

流行現象に関する 消費者視点のテキスト分析

——映画レビューデータによる探索的調査——

佐 藤 圭

1. はじめに

製品やサービス、アイデアなどが世間へ一時的に広く行き渡る現象を、一般に流行 (fashion) と呼ぶ。流行現象の分析においては、医学をはじめ、心理学、社会学、経済学など、様々な知見からのアプローチがある (Solomon, 2013)。しかしながら、とくにマーケティングの分野において、この現象の趨勢を消費者の視点から描き出した研究は、存外少ない。本稿の目的は、流行現象に見られる特徴を、消費者自らによって投稿されたテキスト形式のレビューデータを用いて、探索的に調査することである。

マーケティングの分野で扱われる流行現象について、その参加者を大きく分けると、企業、メディア、そして消費者の3者が考えられる。後述するが、これらのうち、企業およびメディアの行動に着目して、彼らの市場への働きかけやその相互作用によって製品等が社会へ浸透していく過程については、すでに研究がなされている (松井, 2013; 勝又・西本, 2016)。本稿は、残された参加者である消費者に着目し、彼らが流行の最中、どのようなことばでその様子を認識しているのか、その一端を明らかにするための試みである。

ところで、流行は、その現象にかかる時間の長短によって細分化される場

流行現象に関する消費者視点のテキスト分析（佐藤圭）

合がある（Solomon, 2013）。たとえば、非常に短命で一時的な流行のことは、ファッド（fad）と呼ばれる。ファッドのような短期的な熱狂は、事前に発生するかどうかを予測できず、経済的な側面をはじめ、企業活動や消費者にとってはさほど重要なものではないと評価される傾向にある（Kotler and Keller, 2006）。

しかしながら、絶え間なく市場に投入される新しい製品やサービスの数々、目まぐるしい勢いで流れる SNS のタイムライン、あるいは毎分500時間以上の動画コンテンツがウェブ上で共有されるような現在の情報環境において、つい先刻の情報は、現在の情報によって上書きされていく。そのような社会においては、消費者がある特定の製品やサービスを認知し、それらに注意を向ける可能性は極めて低くなるだろう。流行は、少なくともその確率を押し上げる点において、分析対象として魅力的であり、また、その知見を企業やブランドにとってのマーケティング戦略へ組み込もうとする努力は、価値のあることと考えられる。

本稿の構成は、以下の通りである。まず、流行現象に関連した研究をいくつか取り上げ、本稿の調査課題について検討する。つづいて、本稿で利用する分析手法およびデータについて説明する。本稿では、映画作品を分析対象として、流行下における消費者たちの語りの内容をトピックモデルによって明らかにする。その後、分析結果の一部を示したうえで、結びに本稿での発見事項について考察を行う。

2. 関連研究

本節では、おもに流行に関連する研究を紹介し、本稿における調査課題の設定を目標とする。しかし、その前に、まずは流行の区分について確認しておこう。

Solomon によれば、流行（fashion）とは、「ヒット」すれば市場に急速

に広まり、ピークを迎え、そして次第に廃れ忘れられていくプロセス」のことであり、「その受容サイクルの相対的な長さに着目すること」で、流行を類型化できるという (Solomon, 2013, 邦訳書765頁)。このサイクルが極端に長い場合は「クラシック (classic)」と呼ばれ、逆に、極端に短い場合は「ファッド (fad)」と呼ばれる。現象の始まりから終わりまでの経過時間という観点からすると、流行はその中間に位置しているわけであるが、そうであるがゆえに、その特徴が曖昧である。

そのため、現象の両端から流行の特徴に迫りつつ、以下、本稿での調査課題を求めていく。ただし、クラシックについては、対象が消費者に受容された後、なかなか衰退せず社会に定着することから、普及の概念として取り扱うこととする。

2.1 普及からみた流行

はじめに、松井 (2013) による研究を取り上げる。松井は、2000年代から普及したことばの1つである「癒し」ブームに焦点を当てて、その発展プロセスについて検討している。一般に、ブームとは「ある物が一時的に盛んになること」や「急に熱狂的な人気の対象となること」を指し、流行とほぼ同義である。⁽¹⁾しかし、松井は、「癒し」ブームに関して、そのことばの意味が長期にわたって変容しながら社会に普及してきたことを強調している。

松井は、文化的意味をもつ製品・サービスの生産者やそれらを利用する消費者によって共有される認知の観点から、20年分 (1988年から2007年まで) の雑誌記事8000件以上の記事タイトル等を用いて、「癒し」ブームの形成過程を詳細に分析している。その結果、①「癒し」をめぐる主体性の変化、②ジェンダーによる認知の分化、③癒しの世俗化という認知の変化が経時的に

(1) デジタル大辞泉「ブーム」の項目を参照。

流行現象に関する消費者視点のテキスト分析（佐藤圭）

生じていることを明らかにした。このようなことばを巡る意味の変化が生じた理由として、松井は、消費者のニーズ、あるいはそれらへの適応のみならず、企業やビジネス・メディアによる市場への積極的なマーケティング活動の結果としてニーズを創造するという相互作用の存在を挙げている。

次に、勝又・西本（2016）の研究を見てみよう。勝又らは、新製品の普及について、制度論の立場から、企業と消費者の相互作用によって生じる「社会的関心」の変化を、新聞記事に掲載されたことばの変化から読み取ろうとしている。勝又らは、スマートフォン市場を分析対象として、2005年7月から2013年12月までの約7年間にわたる日本経済新聞本誌、日経MJ、日経産業新聞から、記事タイトルに「スマートフォン」もしくは「スマホ」を含む計6000件以上の記事を収集し検証している。収集された各記事の単語を集計し、特定のことばと経過時間などを変数として含む2項選択モデルにより分析した結果、記事中の表現が「スマートフォン」から「スマホ」ということばに置き換わるにつれて、「社会的関心」がその本質的機能（ハードやネットワークの性能など）から補助的機能（ゲームやアプリといったコンテンツやデザインなど）に移り変わっていることを明らかにした。

松井や勝又らの研究に共通して出現し、流行と類似した概念としてよく知られているのが普及理論である（Rogers, 2003）。普及理論では、市場に投入されるイノベーションのライフサイクルにおいて、時間経過とともに、採用する消費者の類型が①イノベータ、②初期採用者、③初期多数派、④後期多数派、⑤ラガードと変化することが指摘されている（Rogers, 2003, 邦訳書228-235頁）。

たしかに、普及理論では、イノベーションの受け手、つまり消費者側におもな視点をおいて、それが社会に普及していくプロセスを記述している。そのため、企業のマーケティング活動やメディアの言説、そして消費者との相互作用によって市場が創り出され、変化するという松井や勝又らの主張と比

較しても、マーケティング研究における流行現象と普及理論をまったく同質的に扱うことは困難であろう。また、普及のプロセスには、流行現象に見られるような対象の新奇性や短命性に関する消費者の心理状態が必須ではないため、流行を普及の一部と見ることは議論が分かれるとも指摘されている(中島, 2013, 103頁)。しかし、それでもなお、流行現象も普及理論も、消費者による採用数の過程はS字曲線を描く点で共通していることから、時間経過によって採用する消費者の類型が異なっていくという普及理論の捉え方は、短期的な流行現象を検討するうえでも、援用され得る視点であると考えられる。

2.2 ファッドからみた流行

さて、ここからは、時間の長短という観点において普及とは対照的な概念であるファッドの視点から、流行について検討する。まずは、Best (2006)の研究を概説する。Bestによると、ファッドはエピソード的(一時的)で、流行はシステマティック(系統的)だとしながらも、両者は明確に二分できるものではなく、「連続的に変化する系列」上にあるものとされる(Best, 2006, 邦訳書32頁)。Bestの分析対象は、医療、経営、科学などの社会制度における熱狂的な広がりであり、Bestはこれを「制度ファッド」と呼んでいる。制度ファッドは、採用者が熱狂している時点で、それが合理的なものであると多くの人々が信じており、これから社会にそれが普及するという錯覚に陥っている。そのため、その制度に懐疑的な人々が制度ファッドの採用者の集団から排斥されてしまい、結果として、ファッドが生じるものと、Bestは主張する。また、ファッドのサイクルを、①発生(emerging)、②急上昇(surging)、③消滅(pursing)と状況によって区分しながら考察している点も独創的であり、それらの期間の特徴は図表1に要約している。

