭から1-7gramの頻度を計算する。図6に N -gram に対し頻度を求める方法を示す。各キャッチフレーズ、および文に対して品詞解析を行い、 N -gram 目までの品詞が一致した場合加算して、 N -gram の頻度を求める。

一般に N -gram においては、 N の数が増大するほど、一致する確率は下がっていく。具体的には、順方向品詞 N -gram に関して、上位10件の品詞列の全体に占める割合は、 $N=1$ のとき99.8% (6,453個)、 $N=2$ のとき82.9% (5,363個)、 $N=3$ のとき63.8% (4,123個)、 $N=4$ のとき40.7% (2,630個) であり、 $N=7$ のときは7.14% (462個) である。以上のことを考慮し、順方向品詞 N -gram については $N=3$ のときを対象とする。後述する逆方向品詞 N -gram については、キャッチフレーズにおいて頻出することが予想される「体言止め」かつ「記号終わり」など終端の特徴を調べるために $N=2$ のときを対象とする。

表2にキャッチフレーズコーパスにおいて、高頻度であった順方向の品詞3gram上位10件の全体に占める割合を示す。また、該当する品詞 N -gram が他コーパスにおいて占める割合を右側に併載する。表よりキャッチフレーズは名詞から始まるものが多いことがわかる。なお、名詞で始まるものは全体の約7割 (4,515/6,466) 占めている。他コーパスとの比較では、キャッチフレーズコーパスでよく出現した「名詞、助詞、名詞」のパターンが、百科事典・ブログコーパスにおいても高い割合で見られることがわかる。会話文コーパ

名詞,	助詞,	名詞,	名詞,	助詞,	...
接頭詞,	名詞,	名詞,	記号,	名詞,	...
名詞,	名詞,	名詞,	名詞,	名詞,	...
名詞,	名詞,	助詞,	名詞,	名詞,	...
名詞,	記号,	名詞,	名詞,	名詞,	...
助詞,	動詞,	動詞,	記号,	名詞,	...
名詞,	助詞,	名詞,	名詞,	名詞,	...
名詞,	名詞,	記号,	名詞,	名詞,	...
⋮	⋮	⋮	⋮	⋮	⋮
1gram	2gram	3gram	4gram	5gram	...

図6 順方向の品詞 N -gram 解析方法

表2 各コーパスの順方向品詞3gramの頻度割合

※CFはキャッチフレーズを表す

品詞 <i>N</i> -gram	CF (%)	百科事典 (%)	ブログ (%)	会話文 (%)
名詞, 助詞, 名詞	21.6	17.9	18.0	2.32
名詞, 助詞, 動詞	12.1	2.12	6.43	0.947
名詞, 名詞, 助詞	8.71	18.8	8.61	2.02
名詞, 助詞, 記号	5.00	3.04	2.21	1.50
名詞, 助動詞, 名詞	3.94	0.508	1.04	0.223
形容詞, 名詞, 助詞	2.74	0.254	0.584	0.0557
名詞, 助詞, 助詞	2.63	3.12	2.95	0.409
名詞, 名詞, 名詞	2.47	15.5	3.73	0.799
名詞, 助詞, 形容詞	2.37	0.333	1.37	0.297
連体詞, 名詞, 助詞	2.24	2.91	3.15	0.910

スについては、特に類似するパターンは見られない。

終端からの解析も同様に行う。逆方向の解析方法を図7に示す。逆方向であるため表中の左側が終端となっている。逆方向に一致する品詞 *N*-gram を、順方向と同様に加算し頻度を算出する。表3にキャッチフレーズコーパスにおいて、高頻度であった逆方向品詞2gram上位10件の全体に占める割合と該当する品詞 *N*-gram が他コーパスにおいて占める割合を示す。逆方向であるため表中の左側が終端となっている。表より、キャッチフレーズコーパスにおいて記号で終わっているものが全体のほとんどを占めていることがわかる(5,107/6,466)。その中でも、名詞が記号の前に来るものが多く、名詞で終わるキャッチフレーズが合計で2,383個と全体の4割近く占めていることがわかる。他コーパスとの比較では、一番多かった「記号, 名詞」のパターンに関して、百科事典コーパスに類似していることがわかる。

