

ことばの総合政策

-- アバウト推論による流行ことば予測研究 --

池田 定博^{*)} 金杉 友子^{**)} 松澤 和光^{**)} 金田 重郎^{*)}

skaneda@mail.doshisha.ac.jp

^{*)}同志社大学大学院・総合政策科学研究科

^{**)}NTTコミュニケーション科学基礎研究所

あらまし：効果的な商品コンセプトや広告コピーの作成は、今日の企業活動にとって、極めて重要である。しかし、どのようなコンセプトやコピーが近い将来流行するかを推定したり、コピーの作成を支援する手法は知られていない。本稿では、アバウト推論に基づく「ことばとことばの距離」を利用して、流行語の背後に存在する社会的な要因を明らかにする手法を提案する。これにより、流行語の社会的要因を、時間的推移に従って分析できるだけでなく、創作した流行語候補が流行するか否かの判断を支援できる。自由国民社「現代用語の基礎知識」と流行語大賞によるプロトタイプ実験の結果、アバウト推論的アプローチにより、一定の推定が可能であることを確認した。

1. はじめに

日本列島総不況といわれる現在、企業の担当者は、商品開発やマーケティング等において、効果的なコンセプトづくりに頭を抱えている。コンセプトや広告コピーを作ることは人間のみが成し得る創作行為である。しかし、思いついた多数のコンセプトや広告コピー案の中から、いったいどのようにして、戦略の柱となるキーワード候補を絞り込んでゆけば良いのだろうか？ 候補選択は、現状では、各部署の決定者によって経験的に行われている。しかし、このような、経験的・曖昧な選択基準では、開発されたコンセプト・コピーが「ありふれたもの」になったり、「二番煎じの商品開発」に終わってしまう恐れがある。

本稿では、これに対し、『流行ことば予測』という新しい手法を提案する。そこでは、まず、過去・現在・将来における流行語の背景となる社会的要因を、「ことば」として表現し、その重要性を数値表現する。そして、今後流行の可能性がある新しい「ことば」と、これら社会的要因との距離を計算し、「近い」と算出されたことばから、流行のキーワードを開発することを狙いとする。

以下、第2章では、流行ことばが社会的背景に基づいているとの仮説を示し、第3章で、この仮説に基づいた、流行ことば作成支援手法を提案する。更に、第4章では、自由国民社発行の『現代用語の基礎知識』によるプロトタイプ作成の結果を報告する。最後に、第5章では、今後の可能性について触れる。

2. 「ショムニ」はなぜ流行ったのか？

「ショムニ」ということばがある。昨年（1998年）ヒットしたフジテレビ系ドラマのタイトルで、98年度の流行語大賞にランクインした新語である。このドラマは、ある商社を舞台に、庶務二課（通称ショムニ）とは名ばかりの、備品管理室に追いやられたOLたちが痛快に活躍する、漫画原作のストーリーである。中でも超ミニの制服で「女の価値は男の数で決まるのよ」をキメ台詞にした女優・江角マキコ扮するキャラクターは共感呼び、高視聴率を得た。

このドラマは、なぜ流行したのだろうか？ もちろん江角マキコら人気の女優や人気漫画をもとに作ったという要因がある。しかし、それだけでは、高視聴率の要因としては、不十分であろう。ヒットの理由を思い付くままに列挙すると、以下のような**要因があるように思われる**。

- 長期にわたる不況で、「いつ閑職に追いやられるか分から」ない不安を抱えて日を過ごし、一方で、幹部の無策に不満を抱きながらも声にはできない、多くの「悩める」社員達からの共感
- 言葉はキツイが、一人一人が個性を持ち、周囲への配慮や思いやりのある、かつての「古き良き男社会」へのノスタルジー
- 男女雇用機会平等法の施行による、女性の職場への進出

いずれの理由ももっともらしい。この例からも、流行の背後に社会的・時代的な要因が隠されていることは疑い得ない。しかし、単に、感覚的に思い付くままに、社会的要因を想起するだけでは、数量的な裏付けとはならない。まさに、上記イタリック体部分のように、「**要因があるように思われる**」に過ぎない。