Bestは、映画のような文化産業について、次の流行りが決まっている産

図表1 制度ファッドのサイクルとその主要な特徴

期間	主要な特徴
発生	<ul style="list-style-type: none"> 理論家や発明家といった創始者、それを広報するプロモーター、さらにそれを受容するトレンドセッターにより、制度ファッドが生まれる。 さらに、メディアの新しい情報を報道するという特性や、媒介中心性の高い個人によって、ネットワーク内外に新奇なものが共有される。
急上昇	<ul style="list-style-type: none"> 採用者には、変化した状況に興奮したり、大勢にしたがうことで安心感を得たり、世間から排除されるという不安や恐怖から解放されるような影響が及ぶ。 非採用者には、採用者から採用しないことへの批判が向けられる。
消滅	<ul style="list-style-type: none"> 採用者が増加しすぎたことにより、革新的な採用者にとっては対象の魅力や価値が低減し、また、メディアは古くなったものを報道しなくなり、結果として、対象の情報が消えていく。

出所) Best (2006) を参考に筆者作成。

業構造になっていることから、ファッドではなく、流行として位置づけている。本稿での分析対象は映画であるため、Best の指摘に従えば、本稿の分析対象はファッドではない。しかし、その一方で、両者の線引きは曖昧でもあることから、上述のファッドのサイクルは、流行を考察する際に部分的に援用できるものと推察される。

次に、大森 (2017) の研究について確認する。大森は、Simmel (1957) によるトリクルダウン理論を援用しながら、腐女子と呼ばれるコミュニティで発生するファッドの形成過程について、腐女子と自認する10名の協力者へのインタビューデータにもとづき検証している。ここで、腐女子とは、「男性同士の恋愛を扱った小説や漫画、アニメ、または二次創作作品を好む女性」と定義されている。大森の分析の結果、腐女子のコミュニティにおけるファッドのサイクルは、①定住型（同一ジャンルに長く居座る）、②ノマド型（早期に別ジャンルの流行りを検索して移行する）、③迷惑住人型（人気が高まったジャンルへ参入してきて迷惑行為を働く）という、ファン同士の

異なる3つのグループ間でのインタラクションによって生み出されていることが明らかとなった。大森の研究もまた、特定の流行について、時間経過によってその現象への参加者が変化していることを示している。

2.3 本稿での調査課題

ここまで、関連研究から流行の特徴について議論してきた。そこでの論点にもとづいて、本稿における調査課題を以下に整理する。

【調査課題1】 流行にパターンは存在するか？

【調査課題2】 流行を採用した消費者たちは何を語るか？

【調査課題3】 流行の経過に応じて、消費者の語りの傾向は変化するか？

調査課題1について、流行は、その受容サイクルにかかる期間の長短で類型がなされていることを説明したが、時間という軸以外において、流行をさらに類型化できるかどうかは、検討の余地が残されているだろう。調査課題2について、これまでのマーケティング研究における流行や普及に関する検討では、おもにメディアデータが分析対象となっていたが、流行を形成する主体の1つとして、流行へ参加した多数の消費者の反応がどのようなものであるか、検証する必要があるだろう。調査課題3について、普及やファッドにおいて、時間経過とともに、主要な概念が変質することや、採用に関わる消費者が変化していくことがこれまでに確認された。では、流行についてはどうだろうか。時間経過によって、採用する消費者が変化することで、彼らの語りに何らかの変化が観察されるだろうか。

本稿では、これら3つの調査課題について、次節以降に説明する分析手法とデータを用いて検討していく。

3. 分析手法

本節では、本稿で利用する分析手法であるトピックモデルについて説明する。本稿で対象となるデータは、次節にて説明する映画作品についてのレビューである。レビューはテキストで表現されており、非構造的である。また、レビューは多数の消費者によって投稿されており、文中で利用される語彙も様々である。そのため、単純に単語の頻度を計上するだけでは、レビュー全体に含まれる意味のまとまり、つまりトピックを捉えることは困難である。

そこで、データに含まれた潜在的なトピックを取り出す手法が必要となる。くわえて、個々のレビューは、新聞や雑誌の記事ほどの文量を見込めないため、文脈情報が不足している短い文書にも対応可能な分析手法を選択する。

3.1 トピックモデル

トピックモデルとは、文書をトピックの混合として捉えた確率的生成モデルである。ここでいうトピックとは、話題やジャンルといった意味をもつ。トピックモデルにおいては、Blei et al. (2003) によって提案された Latent Dirichlet Allocation (以下、LDA) が代表的である。LDA のグラフィカルモデルを図表 2 (A) に示す。

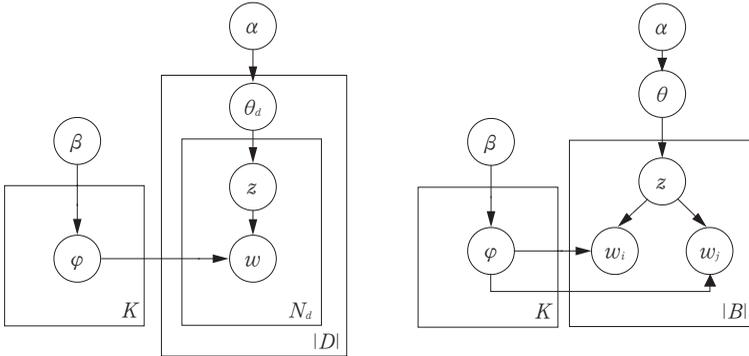
LDA は、与えられた全文書に含まれるトピックと、各文書におけるトピックの混合比を推定することで、文書を要約できるモデルである。トピックは単語分布として表現され、全文書に出現する語彙数の次元をもつ多項分布である。各文書におけるトピックの混合比はトピック分布と呼ばれ、こちらはトピック数の次元をもつ多項分布となる。

LDA では、単語分布とトピック分布を推定するために、各文書に出現する各単語に潜在変数（これがトピックに相当）を割り当てる。つまり、ある

図表2 LDA および BTM のグラフィカルモデル

(A) LDA

(B) BTM



注) D は文書集合, θ_d は文書 d のトピック分布, N_d は文書 d に含まれる単語数, K はトピック数である。それら以外の変数については、本文中で示されている。

出所) Blei et al. (2003) および Yan et al. (2013) を参考に筆者作成。

単語がどのトピックから出現したかを推定することにより、事後的に単語分布とトピック分布を推定する。各単語に割り当てるトピックは、①同一文書内に出現する同一トピックの割合と、②全文書に出現する同一単語に割り当てられた同一トピックの割合を利用することで推定される。

そのため、LDAの問題点の1つとして、ある文書が短く、それに含まれる単語数が少ない場合、上述の①によって文脈を捉えることができず、トピックの推定精度が不安定になることが知られている。本稿で利用するようなレビューデータは、一般に、投稿1件あたりに含まれる文章が短く、推し量れる文脈には限界があると考えられる。したがって、本稿では、Yan et al. (2013) によって提案された Biterm Topic Model (以下、BTM) を利用する。

3.2 Biterm Topic Model

BTMは、たとえばTwitterでのツイートといった、各文書が短い文章で構成されるデータに対して提案されたトピックモデルである。BTMのグラ

流行現象に関する消費者視点のテキスト分析 (佐藤圭)

フィカルモデルを図表 2 (B) に示す。また, BTM における文書の生成過程は以下の通りである。

1. 各トピック z について
 - i. トピック z の単語分布 φ_z をディリクレ分布 $\text{Dir}(\beta)$ から生成。
2. すべての文書に対するトピック分布 θ をディリクレ分布 $\text{Dir}(\alpha)$ から生成。
3. バイターム集合 B における各バイターム b について
 - i. 割り当てるトピック z をトピック分布 θ から生成。
 - ii. 単語分布 φ_z から 2 つの単語 w_i, w_j を生成。

BTM では, まず, 全文書をバイターム集合に変換する。ここで, バイタームとは, ある文章内で近くに出現した単語組のことである。具体的には, ウィンドウと呼ばれる範囲内で出現している複数の単語から, 順序を無視した 2 つの単語組を作る⁽²⁾。この単語組のことをバイタームと呼び, 全文書から抽出された全バイタームをバイターム集合と呼ぶ。BTM では, バイターム集合の各バイタームに対してトピックを推定する。

上述の文書の生成過程から, 1 つのバイターム b の同時確率は,

$$P(b) = \sum_z \theta_z \varphi_{w_i|z} \varphi_{w_j|z} \quad (1)$$

であり, バイターム集合 B の尤度は,

$$P(B) = \prod_{(i,j)} \sum_z \theta_z \varphi_{w_i|z} \varphi_{w_j|z} \quad (2)$$

(2) たとえば, 「ABCD」という順に 4 つの単語が並んでいるとして, ウィンドウを 3 とした場合, まず, 「ABC」から「AB」「AC」「BC」, 次に, 「BCD」から「BC」「BD」「CD」というバイタームがそれぞれ作られることになる。