ここで、キャッチフレーズ特有の品詞列の定量的な分析を行う。具体的には、キャッチフレーズコーパスと他コーパス

...	名詞,	助詞,	名詞,	名詞,	記号
		接頭詞,	名詞,	名詞,	記号
...	名詞,	名詞,	助詞,	名詞,	記号
...	助詞,	名詞,	名詞,	名詞,	記号
...	記号,	名詞,	名詞,	名詞,	記号
		助詞,	動詞,	動詞,	記号
...	名詞,	名詞,	名詞,	名詞,	記号
...	名詞,	記号,	名詞,	名詞,	記号
...	...	...	...	...	...
...	5gram	4gram	3gram	2gram	1gram

図7 逆方向の品詞 *N*-gram 解析方法

表3 各コーパスの逆方向品詞2gramの頻度割合

※CFはキャッチフレーズを表す

品詞 <i>N</i> -gram	CF (%)	百科事典 (%)	ブログ (%)	会話文 (%)
記号, 名詞	29.1	20.7	10.2	6.65
記号, 助動詞	15.7	39.2	41.5	20.2
記号, 動詞	15.3	34.7	14.1	2.95
記号, 助詞	13.0	0.631	12.4	46.7
名詞, 名詞	4.58	0.0175	1.83	0.204
名詞, 助詞	3.19	0.00877	0.686	0.204
記号, 形容詞	2.71	2.07	4.09	0.501
動詞, 助詞	2.66	-	0.229	0.149
記号, 記号	2.20	1.39	5.74	0.316
名詞, 動詞	1.62	0.00877	0.279	0.0929

との違いを明確にするために次式のような頻度比を用いる。

$$S_f = \frac{F_c/N_c}{F_a/N_a} \quad (1)$$

F_c はキャッチフレーズコーパスにおける品詞 *N*-gram の頻度、 F_a は他のコーパスにおける品詞 *N*-gram の頻度である。キャッチフレーズコーパスにおける文の数 N_c 、他コーパスにおける文の数 N_a を用い正規化を行っている。

式(1)によって、順方向において各コーパスと比較を行った際の頻度比 S_f を表4にまとめる。まず、表の左側の百科事典との比較では、名詞が支配的なのではなく、動詞・形容詞から始まるところがキャッチフレーズの特徴であることがわかる。表中央のブログコーパスとの比較では、「動詞, 助詞, 記号」の品詞列の組み合わせが最も頻度比において高かった。このような例としては「贈っ(←動詞)て(←助詞), (←記号)もらって, 大人気」というキャッチフレーズが挙げられる。2番目, 3番目についても文頭に近い場所で記号が現れている。従って、キャッチフレーズはブログと異なり、端的で歯切れのよい構造を持つ傾向が強いことがわかる。

表の右側の会話文コーパスとの比較では左側・中央と比べると、右側は全般的に上位の値が高いことが分かる。特に「形容詞, 名詞, 助詞」の構造を頻繁に用いられることが会話文との大きな違いであると言える。これは、会話文では対象となる名詞は文脈上共有できているが、キャッチフレーズは対象の名詞についてどのようなものか知らせる必要があるという違いがあるからと考えられる。

