

2021 年度（令和 3 年度） 卒業論文

KH Coder を用いた Yahoo ニュースでの
評価を受けやすいコメントの分析

指導教員

村田真樹

村上仁一

鳥取大学工学部 電気情報系学科

自然言語処理研究室

B18T2117A 森 大起

概要

近年 SNS の利用は日常化しており，SNS において高評価を得ることを分析することには社会的な関心が注がれているといえ，チェックシステムなどの応用技術の作成に必要な技術である．

本研究では，この分析に KH Coder を活用する．KH Coder を使用した先行研究では，岩森 [2] や小野ら [3] のような社会学的な研究が多く行われてきたが，開発者の樋口は言語研究の分野での活用にも言及しているため [8]，これを採用する．

本研究では，Yahoo ニュースのコメント欄に注目し，コメント欄での高評価を受けたコメントと低評価を受けたコメントで，どのような点で違いが生まれているのかを考察する．

実験では，まず KH Coder の対応分析で視覚的に高評価や低評価に頻度が偏っている単語を抜き出し，それらの語について数値的に評価する．また偏りの大きい単語の一部を抜き出し，その単語を含むコメント文のいくらかを抽出し比較することで新しい知見を得ようとした．

結果として，高評価コメントでは「皇室」や「本当に」など，低評価コメントでは「コメント」や「笑」などの単語が偏って出現していることなどが分かった．また高評価コメントでは問題提起をするようなコメントが多かったが，低評価コメントでは記事が書かれた対象に対して擁護するようなコメントが目立った．

目次

第1章	はじめに	1
第2章	先行研究	3
	2.1 KH Coder を活用した学生満足度調査における自由記述の解析	3
	2.2 訪問診療実習を通じて得た医学生の学びの解析	3
	2.3 保護者アンケートにおける「自由記述」の活用-学校評価における KH Coder による分析の試み-	4
	2.4 失敗談ツイートのテキスト分析による北海道における観光失敗 リスクの把握	4
	2.5 口コミの言語解析から見た日本人と外国人の浜離宮恩賜庭園に 対する観光体験の相違	5
	2.6 COVID-19 における Twitter の利用傾向に関する探索的研究	5
	2.7 言語研究の分野における KH Coder 活用の可能性	6
第3章	KH Coder の機能	7
	3.1 前処理	7
	3.2 クロス集計	10
	3.3 対応分析	11
	3.4 コロケーション分析	16
第4章	提案手法	18
	4.1 手順1:「入力ファイルの作成」	18
	4.2 手順4:「クロス集計の整理」	19
第5章	実験	21
	5.1 実験データ	21
	5.2 実験結果	21
	5.2.1 対応分析	21
	5.2.2 クロス集計のソート	23
	5.2.3 コロケーション分析	24
	5.2.4 grep コマンドを用いた分析	30
	5.3 追加実験	31
	5.3.1 高評価コメントと低評価コメントで同一の中間語を 用いたコロケーション分析	31

5.3.2 中評価コメントでの中間語を用いたコロケーション 分析	44
第6章 SNS において評価を得るための考察	60
第7章 今後の課題	61
第8章 おわりに	62

表目次

3.1	KH Coder の入力ファイルの例	7
3.2	KH Coder の品詞体系	13
4.1	「総合評価」の分割条件	19
4.2	エクセルファイルの例	19
5.1	実験データ	21
5.2	各評価コメントの数	23
5.3	各評価コメントの例	23
5.4	研究での対応分析の設定	23
5.5	図 5.1 の各単語の比率のソート(1)	24
5.6	図 5.1 の各単語の比率のソート(2)	25
5.7	総合評価が高のみでの「本当に」と、低のみでの「笑」のコロケーション分析でのスコアのトップ 10	26
5.8	高のデータに含まれる”本当に”の左右に多く含まれている語の上位 10 単語	28
5.9	低のデータに含まれる”笑”の左右に多く含まれている語の上位 10 単語	29
5.10	ランダムに抜き出した”本当に”を含む高評価コメント 20 件(1)	33
5.11	ランダムに抜き出した”本当に”を含む高評価コメント 20 件(2)	34
5.12	ランダムに抜き出した”本当に”を含む高評価コメント 20 件(3)	35
5.13	ランダムに抜き出した”本当に”を含む高評価コメント 20 件(4)	36
5.14	ランダムに抜き出した”笑”を含む低評価コメント 20 件(1)	37
5.15	ランダムに抜き出した”笑”を含む低評価コメント 20 件(2)	38
5.16	ランダムに抜き出した”笑”を含む低評価コメント 20 件(3)	39
5.17	総合評価が高もしくは低での”思う”スコア値のトップ 30	40
5.18	高評価コメントの”思う”の左右に多く含まれている語の上位 10 単語	41
5.19	低評価コメントの”思う”の左右に多く含まれている語の上位 10 単語	42
5.20	ランダムに抜き出した”思う”を含む高評価コメント 20 件(1)	45
5.21	ランダムに抜き出した”思う”を含む高評価コメント 20 件(2)	46
5.22	ランダムに抜き出した”思う”を含む高評価コメント 20 件(3)	47
5.23	ランダムに抜き出した”思う”を含む高評価コメント 20 件(4)	48
5.24	ランダムに抜き出した”思う”を含む低評価コメント 20 件(1)	49
5.25	ランダムに抜き出した”思う”を含む低評価コメント 20 件(2)	50
5.26	ランダムに抜き出した”思う”を含む低評価コメント 20 件(3)	51

5.27	総合評価別の”思う”スコア値のトップ 30	52
5.28	中評価コメントの”思う”の左右に多く含まれている語の上位 10 単語..	53
5.29	各評価コメントの”思う”の左右に多く含まれている語の上位 10 単語..	54
5.30	ランダムに抜き出した”思う”を含む中評価コメント 20 件(1)	56
5.31	ランダムに抜き出した”思う”を含む中評価コメント 20 件(2)	57
5.32	ランダムに抜き出した”思う”を含む中評価コメント 20 件(3)	58
5.33	ランダムに抜き出した”思う”を含む中評価コメント 20 件(4)	59

目次

3.1	KH Coder のメニューの例	8
3.2	KH Coder の新規プロジェクトの例	8
3.3	KH Coder の語の取捨選択の例	9
3.4	「こころ」のクロス集計の様子	11
3.5	KH Coder の対応分析のメニュー	11
3.6	KH Coder の対応分析の例	15
3.7	KWIC コンコーダンスの例	16
3.8	コロケーション分析の例	17
5.1	Yahoo ニュースのコメントの対応分析	22

第1章 はじめに

近年の SNS の利用は日常化しており，Twitter,FaceBook,Instagram など様々な SNS が飽和している時代だと言えそうだ．今や SNS は個人利用に留まらず一般企業においてもイメージ戦略や広報活動などのための大切なツールになっている．一方，SNS の使用は炎上などで企業のイメージに取り返しのつかない重大なダメージを与えることもある．また，個人でも SNS の評価によってストレスを感じたり，酷い場合には自殺などが発生するケースもあった．以上のために SNS において高評価を得ることを分析することには社会的な関心が注がれているといえる．また，このような分析は SNS において他人が不快感を得るようなコメントを自動で削除するチェックシステムなどにも活用が可能である．

一方，KH Coder は立命館大学の樋口耕一によって開発されたフリー・ソフトウェアである [1]．主に計量テキスト分析をするために開発され，対応分析やクラスター分析，多次元尺度構成法，自己組織化マップ，共起ネットワーク図など様々な分析方法を兼ね備えている．

今までの KH Coder を用いた先行研究は，岩森ら [2] や小野ら [3] といった特に自由記述のアンケートやレポートからよく使われた語彙を抜き出しニーズや評価を考察するような社会学的な研究が多かった．

このようにこれまでの研究では，KH Coder は社会学的な研究に最もよく使われた．しかし，KH Coder の開発者の樋口は言語研究の分野での活用にも言及している [8]．

本研究では，特に Yahoo ニュースのコメントにおいて高評価コメントと低評価コメントの差や違いを考察していく．また，SNS において高評価や低評価を得るコメントの傾向を分析するために KH Coder を用いる．

本研究では KH Coder を利用して対応分析によって，高評価，低評価コメントによく出現する単語を視覚的に明らかにしようとした．また数値的にも明らかにするため，KH Coder を使用して高評価に対する「高の比率」を定義してソートした．最後に人手で研究に使用したコメント文をランダムにいくつか取り出して考察した．

本研究の主な主張点を以下に羅列する．

- KH Coder の対応分析やクロス集計によって、高評価コメントや低評価コメントに偏って出現している単語が分かった。例えば、高評価コメントに偏って出現している語には「皇室」や「本当に」などがあった。また低評価コメントに偏って出現している語には「コメント」や「笑」などがあった。
- KH Coder のコロケーション分析によって、特定の評価のコメントに偏って出現している語に関して、その周囲にどのような語が出現しているのかをみることで、高評価コメントと低評価コメントの性質を調査した。高評価コメントでは要望や意見を表した語や自分の感想を表現する語などが多く見られた。低評価コメントでは他人のコメントに言及するような語や自分のコメントに対して強く同意を求めるような語が多く見られた。
- KH Coder の grep 分析やその後の追加実験によって、高評価コメントや低評価コメントの特徴を内容的に読み取った。高評価コメントでは問題提起をするようなコメントが多かった一方、低評価コメントでは記事が書かれた対象に対して擁護するようなコメントが目立った。

本論文の構成は以下の通りである。第2章では、KH Coder を使った過去の研究や技術などを紹介する。第3章では、KH Coder に組み込まれている機能の内から本研究に使用するものを紹介する。第4章では、本研究の提案手法について説明をする。第5章では、本研究で行った実験についての説明と、その結果について説明する。第6章では、第5章の結果を受けた考察を行う。第7章では、今後の課題について説明する。第8章では、本研究のまとめをする。

第2章 先行研究

本章では、これまでの KH Coder を使用した先行研究について説明する。

2.1 KH Coder を活用した学生満足度調査における自由記述の解析

岩森ら [2] の研究では、大学などを含めた高等教育機関が学生のニーズを満たし、高等教育機関への満足度を高めるために学生 388 人が書いた学生生活満足度に関する調査の自由記述欄を入力として KH Coder により分析した。

岩森らは入力された学生たちの自由記述を KH Coder によって共起ネットワーク図で描いた。また階層クラスタリングを施したりして、学生たちが高等教育機関のサービスに対してどのようなところに満足をし、どのような点を改善してほしいのかを導いた。

また KH Coder で対応分析を行い、学生たちが日々の学生生活でどのような点に着目しているのかを学年や学科別に調べた。

2.2 訪問診療実習を通じて得た医学生の学びの解析

小野ら [3] は、在宅医療や終末期医療における訪問診療学習の学習効果をより具体的に示し、自由記述式レポートの評価手法として KH Coder が有用かを検討するための分析を行った。

分析は 2015 年度から 2019 年度までに 1 日間の在宅医療実習を行った東京女子医科大学の医学部 5, 6 年生, 80 名を対象として行われた。入力には在宅医療実習を受けた学生に課せられた自由記述式のレポートを元に行い、複合語の強制抽出¹や誤字脱字の修正といったデータ整理をしたものを使用した。

¹強制抽出は、KH Coder において、例えば「卒業研究」という語彙を ChaSen などによって「卒業」と「研究」に分けることなく、「卒業研究」という 1 つの表現として分析をするためのものである。3.1 節参照。

小野らは、共起ネットワーク図や対応分析を利用した分析を行っていた。共起ネットワーク図を利用した分析では、出現回数が300回以下の単語のみを抽出して在宅医療実習における特徴的な単語を導いたり、学生ごとに共起ネットワーク図を作成し学生別の学びの違いを導いたりした。対応分析を利用した分析では、対応分析の図の端の方の単語(特徴を示した語)を利用し、特定年度における学びの変化などを導いていた。

2.3 保護者アンケートにおける「自由記述」の活用-学校評価におけるKH Coderによる分析の試み-

仁科[4]は教育活動における説明責任と外部評価の具体化のために行われている児童や保護者らへのアンケートをKH Coderを使用してより簡単に分析することを考えた。

児童や保護者らへのアンケートには、数値評価が可能な段階評価と、それが難しい自由記述とがある。自由記述には、学校が気付いていない課題や視点が書かれていることもあり、これを明らかにすることは保護者の不満を解消し教員の達成感や自信、モチベーションに繋がる。

仁科は共起ネットワーク図を利用して読み手が理解していなかった新しい注目点を発見した。この結果から仁科は学校評価の質的な向上の一助になる可能性があるとは結論付けた。

2.4 失敗談ツイートのテキスト分析による北海道における観光失敗リスクの把握

吉田ら[5]は、失敗談ツイートの地図上での可視化をしたところ各観光地で特徴的な観光失敗経験があったため、失敗談ツイートに対するテキスト分析を行うことで観光ツイートから定量的に観光失敗に関する情報抽出をするために以下の実験を行った。データには2016～2017年に投稿されたツイッターの位置情報付きツイート約1143万件から、北海道内の失敗談をつぶやいたツイート851885件を実験データとして用いた。

まず、全国の失敗談ツイートと北海道内の失敗談ツイートの各単語においてTF-IDF値を算出し比較する。TF-IDF値を比較し、北海道内の失敗談ツイートの方がTF-IDF値が高い上位50語を北海道内での失敗談ツイートの特徴語とする。

次にKH Coderの共起ネットワーク図を用いて特徴語の関係などを分析した。例えば、「辛い」は「カレー」や「ラーメン」との関係が深いことが分かったので「辛(か

ら)い」であることが分かった。

次に GIS を使用して、北海道の地図上に失敗談をプロットすることで土地ごとの失敗談の違いを分析した。失敗談は、札幌や函館などの大都市に多く、例えば「辛い」は札幌に、「見える」は函館や稚内などに多かったことが分かった。

最後により詳細に特徴語を分析するため、Embedding Projector を使用して特徴語にベクトル上の意味が近い文章を選定し考察を行った。その結果、例えば「凍」の場合「給油口のふたが開かない」など具体的な失敗談ツイートと繋がった。