以上の分析から、流行を生むためのコンセプト作り、コピー作成等において、コンピュータに求められる機能として、以下の2つが必要と考える。

- (1) 来年の「流行ことば」を支える社会的要因が何であることをある程度予測したい。この場合、社会的要因を表わすものは、あくまで「ことば」であるが、種々の粒度（抽象度）が考えられる。たとえば、「経済」と言った大きな概念かもしれないし、「サプライチェーン」と言った、細かいレベルのことばの可能性もある。
- (2) コピーライター等により創作された新しいコンセプトやことばが、どんな社会的要因を人々の心の中に想起させるかを定量的に推定したい。そのためには、上記の社会的要因を表わすことばと、創生された新語との間の距離を測定したい。

流行の可能性や、あるいは、新語の中に隠された社会的要因を定量的に明らかにでき

ば、コピー決定の参考とすることができる。また、クライアントに対して、そのコンセプトやコピーをなぜ選定したのかを、客観的・明示的に説明できる。これは、マーケティング上の大きな利点となる。

以上のような観点から、社会的要因を表すことばをもととして、流行ことばを検証し、さらに、その予測を試みる。実現にあたって、著者らは「アバウト推論」に着目した。アバウト推論では、ことばとことばの意味的な距離を数値的・定量的に扱い得る「概念ベース」技術を開発している。しかし、現状の概念ベース技術では、見出し語のソースとして、国語辞書を利用しているため、普通名詞として定着ずみの一般的な見出し語に、扱いうる用語が限定される。本研究のような、新しいことばを扱うのには、既存の辞書は不向きである。そこで、我々は、自由国民社発行の「現代用語の基礎知識」¹を評価対象として選択した。普段慣れ親しんでいる新しい単語に対応できることと、見出し語が多層の上位概念によりグループ化されているので、背景となる「ことば = 社会的要因」の特定に有効と考えたからである。

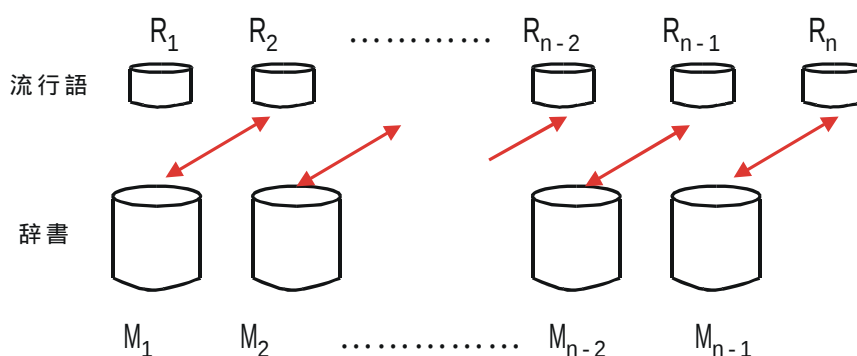


図1 流行ことば予測システム

3. 流行ことばを予測するシステム

本研究では、この数値（スコア）化において以下の手法に着想した。本システムは、広告コピーの作成支援と検証を主目的とするシステムである。但し、過去の何年間かの間の流行語と、各年度において利用されている用語の辞書は、あらかじめ準備されているとする。本システムによる処理ステップを以下に示す。

【STEP1: データの準備】

図1に示すように、過去の流行語の集合を年度毎に準備して R_j ($R_1, R_2, R_3, \dots, R_{n-1}$)とする。各流行語には、その説明文が添付されている。説明文は、距離計算に必要である。また、

¹ 本研究に際して、データ利用の許諾を頂いた、株式会社・自由国民社に深謝いたします。

来年度の流行語の候補を R_n とする。この R_n にも、説明文を付加しておく。また、各年度毎に利用されていたことばの辞書 M_j を準備する。 $j=1,2,\dots,n-1$ である。 j は年度を表わす。来年分の辞書 M_n は空である。

【STEP2: 既知流行語との距離計算】

各年度毎の流行語を前年度の辞書中のすべての見出し語と比較して、距離計算を行う。この距離計算には概念ベース技術を用いる²。これにより、今年を含めて、今年以前の各年度毎の流行語に対して、各年度毎にどのような「ことば = 社会的要因」が影響を与えて来たかを数値的に示す。また、これにより、来年度の流行語に影響を与えるであろう社会的要因を、各年度毎の社会的要因、すなわち、流行語に近い「ことば」の推移から推定する。