となる。

BTMで推定すべきパラメータは、 θ と φ と z であるが、周辺化ギブスサンプリング法を用いることで、まずは z のみを推定し、 θ と φ については事後的に推定可能である。バイトーム b にトピック z が割り当てられる確率は、

$$P(z|z^b, B, \alpha, \beta) \propto (n_z + \alpha) \frac{(n_{w_j|z} + \beta)(n_{w_j|z} + \beta)}{(\sum_w n_{w|z} + M\beta)^2} \quad (3)$$

である。ここで、 α と β はハイパーパラメータ、 M は全文書に出現した語彙数、 n_z はバイトーム b にトピック z が割り当てられた数、 $n_{w|z}$ は単語 w にトピック z が割り当てられた数である。

(3)式によって、バイトーム集合における全バイトームへのトピックが推定できれば、単語分布とトピック分布については、以下のように推定できる。

$$\varphi_{w|z} = \frac{n_{w|z} + \beta}{\sum_w n_{w|z} + M\beta}, \quad (4)$$

$$\theta_z = \frac{n_z + \alpha}{|B| + K\alpha}. \quad (5)$$

なお、各文書 d に対するトピック分布は、その文書内で作られたバイトームのトピック割合の期待値として、以下のように求められる。

$$P(z|d) = \sum_b \frac{\theta_z \varphi_{w_j|z} \varphi_{w_j|z}}{\sum_z \theta_z \varphi_{w_j|z} \varphi_{w_j|z}} \cdot \frac{n_d(b)}{\sum_b n_d(b)}. \quad (6)$$

ただし、 $n_d(b)$ はバイトーム b が文書 d に出現する数である。

ここまで、本稿で用いる分析手法について説明した。次節では、BTMを適用するレビューデータについて説明する。

4. 映画レビューデータ

本節では、本稿で分析に利用するデータについて説明を行う。本稿では、分析対象として映画レビューデータを取り上げる。映画を取り上げる理由は、

流行現象に関する消費者視点のテキスト分析（佐藤圭）

上映開始前に入手できる情報が限定的で、製品仕様が明示的でないために客観的指標が制作費などを除きほぼなく、それゆえ、すでに視聴した消費者によるレビューがその作品の潜在的な消費者の購買意思決定にとって決定的に重要であるという性質をもつカテゴリだからである。さらに、各作品の上映期間に一般的な目安はあるものの、実態としては観客動員数に応じて可変であり、流行した作品を識別しやすいという利点もある。

4.1 映画作品の選定

まず、本稿で分析対象とする具体的な作品を選定する。作品の選定にあたっては、興行収入を基準とする。映画産業においては、国内だけでも年間でおおよそ600本以上の作品が制作されていると言われる（河島, 2020）。国外で制作され、国内で上映される作品も併せると、それよりもさらに多くの選択肢が消費者に与えられることになる。そのような状況において、過去の興行収入上位に名を連ねる映画作品を国内で流行したものと見なしても、大きな問題はないであろう。

図表3は、国内で上映された映画の興行収入における上位20位の作品である。図表3の作品のうち、2021年5月16日時点で映画の上映期間がすでに終了しており、かつ、2012年以降に上映された作品を分析対象として選定した⁽³⁾。結果として、本稿では「君の名は。(以下、作品A)」「天気の子(以下、作品B)」「アナと雪の女王2(以下、作品C)」「ボヘミアン・ラプソディー(以下、作品D)」の4作品を分析対象とする。それぞれの作品のストーリーについては、図表4に要約しているので、適宜参照されたい。

(3) 当該基準でいえば、図表3における順位4の「アナと雪の女王」も対象として含まれることになるが、本作品の上映期間中、Filmarksにおいて一定日数のレビュー投稿の欠損が確認されたため、分析対象外としている。また、2012年以降に上映を開始した作品を対象とする理由は、Filmarksへのレビュー投稿の開始時期が2012年からであることによる。

図表3 国内映画市場における興行収入上位20位までの作品

順位	作品名	興行収入(億円)	公開日
1	劇場版「鬼滅の刃」無限列車編	399.7	2020/10/16
2	千と千尋の神隠し	316.8	2001/07/20
3	タイタニック	262.0	1997/12/20
4	アナと雪の女王	255.0	2014/03/14
5	君の名は。	250.3	2016/08/26
6	ハリー・ポッターと賢者の石	203.0	2001/12/01
7	もののけ姫	201.8	1997/07/12
8	ハウルの動く城	196.0	2004/11/20
9	踊る大捜査線 THE MOVIE 2 レインボーブリッジを封鎖せよ!	173.5	2003/07/19
10	ハリー・ポッターと秘密の部屋	173.0	2002/11/23
11	アバター	156.0	2009/12/23
12	崖の上のポニョ	155.0	2008/07/19
13	天気の子	141.9	2019/07/19
14	ラスト・サムライ	137.0	2003/12/06
15	E. T.	135.0	1982/12/04
15	アルマゲドン	135.0	1998/12/12
15	ハリー・ポッターとアズカバンの囚人	135.0	2004/06/26
18	アナと雪の女王2	133.7	2019/11/22
19	ボヘミアン・ラブソディ	131.1	2018/11/09
20	ジュラシック・パーク	128.5	1993/07/17

注) 2021年5月16日時点のデータである。

出所)「興行通信社 (<http://www.kogyotsushin.com>)」にもとづき筆者作成。

これら4作品について、映画レビューサイト「Filmmarks」より、取得可能なレビューデータをクローリングによりすべて取得した。⁽⁴⁾

(4) Filmmarks (<https://filmmarks.com/>) は株式会社つみきが運営する映画レビューサイトである。また、本稿で扱うレビューデータは、2021年5月16日時点で取得可能なものである。

図表4 分析対象である各作品のおもなストーリー

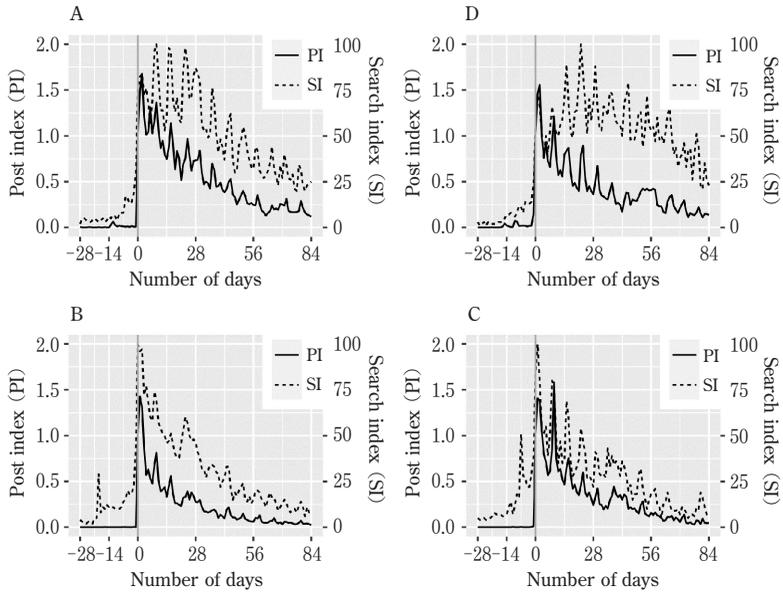
作品名	配給	ストーリー
君の名は。 (作品A)	東宝	千年ぶりの彗星の来訪を1か月後に控えた日本。山深い田舎町で暮らす三葉は、都会へ憧れを抱き憂鬱な日々を過ごしていた。そんなある日、三葉は自分が東京に住む男の子になっている夢を見て、都会の生活を満喫する。一方、東京で暮らす瀧も奇妙な夢を見ていた。
天気の子 (作品B)	東宝	高校1年生の夏に離島から家出し、東京にやってきた帆高。しかし彼の生活はすぐに困窮し、孤独な日々を果てにようやく見つけた仕事は、怪しげなおカルト雑誌のライター業だった。連日雨が降り続ける中、帆高は都会の片隅でひとりの不思議な少女・陽菜と出会う。
アナと雪の女王2 (作品C)	ディズニー	アレンデル王国を治め、城に住むこととなったエルサとアナの姉妹。幸せな日々を過ごしていたふたりだったが、ある日、エルサにだけ不思議な“歌声”が聞こえてくる。そして、エルサとアナはその歌に導かれて、未だかつて訪れたことのない地へと旅立つことに。
ボヘミアン・ラプソディ (作品D)	FOX	1970年、ロンドンで生活する青年フレディー・マーキュリーは、昼間は空港で働き、夜はライブハウスに入り浸る生活を送っていた。そんなある日、ブライアン・メイとロジャー・テイラーのバンドのボーカルが脱退。そこで彼は、自らを売り込みに行くことに。