表4 順方向品詞3gram頻度比

キャッチフレーズ VS. 百科事典		キャッチフレーズ VS. ブログ		キャッチフレーズ VS. 会話文	
品詞 <i>N</i> -gram	頻度比 S_f	品詞 <i>N</i> -gram	頻度比 S_f	品詞 <i>N</i> -gram	頻度比 S_f
動詞, 助詞, 動詞	39.4	動詞, 助動詞, 記号	14.0	形容詞, 名詞, 助詞	49.1
動詞, 助動詞, 助詞	26.5	名詞, 記号, 動詞	11.0	動詞, 名詞, 助詞	33.9
形容詞, 助詞, 記号	24.7	名詞, 動詞, 記号	10.4	名詞, 助動詞, 名詞	17.7
動詞, 助詞, 助詞	22.0	動詞, 形容詞, 名詞	7.92	副詞, 助詞, 名詞	15.4
動詞, 記号, 動詞	17.6	動詞, 名詞, 助詞	6.75	名詞, 動詞, 名詞	15.0

表5 逆方向品詞2gram頻度比

キャッチフレーズ VS. 百科事典		キャッチフレーズ VS. ブログ		キャッチフレーズ VS. 会話文	
品詞 <i>N</i> -gram	頻度比 S_f	品詞 <i>N</i> -gram	頻度比 S_f	品詞 <i>N</i> -gram	頻度比 S_f
名詞, 助詞	363	名詞, 形容詞	23.7	動詞, 名詞	24.2
名詞, 名詞	261	動詞, 助詞	11.6	名詞, 名詞	22.4
名詞, 動詞	185	名詞, 副詞	6.09	動詞, 助詞	17.9
名詞, 助動詞	106	名詞, 動詞	5.81	名詞, 動詞	17.5
記号, 副詞	28.2	助詞, 名詞	5.48	名詞, 形容詞	16.2

また、キャッチフレーズの文頭から3gram分の品詞列が全体としてどのような傾向を持っているのか、表2及び頻度スコア式を基に計算を行った。表2において全体の5割以上(3,318/6,466)を占める上位5位までの品詞列は、百科事典コーパスの42.4% (4,839/11,406)、ブログコーパスの36.3% (1,429/3,947)、会話文コーパスの7.02% (378/5,348) で用いられている。従って、文の前半において、キャッチフレーズの品詞列は百科事典及びブログのものに近いことがわかる。

表5に逆方向において各コーパスと比較を行った際のスコア S_f を示す。まず、表の左側の百科事典との比較では、名詞終わりの体言止めが多いことがわかる。この例としては、「ひと口で幸せな気分になれる、グラスの中の(←助詞)パラダイス(←名詞)」のようなキャッチフレーズが挙げられる。百科事典は基本的に説明文であり「である」のような助動詞で終わることが多く、体言止めが多く見られるキャッチフレーズコーパスとの大きな違いとなったと考えられる。表中央のブログコーパスとの比較では、体言止めが多いことの他に、助詞終わりも特徴として考えられることが分かる。これはキャッチフレーズが「ようこそ日本の、ファーストクラスへ」などのようにイメージ形成のために「へ」や「に」などを多用しているからと考えられる。表の右側の会話文コーパスとの比較では「動詞, 名詞」(語順は「名詞, 動詞」)のパターンが多く見られる。この例として「本気モードをクールに刺激する」がある。日本語における会話において、「～みたいな」などの曖昧な表現が多い一方で、キャッチフレーズにおいては終端で歯切れのよい表現が多いためと考えられる。

以上より、キャッチフレーズにおける前半の品詞列は、「名詞, 助詞, 名詞」などのパターンが多く、百科事典やブログの構造に似ていると言える。さらに、後半の品詞列は名詞で終わることが多く、体言止めが多用されている傾向があることがわかる。

3.3 キャッチフレーズコーパスにおける特徴

ここでは、キャッチフレーズコーパスそのものに含まれる単語、係り受け関係、分野ごとの名詞について分析を行い、その特徴を明らかにする。

3.3.1 キャッチフレーズコーパスにおける単語

キャッチフレーズコーパスにおいて多く用いられる名詞について検討を行う。表6に6,466個のキャッチフレーズから抽出された、高頻度で現れた名詞・固有名詞を示す。表より、