【STEP3: 来年度の流行語の予測】

システムの最終的な目的は、来年度の流行語を予測することである。ひとつのアプローチは、上記の分析から推定された、来年度の流行語の社会的背景となることばからコピーやコンセプトを、人間自身が創生することである。もうひとつの方法は、種々の流行語候補を作成して、これがもし流行するとすると、どのような社会的背景となるかを分析することである。この場合、流行語候補には、それを思い付いたコピーライター自身が企画意図として説明を付す。説明付き流行語候補と、今年の辞書との間で距離を計算して、社会的背景を明らかにする。

4 . プロトタイプ実験

4 .1 プロトタイプ実験の概要

前章で提案したシステムの辞書としては、自由国民社の「現代用語の基礎知識」を選定した。理由は以下のようなものである。

- (1) 見出し語（1998 年版で約 9,000 語）として、現代的なことばのみが選ばれており、本研究には都合がよい。現状の概念ベース用辞書の用語と補完的關係にある。
- (2) 毎年（1984 年開始）この見出し語の中から、「流行語大賞」が選定されており、各年毎に流行したことばとして利用できる。
- (3) 見出し語は、その上位概念として 3 レベルのグループを持ち、見出し語のレベルよりも、より高い概念レベルで、社会的背景を分析できる。ちなみに、最上位の階層は、「経済・経営」「文化・芸術」「国際情勢」「社会・生活」「政治」「スポーツ・趣味」「健康・医療」「情報・産業」「科学・技術」「風俗・流行」である。

² 前述のように、実際の概念ベース用辞書は、市販の国語辞書等を元になっているため、見出し語が必ずしも、流行語や社会的背景とはなっていない。このため、今回の実験では、現代用語の基礎知識の説明文から作成した意味ベクトルの内積を利用している。

但し、前章のシステムを実現するには、1984 年以降のすべての現代用語の基礎知識のデータが必要となる。しかし、現状では、そこまでのデータは入手していないので、以下の方針により、予備的な実験を行うこととした。

- (1) 辞書として、1998 年版の「現代用語の基礎知識」のみを用いる。この辞書を 1997 年度までの年度毎の辞書として投入し、1998 年度の流行語を予測する³。
- (2) 各年度の流行語には、「現代用語の基礎知識 1999 年版(CD-ROM版)」に掲載されている 1994 年から 1998 年までの流行語大賞のことばと説明文を利用する⁴。
- (3) 距離の計算は、説明文の意味ベクトルの内積を利用する。ある見出し語中の説明文に現れた単語は、当該見出し語の説明に何度現れても 1 回であるとした。

説明文の形態素解析には奈良先端科学技術大学院大学の茶笥を利用し、意味ベクトルには、普通名詞、固有名詞、サ変名詞、動詞、形容詞を単語として利用した。但し、意味のない形態素（例えば、「さ」「ない」等）が発生するが、これらはソフト的にフィルタリングして除外してある。

4.2 プロトタイプ実験の結果

図 2 は、1984 年から 1997 年までの間で、流行語大賞から近いとされた見出し語に対す

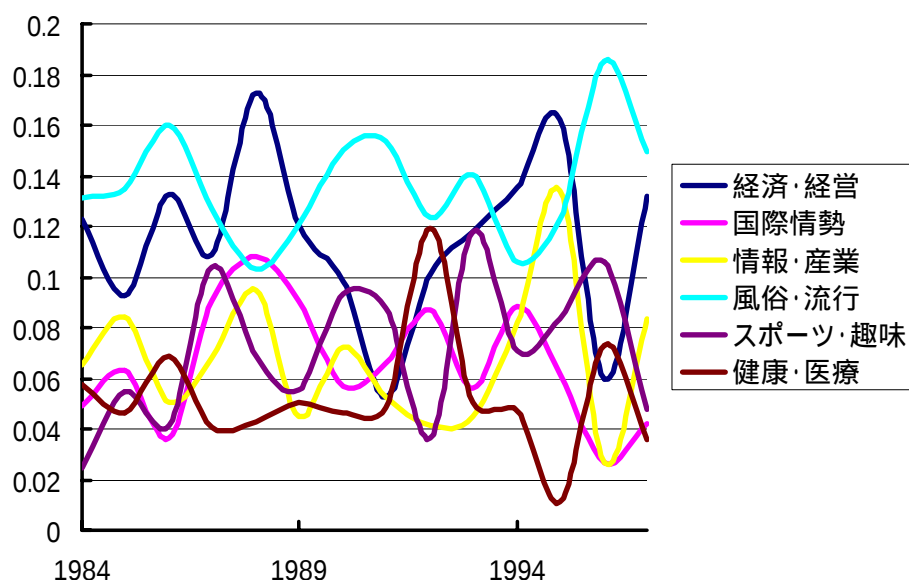


図 2 最上位概念の流行語との平均距離年度推移

³ 1998 年版の現代用語の基礎知識には、1998 年の流行語大賞に該当した単語は含まれていない。従って、1998 年の現代用語の基礎知識を利用して推定する限りは、未知の流行語を予測していることになる。