出所)「映画ナタリー (<https://natalie.mu/eiga>)」にもとづき筆者作成。

4.2 分析対象作品に関するレビュー投稿件数と検索指数の推移

レビューデータを取得した4作品について、それぞれ上映開始28日前(4週間前)から、上映開始84日後(12週間後)までのレビュー投稿件数の動向を確認する。同時に、同作品に関する検索指数の推移についても確認する。⁽⁵⁾ 検索指数を比較する理由について、たとえば、Goel et al. (2010)では、米

図表5 各作品のレビュー投稿件数および検索指数の推移



注) 左上のグラフから反時計回りの順に、作品A、作品B、作品C、作品Dの推移を示している。実線が投稿件数、破線が検索指数をそれぞれ表している。横軸は上映開始日からの経過日数であり、0は上映初日、負数は上映開始前である。縦軸左側がレビュー投稿件数、縦軸右側が検索指数に関する目盛りである。ただし、レビュー投稿件数については、4つの作品を同じ尺度で比較できるように、上映初日の投稿件数が1となるように正規化している。

出所) 筆者作成。

国の長編映画に関する検索指数が当該作品の公開初週の興行収入を説明できることを明らかにしており、本稿においても映画作品との関係性が深いと判断したからである。また、レビューを投稿した採用者は、消費者の中でも、映画を比較的好む集団である可能性が高い。検索指数は、そのような消費者

(5) 検索指数は「Googleトレンド」から取得した。検索条件として、検索クエリには各作品名、検索期間には各作品の上映開始28日前から上映開始84日後までをそれぞれ指定した。なお、本サービスから取得できるデータは、指定された期間において、検索数が最大となった日を100として正規化されている。

流行現象に関する消費者視点のテキスト分析（佐藤圭）

以外の思考や行動も反映されていると見込まれるため、それらを押し量るうえで重要な情報と考えられる。各作品のレビューの投稿件数および検索指数を可視化したものが図表5である。⁽⁶⁾

さて、図表5について、まずは投稿件数の推移から確認してみよう。投稿件数の推移は、グラフ中の実線で表記されている。ただし、投稿件数については、すべての作品を同様の尺度で表現できるように、上映初日に投稿されたレビュー件数を1として正規化している。投稿件数の推移について確認すると、上映開始後、概ね、初めての週末で投稿件数がピークを迎え、その後、増減を繰り返しながら減少していくことがわかる。なお、頻繁に増減を繰り返している理由は、図表6に示すように、各週末に作品を視聴してレビューを投稿している消費者が多い傾向にあるからと考えられる。なお、本稿で扱うデータはレビューということで、原則、作品を視聴した後に消費者から情報発信がなされるため、上映開始前の投稿についてはほぼ存在しない。

図表6 上映開始後における各作品の曜日別レビュー投稿件数の割合

	月曜日	火曜日	水曜日	木曜日	金曜日	土曜日	日曜日
作品A	15.1%	12.2%	12.4%	12.4%	13.3%	15.9%	18.6%
作品B	14.8%	10.6%	11.3%	10.4%	13.9%	18.9%	20.1%
作品C	14.9%	12.0%	11.7%	10.2%	12.9%	17.1%	21.3%
作品D	14.4%	11.1%	12.1%	10.6%	13.4%	17.7%	20.6%

注) 各作品について、上映開始から84日後までのレビュー投稿件数が集計対象である。
出所) 筆者作成。

(6) たとえば、石井(2021)は、数理モデルの観点からヒットした映画作品の動向を検討している。その検討の中で、石井は、特定の作品に関連するTwitterやブログでの投稿件数の推移を上映開始前後の期間で可視化しているが、それらは本稿で示したものと同様の傾向で推移している。つまり、本稿で扱うレビューデータのみが、他のSNS等で投稿されるデータとまったく異なる性質を持っているとは考えにくく、本稿の調査結果が一般化可能であることを示唆している。

つづいて、検索指数について確認してみよう。検索指数の推移は、グラフ中の破線で表記されている。上述したように、検索指数は、レビューの投稿者以外の幅広い範囲における消費者たちを含んだ、各作品への注目度と捉えることができる。したがって、レビュー投稿件数とは、やや異なる推移をしていることがわかる。

上映開始前について、すべての作品に共通して、上映初日に向かって検索指数が徐々に高まっていることが見て取れる。また、作品Bおよび作品Cについては、上映開始前の期間において、検索指数が部分的にスパイクしている様子が確認できる。これは、上映開始前の作品に関する情報が公開された結果、消費者たちの注目が一時的に集まったものと推測される。

一方、上映開始後の趨勢は、作品Aおよび作品D、作品Bおよび作品Cで異なる傾向を示している。前2作品については、上映開始28日後付近を目安に検索指数が最大となっていることに対して、後2作品については、上映開始と同時に検索指数が最大となっている。これは、前2作品への注目度が、後2作品と比較して、上映開始後に複合的な影響を受けた消費者たちによって徐々に高まっていったものと推測される。

ここで、上映開始前後の検索指数の様子から、作品Bおよび作品Cを注目先行型、作品Aおよび作品Dを注目遅行型と呼ぶことにしよう。さらに、これら2つの型と投稿件数との関係を確認してみると、注目先行型よりも注目遅行型の方が比較的長期にわたって作品に対する議論が続いていることがわかる。具体的には、各作品の上映初日の投稿件数を1としたとき、上映開始から84日後の投稿件数は、作品Bは約0.02、作品Cは約0.04であることに対して、作品Aは約0.12、作品Dは約0.14となっている。これはつまり、同一カテゴリであっても異なるパターンの流行現象が存在する可能性を示唆しているものと考えられる。

それでは、消費者たちはレビューにおいて何を語っているのだろうか。

流行現象に関する消費者視点のテキスト分析（佐藤圭）

次節では、まず、各作品の上映期間全体において推定されたトピックを概観する。さらに、本節で得た発見事項にもとづき、各作品の上映期間を、Ⅰ期（上映前期：上映開始～28日後）とⅡ期（上映後期：上映開始29日以降）に分類し、それぞれの期間におけるトピックの特徴的な動向を検証する。

5. BTM によるトピックの推定結果

本節では、BTM と映画レビューデータを利用して、流行現象下で消費者がどのようなことを語っているのか、トピックの推定を行い、その結果の一部を示す。まず、レビューデータに対する前処理の手順と BTM のパラメータの設定について述べる。その後、各作品のレビューデータ全体からどのようなトピックが発見されたのかについて記述し、さらに流行のパターン（注目先行型・注目遅行型）や流行の経過時期（Ⅰ期・Ⅱ期）によるトピックの特徴についても比較する。

5.1 レビューデータの前処理

前節で示した通り、各作品の分析期間は、上映開始から84日後までの85日間とする。この期間に投稿されたレビューをそれぞれ分析対象とする。具体的には、作品 A は15,126件、作品 B は18,258件、作品 C は10,392件、作品 D は25,743件の投稿が確認された。

次に、これらのレビューを対象として、形態素解析器 Mecab により形態素解析を実施した。形態素解析の際、「名詞」「動詞」「形容詞」以外の品詞や、「する」「思う」「できる」といったそれだけでは意味を持たないことばについては除去した。なお、本稿では、Mecab のシステム辞書として、新語や固有表現の抽出に適している mecab-ipadic-NEologd を利用した（佐藤他, 2017）。

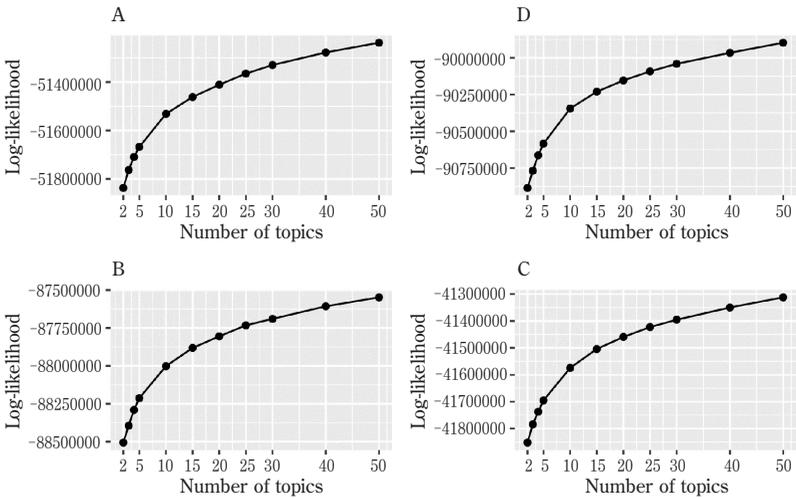
効率的な推定のため、各作品において、分析対象のレビューの中で出現頻

度が1回の単語は削除した。また、1件のレビューにつき、含まれる単語が10個に満たない投稿は、ノイズとなり得るため、削除した。最終的に、作品Aは12,148件、作品Bは15,274件、作品Cは8,524件、作品Dは20,777件のレビューデータが残った。これらのデータに対して、BTMを適用する。