2.5 口コミの言語解析から見た日本人と外国人の浜離宮恩賜庭園に対する観光体験の相違

安原ら [6] は、都立庭園の日本人と外国人の利用者がどのような訪問体験を有して、庭園にどのような評価を与えたのかを明らかにし、都立庭園の利用者ありきの保存管理および活用の今後の展開のために以下の実験をした。

実験のデータには旅行口コミサイト「tripadvisor」のデータの内から浜離宮恩賜公園に寄せられた訪問時期の記載のあるコメントを日本語で 682 件、英語で 753 件収集した。これらのデータには、分析上不必要な顔文字や記号を修正し、同一語の表記を統一した。

まず、KH Coder で日本語と英語のコメントで単語ごとの出現頻度上位 100 語を出力し比較した。この結果、庭園に関して日本人は「広い」とコメントしている人が多かったのに対し、外国人は「small」とコメントしている人の方が多いことなどが分かった。

また次に KH Coder で日本語と英語で共起ネットワーク図をそれぞれ作成し比較をした。結果として、日本人の周遊ルートには「新橋」「徒歩」「駅」など周囲の街から来ている人が多いのに対して、外国人は「Tsukiji」とあり築地市場に通っている人がいることが分かった。

2.6 COVID-19 における Twitter の利用傾向に関する探索的研究

渡邊ら [7] は、2020 年の新型コロナウイルス感染症の蔓延に関するツイートを分析することで、新型コロナウイルスへの今後の対応や取り組みにも活用が可能になると考え、以下の実験を行った。

実験データには新型コロナウイルス関連のツイートを使用し、感染が下げ止まった2020年6月19日および20日のものと、感染者数が増加した2020年7月10日および11日のものを使用した。

まずKH Coderで出現回数の多い単語を確認する。その結果、例えば6月から7月にかけて「マスク」は減少したが、「検査」「増える」「拡大」「陽性」などの単語が増加したことが分かった。

次にKH Coderで多次元尺度構成法と共起ネットワーク図を利用して、これらの単語の関係を調べた。

結論として、感染が増加している時期には新型コロナウイルスへの懸念や猛威を表すツイートが多くなったことなどが判明した。

2.7 言語研究の分野におけるKH Coder活用の可能性

樋口 [8] はKH Coderを言語研究の分野において活用するために以下のことを提案した。

まず一つは分析に「その他」の品詞の単語を含めることである。KH Coderでは通常、動詞や名詞などの内容語のみを分析の対象とする。言語研究の分野では内容語だけでは不十分な場合もあり、助詞や助動詞などの機能語を「その他」の単語を加えることで分析することができる。

しかしこのままでは、ある助詞や助動詞を取り出しても品詞が「その他」にまとまってしまって細かい分析ができないので、次にKH Coderの品詞体系の編集をする必要がある。KH Coderのconfigファイル上のテキストに登録することで品詞体系を編集することができる。例えば、「から」の品詞を格助詞と接続助詞に分けて傾向を調べられる。

より細かい分析を行うために形態素解析の結果を手作業で修正したり、KH Coderへ入力するための多数のテキストファイルを1つのファイルにまとめたりすることができる。

第3章 KH Coderの機能

3.1 前処理

KH Coderでは、様々な機能を効率的に行うために共通する処理部分を最初に行うことで後の処理でこれを省略することができる。これを「前処理」という。

KH Coderの入力ファイルは主にテキストファイルかエクセルファイルで作成される。表3.1には、エクセルファイルでの例について載せる。エクセルファイルを入力ファイルとして使用する場合、1行目にその列にどのような内容が書かれているかを示す変数名を書き入れ、2行目からその列の内容を入れる。この表の場合、変数名は「テキスト」「部」「章」「章(ラベル)」のことである。

KH Coderを立ち上げると図3.1のようなウィンドウが立ち上がる。これを「メニュー」という。

このメニューから、プロジェクト→新規とすると図3.2のウィンドウが立ち上がる。図3.2の分析対象ファイルに入力ファイルを、分析対象とする列にテキスト文を指している変数名を入れることでプロジェクトを作成することができる。例えば表3.1の場合、変数名「テキスト」を入れる。この状態で「OK」を押すと、メニューに図3.1のように表示される。

また、ここで分析対象とした列以外の情報を外部変数と呼ぶ。表3.1の場合、「テキスト」の列を分析対象としているので外部変数は表のそれ以外の部分全てである。つまり外部変数は変数名「部」「章」「章(ラベル)」とこれらの変数名を含む列の全ての情報のことである。この外部変数は対応分析やクロス集計などの様々な機能の中で利用可能な変数である。

表 3.1: KH Coder の入力ファイルの例

テキスト	部	章	章 (ラベル)
私はその人を常に...	[1] 上__先生と私	1.01	上・一
私とその掛茶屋で...	[1] 上__先生と私	1.02	上・二
...

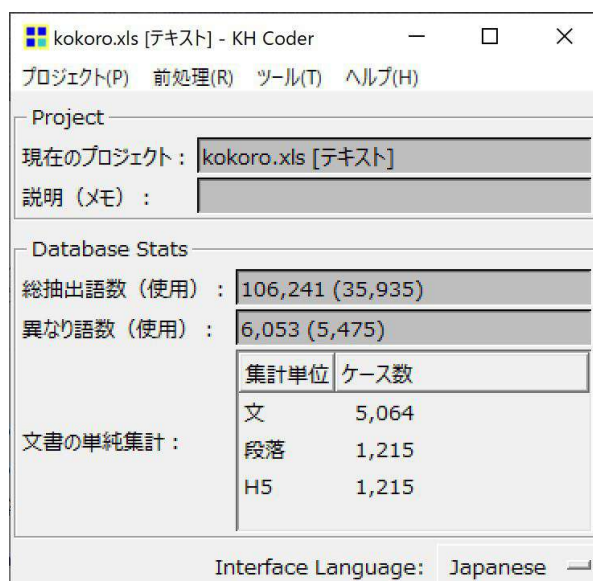


図 3.1: KH Coder のメニューの例

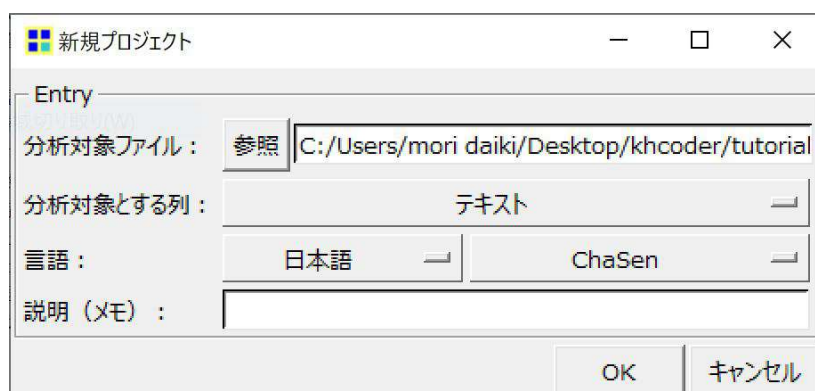


図 3.2: KH Coder の新規プロジェクトの例

KH Coder の処理に利用する品詞や単語の取捨選択は、メニューから前処理→語の取捨選択とすることで図 3.3 のウィンドウからできる。「force pick up」と「force ignore」にある「- - cell - -」は、分析するテキスト間の分割記号である。「force pick up」は強制抽出のことで、例えば「二人」といった単語を「二」と「人」のように分割しないで、「二人」として分析するためのものである。また「force ignore」は、分析に必要な語を単語レベルで排除するためのものである。

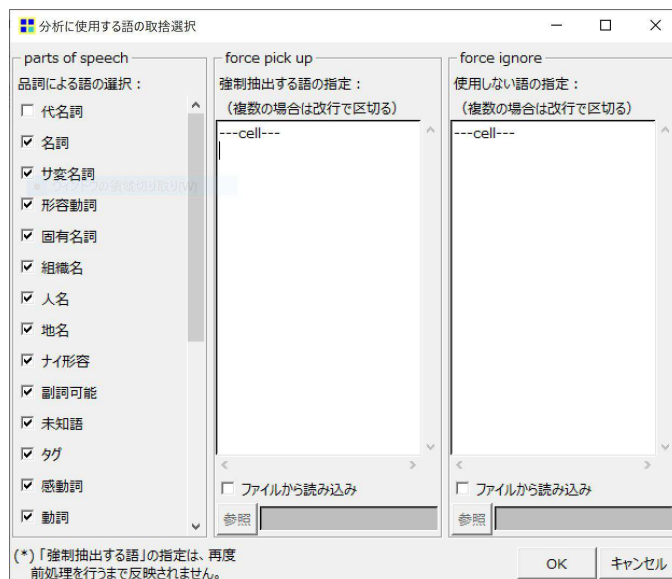


図 3.3: KH Coder の語の取捨選択の例

3.2 クロス集計

クロス集計は、コーディングルール・ファイルによって決められたコードごとに文書を集計する機能である。

コーディングルール・ファイルとは、クロス集計をする際のルールが書かれたファイルである。コーディングルール・ファイルは以下のようなになる。*から始まる部分が、このファイルの一つのコードとなる。このコードは、いくつかの単語を保有することができる。例えば、「人の死」による分析では「死後」であったり「死病」であったりする語を使用することがわかる。

*人の死

死後 or 死病 or …

*恋愛

愛 or 恋 or …

…

KH Coder 上ではメニューからツール→コーディング→クロス集計とすることで呼び出すことができる。呼び出したら、コーディングルール・ファイルのところに上記のように書き入れたファイルを挿入し、セル内容を「度数のみ」、コーディング単位とクロス集計の部分を分析内容に応じ任意のものにすることで図 3.4 のような結果を得れる。

図 3.4 について説明する。各マスの値はその部の文書の中でそのコードを持つ文書の数を示している。この結果から例えば「上_先生と私」には、「人の死」に関する内容が 37 段落で見られたとすることができ、合計では 626 の段落が存在した。なおカイ二乗 (χ^2) の数値の後ろについているアスタリスクは、各コードに付与された文書の割合が統計学的にどの程度有意であるのかを簡単に示している。アスタリスクが 2 つついていると 1% の有意水準、1 つだと 5% の有意水準である。



図 3.4: 「こころ」のクロス集計の様子

3.3 対応分析

対応分析は、各単語間の関係と単語の出現傾向を二次元分布図上にプロットする分析法である。

KH Coder ではメニューからツール→抽出語→対応分析とすることで呼び出すことができる。呼び出した結果が図 3.5 である。

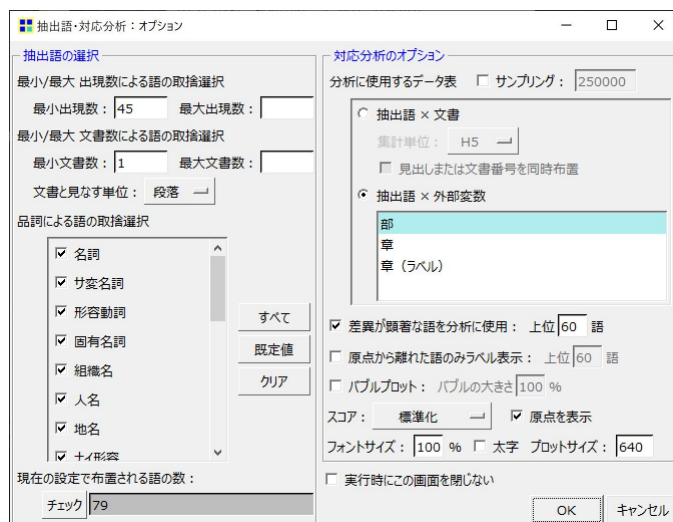


図 3.5: KH Coder の対応分析のメニュー

まず図 3.5 の左の「抽出語の選択」の方から説明する。

「最小/最大 出現数による語の取捨選択」では、全文書中における出現回数にて単語を抽出する条件である。例えば、最小出現数が45なら入力した全ての文書中で30回しか使用されていない単語は分析には使用されない。

一方、「最小/最大 文書数による語の取捨選択」では、その単語が使用された文書の総数により単語を抽出する条件である。例えば、最小文書数が3ならその単語が1つの文書でしか使われていないとき、その単語は分析には使用されない。

なお、ここで文書とみなす単位には「段落」や「文」、「H1」などがある。「段落」と「文」はユーザが特別な分割をせずに使用することができ、「段落」は改行で区切られた部分、「文」は「。」(句点)によって区切られた部分を分析単位とする。また一方、KH Coderの入力テキストファイルではユーザによって特別に「H1」から「H5」までを割り振ることができる。以下の例が入力テキストファイルの例である。この例のとき、例えば「H1」は<H1>から次の<H1>まで、すなわち「上」と「中」がそれぞれ一つの分析単位とする。「H2」と「H3」についても考え方は同様である。またこの例のように、通常「H1」から「H5」は数字が大きくなるほど、小さい分割単位とする。しかし、入力エクセルファイルでは自動的に行ごとに「H5」が割り振られ、「H1」から「H4」を使用することはできない。

```
<H1> 上 </H1>
<H2> 章題 1</H2>
<H3> 文 1</H3>
.....
<H3> 文 2</H3>
.....

<H2> 章題 2</H2>
.....

<H1> 中 </H1>
.....
```

「品詞による語の取捨選択」では、分析に使用する品詞の単語を選択することができる。入力テキストはChaSenによって形態素解析されているため、「品詞による語の取捨選択」での品詞はChaSenの品詞との互換性がある。具体的な対応表は表3.2に

表 3.2: KH Coder の品詞体系

KH Coder 内での品詞名	ChaSen の出力における品詞名
名詞	名詞-一般 (漢字を含む 2 文字以上の語)
名詞 B	名詞-一般 (平仮名のみの語)
名詞 C	名詞-一般 (漢字 1 文字の語)
サ変名詞	名詞-サ変接続
形容動詞	名詞-形容動詞語幹
固有名詞	名詞-固有名詞-一般
組織名	名詞-固有名詞-組織
人名	名詞-固有名詞-人名
地名	名詞-固有名詞-地域
ナイ形容	名詞-ナイ形容詞語幹
副詞可能	名詞-副詞可能
未知語	未知語
感動詞	感動詞またはフィラー
タグ	タグ
動詞	動詞-自立 (漢字を含む語)
動詞 B	動詞-自立 (平仮名のみの語)
形容詞	形容詞 (漢字を含む語)
形容詞 B	形容詞 (平仮名のみの語)
副詞	副詞 (漢字を含む語)
副詞 B	副詞 (平仮名のみの語)
否定助動詞	助動詞「ない」「まい」「ぬ」「ん」
形容詞 (非自立)	形容詞-非自立 (「がたい」「つらい」「にくい」等)
その他	上記以外のもの