⁴ 1998 年度流行語大賞の対象となっている新語は、1999 年版現代用語の基礎知識に始めて掲載された。

表 1 最上位概念相互の相関一覧

	経済・経営	文化・芸術	国際情勢	社会・生活	政治	スポーツ・趣味	健康・医療	情報・産業	科学・技術	風俗・流行
経済・経営	1									
文化・芸術	-0.39353744	1								
国際情勢	0.286094692	-0.26056539	1							
社会・生活	-0.38646938	0.202678107	-0.4099075	1						
政治	0.16951184	-0.10229383	-0.44429349	-0.21869022	1					
スポーツ・趣味	-0.22199914	-0.35837963	0.318462392	-0.28776911	-0.37222833	1				
健康・医療	-0.28118191	0.20288349	-0.20436345	0.452936286	0.100561511	-0.58102594	1			
情報・産業	0.636204656	-0.13128267	0.037832203	-0.29506101	0.129603315	-0.39150675	-0.35466862	1		
科学・技術	-0.14200676	-0.33601086	0.266856041	-0.21301528	-0.18747622	0.002896877	0.163084149	0.049444111	1	
風俗・流行	-0.66314263	0.259474729	-0.76556971	0.48989605	0.006215377	0.187724878	0.124394541	-0.49867161	-0.2266854	1

る最上位概念（10 個の概念中 6 個のみ示す）を割合で示したものである。但し、各流行語に対して、20 位まで選んでいる。これによれば、1998 年度では、「経済・経営」が増加しつつあり、また、「風俗・流行」も下がりつつはあるが、かなり高いスコアを持っている。この 2 概念が 1998 年の決め手と予想してもよいであろう。現代用語の基礎知識では、この 10 種類の最上位階層の下にも、見出し語を含めて 3 段階の階層を持つが、どの階層のどのようなことばで、周期的カーブが出現するかは今後の研究課題としたい。

図 2 を見る限り、最上位概念は、相互に無関係とも見える。しかし、最上位 10 概念相互の相関係数を計算してみると（表 1）かなり強い相関を示す場合があり、相互がバラバラに動いていないことを想像させる。具体的には、以下の事が言える。

- ・「情報・産業」と近い流行語は「経済・経営」とも近い。「風俗・流行」への距離は、その逆である。
- ・「風俗・流行」と「国際情勢」は相反している。そして、意外にも「健康・医療」と「スポーツ・趣味」も相反している。

次に、1999 年度版「現代用語の基礎知識」に掲載された新語である（1998 年度流行語大賞）「ショムニ」について、1998 年度版「現代用語の基礎知識」との距離を計算した。結果を表 2 に示す。「現代用語の基礎知識」の用語の中で、「ショムニ」との距離が近い上位 16 個を示した⁵。この場合、「経済・経営」の個数が多く（6/16）、また、上位 10 位中 5 個である。ショムニは、風俗・流行に近いことばであると同時に経済・経営に近い。しかし、意外にも国際情勢に近いのは、興味ある現象である。これは、相関係数の表で見たように、もともと TV ドラマという風俗・流行のことばであるが、図 2 のグラフで上昇傾向にある逆相関の経済・経営、国際情勢の影響を受けていると考えることができる。

以上見てきたように、今回の実験の範囲でも、1998 年の流行語大賞「ショムニ」はかなり良い予想との一致を見た。図 2 のカーブからも、風俗・流行と経済・経営にまたがるような流行語が生まれることが暗示されていたとすれば、ショムニの流行は妥当である。