名詞については「人」、「あなた」、「私」、「心」などの人を意識したメッセージ性が強い単語が高頻度で現れていることがわかる。一方、固有名詞では地名が多く見られた。

3.2.2の品詞 *N*-gramの解析の逆方向解析において、体言止めがキャッチフレーズの一つの大きな特徴であることが分かった。表7において具体的な例、文末に名詞が来る体言止めのケースを示す。表から「スタイル」「暮らし」など、実際に行動を起こした後の未来をイメージさせるような名詞が多用されていることがわかる。

3.3.2 キャッチフレーズコーパスにおける係り受け関係

前後からの *N*-gramを用いた品詞解析や含まれる単語解析の他に、係る単語、係られる単語を特定することも、キャッチフレーズの特徴を知る上で重要である。ここでは、構文解析器CaboCha[14]を用いて、キャッチフレーズコーパスにおける係り受け関係を分析する。表8に高頻度で現れた名詞・固有名詞を示す。表より、係る側として「この」、「その」、「もっと」のような強調する役割を果たすと考えられる指示語・副詞、「あなたの」のような関係性を強調するもの、「新

表6 高頻度で現れた名詞・固有名詞(上位10件)

名詞	頻度	固有名詞	頻度
人	217	日本	49
肌	197	東京	16
こと	191	イタリア	11
あなた	183	ニッポン	10
夏	173	パリ	9
私	151	ドコモ	7
心	144	銀座	7
自分	122	九州	6
日	119	ハワイ	6
世界	108	フランス	6

表7 高頻度で現れた体言止めにおける単語(上位10件)

単語	頻度
スタイル	30
暮らし	22
時間	21
夏	19
空間	19
デザイン	16
キッチン	16
幸せ	15
贅沢	14
旅	13

表8 高頻度で現れた係る側・係られる側文節（上位10件）

係る側	頻度	係られる側	頻度
この	143	ある	222
その	108	なる	156
もっと	105	いい	102
あなたの	94	ない	78
ある	92	ために	72
新しい	74	こと	62
いい	74	する	46
夏の	61	夏の	44
美しい	58	あなたの	40
大人の	47	あります	34

しい」,「いい」,「美しい」などのポジティブなもの,「夏」や「大人の」などの特定のイメージを喚起する単語が高頻度で現れていることがわかる。係られる側では,「ある」,「なる」などの係る側を現実にする単語が多く現れていることがわかる。

3.3.3 キャッチフレーズコーパスにおける分野ごとの特徴

ここでは, キャッチフレーズコーパスにおける分野ごとの特徴の違いを検討する。具体的には, 分野ごとに使用される名詞の頻度を取得する。表9に各分野において高頻度で出現した名詞を示す。まず食品分野では人々の食生活において, 表現に欠かせない「味」や「人」などが頻繁に利用されていることがわかる。美容分野に関しては, 表6に示したキャッチフレーズコーパス全体においても, 非常に高頻度で「肌」が現れていることがわかる。具体的に「肌」と表現することで対象を明確にして,「美」などの目標を提示することによって, 購買意欲をかき立てていることが推察される。生活用品分野においては,「肌」のような, 非常に高頻度で現れる名詞は存在しなかった。一方で,「自分」,「人」,「私」のような単語が頻度に関して上位になっている。これは人の内面を捉えることと共に, 高頻度だった「部屋」などの単語を取り入れることで, 消費者に対して気づきを与えているためと考えられる。ファッション分野では,「夏」,「日」といった状況に関する単語,「私」のような内面に関する単語,「服」,「スタイル」といったイメージの湧く単語がよく現れている。与えられた状況に関して, 対象となる消費者の心を捉えて, 行動に移すことを意図して作られていることが考えられる。デジタル関連製品分野においては,「ケータイ」や「カメラ」などの具体的なモノと「あなた」や「世界」といった外界との繋がりを強調するような名詞が選ばれていることがわか