なる。

次に図 3.5 の右の「対応分析のオプション」について説明する。

「対応に使用するデータ表」には「抽出語 × 文書」と「抽出語 × 外部変数」の 2 種類がある。「抽出語 × 文書」では抽出語と文書の対応分析をすることができる。ここでの文書は上記で述べたもので、「段落」や「文」、ユーザによって割り振られた「H1」から「H5」のことである。「抽出語 × 外部変数」では、エクセルファイルなどで決められた外部変数¹によって対応分析することで抽出語と外部変数の対応分析をすることができる。

またその他の調整オプションとして、「差異が顕著な語を分析に使用」や「バブルプ

¹詳しくは、3.1 節にて説明

ロット」などがある。

「差異が顕著な語を分析に使用」は、特徴のない語よりも入力データ表で出現割合が大きく異なった語を分析に使用オプションである。また、KH Coderでは原点付近に単語が集中しやすいので「原点付近を拡大」することも可能である。

「バブルプロット」は、二次元プロット上に出現回数の多い語ほど大きな円で描写するオプションである。出現回数がいくらであれば、どのくらいの円になるのかは凡例に表示される。

対応分析の結果は、図 3.6 である。この図の設定は、図 3.5 の通りである。点線が交差している点がこの図の原点である。

図 3.6 では、原点からの方向と距離が単語の傾向に関係する。例えば図の右上の「先生」と「持つ」では、「持つ」の方が「[1] 上_先生と私」に近いので偏りが大きいと考えてしまうが、実際は原点から見て「[1] 上_先生と私」の方向へ最も遠くにあるのは「先生」の方なので偏りが大きいと見る。

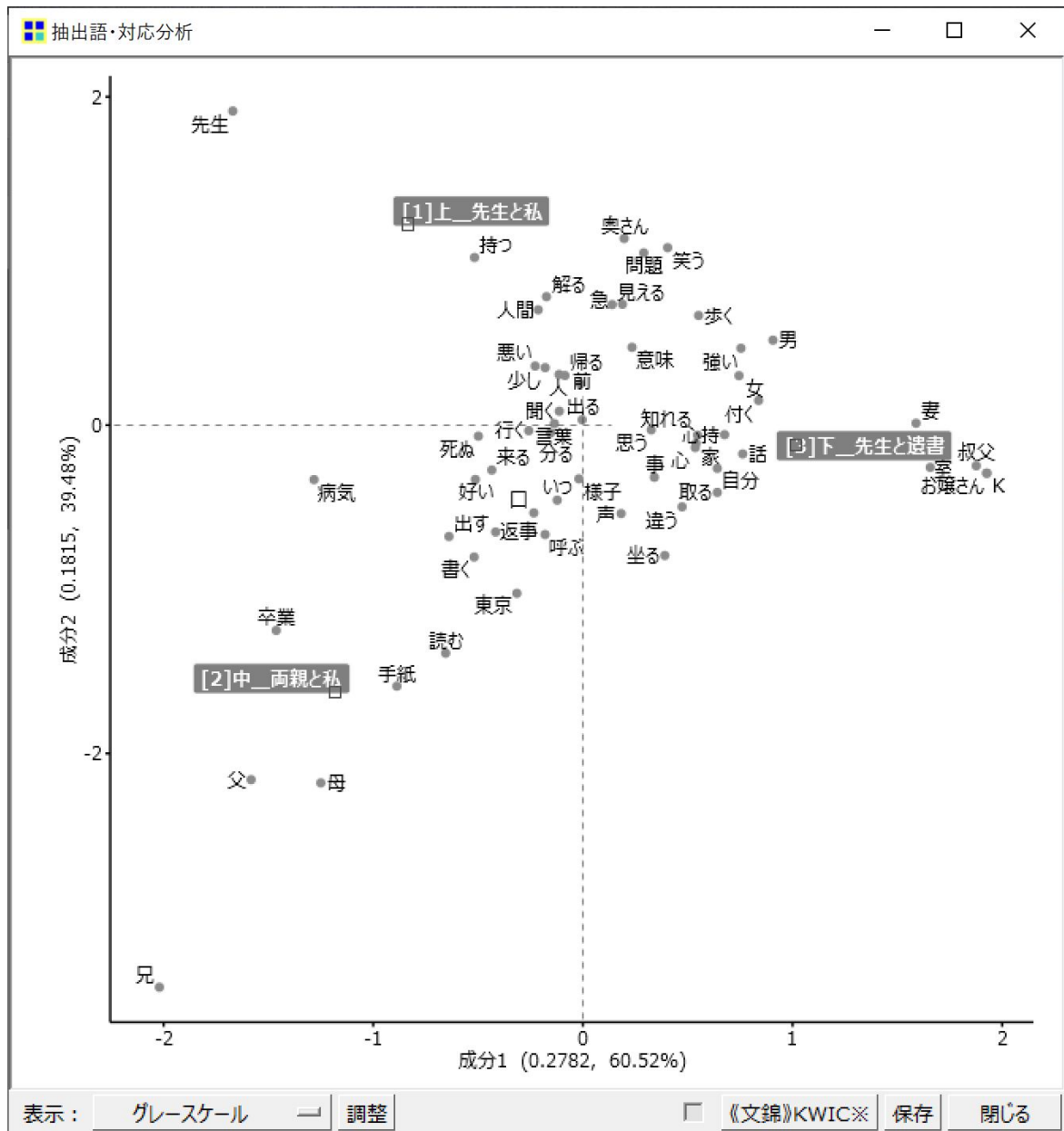


図 3.6: KH Coder の対応分析の例

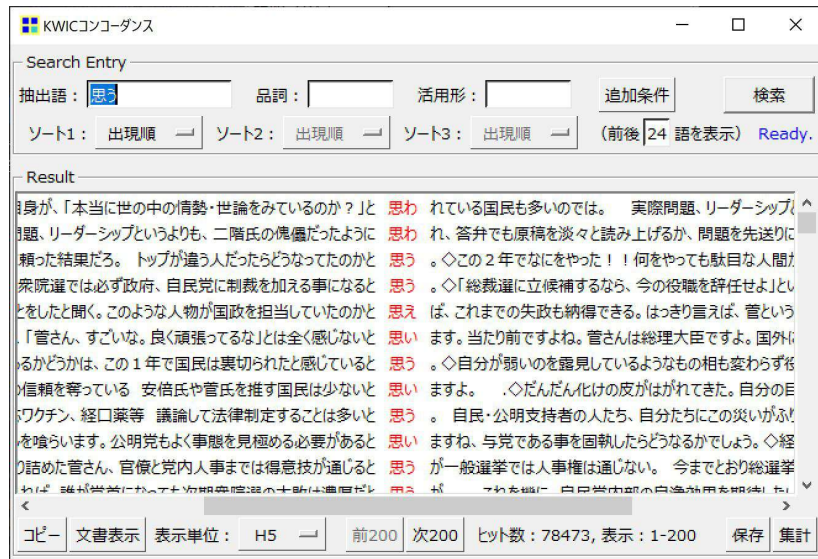


図 3.7: KWIC コンコーダンスの例

3.4 コロケーション分析

コロケーション分析はKWIC コンコーダンスにて左右に出現した語の出現数などを集計する分析である。

KWIC コンコーダンスはKH Coder ではメニューからツール→抽出語→KWIC コンコーダンスとすることで呼び出すことができる。KWIC コンコーダンスは図 3.7 のようになる。図の抽出語の欄に分析対象の語を入れて、検索することで図の Result 表示を出力する。Result には、抽出語の前後数語 (図では 24 語) のテキストが表示される。Result に抽出されるテキストは抽出語として入力した語の活用も含めて集められる。また、Result に存在する「◇」は Excel ファイル中のセルの区切りを示されている。

しかし、KWIC では無数のテキストを参照するため抽出語の特徴や特性が分かりづらくなっている。そのため図 3.7 の右下の「集計」を押すと、図 3.8 のようなコロケーション分析の結果が得られる。

コロケーション分析では、入力テキスト中で抽出語の周囲によく現れる語を表示する。分析結果では、その語が抽出語の左右に出現する回数が表示されている。例えば、図では「ない」は「思う」の 1 つ左側に 57,231 回、1 つ右側に 114 回出現していることがわかる。

通常コロケーション分析では各単語は「スコア値」によって降順でソートされた状態で表示されている。スコア値 $f(w)$ は抽出語を中心として、その語が 1 つ左側に現れ

コロケーション統計

Node Word
 抽出語: 思う 品詞: 活用形: ヒット数: 78473

Result

N	抽出語	品詞	合計	左合計	右合計	左5	左4	左3	左2	左1	右1	右2	右3
1	ない	否定助動詞	14855	10676	4179	1688	1438	1713	5723	114	1936	601	457
2	いい	形容詞 (非自立)	2339	2268	71	89	202	107	1859	11	0	12	17
3	良い	形容詞	2472	2305	167	164	325	438	1365	13	0	9	25
4	人	名詞C	2761	1480	1281	564	521	299	96	0	261	236	276
5	良い	形容詞 (非自立)	1350	1310	40	57	99	204	946	4	1	18	3
6	ん	否定助動詞	928	233	695	62	43	26	102	0	132	332	73
7	必要	形容動詞	1380	1290	90	133	441	628	88	0	20	6	10

コピー フィルタ設定 ソート: スコア 集計範囲: 左5 右5

図 3.8: コロケーション分析の例

る回数を l_1 , 1つ右側に現れる関数を r_1 として式 3.1 のように表せる.

$$f(\omega) = \sum_{i=1}^5 \frac{(l_i + r_i)}{i} \quad (3.1)$$

第4章 提案手法

本研究では以下の手順で，高評価コメントと低評価コメントの違いや差を調査する．

手順1 Yahoo ニュースのコメントから KH Coder の入力ファイルを作成する．

手順2 KH Coder で高，中，低の対応分析¹にかける．

手順3 対応分析で抽出された語でコーディング・ファイルを作成する．

手順4 コーディング・ファイルを用いて，KH Coder でクロス集計²を行い，エクセルファイルで整理する．

手順5 ソートの結果から，特に高，低の特徴が抜き出せた語について KH Coder でロケーション分析³をする．

手順6 同様に高，低の特徴が抜き出せた語のコメントを grep コマンドでランダムに抜き出し考察する．

4.1 手順1：「入力ファイルの作成」

本研究では，入力ファイル⁴はエクセルで作成する．エクセルファイルには，Yahoo ニュースから「コメント文」，「GOOD が押された数」，「BAD が押された数」，「何時間前にコメントが投稿されたのか」などを記述する．

「何時間前にコメントが投稿されたのか」の表記は，○分前であったり，○時間前であったりと様々なので， n 時間前の時に $(n+1)$ と定義する「時間評価」を作成する．この時，○秒前や○分前は「時間評価」が1，○日前は「時間評価」が25とする．

¹3.3 節にて説明，単語と単語もしくは単語と各評価との関係を分かりやすい図を作成する KH Coder の機能

²3.2 節にて説明，コーディング・ファイルの規則に従い集計をする KH Coder の機能

³3.4 節にて説明，ある単語の左右に出現する語の出現回数などを表示する KH Coder の機能

⁴KH Coder の入力ファイルに関しては，3.1 節を参照

「GOOD が押された数」から「BAD が押された数」を引き、「時間評価」で割ることで「数値評価」を作成する。つまり、「数値評価」は式 4.1 の通りである。

$$\text{数値評価} = \frac{\text{GOOD が押された数} - \text{BAD が押された数}}{\text{時間評価}} \quad (4.1)$$

この「数値評価」によって、各コメントの「総合評価」を決定する。「総合評価」では、高中低の三段階に分割する。分割の条件は、表 4.1 に記した。

表 4.1: 「総合評価」の分割条件

総合評価	条件
高	数値評価が 5 以上
中	数値評価が 0 以上 5 未満
低	数値評価が 0 未満

以上のように、エクセルファイルを作成した。表 4.2 に作成したエクセルファイルの例を載せた。なお、COMMENT については例の一部を省略した。

表 4.2: エクセルファイルの例

COMMENT	GOOD	BAD	TIME	時間評価	good-bad	数値評価	総合評価
下村に出馬を…	37090	2355	5 時間前	6	34735	5789.17	高
下村は目の前…	2	1	2 時間前	3	1	0.33	中
…	…	…	…	…	…	…	…

4.2 手順 4: 「クロス集計の整理」

本研究では、手順 3 の対応分析で出力された単語を元にコーディング・ファイル⁵を作成する。コーディング・ファイルは、テキストファイルで作成する。コーディング・ファイルの各コードは、各単語と一対一で作成する。たとえば、手順 3 の対応分析で「思う」が出力されたら以下のようにコーディング・ファイルを作成する。

<p>* 思う 思う</p>

⁵3.2 節のコーディングルール・ファイルと同様のもの

*絶対

絶対

...