5.2 BTMのパラメータ設定

ハイパーパラメータについては、Yan et al. (2013) に依拠し、 $\alpha=50/K$ 、 $\beta=0.01$ とした。ギブスサンプリングのイテレーションについては1000回、ウィンドウのサイズについては10とした。トピック数については、 $K=\{2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50\}$ と設定してそれぞれ推定し、各トピック数におけるモデルの対数尤度を比較した。図表7を確認する限り、候補のトピック数において、対数尤度は上昇し続けたが、対数尤度の上昇率が緩やかにな

図表7 モデルの対数尤度とトピック数



注) 各グラフにつき、横軸はトピック数、縦軸はモデルの対数尤度を示している。
出所) 筆者作成。

流行現象に関する消費者視点のテキスト分析（佐藤圭）

る点と、推定された各トピックの解釈の容易性を考慮して、本稿では作品一律にトピック数を30と設定した。なお、推定にはRおよびBTMパッケージを利用した。

5.3 推定されたトピック

推定されたトピックの様子を概観するために、各作品について、トピック分布の確率降順に1位、10位、20位の3件のトピックの単語分布を確認する。また、各トピックから生成されている確率が高いと推定された実際の投稿データについても、確率降順に3件取り上げて、併記する。結果については、図表8-aから図表8-dに掲載している。

さて、図表8-aから図表8-dについて、各トピックの単語分布と投稿事例を照らし合わせながら、掲載されているトピックを順に解釈していこう。ちなみに、当該図表における各トピックの単語分布は、そのトピックと関係の強い順に10件の単語を示している。

まず、図表8-a（作品A）についてである。「a01」トピックは、作品の良し悪しや注目すべきポイントを説明している総合的な評価であることが伺える。「a10」および「a20」トピックは、作中で印象に残った演出やストーリーの説明であることがわかる。次に、図表8-b（作品B）についてである。「b1」および「b10」トピックは作中の演出、「b20」トピックは作品のストーリーに関する説明である。つづいて、図表8-c（作品C）についてである。「c1」トピックは作品の演出、「c10」トピックは作品のストーリーを説明しており、「c20」トピックは作品を視聴するにあたっての期待を捉えたトピックであると言えよう。最後に、図表8-d（作品D）についてである。「d01」トピックは作品を視聴している状況についての説明であり、「d10」トピックは作中のシーンの説明、「d20」トピックは対象についての感情に関するトピックと考えられる。

図表 8-a 推定されたトピックの一部 (作品A)

a01 (0.109)	見る (0.023)	投稿事例
	映画 (0.021)	なぜヒットしたのか。泣かせよう、中学生にウケさせようという魂胆だけが先走り無理やりうすっぺらい感動物語にしたようだ。つじつまが合わないところも多少まあ中学生は大好きな内容だった
	良い (0.015)	
	ない (0.013)	
	感じ (0.013)	よかった！ほーんまに感動した！運命的すぎて、非現実的などこあるけど、それでもやっぱり引き込まれた！最初の方ハテナ多すぎて、分からなかったけど！総合的にほんまによかった！
	綺麗 (0.012)	
	好き (0.011)	
	いい (0.011)	見る前から好きじゃないだろうなあって思ってたけど一応見てみたら、やはり好きじゃなかった。つまらない訳ではないけど、凡作だと思う。いくらなんでも色々強引すぎる。有無を言わせぬ面白さがあれば気にならないけど、なんで？どうして？が先行してしまうくらいの物語だったし。
	すごい (0.010)	
観る (0.010)		
a10 (0.037)	美しい (0.034)	投稿事例
	綺麗 (0.029)	光がすごく綺麗。空の光、日の光、水の光、涙の光。私には光の粒がこれから先のこと、まだ先に道は続いているんだって言うてるような気がした。
	映像 (0.021)	
	描写 (0.020)	
	風景 (0.017)	
	シーン (0.014)	なんてピュア彗星の色使いや東京の街に降る雪、夜景や空の移り変わり、パノラマ風景めっちゃ良い感じ
	見る (0.013)	
	作品 (0.012)	
	彗星 (0.012)	噂以上の映像美に度肝を抜かれました。日常風景が流れる一コマコマにこだわった美しさがあったと感じました。
描く (0.012)		
a20 (0.017)	彗星 (0.027)	投稿事例
	三葉 (0.018)	(前略)・糸守湖は隕石の落下によって1200年前に出来た(ティアマト彗星は1200年周期)・宮水家特有の能力なのか、他人の人生と入れ替わりは、おばあちゃんも三葉の母も経験している・瀧と三葉には3年の時差このことから、もしかして過去の隕石の落下時にも同じ場面があったのでは…？(後略)
	隕石 (0.011)	
	瀧 (0.011)	
	世界 (0.011)	(前略)そして口噛み酒は神への奉納ではなく悲劇を避けるため、「つながる」ためのトリガーとして登場します。現代もさまざまな災害や事件を忘れないで後世に紡いでいくことをじんわり感じました
	落ちる (0.010)	
	救う (0.010)	
	糸守 (0.009)	糸守1000年の歴史。なんとか後世に伝えようとする。文字よりも長く残る方法で。彗星を龍として。彗星を紐として。割れる彗星を、舞いのしぐさにした。形に刻まれた意味は、いつか必ずまたよみがえる。
	災害 (0.009)	
ない (0.008)		

注) トピック分布の上位1位, 10位, 20位のトピックについて記載している。なお, 括弧内の数値はトピック分布および単語分布における確率を表している。

出所) 筆者作成。

流行現象に関する消費者視点のテキスト分析（佐藤主）

図表 8-b 推定されたトピックの一部（作品 B）

b01 (0.088)	映像 (0.027)	投稿事例
	見る (0.022)	Aug 17th. 感動してしまった。映像もほんとに綺麗で音楽も最高でした。君の名はよりストーリーが難しくなくてみやすかった
	観る (0.020)	
	綺麗 (0.020)	
	君の名は (0.020)	
	音楽 (0.019)	
	映画 (0.019)	
	良い (0.018)	
	作品 (0.018)	
ストーリー (0.016)	ただただ映像の美しさに圧倒されました！またこまかく映像を見たい。テロップ見てもう一度見たくなった四葉ちゃんはどこにいたの??	
b10 (0.039)	雨 (0.056)	投稿事例
	描写 (0.029)	雨粒や水滴、雲の切れ間から差す陽の描写は本当にきれい！ストーリーはガバガバ…観ながら長くなって思ってしまった
	美しい (0.027)	
	映像 (0.025)	ねえ、今から晴れるよ！ただただ只管に映像が綺麗！降り注ぐ雨と雲間から差し込む光の描写が良い！
	綺麗 (0.024)	
	光 (0.014)	
	空 (0.013)	
	東京 (0.013)	
	表現 (0.013)	セカイ系つよめ,, (´・ω・`) けど陽が射す描写とか、雨の水滴とか、細部まで忠実な背景とか流石すぎて。
水 (0.011)		
b20 (0.017)	帆高 (0.031)	投稿事例
	東京 (0.030)	高校1年生の夏、帆高は離島から逃げ出して東京に行くが、暮らして困ってうさんくさいオカルト雑誌のライターの仕事を見つける。雨が降り続くある日、帆高は弟と二人で生活している陽菜という不思議な能力を持つ少女と出会う。 (前略)そして、雨が続く東京は…特殊能力を持つ陽菜と出会った帆高が、世界を変えるオカルト的で、王道のボーイミーツガールストーリー。オカルト雑誌のライター須賀（#小栗旬）とその助手夏見（#本田翼）が魅力的なキャラでした。
	陽菜 (0.029)	
	家出 (0.023)	
	少年 (0.022)	
	少女 (0.016)	
	出会う (0.016)	
	晴れ女 (0.014)	
	雨 (0.013)	
晴れ (0.013)		