る。最後に, ハウジング分野においては対象を指すときに必要な「空間」という単語, 住まう上で欠かせない「キッチン」といった名詞がよく現れていることがわかる。

以上のことから, 分野ごとに高頻度で現れる名詞が違うことがわかり, 特に目的が明白である美容などの分野に関しては, 特定の名詞がよく現れる傾向が見出された。

4. ま と め

本論文では, 統計的処理に基づく言語工学的及びユーザによる感性評価の2つの観点からキャッチフレーズの分析を行った。具体的には, 一つ目の言語工学的分析については, 統計的分析を行うため6,466個の膨大な数のキャッチフレーズを含むキャッチフレーズコーパス及び, 比較対象として百科事典コーパス, ブログコーパス, 会話文コーパスを利用した。また, 分野ごとの特徴を見出すためにキャッチフレーズコーパス内のキャッチフレーズの特徴分析も行った。二つ目の感性工学的分析については, 面白いキャッチフレーズの特徴, 印象に残るキャッチフレーズ中の単語について主観評価実験を行った。その結果, 以下のような知見が得られた。

言語工学的分析については, キャッチフレーズは順方向に関して「名詞, 助詞, 名詞」「名詞, 助詞, 動詞」のような品詞列が多いことがわかった。これに関して, 他コーパスと比較により, キャッチフレーズの順方向の品詞列は百科事典やブログに近いことが明らかになった。一方, 逆方向の品詞列については他コーパスに比べ, 体言止めが多く構造が大きく異なることがわかった。また, キャッチフレーズコーパスにおけるテーマごとの分析の結果,「味」,「肌」など分野ごとに出現しやすい名詞が明確に存在することが示された。

今後の課題として, トрендなどを様々な時代のキャッチフレーズを分析することによって明らかにすることなどが挙げられる。また, これらの得られた知見を利用して, キャッチフレーズ自動生成システムに応用していくことも出来ると考えられる。

参 考 文 献

- [1] 鶴月 洋：広告文の歴史 キャッチフレーズの100年, 日本経済新聞社, 1965.
- [2] 北村 日出夫, 山路 龍天, 田吹日出硯：広告キャッチフレーズ

表9 各分野における高頻度で現れた名詞

食品分野 総数: 495 個		美容分野 総数: 451 個		生活用品分野 総数: 1,133 個		ファッション分野 総数: 1,066 個		デジタル製品 関連分野 総数: 671 個		ハウジング分野 総数: 429 個	
単語	頻度	単語	頻度	単語	頻度	単語	頻度	単語	頻度	単語	頻度
味	33	肌	173	自分	39	夏	63	ケータイ	33	空間	36
人	30	感	31	こと	31	日	36	写真	31	キッチン	33
心	23	美	27	人	31	私	35	あなた	26	暮らし	25
夏	19	夏	23	私	30	服	33	世界	25	窓	25
幸せ	14	白	20	部屋	29	スタイル	33	カメラ	24	家	21