作成したコーディング・ファイルを使用し、3.2節のクロス集計を行う。

クロス集計の結果を元に、エクセル上で単語の「高の比率」でソートをする。例えば「思う」の「高の比率」の計算は、以下のように行う。

まず、「思う」のコメントがどれだけ高評価に偏りがあるのかを知りたいので式4.2のように「高の出現率」を求める。

$$\text{「思う」の「高の出現率」} = \frac{\text{「思う」の「高の度数」}}{\text{「思う」の「合計度数」}} \quad (4.2)$$

しかし、現時点では高評価コメントと低評価コメントが同じくらい取れるかが分からないので「高の出現率」だけでは総コメント量が多い方の評価に偏りが出る可能性がある。そのため、集めたコメント文全体で高評価コメントがどれくらいの出現率があるのかを知りたい。これを「高の出現率」とし、式4.3で求める。

$$\text{「高の出現率」} = \frac{\text{高評価コメントの度数}}{\text{コメント文全体の合計度数}} \quad (4.3)$$

式4.2と式4.3より、式4.4のように「高の補正出現率」をとる。

$$\text{「思う」の「高の補正出現率」} = \frac{\text{「思う」の「高の出現率」}}{\text{「高の出現率」}} \quad (4.4)$$

以上より、「高の補正出現率」を求めたら同様の方法で「低の補正出現率」を求めて式4.5のように「高の比率」を求める。

$$\text{「思う」の「高の比率」} = \frac{\text{「思う」の「高の補正出現率」}}{\text{「思う」の「低の補正出現率」} + \text{「思う」の「高の補正出現率」}} \quad (4.5)$$

以上のように「高の比率」を求めたら、「高の比率」によりエクセル上で降順のソートをする。

第5章 実験

5.1 実験データ

本研究の実験データは、コメント数が1,000件以上のYahooニュースの記事を対象に9月から10月にかけて119記事分収集した。本研究では全ての記事コメントを使用した。実験データの具体的な件数については、表5.1の通りである。

また手順2を元に高、中、低を分割した。表5.2がそれぞれのコメント件数である。手順2で分けた各評価のコメントの例を表5.4に示す。

5.2 実験結果

5.2.1 対応分析

KH Coderの対応分析では、表5.4の設定で行った。

対応分析の結果は図5.1の通りであった。対応分析は原点からの方向と距離が関係するので、「皇室」や「結婚」といった語は高評価のコメントに多かったことなどが分かる。同様に「コメント」や「批判」が低評価コメントに、「開催」や「定年」が中評価コメントに傾向があることも分かる。しかし対応分析の性質より、原点付近に語が集中しすぎて中評価の点と表記ですら大きくズレてしまっており、視覚的には分かりづらい図となった。

表 5.1: 実験データ

Yahoo ニュース	合計数
総記事数	119 件
総コメント数	360,317 件

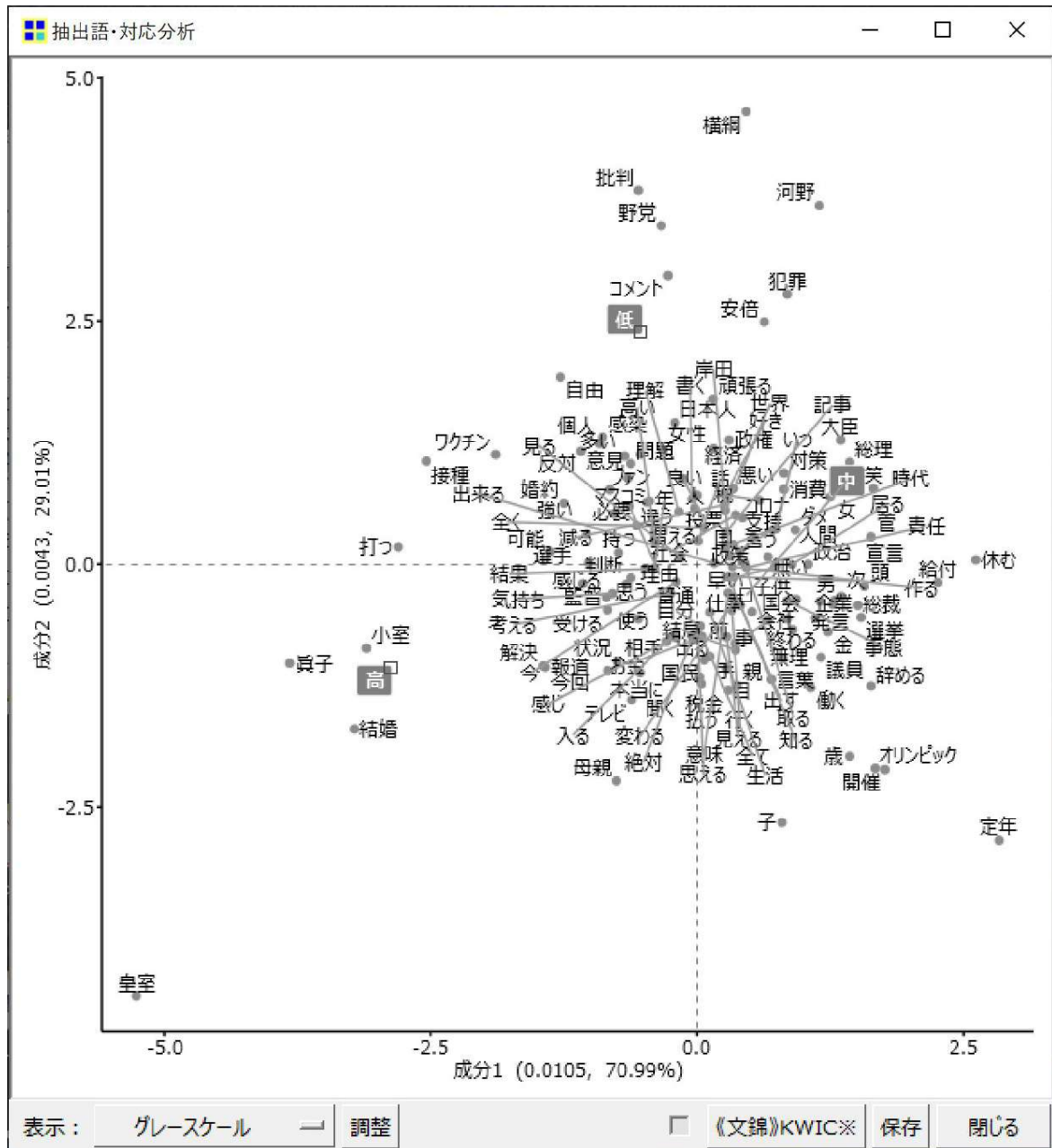


図 5.1: Yahoo ニュースのコメントの対応分析

表 5.2: 各評価コメントの数

評価	コメント数
高	22,540 件
中	291,061 件
低	46,716 件

表 5.3: 各評価コメントの例

評価	コメント
高	話しは聞くけど行動には起こさない、岸田さん出番ですよ。
中	騒がなってないところなんです。
低	餃子の一件の些細なことでそこまで言われるのか(笑)

5.2.2 クロス集計のソート

KH Coder でのクロス集計の結果を「高の比率」でソートしたものが表 5.5 と表 5.6 である。表 5.5, 5.6 は、左上から左下へ、左から右へと「高の比率」の降順でソートされている。

表 5.5 より、高評価コメントに傾向がある語が分かる。ここで「皇室」や「母親」などは、皇室に関する記事や母親の虐待に関する記事などの特定の記事のみに多用されている可能性のある単語である。そのためこのような可能性のある単語は無視すれば、記事全体では「本当に」や「今回」、「状況」などの語が高評価コメントに偏って出現していることが分かった。

また表 5.6 から、低評価コメントの出現傾向が高い語が分かった。記事全体では「笑」や「コメント」などが低評価コメントに偏って出現していることが分かった。

表 5.4: 研究での対応分析の設定

最小出現数による語の取捨選択	4000
最小文書数による語の取捨選択	1
使用した品詞	名詞, サ変名詞, 形容動詞, 人名, ナイ形容, 副詞可能, 未知語, タグ, 感動詞, 動詞, 形容詞, 副詞, 名詞 C
分析に使用するデータ表	総合評価 (外部変数)
差異が顕著な語を分析に使用	上位 150 語

表 5.5: 図 5.1 の各単語の比率のソート (1)

単語	高	低	単語	高	低	単語	高	低
皇室	0.85680	0.14320	言葉	0.62665	0.37335	接種	0.59672	0.40328
母親	0.74285	0.25715	監督	0.62506	0.37494	強い	0.59405	0.40595
結婚	0.74160	0.25840	全て	0.62450	0.37550	生活	0.59284	0.40716
子	0.68161	0.31839	婚約	0.62422	0.37578	開催	0.59100	0.40900
解決	0.67926	0.32074	手	0.62033	0.37967	高い	0.58999	0.41001
報道	0.67454	0.32546	目	0.61527	0.38473	見る	0.58734	0.41266
本当に	0.66812	0.33188	持つ	0.61432	0.38568	話	0.58596	0.41404
今回	0.66290	0.33710	入る	0.61399	0.38601	必要	0.58573	0.41427
状況	0.66187	0.33813	国民	0.61317	0.38683	反対	0.58506	0.41494
打つ	0.64989	0.35011	お金	0.61273	0.38727	作る	0.58403	0.41597
テレビ	0.64947	0.35053	絶対	0.61157	0.38843	理由	0.58302	0.41698
判断	0.64780	0.35220	年	0.61156	0.38844	増える	0.58164	0.41836
受ける	0.64291	0.35709	見える	0.61117	0.38883	出す	0.58138	0.41862
思える	0.63962	0.36038	気持ち	0.60899	0.39101	ファン	0.58126	0.41874
感じ	0.63922	0.36078	自分	0.60886	0.39114	違う	0.57920	0.42080
可能	0.63897	0.36103	結果	0.60881	0.39119	知る	0.57858	0.42142
親	0.63821	0.36179	払う	0.60869	0.39131	無理	0.57823	0.42177
選手	0.63793	0.36207	使う	0.60633	0.39367	個人	0.57619	0.42381
相手	0.63602	0.36398	歳	0.60441	0.39559	仕事	0.57473	0.42527
感じる	0.63552	0.36448	前	0.60321	0.39679	開く	0.57439	0.42561
普通	0.63135	0.36865	思う	0.60279	0.39721	働く	0.57398	0.42602
税金	0.62979	0.37021	今	0.60243	0.39757	出来る	0.57224	0.42776
結局	0.62920	0.37080	行く	0.59922	0.40078	変わる	0.57190	0.42810
出る	0.62689	0.37311	考える	0.59741	0.40259	次	0.56702	0.43298

5.2.3 コロケーション分析

5.2.2 節で高評価コメントに傾向があった「本当に」と低評価コメントに傾向があった「笑」についてコロケーション分析を行った。

高だけのデータで「本当に」に関する各単語のスコア値のトップ10と、低だけのデータで「笑」に関する各単語のスコア値のトップ10を表5.7にまとめた。太字のものは、表5.7の左右の表で共通した単語である。

表5.7を見ると「ない」と「思う」、「日本」、「人」がトップ10で共通して出現していて特徴の薄い語であることがわかる。細字の単語を見ると、「本当に」の方は「良い」や「必要」、「残念」などの語が、「笑」の方は「笑」や「ん」、「言う」と言った語がスコアが高いようだった。以上を元により詳細に見ていきたい。

表 5.6: 図 5.1 の各単語の比率のソート (2)

単語	高	低	単語	高	低	単語	高	低
責任	0.56681	0.43319	社会	0.53875	0.46125	政策	0.49095	0.50905
理解	0.56646	0.43354	好き	0.53825	0.46175	経済	0.49022	0.50978
取る	0.56499	0.43501	世界	0.53811	0.46189	税	0.48826	0.51174
時代	0.56498	0.43502	マスコミ	0.53611	0.46389	いつ	0.48810	0.51190
オリンピック	0.56450	0.43550	議員	0.53440	0.46560	総裁	0.48571	0.51429
ワクチン	0.56401	0.43599	良い	0.53302	0.46698	投票	0.48472	0.51528
会社	0.56247	0.43753	書く	0.53161	0.46839	対策	0.47509	0.52491
全く	0.56150	0.43850	無い	0.52767	0.47233	政権	0.46945	0.53055
自由	0.56069	0.43931	早い	0.52475	0.47525	宣言	0.46738	0.53262
終わる	0.56056	0.43944	減る	0.52464	0.47536	選挙	0.46711	0.53289
国	0.55414	0.44586	金	0.52327	0.47673	頑張る	0.46625	0.53375
意味	0.55388	0.44612	事態	0.52177	0.47823	頭	0.46412	0.53588
問題	0.55348	0.44652	日本人	0.52050	0.47950	コメント	0.43880	0.56120
事	0.55177	0.44823	企業	0.51976	0.48024	消費	0.43869	0.56131
定年	0.55030	0.44970	発言	0.51592	0.48408	批判	0.43852	0.56148
多い	0.54832	0.45168	コロナ	0.51379	0.48621	野党	0.43023	0.56977
感染	0.54685	0.45315	国会	0.51255	0.48745	女	0.42512	0.57488
人	0.54231	0.45769	辞める	0.51190	0.48810	総理	0.40333	0.59667
子供	0.54213	0.45787	悪い	0.50695	0.49305	大臣	0.39441	0.60559
言う	0.54184	0.45816	男	0.50603	0.49397	笑	0.38139	0.61861
女性	0.54074	0.45926	居る	0.50310	0.49690	犯罪	0.37207	0.62793
意見	0.54047	0.45953	支持	0.49833	0.50167	給付	0.36013	0.63987
人間	0.54028	0.45972	ダメ	0.49566	0.50434	横綱	0.30493	0.69507
記事	0.53980	0.46020	政治	0.49433	0.50567	休む	0.25491	0.74509

表 5.8 で、高評価のコメントにおいての「本当に」の左右にどのような語が出現したかを以下のように整理した。

- 表の右 3, 右 4 に「欲しい」や「ほしい」といった語が出現しているのは、何らかの要望や意見を表しているのではないか。
- 表の右 1 には「困る」や「酷い」, 「残念」といったような自分の感情や感想を表す語が多いように思った。
- 表の左 4, 右 1, 右 2, 右 3, 右 4 に「良い」, 「いい」, 「悪い」といった評価を示す語が多くみられた。

以上のまとめから、「本当に」がある高評価コメントでは自分の意見や感想を述べるケースが多いように思われた。その中では自分なりに、なんらかの評価を施すようなコメントもあった。

表 5.7: 総合評価が高のみでの「本当に」と、低のみでの「笑」のコロケーション分析でのスコアのトップ 10

高のみの「本当に」			低のみの「笑」		
抽出語	品詞	スコア	抽出語	品詞	スコア
ない	否定助動詞	51.717	笑	人名	122
思う	動詞	33.517	ない	否定助動詞	40.617
良い	形容詞	26.033	ん	否定助動詞	15.283
必要	形容動詞	23.783	言う	動詞	9.15
日本	地名	23	思う	動詞	8.483
人	名詞 C	21.467	人	名詞 C	6.717
結婚	サ変名詞	18.7	自分	名詞	5.3
国民	名詞	17.95	日本	地名	5.067
残念	形容動詞	15.833	分かる	動詞	4.383
眞子	人名	12.083	ワクチン	名詞	4.05