注) トピック分布の上位1位, 10位, 20位のトピックについて記載している。なお、括弧内の数値はトピック分布および単語分布における確率を表している。

出所) 筆者作成。

図表 8-c 推定されたトピックの一部 (作品 C)

c01 (0.094)	映像 (0.034)	投稿事例
	エルサ (0.030)	ディズニー版『ハウルの動く城』。たしかに、映像も音楽もすばらしく、一見の価値もあるし、一聴の価値もある。しかし、ストーリーは支離滅裂で、超感覚的に突き進み、『ハウルの動く城』を思い起こしてしまった。
	歌 (0.020)	
	前作 (0.019)	え、よかった～ストーリーも映像も、面白いし楽しいし綺麗だし、純粋に引き込まれた何より松たか子の歌声。圧倒的。
	良い (0.018)	
	オラフ (0.017)	ものすごい迫力と映像美！松たか子のすばらしい歌声だけでも観に行く価値あり。
	ストーリー (0.016)	
	シーン (0.016)	
綺麗 (0.015)		
美しい (0.013)		
c10 (0.039)	エルサ (0.064)	投稿事例
	前作 (0.035)	エルサになぜ力を与えられたのか、アナとエルサの親はなぜ旅に出たのかという1の大きな謎がきちんと解明されるのが良かった～！続編の意味ってあるのかなと思ってたけどちゃんとあった！
	力 (0.026)	
	魔法 (0.024)	前作では描かれていなかった出生の秘密や両親の死について詳しく描かれていた。相変わらず綺麗なエフェクトと歌も良くて楽しめた。
	アナ (0.020)	
	物語 (0.014)	力の秘密、出生の秘密、国の隠された歴史…なんだかワードだけ抜き出すと完全に英雄物語で、実際英雄譚でとてもワクワク観れた！オラフですらシリアスし出す終盤、何もわかってないはずなのに「どうしたらいい？」「わかった」のクリストフのイケメンぶり好き。
	両親 (0.014)	
	姉妹 (0.014)	
	ストーリー (0.013)	
謎 (0.012)		
c20 (0.023)	前作 (0.041)	投稿事例
	続編 (0.036)	2014年アナと雪の女王の続編前作を超越するエルサの魔法の秘密が明らかになる・2014年アナと雪の女王は、社会現象をもたらした金字塔・松たか子さんの Let It Go も話題!! □ (後略)
	アナと雪の女王 (0.030)	
	作品 (0.028)	(前略) 大ヒットディズニーアニメ「アナと雪の女王」の続編。前作が大ヒットし過ぎたので、続編はどうか～と思っていたけど、わたし的には残念だった。・美しい画像、可愛いオラフ、良い面もあるのだけど、ストーリーがイマイチで何回か意識が飛んでしまった。(後略)
	ディズニー (0.023)	
	ヒット (0.014)	あんなに大ヒットしたアナと雪の女王の2なんて、きっと売れたからとりあえず続編を作ったんでしょ？大人の事情でしょ？なんて、観る前は斜に構えていました。売れたからとりあえず続編を作った？そんなこと見終わった後にきっと誰も言えないはず。私は感極まって号泣でした。(後略)
	観る (0.013)	
	期待 (0.013)	
	鑑賞 (0.012)	
	作る (0.012)	

注) トピック分布の上位1位, 10位, 20位のトピックについて記載している。なお、括弧内の数値はトピック分布および単語分布における確率を表している。

出所) 筆者作成。

推定されたトピック全体の傾向として、レビューに関する局所的なトピックが推定されていることがわかる。また、一部の投稿事例においては、文脈情報の不足によりトピック分布の推定が不安定になっているものもある。しかしながら、各トピックについては、トピックの単語分布および当該トピックと関連の強い投稿事例を併せて確認することで、概ね、その意味するところを解釈可能である。

ただし、作品ごとに個別具体のトピックを追跡したところで、流行下における消費者の語りを捉えるには不十分である。そこで、全作品に共通した解釈ができるよう、類似していると筆者によって判断されたトピックをさらに類型化する。この類型化されたトピックの集合を本稿では「トピック群」と呼ぶことにする。

本稿の推定結果からは、7つのトピック群が見い出された。その7つとは、①テーマ説明（作品のあらすじや、作品が取り扱うテーマや題材などの説明）、②シーン説明（①よりも局所的で、作品中でおもに動作などをとまなう個別場面の説明）、③演出説明（作品で利用されている音楽や映像の美しさ、演者や演技などに関する説明）、④視聴感情（作品視聴時の消費者の心の揺れであり、おもに泣くことや共感するといった感情が変化したことへの気づき）、⑤視聴期待（視聴作品に対する興味や関心）、⑥視聴状況（視聴した場所や視聴時の周囲の状況など）、⑦総合評価（作品に対する良し悪しや作品に対する視点などを検討）、である。ここで、作品Aについてのみ、各トピック群へ全トピックを割り当てた様子を、図表9に示しておく。⁽⁷⁾

(7) 各トピック群へのトピックの割当件数については、作品A・作品B・作品C・作品Dの順にそれぞれ、①は7件・4件・8件・4件、②は4件・6件・3件・6件、③は7件・5件・6件・2件、④は1件・3件・3件・4件、⑤は2件・1件・2件・1件、⑥は2件・2件・1件・2件、⑦は6件・8件・6件・10件であった。なお、すべての作品について、1件のみ、レビュー言語が英語か否かを判断する役割しか持たないトピックが確認された。当該トピックについては、トピッ

流行現象に関する消費者視点のテキスト分析（佐藤主）

図表 9 作品 A のトピック群

① テーマ説明	t02	人, 忘れる, 自分, 映画, ない, 夢, 感じる, 見る, 作品, しれる
	t07	作品, 新海誠, 描く, 秒速, 二人, 距離, 秒速5センチメートル, 星, 男女, 新海
	t11	三葉, 瀧, 東京, 田舎, 都会, 入れ替わる, 住む, 男子, 高校生, 女子高生
	t17	彗星, 三葉, 隕石, 瀧, 世界, 落ちる, 救う, 糸守, 災害, ない
	t21	展開, 入れ替わり, 入れ替わり, ストーリー, 話, 男女, ない, 良い, 物語, 感じ
	t25	入れ替わる, 入れ替わり, 2人, お互い, 三葉, ない, 瀧, 知る, 夢, 二人
	t28	結び, 人, 組紐, 糸, 三葉, 結ぶ, 言葉, 彗星, 繋がる, 運命
② シーン説明	t10	先生, 言の葉の庭, 作品, 出る, 新海誠, 嬉しい, 秒速, 笑, あと, 見る
	t12	シーン, 電車, ない, 瀧, 髪, 要る, 扉, あと, ドア, 三葉
	t20	三葉, 瀧, シーン, 名前, 書く, 忘れる, 2人, お互い, 最後, 会える
	t23	三葉, シーン, 先輩, 瀧, 言う, おっばい, 好き, 揉む, 泣く, 奥寺
③ 演出説明	t03	東京, 風景, 田舎, 新宿, 都会, 出る, 綺麗, 見る, 描く, 行く
	t13	声, 神木くん, 良い, 声優, 可愛い, 演技, 三葉, 長澤まさみ, すごい, 上白石萌音
	t15	RADWINPS, 曲, 音楽, RAD, 映画, 良い, 流れる, 好き, 歌詞, 家
	t16	作品, 新海誠, 監督, 新海, 新海監督, 言う, ない, 作る, ファン, RADWINPS
	t18	映像, 音楽, ストーリー, 映画, 良い, 作品, 綺麗, 観る, 美しい, ない
	t24	映像, 見る, 観る, 映画, 綺麗, 音楽, 鳥肌, すごい, ストーリー, 作品
t30	美しい, 綺麗, 映像, 描写, 風景, シーン, 見る, 作品, 彗星, 描く	
④ 視聴感情	t09	映画, 観る, 見る, 心, 最後, 作品, 感じ, 良い, 涙, シーン
⑤ 視聴期待	t05	観る, 作品, 期待, 見る, 良い, ない, 行く, 映画, 面白い, 評価
	t14	観る, 見る, 映画, 作品, 鑑賞, 読む, ネタバレ, 君の名は, 新海誠, 書く
⑥ 視聴状況	t04	見る, 観る, 映画, 映画館, 行く, 作品, 人, 良い, ない, 観
	t29	鑑賞, 上映, 1, 観る, 満席, 2, 公開, 劇場, スクリーン, 席
⑦ 総合評価	t01	見る, 映画, 良い, ない, 感じ, 綺麗, 好き, いい, すごい, 観る
	t06	アニメ, 映画, 見る, 日本, ない, 感じる, ジブリ, 観る, 感じ, 作品
	t19	映画, ない, 作品, 言う, 観る, 自分, 新海誠, アニメ, 気持ち, 恋愛
	t22	作品, 新海誠, 観る, 新海, 映画, 見る, 感じ, 好き, 良い, 感じる
	t26	作品, 映画, ない, 感じる, 物語, 世界, 描く, 人, 新海誠, 見る
t27	ない, シーン, 感じる, 良い, 三葉, 気, 描写, 描く, 部分, 作品	

注) 1件のみ、レビュー言語が英語か否かを判断する役割しか持たないトピックが確認された。当該トピックについては、トピック群の類型化の対象外としている。なお、各トピックの名称番号はトピック分布の順位を表すものではない。