⁵ 上位 20 位を出力しているが、その中には他流行語大賞があり、その部分は割愛した。

表2 「ショムニ」における上位16の見出し語とスコア⁶。

大分類	中分類	小分類	見出し語	スコア
風俗・流行	ワードウォッチング用語	失業圏する	クサナギ君	109.682
経済・経営	労働運動用語	労働組合の種類と形態	企業別	109.370
情報・産業	マーケティング用語	環境変化とマーケティング	従業員満足	85.498
スポーツ・趣味	相撲用語	大相撲の技術・ルール	死に体	83.771
健康・医療	健康問題用語	医療環境	脳内モルヒネ	81.111
経済・経営	経済社会統計用語	産業統計	事業所・企業センサス	80.143
経済・経営	労働運動用語	労働組合の種類と形態	本部／支部／分会	78.980
経済・経営	労働問題用語	労働時間・休日・休暇等の労働条件	連続有給休暇／リフレッシュ休暇	77.336
国際情勢	中国問題用語	最近の中国の情勢	工場長責任制	76.354
国際情勢	中国問題用語	最近の中国の情勢	工業企業法	74.472
経済・経営	経営問題用語	環境変化と企業	C!	73.577
科学・技術	原子力用語	最新キーワード	廃炉／原子炉廃止措置	73.299
風俗・流行	若者用語	精神状態	いけてる	72.548
経済・経営	労働運動用語	労働組合の種類と形態	管理職ユニオン	72.548
風俗・流行	子ども文化用語	マスコミ文化	「金田一少年の事件簿」	72.180
文化・芸術	現代映画用語	映画作品のジャンル	スペース・オペラ	71.440

4.3 実験から明らかとなった課題

今回の実験では、流行語説明文から生成された意味ベクトルと、辞書説明文から生成された意味ベクトルが共有する要素の数はあまり多くなかった。従って、スコアが、個々のベクトル毎に大きく揺らいでいる可能性がある。今後は、意味ベクトルのマーカー伝播にアバウト推論における概念ベース辞書を用いる等、実質的な共有要素数を増加させて、意味ベクトルのスパース性を補填する工夫が必要であろう。また、説明文に現れた各単語についても、何らの重みづけをしていない。これらの対策により、図2に示したカーブがより、現実と合致したものとなると予想している。更に、見出し語を含めて4段階存在する現代用語の基礎知識のどの階層が、流行語予測に適するかを見極めてゆくことも今後の課題である。

また、これは予想されたことではあるが、流行語側の説明文が少ないと、仮にある見出し語に近いとされても、それを信用することができない。今回の実験では、例えば、1998年度流行語大賞の「だっちゅーの」についても評価をしているが、説明文が短すぎて、十分な結果を得ていない。どの程度の長さがあれば、推計に耐えるのかも、ある程度は定量的に明らかにする必要がある。

5. おわりに

本研究の具体的な活用場として、広告代理店の広告制作を考えている。特に商品コンセプトを消費者に伝える広告コピーの開発において「何を言うか」の部分にヒントを提供し、また、できあがったいくつかのキャッチコピーに対して評価基準を与えることが可能となる。

東京コピーライターズクラブ会長・秋山晶氏によると最近のヒット商品のキーワードは「大きい小さいはもういらぬ。その商品を持ったらどのように幸せになれるかという理

⁶ スコアは正規化されたベクトルの内積であるが、読みやすくするために定数を乗じている。

由のある商品。」であるという。氏が手がけ、昨年ヒットした商品にトヨタ自動車の「プログレ」があるが、そのコンセプトは『小さな高級車』である。なるほど、ヒットのキーワードが商品と相重なった売り文句である。しかし、このように、明確なコンセプトと商品がマッチして成功する例はレアケースであり、売り文句の提案側にも採用側にも正しい基準と正確な決断力が要求される。本提案の手法は、例えば、このような場面で作り手と売り手を手助けできるツールとして進化させていきたい。

【文献】

- (1) 1998 年版現代用語の基礎知識、自由国民社
- (2) 1999 年版現代用語の基礎知識、自由国民社
- (3) 長尾真編「自然言語処理」岩波書店
- (4) 松澤和光、湯川高志、笠原要、藤本和則「アバウト推論技術」NTT R&D Vol.45 No.11, 1996
- (5) 笠原要、松澤和光、石川勉「国語辞書を利用した日常語の類似性判別」情報処理学会論文誌、第 38 巻、第 7 号、pp.1272-1282、平成 9.
- (6) 阿部明典「欠如節を生成する推論法」電子情報通信学会論文誌 D-II, Vol. J81-D-II, No.6, pp.1285-1292, 1998.