次に表 5.9 で、低評価のコメントにおける「笑」の左右にどのような語が出現したかを整理した。

- 表の左 1, 左 2 に「よね」, 「かよ」, 「たよ」, 「いよ」などの同意を求める呼びかけが散見した。
- 表の左 2 には「な」, 「あ」といった自分の感想を強調するような語があった。
- 表の右 2 には「ネット」, 左 2 には「コメント」があるのは、他のコメントに言及していると思った。
- 表で「笑」が左 1, 左 4, 左 5, 右 1, 右 4, 右 5 で見られることから、「笑」の周囲で「笑」が多用されがちであると思った。
- 表の右 3, 右 4, 右 5 で「自分」が頻発してるので「笑」の後には自分の考えなどが続きがちと思った。

以上から、「笑」がある低評価コメントでは他人に自分の考えを分かってもらおう、もっと言えば押し付けようとするコメントが多い可能性があるとおもった。また中には他のコメントに対して、いらぬ言及をするケースもあると思われる。また”笑”を連発するケースもあり、他人を嘲笑するようなケースもあると思われる。

以下 5.2.3 節のまとめとして、「笑」がある低評価コメントでは他人を嘲笑したり自分の意見の同意を無理に求めたりするような語が多くあったが、「本当に」がある高評価コメントではそのような語は見受けられなかった。また表 5.7 のように「ない」な

どのマイナスの意味の語は低評価コメントでも高評価コメントでもあり，評価には直接関係しないと思われる．

表 5.8: 高のデータに含まれる”本当に”の左右に多く含まれている語の上位 10 単語

左 1		左 2		左 3		左 4		左 5						
抽出語	品詞	抽出語	品詞	抽出語	品詞	抽出語	品詞	抽出語	品詞					
本当に	副詞	3	ない	否定助動詞	22	ない	否定助動詞	32	ない	否定助動詞	16	ない	否定助動詞	20
皇室	名詞	1	人	名詞 C	13	思う	動詞	11	思う	動詞	6	思う	動詞	8
ない	否定助動詞	0	ワクチン	名詞	7	眞子	人名	11	言う	動詞	6	人	名詞 C	5
人	名詞 C	0	日本	地名	7	言う	動詞	5	日本	地名	4	言う	動詞	5
ワクチン	名詞	0	国民	名詞	5	見る	動詞	5	いい	形容詞 (非自立)	4	日本	地名	3
日本	地名	0	見る	動詞	4	結婚	サ変名詞	4	ん	否定助動詞	4	国民	名詞	3
国民	名詞	0	結婚	サ変名詞	4	小室	人名	3	使ら	動詞	4	子供	名詞	3
見る	動詞	0	人間	名詞	4	フアン	名詞	3	小室	人名	3	皇室	名詞	2
結婚	サ変名詞	0	自民党	組織名	4	人	名詞 C	2	出る	動詞	3	気持ち	名詞	2
人間	名詞	0	考える	動詞	3	若い	形容詞	2	人	名詞 C	2	人間	名詞	2
右 1			右 2			右 3			右 4			右 5		
抽出語	品詞	右 1	抽出語	品詞	右 2	抽出語	品詞	右 3	抽出語	品詞	右 4	抽出語	品詞	右 5
良い	形容詞	25	ない	否定助動詞	18	ない	否定助動詞	18	ない	否定助動詞	21	思う	動詞	20
必要	形容動詞	21	思う	動詞	17	思う	動詞	15	思う	動詞	19	ない	否定助動詞	9
残念	形容動詞	15	悪い	形容詞	10	人	名詞 C	9	人	名詞 C	12	人	名詞 C	9
日本	地名	14	結婚	サ変名詞	5	PTSD	未知語	6	ん	否定助動詞	11	日本	地名	7
国民	名詞	12	疲れ	名詞	5	いい	形容詞 (非自立)	5	いい	形容詞 (非自立)	4	自分	名詞	3
結婚	サ変名詞	11	皇室	名詞	4	国	名詞 C	4	ほしい	形容詞 (非自立)	4	国民	名詞	3
困る	動詞	11	人	名詞 C	4	欲しい	形容詞 (非自立)	4	欲しい	形容詞 (非自立)	3	危険	形容動詞	2
酷い	形容詞	9	国民	名詞	3	日本	地名	3	愛す	動詞	3	気持ち	名詞	2
気の毒	形容動詞	9	気の毒	形容動詞	3	必要	形容動詞	3	結婚	サ変名詞	3	若い	形容詞	2
恐ろしい	形容詞	9	国	名詞 C	3	愛す	動詞	3	幸せ	形容動詞	3	苦しむ	動詞	2

表 5.9: 低のデータに含まれる”笑”の左右に多く含まれている語の上位 10 単語

左 1		左 2			左 3			左 4			左 5			
抽出語	品詞	左 1	抽出語	品詞	左 2	抽出語	品詞	左 3	抽出語	品詞	左 4	抽出語	品詞	左 5
笑	人名	59	ない	否定助動詞	16	ない	否定助動詞	36	ない	否定助動詞	31	ない	否定助動詞	18
ない	否定助動詞	6	ん	否定助動詞	10	ん	否定助動詞	16	思う	動詞	8	思う	動詞	15
ん	否定助動詞	3	思う	動詞	5	思う	動詞	7	言う	動詞	7	思う	動詞	6
枝野	地名	2	よね	人名	5	多い	形容詞	5	分かる	動詞	6	笑	人名	5
よね	人名	1	怖い	形容詞	4	出る	動詞	4	ん	否定助動詞	6	日本	地名	3
かよ	人名	1	コメント	サ変名詞	4	厳しい	形容詞	3	関係	サ変名詞	6	分かる	動詞	3
分かる	動詞	1	多い	形容詞	3	早い	形容詞	3	笑	人名	4	意味	サ変名詞	3
輩	名詞 C	1	な	感動詞	3	欲しい	形容詞 (非自立)	3	見る	動詞	4	悪い	形容詞	3
たよ	人名	1	あ	感動詞	3	いい	形容詞 (非自立)	3	ヤバ	未知語	3	入れる	動詞	3
いよ	人名	1	共産	組織名	2	無い	形容詞	3	言える	動詞	3	気	名詞 C	3
右 1		右 2			右 3			右 4			右 5			
抽出語	品詞	右 1	抽出語	品詞	右 2	抽出語	品詞	右 3	抽出語	品詞	右 4	抽出語	品詞	右 5
笑	人名	59	普段	副詞可能	5	ない	否定助動詞	5	自分	名詞	10	人	名詞 C	7
違う	動詞	1	ワクチン	名詞	4	言う	動詞	4	人	名詞 C	7	良い	形容詞	6
裏	名詞 C	1	日本	地名	3	ワクチン	名詞	3	言う	動詞	6	笑	人名	5
ギタリスト	名詞	1	ネット	名詞	3	自分	名詞	3	ない	否定助動詞	4	国民	名詞	5
)STU	未知語	1	菅	人名	3	税金	名詞	3	他	名詞 C	4	日本	地名	4
特に	副詞	1	今	副詞可能	3	打つ	動詞	3	笑	人名	4	自分	名詞	3
本当に	副詞	1	共産	組織名	2	一般人	名詞	3	ティンヤツ	未知語	4	ない	否定助動詞	3
ない	否定助動詞	0	社民	組織名	2	政治	名詞	3	感染	サ変名詞	3	独身	名詞	3
ん	否定助動詞	0	最近	副詞可能	2	政権	名詞	3	日本	地名	3	関心	名詞	3
枝野	地名	0	企画	サ変名詞	2	今	副詞可能	2	確か	形容動詞	3	見る	動詞	3

5.2.4 grep コマンドを用いた分析

まず「本当に」を含む高評価コメントについて見る。まず表 5.10, 表 5.11, 表 5.12, 表 5.13 が「本当に」を含む高評価コメントの一部をランダムに抜粋したものである。

表ではコメント 10, 13 のように何かに問題提起をするようなコメントが散見された。またコメント 9 のように普段言いづらいようなことを代わりに言ってくれたり、思っただけでも言葉にし辛いようなことには高評価がつきやすいように思った。またコメント 5 ではみんなが同意できるようなコメントであるので高評価がついているように感じ、同様に多くの人同意可能なコメント 8 やコメント 19 のような何かの感動や感謝を述べたようなものにも高評価が多いように感じた。

ここで以上の grep 分析の結果と 5.2.3 節の結果と比べてみたい。「欲しい」や「ほしい」を含むコメントは見当たらなかったが、表 5.11 のコメント 10 や表 5.12 の 13 のように何らかの問題提起を自らの意見として述べたコメントが高評価になっているので、要望や意見などのコメントは高評価になりやすいと思われる。また「酷い」という語は表 5.10 のコメント 5 に現れており、単に自分の感情や感想にとどまらず他の人でも同意できそうなコメントに使用されていた。また「良い」は表 5.10 のコメント 3 に含まれているが、何らかに評価をしているというよりは疑問や疑惑を強調するような使用のされ方をしている。このような他人が同意可能な疑問や疑惑を強調する例も表 5.10 のコメント 1, 2, 3, 4 などにみられた。

次に「笑」を含む低評価コメントについて見る。表 5.14, 5.15, 5.16 が「笑」を含む低評価コメントの一部を抜粋したものである。ただし、「笑う」など ChaSen で「笑」とは別にカウントされているものは意図的に省いてある。

表では、全体的に軽い言い方が目立っているように見えた。例えば、コメント 6 の「羨ましい〜」やコメント 8 の「めんどくせーな」など伸ばし棒を使用したコメントである。またコメント 4 やコメント 6, 8 では他人を茶化しているようなコメントが目立った。またコメント 13 のような周囲が納得しがたいのではないかとといったコメントも存在し、中にはコメント 5 のように予言のようなコメントも存在した。

以上の grep 分析の結果と、5.2.3 節の内容と比べてみた。「笑」を含む低評価コメント群で「笑」や「自分」はそこまで多用されてはいないことが分かった。一方、表 5.14 のコメント 3, 表 5.15 のコメント 10, 表 5.16 のコメント 20 などのように他のコメントに言及しているようなコメントは、「本当に」を含む高評価コメントよりも大分多く存在した。また「よね」や「たよ」、「な」などは表 5.14 のコメント 2, 表 5.16 のコメント 12 や 20 にみられて、やはり自分の意見の補強や同意を求めるように使用されていた。

5.3 追加実験

5.3.1 高評価コメントと低評価コメントで同一の中間語を用いたコロケーション分析

中間語(高評価コメントと低評価コメントで同じくらいの頻度で現れる語),特に「思う」という語についてコロケーション分析を行ってみた.

表5.17に評価が高,低での「思う」のスコア値の上位30語の単語を示す.なお,太字は高評価コメントにも低評価コメントにも現れている単語の中で順位が5位以上離れている単語であり,下線はどちらかのコメントにしかない単語である.

まず表5.17を見ると,下線は表の下部に固まっており,高評価コメントと低評価コメントで「思う」の周りに現れる単語はあまり差がない.

一方太字は表の中ほどの単語に多く現れており,どのような単語が現れやすいかは違いがある.高評価コメントでは「国民」,「本当に」,「ぬ」,「ワクチン」が,低評価コメントでは「違う」,「見る」,「欲しい」が現れやすいことが分かった.

また表 5.18 は高評価コメントでの「思う」の左右に現れる語の上位 10 語で、表 5.19 は低評価コメントでの「思う」の左右に表れる語の上位 10 語である。なお、太字は表 5.17 で高評価コメントの方がスコア順位が 5 位以上高かった単語で、下線は同様に低評価コメントの方がスコア順位が 5 位以上高かった単語のものである。

これらの表を見ると、下線は左 2 から左 5 までに存在していることがわかる。「違う」と「欲しい」は表の左 2 にのみ存在しており「違うと思う」や「欲しいと思う」などというように使用されていると見られる。一方、「見る」は表の左 3 から左 5 にのみ存在しており「見ないと思う」や「見る必要ないと思う」などと言ったように使用されていると見られる。

表 5.10: ランダムに抜き出した”本当に”を含む高評価コメント 20 件 (1)

1	本当に複雑性 PTSD なのでしょうか？
2	<p>金銭問題に国民への説明断固拒否、 更には新たに大学を卒業していないという疑惑が出てきていますが 本当に大丈夫なのでしょうか？次から次へと疑惑が出てきてます。 我々日本国民は、幼い頃から眞子様の成長を見守り、 国民と共に成長してきただけに心配でなりません。 小室氏が一般女性の方とご結婚されるのであれば、 このようなやり方もあるのでしょうか、立場ある相手、 ましてや皇族という（天皇家は世界一で最も長く継承される王族です） 伝統ある家系の方とのご結婚です。 天皇家の面子を壊すような方が果たして 良識ある人間とは到底思えないのです。 もし小室氏にお子さんができて、 相手方がこのような方でも心からご結婚をお許しになるのでしょうか？ よく考えていただきたい。</p>
3	<p>佳代のしていたことは後妻業そのものですよね。 あのまま元婚約者の方が気づかずに貢続けていたら、 最後は自殺に見せかけた殺人事件になっていたかもしれません。 そうなる前に気付かれて本当に良かった。</p>
4	<p>長男は就職してワクチン打ちましたが、次男は打っていません。 もし、頑張って入った大学が義務化でもしたら、打つしかないと言います。 私は、万が一のことが恐くて、無人島にでも連れていきたいくらいですが、 何事にも確信はないし、子供の人生に責任を取る勇気もない。 子宮頸がん被害者のビデオを見て、当たってしまった若い彼女たちが、 本当に気の毒でした。 当事者以外はいかにも他人事、政府もなにもしてくれない。 こちらもまた被害者が出ないよう祈ります。</p>
5	<p>右折はショートカットが多いです。 信号のない交差点で左折する時、 右側から車来ないのを確認して行こうとすると、 左側からやってきた車が鼻先かすめるようなショートカットで 右折してくるから行けないんですよね。 本当に酷いショートカットする奴がいるから気をつけないと。</p>