- ズ, 有斐閣, 1981.
- [3] 深川 英雄: キャッチフレーズの戦後史, 岩波新書, 1991.
- [4] 阿久津 聡, 石田 茂: ブランド戦略シナリオ, ダイアモンド社, 2002.
- [5] Chiranjeev Kohli, Lance Leuthesser, Rajneesh Suri: Got slogan? Guidelines for creating effective slogans, Business Horizons, Vol.50, No.5, pp.415-422, 2007.
- [6] Tina Lowrey, L. Shrum: Phonetic Symbolism and Brand Name Preference, Journal of Consumer Research, Vol.34, pp.406-414, 2007.
- [7] 新井 恭子: 関連性理論における「広告のことは」の分析, 東洋大学経営学部 経営論集 (68), pp.79-91, 2006.
- [8] Patricia Wetzel: 広告にみる日本の言語世界, 名古屋大学大学院国際言語文化研究科 国際シンポジウム「異文化としての日本」記念論文集, pp.73-82, 2009.
- [9] 松平 智史, 萩原 将文: 電子化辞書と遺伝的プログラミングを用いたキャッチコピー作成支援システム, 電気学会論文誌C (電子・情報・システム部門誌), Vol.124, No.1, pp.164-169, 2004.
- [10] 松平 智史, 萩原 将文: 対話型遺伝的プログラミングと電子化辞書を用いたキャッチコピー作成支援システム, 電気学会論文誌C, Vol.125, No.4, pp.616-622, 2005.
- [11] 西原 陽子, 砂山 渡, 谷内田 正彦: 聴講者の興味をひく研究発表タイトルの作成支援, 言語処理学会年次大会発表論文集, Vol.13, pp.448-451, 2007.
- [12] 西原 陽子, 砂山 渡, 谷内田 正彦: 多くの興味をひく研究発表タイトルの作成支援, 第21回人工知能学会全国大会, 2H4-5, 2007.
- [13] 中野 俊亮, 鬼沢 武久: 目を引くキャッチフレーズ生成に関する基礎的研究, 日本感性工学会大会予稿集, Vol.9, B71, 2007.
- [14] 中野 俊亮, 鬼沢 武久: ユーザ対話による意外性を持つキャッチフレーズ作成支援, 全国大会講演論文集, 情報処理学会, Vol.70, No.2, pp.2-201-2-202, 2008.
- [15] 越川 靖子: ブランド・ネームにおける語感の影響に関する一考察 ―音象徴に弄ばれる私達―, 商学研究論集, 第30号, 2009.
- [16] 山田 理英: 脳科学から広告・ブランド論を考察する, 評言社, 2007.
- [17] 久野 寧子: カタログチラシ キャッチコピー大百科, ピエ・ブックス, 2008.

- [18] Wikipedia 日本語 アーカイブ, <http://dumps.wikimedia.org/jawiki/>
- [19] KNB コーパス (Kyoto-University and NTT Blog コーパス), 京都大学大学院情報学研究科黒橋研究室, <http://nlp.kuee.kyoto-u.ac.jp/kuntt/>, 2009.
- [20] インタビュー形式による日本語会話データベース, 北九州市立大学国際環境工学部情報メディア工学科上村研究室, <http://www.env.kitakyu-u.ac.jp/corpus/texts/index.html>, 1996.
- [21] 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.
- [22] Taku Kudo, Yuji Matsumoto: Japanese Dependency Analysis using Cascaded Chunking, in Proc. of Natural Language Learning, pp.63-69, 2002.
- [23] 北 研二: 確率的言語モデル, 東京大学出版会, 2001.

山根 宏彰 (非会員)

2010 年慶大・理工・情報工卒. 2011 年同大大学院修士修了. 現在同大大学院博士課程在学中. 感性工学・自然言語処理の研究に従事.

萩原 将文 (正会員)

1982 年 慶大・工・電気卒. 1987 年 同大学院博士課程了. 工博. 同年同大助手. 以来, ソフトコンピューティング, 言語工学, 感性工学の研究に従事. 現在, 同大教授. 1991-92 年度スタンフォード大学訪問研究員. 1986 年丹羽記念賞, 1987 年電子情報通信学会学術奨励賞, 1990 年 IEEE Consumer Electronics Society 論文賞, 1994 年安藤博記念学術奨励賞, 1996 年日本ファジィ学会著述賞, 2003 年日本感性工学会技術賞, 2004 年日本感性工学会論文賞, 2005 年日本知能情報ファジィ学会貢献賞受賞. 電子情報通信学会, 情報処理学会, 日本知能情報ファジィ学会, 人工知能学会, 電気学会, 日本神経回路学会, 日本感性工学会, 日本デザイン学会, 言語処理学会各会員. IEEE シニアメンバ.