出所) 筆者作成。

それでは、最後に、流行下におけるこれらのトピック群の推移について、分析を行う。

ク群の類型化の対象外としている。

5.4 流行の経過時期によるトピックの変動

本節の最後として、前項で集約されたトピック群を利用して、流行下の消費者の語りの動向について検証を行う。その際、前節で示された流行の時期（Ⅰ期・Ⅱ期）を対象に、それぞれの期間において、各トピック群に特定の傾向が確認されるかどうかを検証する。これは、関連研究で示されたように、流行の経過時間によって、採用する消費者に何らかの変化が見られるかどうかを確認するためである。

まず、各作品における日次のトピック分布は、当該日に投稿されたレビューのトピック分布の平均として算出した。その後、トピック群の類型にしたがって、それぞれのトピック分布の平均を合算したものを、日次のトピック群の比率とした。

検証方法としては、日次ベースにおけるトピック群の比率が一定割合で増減していることを仮定して、目的変数には日次のトピック群の比率を対数変換した数値を、説明変数には経過日数（1から始めて、日数が経過するごとに1を加算）をそれぞれ設定し、前節のⅠ期とⅡ期の期間に分けたうえで、各作品について回帰分析を実施した。図表10がその結果である。

図表10から、以下の2点が読み取れる。まず1点目は、注目遅行型（作品A・作品D）はⅠ期において「視聴期待」の比率が上昇することに対して、注目先行型（作品B・作品C）はそれが減少することである。つまり、注目遅行型は、上映開始後に消費者の期待がますます膨らんでおり、逆に、注目先行型は、上映開始後から消費者たちの期待が萎んでいくと解釈できる。この結果は、注目の遅行および先行という現象と、直感的に一致している。

また、「視聴期待」については、その種類について検討の余地が見られる。図表11は、作品Aにおける「視聴期待」のトピック群として割り当てられたトピックと関係の強い投稿を一部抜粋したものである。図表11の（1）および（2）は上映開始前の作品に対する世間の評判，（3）および（4）は

図表10 各作品のⅠ期およびⅡ期におけるトピック群の傾向

Ⅰ期	テーマ説明	シーン説明	演出説明	視聴感情	視聴期待	視聴状況	総合評価
作品A	-0.0031* (0.031)	-0.0021 (0.235)	-0.0019* (0.031)	-0.0002 (0.910)	0.0036* (0.019)	0.0008 (0.676)	0.0027* (0.000)
作品B	0.0021* (0.019)	0.0001 (0.931)	0.0006 (0.471)	-0.0039* (0.000)	-0.0100* (0.000)	-0.0079* (0.002)	-0.0012 (0.055)
作品C	-0.0005 (0.710)	0.0016 (0.512)	-0.0002 (0.779)	0.0002 (0.819)	-0.0038* (0.019)	-0.0066 (0.181)	0.0003 (0.780)
作品D	0.0000 (0.983)	-0.0039* (0.000)	-0.0003 (0.725)	-0.0009 (0.171)	0.0037* (0.002)	-0.0121* (0.000)	0.0023* (0.000)

Ⅱ期	テーマ説明	シーン説明	演出説明	視聴感情	視聴期待	視聴状況	総合評価
作品A	-0.0018* (0.025)	-0.0027* (0.020)	-0.0002 (0.662)	0.0000 (0.926)	0.0013 (0.164)	0.0023 (0.065)	0.0002 (0.460)
作品B	0.0002 (0.774)	-0.0014 (0.217)	-0.0005 (0.528)	-0.0012* (0.035)	0.0033* (0.010)	0.0015 (0.304)	0.0003 (0.412)
作品C	-0.0023* (0.030)	-0.0013 (0.256)	-0.0005 (0.52)	0.0005 (0.543)	0.0016 (0.262)	-0.0015 (0.744)	0.0008 (0.169)
作品D	-0.0006 (0.260)	-0.0007 (0.141)	0.0006 (0.130)	-0.0008* (0.037)	0.0010 (0.071)	0.0017 (0.161)	0.0001 (0.524)

注) Ⅰ期は29日間、Ⅱ期は56日間である。各項目について、上段は経過日数に対する回帰係数、下段括弧は回帰係数のp値を表している。なお、「*」が記載されている白色の項目は、有意水準5%で統計的に有意と認められた回帰係数を表している。

出所) 筆者作成。

上映開始後の作品に対する世間の評判、(5)と(6)は自身の経験によって高められた作品に対する評判が、それぞれ記述されている⁽⁸⁾。つまり、ある対象についての期待というものは、「周囲の思考・行動にもとづく期待」と「自己の経験にもとづく期待」とに分類でき、さらに前者は「事前評判」と「現在評判」に分けて捉えられる可能性を示唆している。ただ、残念ながら、これらの3つの期待の関係性についてまで、本稿の分析では検討が及ばない。

次に、2点目は、注目遅行型はⅠ期において「総合評価」の比率が上昇することである。ここで、「総合評価」の比率が上昇するといっても、必ずし

(8) 図表11の(5)と(6)について、作品Aは上映開始の約2ヶ月前(2016年6月18日)に、同名タイトルの原作小説が発売されている。

図表11 作品Aにおける「視聴期待」の一例

(1)	君の名は。前評判, 周りの評価もかなり上々な本作。ハードルを上げていったんですが期待通りの作品でした。
(2)	新海誠の best 盤, といった印象。良作ではある。が, 惜しい部分が多すぎた。前評判が高いなか観に行ったため, 自分の中のハードルが上がり過ぎていた感は否めない
(3)	周りの評価が良すぎて, ハードル上がっちゃって, いざ見たら大したことないんじゃないかという疑念を持ちつつ見ました。ごめんなさい。最高でした。
(4)	旦那さまと観に行きました。世の中が絶賛している中, 私は…正直特になにも感じませんでした, ごめんなさい (・_・; 話題になっていたから観たものの何故こんなに話題になっているのかわかりませんでした。でも旦那さまは, 4.0だそうです。
(5)	小説読んでから映画を観た。小説を読む前に映画を観てたら, また違う面白いになってたのかな。あー, ブルーレイ買うかー。誰かを探して誰かを待ってる。
(6)	公開が待てず, 先月で原作を読んでしまったので内容は承知済み。あまりにも面白くて, 読んだの失敗だったかなあー, 公開されるまで待てば良かったかなあーと少し後悔したが, それでも尚, 素晴らしかった。(後略)

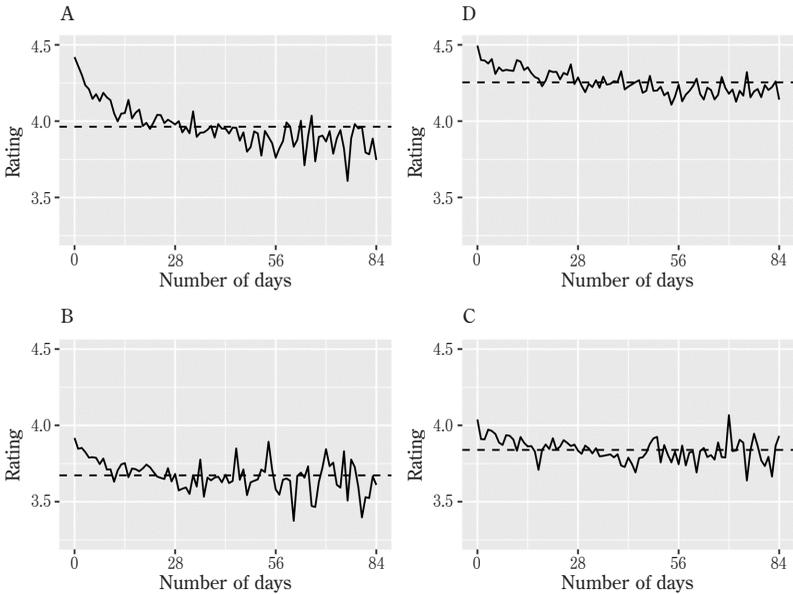
出所) 筆者作成。

も「良い評価」ばかりが投稿され続けるわけではない。図表12は, 各作品のレビューに紐付いている5点満点のスコアを, 日次で集約してその平均値の推移を表したものである。図表12を見る限り, 各作品のスコアは, 上映開始後, 緩やかに低下している傾向が見て取れる。また, 図表13は, 作品Aにおける「総合評価」に含まれるトピックと関係の強い投稿を一部抜粋したものである。図表13では, その作品をどのように解釈し, 楽しめばよいかのヒントが記述されている。つまり, 注目遅行型は, 上映初期から, 作品に対する多様な視点が視聴した消費者によって提供されており, これが対象の魅力となって注目をより強化している可能性を示唆している。ただし, こちらについても, 本稿での分析手法ではそれらを分離して検証することができないため, 詳細については今後の検討課題とする。