表 5.11: ランダムに抜き出した”本当に”を含む高評価コメント 20 件 (2)

6	<p>以前から、Youtube でホームレスの方々に食料や衣類を提供している人がいる。現在、何に困っているとか聞いて、出来る範囲で支援してあげているよ。その人達は、もう何年もやっていて、誰に言われたわけでも無くやっている。登録者数も多いし、再生回数も多い。視聴者は彼らに支援してあげて欲しいと応援している。これ、国内、海外でやっているよ。中には、難民の方達や恵まれない地域に住む人に炊き出しをやっている人もいる。本当に尊敬する人達だよ。結局、ゆたぼんがやっている事って、その場限りの対応なんだよね。一度、話に行ったからOK的な。褒められるべきは、何年も支援し続けている人達。勘違いするなど言いたいわ。子供を使用する亀田も最低だわ。</p>
7	<p>人間って本当にどうしようもないな。悪魔が強すぎる。人間が弱すぎる。平和ボケしてこの現実気づかない見て見ぬふり自分さえ良ければ良い。悪魔を野放し…。読むのに勇気がいりました。この警察官の方の罪はきえないけれど、勇気のある告発は今何よりも必要でウィグル人を救う一歩なのは間違いない。</p>
8	<p>もう、ミサイル(飛翔体)は撃たなくていいです。撃っても、対応はそう変わりませんし。もはや撃つのが当たり前のかんしゃく持ち程度と考えられていますよ。それよりも、残骸で海を汚されることと、日本が遺憾の意を表明することが、本当に迷惑&残念でなりません。ミサイルを撃つ度に、制裁措置を段階的に引き上げるなど、国として強くの態度を見せることで。他国への牽制にもなります。遺憾の意を表明で改善された事ってありましたか？</p>
9	<p>ウィルスは人に移れば移るほどに強くなり、それが変異に繋がる。つまり、これ以上ストレスフルな毎日が続けさせないために、こんなのもう終わりにするために、みんなができるだけ意識して、家にいなきゃならないと思う。来年の夏も何処にも行けないなんて、、もう勘弁。もう本当に終わりにしたい。。みなさんも同じ気持ちでしょうだから、ここは最後の我慢にして終わりにしましょう！</p>
10	<p>本当に見直しが必要ですよね。ガッポリ貰うものは貰っておいて、カゴの中の鳥…と言うのは調子が良過ぎます。ご本人達が、皇室なんて嫌！と思われているので、望み通りもう無くしてしまっても良いのでは？</p>

表 5.12: ランダムに抜き出した”本当に”を含む高評価コメント 20 件 (3)

11	どれも酷い殺され方で本当に人間かなと思う。
12	バラダンで武井壮が大麻は悪くない！と肯定してる芸能人の方がいますが、いくら肯定しても日本では犯罪なので、許されるものではないと言っていましたね 本当にその通りだと思います
13	本当にね。 一人の我儘な娘の強行結婚でこんな疑惑だらけで問題一つも解決もしない人に税金使い使われ続けてるのに誰も止めない止められないし、結婚して一般人になろうと A 宮家とは縁は切れないでしょ？どーするの？ そして数々の疑惑も有耶無耶にされちゃうの？おかしくない？本当に…こうなってしまった A 宮さんの責任は重くないですか？
14	その世代が日本の中で一番人口が少ないからです。支払う額が少なくて済む、一番体裁よく給付出来るからという腐った考えです。もう自公民では日本経済が破綻します！どなたか救世主が本当に必要です！
15	本当にひどい輩（小室）とひどい一家（秋篠宮）
16	本当にお金持ちなんだと思います。だから庶民の感覚では理解できないとも思います。弟さんは大学に全身 Vuitton で来られてましたし、聞いたところお姉さんからもらったとも仰っていたので、桁違いのお金持ちなんだと思います。そのお姉さんがウノさんだと知ったのは後日で腰抜かしました。知ってしまったので、仲良くなろうとは思いませんでした。すごく性格の良い人でした。育ちの良さがあるんでしょうね。

表 5.13: ランダムに抜き出した”本当に”を含む高評価コメント 20 件 (4)

17	<p>お金の余裕のレベルが違いますしね。 平日と週末で分担してるっていうならそれもアリなのかも。 お父さんと子供がとても仲良しとかなら尚更。 確かに、おうちの事情って本当に外からはわからないので、 一概に長谷川さんのことを悪く言うのもちょっと違うかなーと思いました。 夫婦のことって、DV とかは別として 100:0 でどちらが悪いって あまりないと思うんですよね。</p>
18	<p>原作改変は本当に実力がある人にしか許されない領域ですよ。 ビューティフルドリーマーは最高でした…</p>
19	<p>白鵬さんありがとうまたご苦労様でした。 貴方のおかげで相撲が面白く楽しかったです。 横審という馬鹿の集団に負けずに頑張った事私達は評価します。 格闘技ですから勝ったら大きな声で喜びを表しても良いと思います。 休場が多かったのは晩年で仕方がないと思います。 横審の連中の現役の頃はほぼ全員情けない相撲取りばかりでした。 数多くの方は白鵬さんの事を評価しています。 今や相撲は日本の国技ではなくモンゴル相撲と 名前を変えても良いぐらいモンゴルの人達が頑張ってくれました。 モンゴルの人がいなければ今頃相撲は衰退していたと思う。 最後に本当にご苦労様でした。</p>
20	<p>本当の困窮者は住民税非課税世帯ではない。 パートやアルバイト、フリーランスで 年収 150 万や 200 万円という人は物凄く多い。 150 万を越えて非課税になる自治体は、ほぼ無い。 そこから住民税、健康保険料、年金等払ったら残りは月 8 万とかザラ。 家賃、光熱費払った残りで安い物を買って食べる。 一方で非課税だと払う物がそもそも安くなる上に各種手当の対象となる。 その差は果てしなく大きい。 この非課税を僅かに超えた人達で保険料も年金も真面目に払ってる人は 本当に支援しないといけない。 若年層や女性にはこのような人が沢山居ます。</p>

表 5.14: ランダムに抜き出した”笑”を含む低評価コメント 20 件 (1)

1	税金のこともあるから、現金手渡しなのはちょっと気になるけどな笑 脱税しようとしてねーか
2	昔の車ならいざ知らず、最近の車ってホントに今のレベルの法定点検を 2年ごとに受けないといけないの？日本の車検制度も見直し必要でしょ。 話逸れるけど、免許更新のときに教官が質問したんだけど、 自分の車のボンネットを一度も開けた事ないっていう人が そこそこいてホントびっくりした。 日常点検くらい自分でしましよや。 ... といっても先日近所で休日にタイヤ交換してる人がいて、 自分でやるのはいいんだけど、ボルト締めるのにレンチの上に乗って 思いっきり全体重かけて締めてるのを見た。 ボルトの頭が飛ぶんじゃないかと思って ヒヤヒヤして遠くから見てた(笑) ユーザー側も「ほんとに乗るだけ」の人が多いよね。
3	>dey***** 安倍の言動や人間性を見聞きして、そういう想像力すら働かないのか。 おめでたい方なこと(笑)
4	偉そう笑 なんの理由もなく笑った 批判するにもそれなりの理由は並べた方がいいかと 小学生の意見みたい
5	財務省の現役次官が財政破綻の寄稿をする裏を読まなければならない、 現役次官であれば国家財政の隅から隅まで熟知してる、 これは日本の財政破綻が近い将来避けられないとの思いから 財務省の責任回避の為の予防線だと思う。 間違いなく近い将来国内債務調整の名の下に 庶民はとんでもない増税かインフレを味わうだろ、 その時財務省は何年も前から言ったじゃないか！ってなるから、 まっ楽しみにしとくわ笑

表 5.15: ランダムに抜き出した”笑”を含む低評価コメント 20 件 (2)

6	<p>旦那さんを知らないんだけど、めっちゃ頼りに旦那さん、父親なのかなぁと想像した。もし旦那が主夫になると言っても、もし自分が高給取りでも、絶対に絶対に、旦那に子育てや家事は任せられない！任せたくない笑 羨ましい～</p>
7	<p>いつからこの国は共産主義国家になったんだ？際限なくばら撒いてもらわなければやっていけないなどというのであれば、共産党にでも入れたらどうだ（笑）。破綻するだろうが。</p>
8	<p>日本の組織、めんどくせーな笑</p>
9	<p>何百万もくれんなら中流 (笑) かな?! イメージだけの人寄せパンダ的な根拠ない政策で誰もが、投票すると思ってんの？学生レベルの考え方。</p>
10	<p>このコメ主のようにこんなことに必死に反論する人間や、それに賛同して加勢しちゃうような人間が溢れているつまり当時のクレームはやっぱり正しかったということでは 笑</p>

表 5.16: ランダムに抜き出した”笑”を含む低評価コメント 20 件 (3)

11	設計ミスでしょうが施工ミスですね。どこの会社が施工したの？ 恥ずかしくて言えないか (笑) こんなことも気付かない施工店。情けない。
12	高山、鳥谷に筋肉増加でホームラン増やせなんて言っていないよ 強くスイングしなさいって指導してたんだが？ 金本、片岡叩く人って高山、藤浪壊されたって 非難するけど本人の実力だろ？ 伊藤隼太なんか掛布はじめみんな指摘してたよ笑 首傾けて打つなんてプロでは一流になれんやろ 指摘指導するの当然じゃない？ 最後まで治らなかったけどでも現場に片岡はないなあって思う
13	大麻が有害という誤解も広まっているけどね。 結局、真意を見抜けない分からない奴は永久に分からない。 嘘だらけの日本、令和3年へようこそ (笑)
14	> 基本的に下手な人の代表的ドライブの一つ 街中で普通の人ほとんどですが、 こういうド下手な人も多く混合しているので。 めちゃくちゃ嫌いな運転の代表例で 右手だけの12時の位置の片手ハンドルで 偉そうに運転している人やたら多い。 カッコええつもりかこれもド下手な代表例。 所詮は見た目で判断してるだけ。笑 運転で必要なのは予測。
15	ハルイチさんが見た目にこだわるタイプっぽいじゃんね。 京子が結婚直後に「男性は味っていうよりも盛り付けとか 見た目のおいしさが重要なのかなとったりします」って いいともでいっていた。京子さんも行き詰まっていたのかもね (笑)
16	貰えない世帯がギャーギャー騒いでる笑
17	世間知らずなんだからこの人にまともな事聞いちゃダメだって。 茶化すぐらいで丁度いいんじゃない？ 総理時代からそれくらいの事しか答えられない ガッカリ番長なんだから (笑)
18	(笑) 延長しかやる事ないので頭がいいね
19	相変わらずこの手のマスゴミはハレンチだな。 こんな仕事して恥ずかしくないのかね。自分なら死にたくなる (笑)
20	で、感情論と感傷で輩達のコメントは終わりのな (笑) 可哀想に 鬼畜だ 死の親ガチャ 女は男を選んだ 生まれ変わったら幸せに (世界が日本国民大丈夫？と懸念) 日本国民総統合失調症

表 5.17: 総合評価が高もしくは低での”思う”スコア値のトップ 30

高評価コメント			低評価コメント		
抽出語	品詞	スコア	抽出語	品詞	スコア
ない	否定助動詞	773.333	ない	否定助動詞	980.95
人	名詞 C	118.583	いい	形容詞 (非自立)	202.55
良い	形容詞	105.383	良い	形容詞	174.317
いい	形容詞 (非自立)	92.7	人	名詞 C	142.333
良い	形容詞 (非自立)	58.9	良い	形容詞 (非自立)	102.733
必要	形容動詞	51.133	必要	形容動詞	75.683
今	副詞可能	45.767	ん	否定助動詞	68.5
国民	名詞	45.017	今	副詞可能	60.617
ん	否定助動詞	41.45	無い	形容詞	52.55
無い	形容詞	40.983	問題	ナイ形容	49.55
多い	形容詞	40.967	多い	形容詞	49.15
自分	名詞	36.75	自分	名詞	45.417
問題	ナイ形容	34.733	違う	動詞	44.8
本当に	副詞	34.317	見る	動詞	42.75
ぬ	否定助動詞	30.4	欲しい	形容詞 (非自立)	42.167
ワクチン	名詞	29.55	国民	名詞	39.6
言う	動詞	28.367	悪い	形容詞	37.233
日本	地名	27.383	日本	地名	34.983
打つ	動詞	26.05	本当に	副詞	33.783
違う	動詞	25.833	言う	動詞	32.667
悪い	形容詞	25.267	ワクチン	名詞	28.95
見る	動詞	24.067	出来る	動詞	28.9
出来る	動詞	23.967	ぬ	否定助動詞	28.733
ほしい	形容詞 (非自立)	23.85	子供	名詞	26.933
欲しい	形容詞 (非自立)	23.017	ほしい	形容詞 (非自立)	23.117
結婚	サ変名詞	21.483	思う	動詞	21.8
思う	動詞	19.933	別に	副詞	21.55
疑問	形容動詞	19.917	接種	サ変名詞	20.683
高い	形容詞	17.3	難しい	形容詞	20.6
出る	動詞	16.633	変わる	動詞	20.417

表 5.18: 高評価コメントの”思う”の左右に多く含まれている語の上位 10 単語

左 1		左 2		左 3		左 4		左 5	
抽出語	品詞	抽出語	品詞	抽出語	品詞	抽出語	品詞	抽出語	品詞
何とも	副詞	ない	否定助動詞	ない	否定助動詞	ない	否定助動詞	ない	否定助動詞
今	副詞可能	いい	形容詞 (非自立)	必要	形容動詞	人	名詞 C	人	名詞 C
強い	形容詞	良い	形容詞	問題	ナイ形容	必要	形容動詞	言う	動詞
改めて	副詞	良い	形容詞 (非自立)	良い	形容詞	言う	動詞	国民	名詞
最近	副詞可能	無い	形容詞	良い	形容詞 (非自立)	意味	サ変名詞	良い	形容詞
ない	否定助動詞	多い	形容詞	人	名詞 C	いい	形容詞 (非自立)	必要	形容動詞
良い	形容詞	違う	動詞	出来る	動詞	見る	動詞	本当に	副詞
本当に	副詞	欲しい	形容詞 (非自立)	大事	形容動詞	良い	形容詞	見る	動詞
心から	副詞	ほしい	形容詞 (非自立)	変わる	動詞	責任	名詞	思う	動詞
毎回	副詞可能	疑問	形容動詞	無理	形容動詞	問題	ナイ形容	可能	形容動詞
回数	14	回数	680	回数	229	回数	170	回数	202
品詞	副詞	品詞	否定助動詞	品詞	否定助動詞	品詞	否定助動詞	品詞	否定助動詞
抽出語	ない	抽出語	いい	抽出語	必要	抽出語	人	抽出語	人
品詞	副詞可能	品詞	形容詞 (非自立)	品詞	形容動詞	品詞	名詞 C	品詞	名詞 C
回数	13	回数	155	回数	72	回数	63	回数	79
抽出語	良い	抽出語	良い	抽出語	問題	抽出語	必要	抽出語	言う
品詞	形容詞	品詞	形容詞	品詞	ナイ形容	品詞	形容動詞	品詞	動詞
回数	11	回数	134	回数	59	回数	56	回数	33
抽出語	良い	抽出語	良い	抽出語	良い	抽出語	言う	抽出語	国民
品詞	形容詞 (非自立)	品詞	形容詞	品詞	形容詞	品詞	動詞	品詞	名詞
回数	10	回数	85	回数	58	回数	26	回数	29
抽出語	無い	抽出語	良い	抽出語	良い	抽出語	意味	抽出語	良い
品詞	副詞可能	品詞	形容詞	品詞	形容詞 (非自立)	品詞	サ変名詞	品詞	形容詞
回数	8	回数	63	回数	29	回数	26	回数	20
抽出語	多い	抽出語	人	抽出語	人	抽出語	いい	抽出語	必要
品詞	否定助動詞	品詞	名詞 C	品詞	名詞 C	品詞	形容詞 (非自立)	品詞	形容動詞
回数	7	回数	51	回数	27	回数	25	回数	20
抽出語	違う	抽出語	出来る	抽出語	動詞	抽出語	見る	抽出語	本当に
品詞	形容詞	品詞	動詞	品詞	動詞	品詞	動詞	品詞	副詞
回数	4	回数	41	回数	23	回数	24	回数	20
抽出語	欲しい	抽出語	大事	抽出語	良い	抽出語	良い	抽出語	見る
品詞	副詞	品詞	形容動詞	品詞	形容動詞	品詞	形容詞	品詞	動詞
回数	4	回数	40	回数	23	回数	21	回数	19
抽出語	ほしい	抽出語	変わる	抽出語	責任	抽出語	責任	抽出語	思う
品詞	副詞	品詞	動詞	品詞	動詞	品詞	名詞	品詞	動詞
回数	4	回数	37	回数	22	回数	21	回数	19
抽出語	疑問	抽出語	無理	抽出語	無理	抽出語	問題	抽出語	可能
品詞	副詞可能	品詞	形容動詞	品詞	形容動詞	品詞	ナイ形容	品詞	形容動詞
回数	4	回数	35	回数	20	回数	20	回数	17
抽出語	ない	抽出語	人	抽出語	人	抽出語	ない	抽出語	ない
品詞	否定助動詞	品詞	名詞 C	品詞	名詞 C	品詞	否定助動詞	品詞	否定助動詞
回数	196	回数	53	回数	46	回数	56	回数	78
抽出語	人	抽出語	人	抽出語	人	抽出語	ない	抽出語	ない
品詞	名詞 C	品詞	名詞 C	品詞	名詞 C	品詞	否定助動詞	品詞	否定助動詞
回数	31	回数	37	回数	45	回数	32	回数	36
抽出語	ぬ	抽出語	ワクチン	抽出語	ワクチン	抽出語	ワクチン	抽出語	人
品詞	否定助動詞	品詞	名詞 C	品詞	名詞	品詞	名詞	品詞	名詞 C
回数	25	回数	29	回数	31	回数	28	回数	28
抽出語	今	抽出語	ワクチン	抽出語	ワクチン	抽出語	今	抽出語	ワクチン
品詞	副詞可能	品詞	名詞	品詞	名詞	品詞	副詞可能	品詞	名詞
回数	9	回数	12	回数	25	回数	27	回数	28
抽出語	今	抽出語	日本	抽出語	日本	抽出語	国民	抽出語	今
品詞	動詞	品詞	地名	品詞	地名	品詞	名詞	品詞	副詞可能
回数	8	回数	12	回数	21	回数	24	回数	24
抽出語	自分	抽出語	今	抽出語	人	抽出語	人	抽出語	国民
品詞	名詞	品詞	副詞可能	品詞	副詞可能	品詞	名詞 C	品詞	名詞
回数	5	回数	11	回数	20	回数	21	回数	22
抽出語	国民	抽出語	自分	抽出語	自分	抽出語	日本	抽出語	接種
品詞	名詞 C	品詞	名詞	品詞	名詞	品詞	地名	品詞	サ変名詞
回数	3	回数	9	回数	18	回数	19	回数	19
抽出語	幸せ	抽出語	国民	抽出語	国民	抽出語	接種	抽出語	小室
品詞	形容動詞	品詞	名詞	品詞	名詞	品詞	サ変名詞	品詞	人名
回数	2	回数	8	回数	12	回数	15	回数	19
抽出語	必要	抽出語	子供	抽出語	子供	抽出語	小室	抽出語	自分
品詞	形容動詞	品詞	動詞	品詞	名詞	品詞	人名	品詞	名詞
回数	2	回数	7	回数	12	回数	14	回数	19
抽出語	気持ち	抽出語	個人	抽出語	個人	抽出語	自分	抽出語	結婚
品詞	名詞	品詞	人名	品詞	名詞	品詞	名詞	品詞	サ変名詞
回数	2	回数	5	回数	11	回数	14	回数	19
抽出語	行動	抽出語	本当に	抽出語	本当に	抽出語	結婚	抽出語	思う
品詞	サ変名詞	品詞	副詞	品詞	副詞	品詞	サ変名詞	品詞	動詞
回数	2	回数	5	回数	11	回数	14	回数	19

表 5.19: 低評価コメントの“思う”の左右に多く含まれている語の上位 10 単語

左 1		左 2		左 3		左 4		左 5	
抽出語	品詞	抽出語	品詞	抽出語	品詞	抽出語	品詞	抽出語	品詞
何とも	副詞	ない	否定助動詞	ない	否定助動詞	ない	否定助動詞	ない	否定助動詞
ない	否定助動詞	いい	形容詞 (非自立)	必要	形容動詞	人	名詞 C	人	名詞 C
今	副詞可能	良い	形容詞	問題	ナイ形容	必要	形容動詞	言う	動詞
全く	副詞	良い	形容詞 (非自立)	良い	形容詞	良い	形容詞	良い	形容詞
毎回	副詞可能	無い	形容詞	人	名詞 C	問題	ナイ形容	見る	動詞
嬉しい	形容詞	違う	動詞	人	形容動詞	見る	動詞	自分	名詞
嬉しい	副詞	欲しい	形容詞 (非自立)	大事	形容動詞	見る	形容詞 (非自立)	思う	動詞
心から	副詞可能	多い	形容詞	見る	動詞	いい	形容詞 (非自立)	思う	動詞
最近	副詞可能	多い	形容詞	良い	形容詞 (非自立)	考える	動詞	必要	形容動詞
いい	形容詞 (非自立)	ほしい	形容詞 (非自立)	話	サ変名詞	国民	名詞	国民	名詞
本当に	副詞	悪い	形容詞	関係	サ変名詞	本当に	副詞	出来る	動詞
回数	18	回数	799	回数	240	回数	224	回数	258
品詞	副詞	品詞	否定助動詞	品詞	否定助動詞	品詞	名詞 C	品詞	否定助動詞
抽出語	何とも	抽出語	ない	抽出語	必要	抽出語	人	抽出語	人
品詞	副詞	品詞	形容詞 (非自立)	品詞	形容動詞	品詞	名詞 C	品詞	名詞 C
回数	17	回数	354	回数	116	回数	73	回数	91
品詞	否定助動詞	品詞	形容詞	品詞	ナイ形容	品詞	形容動詞	品詞	動詞
抽出語	ない	抽出語	良い	抽出語	問題	抽出語	良い	抽出語	良い
品詞	否定助動詞	品詞	形容詞 (非自立)	品詞	形容詞	品詞	形容詞	品詞	形容詞
回数	15	回数	239	回数	73	回数	64	回数	33
品詞	副詞可能	品詞	形容詞 (非自立)	品詞	形容詞	品詞	名詞 C	品詞	動詞
抽出語	今	抽出語	無い	抽出語	人	抽出語	問題	抽出語	見る
品詞	副詞	品詞	形容詞	品詞	名詞 C	品詞	ナイ形容	品詞	動詞
回数	12	回数	76	回数	54	回数	38	回数	28
品詞	副詞可能	品詞	形容詞	品詞	形容動詞	品詞	名詞	品詞	名詞
抽出語	毎回	抽出語	違う	抽出語	大事	抽出語	見る	抽出語	自分
品詞	形容詞	品詞	動詞	品詞	形容動詞	品詞	動詞	品詞	名詞
回数	7	回数	71	回数	32	回数	34	回数	22
品詞	形容詞	品詞	形容詞 (非自立)	品詞	形容動詞	品詞	形容詞 (非自立)	品詞	動詞
抽出語	嬉しい	抽出語	欲しい	抽出語	見る	抽出語	いい	抽出語	思う
品詞	副詞	品詞	形容詞	品詞	動詞	品詞	形容詞 (非自立)	品詞	動詞
回数	6	回数	66	回数	31	回数	26	回数	22
品詞	副詞可能	品詞	形容詞	品詞	動詞	品詞	動詞	品詞	動詞
抽出語	心から	抽出語	多い	抽出語	良い	抽出語	考える	抽出語	必要
品詞	副詞可能	品詞	形容詞	品詞	形容詞 (非自立)	品詞	動詞	品詞	形容動詞
回数	5	回数	62	回数	31	回数	24	回数	20
品詞	副詞可能	品詞	形容詞	品詞	サ変名詞	品詞	名詞	品詞	名詞
抽出語	最近	抽出語	ほしい	抽出語	話	抽出語	国民	抽出語	国民
品詞	形容詞 (非自立)	品詞	形容詞 (非自立)	品詞	サ変名詞	品詞	名詞	品詞	名詞
回数	4	回数	38	回数	30	回数	23	回数	20
品詞	副詞	品詞	形容詞	品詞	サ変名詞	品詞	副詞	品詞	動詞
抽出語	本当に	抽出語	悪い	抽出語	関係	抽出語	本当に	抽出語	出来る
品詞	副詞	品詞	形容詞	品詞	サ変名詞	品詞	副詞	品詞	動詞
回数	4	回数	31	回数	29	回数	22	回数	19
品詞	副詞	品詞	形容詞	品詞	サ変名詞	品詞	副詞	品詞	動詞
抽出語	何とも	抽出語	ない	抽出語	必要	抽出語	人	抽出語	人
品詞	副詞	品詞	形容詞 (非自立)	品詞	形容動詞	品詞	名詞 C	品詞	名詞 C
回数	294	回数	55	回数	60	回数	67	回数	93
品詞	否定助動詞	品詞	否定助動詞	品詞	否定助動詞	品詞	否定助動詞	品詞	否定助動詞
抽出語	ない	抽出語	ない	抽出語	人	抽出語	日本	抽出語	人
品詞	否定助動詞	品詞	否定助動詞	品詞	名詞 C	品詞	地名	品詞	名詞 C
回数	35	回数	50	回数	40	回数	36	回数	49
品詞	名詞 C	品詞	否定助動詞	品詞	名詞 C	品詞	地名	品詞	名詞 C
抽出語	ぬ	抽出語	人	抽出語	今	抽出語	人	抽出語	自分
品詞	否定助動詞	品詞	名詞 C	品詞	副詞可能	品詞	名詞 C	品詞	名詞
回数	22	回数	33	回数	35	回数	33	回数	33
品詞	否定助動詞	品詞	副詞可能	品詞	副詞可能	品詞	副詞可能	品詞	名詞
抽出語	ん	抽出語	自分	抽出語	自分	抽出語	今	抽出語	ワクチン
品詞	否定助動詞	品詞	名詞	品詞	名詞	品詞	副詞可能	品詞	名詞
回数	21	回数	21	回数	28	回数	31	回数	32
品詞	名詞	品詞	副詞可能	品詞	名詞	品詞	副詞可能	品詞	名詞
抽出語	気持ち	抽出語	自分	抽出語	日本	抽出語	ワクチン	抽出語	日本
品詞	名詞	品詞	名詞	品詞	地名	品詞	名詞	品詞	地名
回数	11	回数	13	回数	28	回数	28	回数	28
品詞	動詞	品詞	動詞	品詞	地名	品詞	名詞	品詞	地名
抽出語	浮かぶ	抽出語	見る	抽出語	子供	抽出語	自分	抽出語	今
品詞	動詞	品詞	動詞	品詞	名詞	品詞	名詞	品詞	副詞可能
回数	11	回数	11	回数	25	回数	26	回数	24
品詞	名詞 C	品詞	地名	品詞	名詞	品詞	名詞	品詞	副詞可能
抽出語	壺	抽出語	ワクチン	抽出語	ワクチン	抽出語	子供	抽出語	子供
品詞	形容動詞	品詞	名詞	品詞	名詞	品詞	名詞	品詞	名詞
回数	6	回数	11	回数	22	回数	22	回数	23
品詞	形容動詞	品詞	名詞	品詞	名詞	品詞	動詞	品詞	名詞
抽出語	必要	抽出語	ワクチン	抽出語	小室	抽出語	言う	抽出語	国民
品詞	名詞	品詞	名詞	品詞	人名	品詞	サ変名詞	品詞	名詞
回数	4	回数	9	回数	17	回数	22	回数	23
品詞	名詞	品詞	名詞	品詞	副詞	品詞	サ変名詞	品詞	名詞
抽出語	記事	抽出語	国民	抽出語	本当に	抽出語	接種	抽出語	思う
品詞	名詞	品詞	名詞	品詞	副詞	品詞	サ変名詞	品詞	動詞
回数	3	回数	8	回数	16	回数	17	回数	22
品詞	名詞	品詞	名詞	品詞	副詞	品詞	組織名	品詞	動詞
抽出語	部分	抽出語	W	抽出語	国民	抽出語	自民党	抽出語	接種
品詞	名詞	品詞	未知語	品詞	名詞	品詞	組織名	品詞	サ変名詞
回数	3	回数	7	回数	15	回数	16	回数	19
品詞	名詞	品詞	未知語	品詞	名詞	品詞	組織名	品詞	サ変名詞