なお, II期については, 注目遅行型, 注目先行型の双方において, 顕著な

流行現象に関する消費者視点のテキスト分析（佐藤圭）

図表12 各作品のレビュースコアの推移



注) 左上のグラフから反時計回りの順に、作品A、作品B、作品C、作品Dの推移を示している。横軸は上映開始日からの経過日数であり、0は上映初日である。縦軸はレビュースコアの日次平均を表している。グラフ中の破線は、分析期間全体におけるスコアの平均値を示している。

出所) 筆者作成。

図表13 作品Aにおける「総合評価」の一例

嫌いじゃなかったけど、大事な部分を「なぜか」「たまたま」で済ませてるのは許せない。宇多丸さんの「セカイ」vs「世界」という視点の評論が面白い。

完璧すぎてもう感動。あの映画ができなかったこと、あのゲームができなかったこと。フィクションのノンフィクション化をフィクションでするっていう神業。そして僕自身が生きてる世界感もこうで、望むべき世界感もこう。自信が湧いた。映画としてのトレンドにするにはもったいなすぎる歴史的作品あと6000回見たい。

この数十年、精神と肉体の乖離をテーマにした物語は多いが、本作ほど一般性に適応した挑戦的な作品は稀。この点においてはクリストファー・ノーランのインターステラーの試みを遥かに超えている。新海誠の作品は思春期の童貞が抱えるような、過剰な自我を肯定的に受け入れていこうとする、彼以前には無い作調が特徴だった。(後略)

出所) 筆者作成。

特徴は確認されなかった。

ここまで、流行に関する探索的分析とその結果について見てきた。次節にて、これらの結果と調査課題についての関係を整理し、今後の課題について検討する。

6. 結びにかえて

本稿では、流行現象について、消費者の視点から探索的に検証してきた。これは、関連する研究において、流行の形成に関わる主要な主体として、企業、メディア、消費者という3者がある中で、消費者に重点を置いた分析が不足していたためであり、本稿はその視点からの知見を補うための実験的な試みである。

本稿では、流行の対象として映画を取り上げ、レビューデータを併用することで、テキストから消費者のトピックを抽出し、流行下で消費者がどのようなことを語っているのか、確認した。以下、分析結果と本稿での調査課題について要約しておこう。

まず、調査課題1について、消費者による対象への投稿件数と注目度の双方を用いることで、流行の継続性を識別できる可能性を示した。具体的には、流行には注目先行型と注目遅行型の2つのパターンが存在し、前者よりも後者の方が、消費者間の議論が比較的持続する傾向が確認された。

次に、調査課題2について、消費者は流行下において、対象に関する客観的な情報から主観的な説明にいたるまで、幅広い観点から議論を行っていることが明らかとなった。本稿では、その具体的な内容として、テーマ説明、シーン説明、演出説明、視聴感情、視聴期待、視聴状況、総合評価といったトピック群が見いだされた。

そして、調査課題3について、とくに、流行が生じ始めた初期の段階において、トピックの推移に特定の傾向が見受けられた。本稿では、流行の開始

流行現象に関する消費者視点のテキスト分析（佐藤圭）

28日後までの期間において、注目遅行型は消費者の期待が徐々に増加することに対して、注目先行型はその期待が徐々に減少することが示された。さらに、期待については3つの類型化への可能性が提示され、これらの期待の性質と関係性については、今後検討されるべきであろう。

また、同じく、流行が生じ始めた初期の段階において、注目遅行型は総合評価の議論が徐々に増加することが確認された。これは、各作品のレビュースコアが時間とともに緩やかに減少する傾向にあることから、流行の持続にとって、良い評価ばかりが得られることのみが重要ということではなく、対象をどのように解釈し楽しむか、といった多様で魅力的な視点が採用した消費者から提供されることもまた重要であることを示唆している。

ただ、流行がしばらく経過した後の期間における消費者のトピックには、特徴的な動向が見受けられなかった。関連する研究では、流行の経過時間に応じて消費者の層が変化することが指摘されていたため、流行の期間によって消費者の語りが変化することが期待されたが、本稿では、流行の初期における消費者の特徴のみが明らかとなったことに留まった。それ以外の時期に関する消費者の特徴については、今後の検証課題としたい。

さて、本稿での発見事項は限定的ではあるが、ここから実務上のインプリケーションを指摘するならば、流行は、その現象の初期段階において、消費者の注目度と期待の流入を適切にモニタリングすることで、今後、その現象がさらに持続するか否かを検証できるかもしれない。また、そのような潜在性のある現象については、これもまた初期段階において、対象についての多様な視点（楽しみ方や利用方法など）を企業や採用者が提供することで、消費者の議論をより活性化できるかもしれない。つまり、これは、Choi and Varian (2012) が指摘する「現在の予測 (predicting the present)」について、さらに精度を高めることによって、企業のマーケティング戦略への選択肢の追加を示唆するものと考えられる。

最後に、本稿の限界と今後の課題について述べておこう。1点目に、本稿では消費者視点に注力したため、流行の参加者である残りの2者との相互作用が確認できていない。企業のマーケティング努力やメディアによる情報発信の影響も含めて、今後、複合的に流行現象を検証していく必要がある。2点目として、流行しなかったものについての特徴が検討できていない。ただし、流行しなかったものについては、そもそも、それに対する消費者の反応も少ないという「二重の危険」を抱えているため、どのように解決すべきであるかは、より深い検討が必要である。3点目は、対象の限定性についてである。本稿で取り上げた対象は映画のみであるうえに、シングルデータソースである。これ以外のカテゴリや類似するデータ、あるいは国内のみならず国外の傾向にも目を向けるといった多角的な分析が必要となるであろう。いずれにせよ、これらの事実を明らかにするためには、流行現象に関するさらなる研究蓄積が必要であるように思われる。

参考文献

- 石井晃 (2021) 「社会物理」鳥海不二夫編『計算社会科学入門』丸善出版, 239-264頁。
- 大森友子 (2017) 「なぜ、ファッド群が起こるのか—腐女子の質的研究—」『日本マーケティング学会カンファレンス・プロシーディングス』第6巻, 49-60頁。
- 勝又壮太郎・西本章宏 (2016) 「市場創造と成熟過程における社会的関心の推移—新聞記事から読み解く市場の変質」『消費者行動研究』第22巻第1・2号, 27-48頁。
- 河島伸子 (2020) 『コンテンツ産業論第2版—文化創造の経済・法・マネジメント—』ミネルヴァ書房。
- 佐藤敏紀・橋本泰一・奥村学 (2017) 「単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討」『言語処理学会第23回年次大会発表論文集』, 875-878頁。
- 中島純一 (2013) 『メディアと流行の心理』金子書房。
- 松井剛 (2013) 「言語とマーケティング: 「癒し」ブームにおける意味創造プロセス」『組織科学』第46巻第3号, 87-99頁。
- Best, J. (2006) *Flavor of the month: Why smart people fall for fads*. University of California Press. (邦訳書 林大訳『なぜ賢い人も流行にはまるのか: ファッドの社会心理学』白揚社, 2009年。)
- Blei, D. M., A. Y. Ng and M. I. Jordan (2003) "Latent dirichlet allocation," *Journal of*

流行現象に関する消費者視点のテキスト分析 (佐藤圭)

machine learning research, 3, pp. 993-1022.

Choi, H. and H. Varian (2012) "Predicting the present with google trends," *The economic record*, 88, pp. 2-9.

Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock and D. J. Watts (2010) "Predicting consumer behavior with web search," *Proceedings of the national academy of sciences*, 107 (41), pp. 17486-17490.

Kotler, P. and K. L. Keller (2006) *Marketing management, 12th edition*. Prentice-Hall. (邦訳書 恩蔵直人監訳『コトラー&ケラーのマーケティング・マネジメント』丸善出版, 2014年。)

Rogers, E. M. (2003) *Diffusion of innovations, 5th edition*. Free Press. (邦訳書 三藤利雄訳『イノベーションの普及』翔泳社, 2007年。)

Simmel, G. (1957) "Fashion," *American journal of sociology*, 62(6), pp. 541-558.

Solomon, M. R. (2013) *Consumer behavior: Buying, having, and being, 10th edition*. MA: Pearson. (邦訳書 松井剛監訳『ソロモン消費者行動論』丸善出版, 2015年。)

Yan, X., J. Guo, Y. Lan and X. Cheng (2013) "A biterm topic model for short texts," In *Proceedings of the 22nd international conference on World Wide Web*, pp. 1445-1456.