さらに、より深い考察を得るため grep でいくつかコメントを取得して考えてみたい。表 5.20, 表 5.21, 表 5.22, 表 5.23 は高評価コメントの中での「思う」を含むコメント群である。また同様に、表 5.24, 表 5.25, 表 5.26 は低評価コメントの中での「思う」を含むコメント群である。

記事対象に対する批判は高評価コメント (コメント 3, 10, 19) でも低評価コメント (コメント 4, 5, 6, 15) でも見られていた。ただ高評価コメントの方はあくまでも公人に対する批判のみであるのに対し、低評価コメントでコメント 6 やコメント 15 のように一般の人に対してや一部地域に対して行ったものが目立った。

また、他コメントにたいして言及したコメントも高低どちらでも見られた。しかし高評価コメントのコメント 8 では他のコメントに対して一部同意をし新たな発見や気付きを加えるものだったが、低評価コメントのコメント 1 では他のコメントに対して非難を加えてコメントの流れを変えようとするものだった。

低評価コメントの中ではコメントの意見の流れに逆行するような表現が多く見られた。例えば、低評価コメントのコメント 3 の「別に総批判することではないかなど。」やコメント 4 の「逆に、ワクチンも治療費も有償にしてもらいたい。」、コメント 10 の「だったら GOTO 再開してほしい。」などがあった。

また低評価コメントの中ではヘイトや蔑視であると思われるようなコメントもあり、コメント 14 やコメント 16 では特定の地方に対して、コメント 15 は障害者に対して、コメント 19 は出所した人物に対して見られた。

一方、高評価コメントでは新しい気付きや考えを与えるようなコメントも見られた。例えば、高評価コメントのコメント 13 やコメント 18, コメント 19 などがあった。

まとめると、高評価コメントには以下のような傾向がみられた。

- 批判は公人に対してのみ
- 他コメントへの言及は同意をするもの
- 新しい気付きや考えを与えるようなコメントも多い

また、低評価コメントには以下のような傾向がみられた。

- 批判は一般の人に対するものや一部の地域に対するものもあった
- 他コメントを非難するものがあった
- コメントの意見の流れに逆行するようなコメントがみられた
- ヘイトや蔑視するようなものもあった

5.3.2 中評価コメントでの中間語を用いたコロケーション分析

高と低評価で同じくらい出現する中間語である「思う」を中評価のみのコメント群を使用してコロケーション分析をした。まず表 5.27 が低, 中, 高評価のコメント群を使用したスコア値のトップ 30 の表である。これに関しては, 特に大きな特徴は見られなかった。順位の違いについても高評価コメントと低評価コメント, 両コメントに対して大きな差を持つものは少なかった。ただ, 「今」が中評価コメントでのみ順位が低くなっていることが気になった。

次に表 5.28 が中評価コメントの「思う」の左右に含まれている語の上位 10 単語を表したものである。表 5.18, 表 5.19 と比較するため, 表 5.29 を作成した。表 5.29 を見てみると, 低評価と高評価と比べて中評価コメントで「思う」の左右に登場する語彙に差が見られなかった。

表 5.20: ランダムに抜き出した”思う”を含む高評価コメント 20 件 (1)

1	<p>ぜひこのシステムをもっと推し進めて頂きたい。 一般の監視カメラも犯罪が起これば捜査に協力している。 善良な人を守ることを最優先にしてほしい。 出所者も何も悪いことをしなければ感知されていても 関係ないのだから堂々と生活すればいい。 人権を声高に叫んで阻止しようとしている識者は もっと大きな問題に目を向けたほうが良いと思う。</p>
2	<p>正直って俳優って一度付いてしまった浮気の印象は一生続くでしょう。 だから映画見ても彼の私生活しかイメージしか伝わらないと思う。 アクションとか悪役とかに挑戦した方がいいでしょう。 それでも見たくない人もいるかもしれません</p>
3	<p>そりゃ、「返さなくてもいい」という言葉を隠し撮りしたり、 借金を踏み倒した親子だと思われたくないから返しませんと言ったり、 金目当ての婚約なのに、 親子になることを信じて金を払ってくれた相手に 感謝の気持ちもないよね。</p>
4	<p>うちの父親も 70 歳まで勤めあげた事務職を退職し、 それでもまだ働きたいと言ってコンビニで働き始めました。 畑違いの仕事で大変だろうけど、家にいたらすぐに衰えるだろうからと 言う父親の言葉を半分納得しながら見守ってます。 この記事を見てもらいたいと思いました。 父親より更に先輩の言葉は私以上に響くと思います。 健康には気をつけながらも、楽しく働ける事が一番ですね。</p>
5	<p>約十年前にもウイグルの弾圧問題は日本でも報道されていたが、 その行為について是正する動きはどの国にも見られなかった。 その頃からなにも変わったように思えないので、 世界で助けるべき問題だと思う。</p>

表 5.21: ランダムに抜き出した”思う”を含む高評価コメント 20 件 (2)

6	<p>これを機にインバウンドに偏った取り組みを止めたら良い。 国内で需要と供給を回すべきだ。 安易に外国人観光客に頼るべきではない。 マナーを守らない団体客が大挙して訪れていた ここ数年は異常だった。。 おまけに大声で話しながら我が物顔で歩いていてね。 日本の景観が台無しだったよ。 ここ数年、企業は過去最高益を出しても給与に反映させて来なかった。 日本の給与水準は 20 年変わっていない。さらにデフレ。 これでは国力も上がらないよ。 給与水準を上げて購買意欲を高めて行かないと。 首相も変わるし、国民が一丸となって 経済の復興に取り組むべきだと思います。</p>
7	<p>卒業じゃないとすると、なんかやったんだろうとか思っちゃう 甘くない対応するときほだいたい飲酒とか喫煙なのかな</p>
8	<p>>嫌なら記事を読んだり、コメントを書き込むべきではないと思うけど。 報道も見なければ良いですよ。 あの PTSD の発表までは、 まさか内親王ご自身がエゴサしてるなんて思いもよらなかった。 なんせ「品位」のある方ですから。 いや、実際してないのかもしれないけど、 国民の誹謗中傷が原因で って言われて、 もしかして、このヤフコメとかガルチャンとか見てくれてるの !? って、 かえって内親王や小室さんに言いたいこと 書き込む人増えるんじゃない? って思った。</p>
9	<p>副反応だけで打つ打たないを決めるのは ちょっとどうなんだろうと思いますけどね。 目的は副反応の先にある、コロナの発症防止や重症化防止なわけで、 その効果があると思えば打てばいいし、 ないと思うなら打たなければいいと思いますが。</p>
10	<p>杏さんの稼ぎが多いとしても、こども 1 人 1 万って。。 あんな別れ方したのに、申し訳ないと思ってないんですかね。 もし本当なら、父親になる資格ないです。</p>

表 5.22: ランダムに抜き出した”思う”を含む高評価コメント 20 件 (3)

11	<p>被害者と同じ目に合わせると刑罰が、人間には必要なものではないかと思う。 役にも立たない性善説に立つ事が、被害者を次々と生む。 未来の被害を防ぐ為に厳罰を与えることのほうが、よほど人権尊重だ。</p>
12	<p>結局、ワクチン効果はあるのでしょうか？ 抗体値が下がっても感染した時には、ワクチンが反応して対抗してくれるから、 2回接種した方が良いニュースを以前読みました。 なのに、抗体値が下がってきたから、3回目接種…？ コンサート会場で、接種者と未接種者の席を分けて、 感染の予防効果を検証と言うニュースを見たのが、先月。 検証と言う事は、判っていないという事…？ ワクチン接種者で、コロナ後遺症になった人の 割合調査を始めるとニュースで見たのも最近。 以前、後遺症にもワクチンは効果あるらしい？事を言っていた様な…？ 高齢者や基礎疾患の人に、ワクチン接種を呼びかけていましたが、 高齢者や基礎疾患の人はワクチン接種で命を落としやすい対象でもあると思います。 ワクチン接種で亡くなった時に、高齢者でしたや、 基礎疾患があったので亡くなりました何て、本末転倒な気がします。 本当の事を知りたいです。</p>
13	<p>アベノミクスの理屈は、我々庶民に税金等の負担をかけて、 その分、大企業の税金等の負担を減らすと、大企業が潤って、 やがては給料として我々に戻ってくるという理屈。 そして、結果、大企業は利益を貯め続けるだけで、殆ど吐き出さなかった。 つまり、我々の財産がどんどん大企業に流れて、大企業は潤うけど、 我々はどんどん失っていく。要するにコレが格差拡大だ。(+-+) で、そんなもんいつまで続ける気だ！フザけるな！って話で、 今回の選挙は格差是正がテーマな訳だが、 このフザけた仕組みを作った自民党が善人面して「是正します！」って言ったら、納得なの?? 根本のアベノミクスを否定しないで、「修正します！」って言ったら、納得なの?? 更に言うと、俺は、納得するとか、しないとか以前に、 また騙されるんじゃないの??って思うけどね。(+-+)</p>
14	<p>いやいや、最初から KK 一家と AKNM 家は繋がってたんだと思うよ。 子達よりも先に。ただの結婚話じゃないとおもうよ、これは。</p>
15	<p>最終的には個人の判断に委ねる。としているのだから、それまでかと思うが、 問題は周囲の反応で 打たない人は圧倒的に不利な状況に追い込まれる。 この感覚は、あの人はコロナに感染したから 本人は愚か、 その家族や友人にも近づかないように…そんな風当たりが現実に存在していた。 最近、特に気になったのは 高齢者のグループが紅葉狩りのハイキングに来ていて、 多くの方がノーマスクでお喋り三昧。こう云う方々は、既にワクチン接種が完了して 伝染る確率も重症化のリスクも低くなったと 思い込んで自由を手に入れたと勘違いしている。 確かにワクチン接種の進捗が向上した事で新規感染者が一気に減少した要因も あるかと思うが、ワクチン接種しかりではないことを周知する事も 政府や自身体或いは報道機関の仕事ではないかとおもう。</p>

表 5.23: ランダムに抜き出した”思う”を含む高評価コメント 20 件 (4)

16	<p>急いては事を仕損じるという諺があります。 取り返しのつかなくなる前にお考え直された方がよろしいかと思えます。</p>
17	<p>古臭くないですよ。私は母親の立場ですが 将来娘が14歳年上のバツイチ子持ちとデキ婚は受け入れられないですね。 もちろん大喜びするお母さんも大勢いるとはおもいますが、私は無理です。 夫も大喜びで祝福はしないと思います。 あ、もちろん大喜びで「よくやった！」なんて 褒めるお父さんもたくさんいることでしょう。 家庭ごとにしあわせの価値って違うことは理解していますが、 私は無責任だな、と第一印象は悪いものだと思います。</p>
18	<p>朝青龍がいたころは白鵬は善玉として振舞ってた 白鵬一人横綱になってからも しばらくは善玉っぽかったけど一人横綱が続き、 白鵬自身、相撲界に危機感を抱いたのではないのでしょうか。 「自分は外国人力士だし、後進の強い力士の壁になろう」て思いから ヒールになったように感じます だってここ数年の、身体ボロボロながらも、 なりふり構わず勝ちに行く白鵬に勝てば、 それだけで、後輩力士はめっちゃヒーローになれたんですよ。 白鵬がそういう機会を用意し続けたのに 応えてくれる力士が出てこない その期間が長くなりすぎて 強い力士が出てこないがために 白鵬はただのヒールみたいになりつつ でも、腐らずに 勝ち続けようとしてきたんですよ 横審、それに応えられずに なんて文句言うとなねん あほか いつまでも上から目線で「白鵬の活躍に、一定の評価与える」じゃなくって 「ごめんなさい、ちゃんとやります」だろ</p>
19	<p>ご心配のことと思います。 でも皇室の方は、この状況でお見舞いに行けるのですね。 一般人は病院も施設も面会できません。 カゴの鳥と嘆くより、皇室の沢山の特権を感謝なさって下さい。</p>
20	<p>同様の谷間に騙されたことある 寄せ上げで実はBくらいだったと思う (泣</p>

表 5.24: ランダムに抜き出した”思う”を含む低評価コメント 20 件 (1)

1	<p>> まともな意見を書くとき青ポチの嵐。 与党に権力を与えすぎるとこうなるのです。法律一つで社会は徐々に変わっていく。だから選挙で権力のチェックが必要なのです。自分の意見と異なる意見はまともでないと言っている時点であなたもまともじゃないと思いますよ。いくら青ポチがついても逮捕されたり監禁されたりしていないなら、それは民主主義だからでしょ？</p>
2	<p>なんだかんだ言っても悪いのは小室家だから。 解決金ではなく慰謝料だろ？ 1度も謝ってないし悪いとも思ってない。それが原因で眞子さまが被害にあった様なもの。</p>
3	<p>好転望むから、仮に感染しても軽症であって欲しいからワクチン接種した。そういう方がほとんど。 まだ早いけど経済回す一手段としては認めるべきではあくまで時期が重要だと思います。 別に総批判することではないかなど。</p>
4	<p>結局、国民が負担することになるので、無駄な出費だと思う。逆に、ワクチンも治療費も有償にしてもらいたい。受益者負担が原則でしょう。</p>
5	<p>単純に X さん、母ちゃんのことをまだ好きなんだろうね。どうしても会いたいみたいだし。それで世の中騒がせてって感じかな。前から痛い奴だとは思ってたけど、何となく気持ちが分からないでもない。400万ぼっち本当はどうでも良いんだよ。坊主は関係ないから会わんでよろしい。X さんはお前ではなく母ちゃんに会いたいんだよ</p>
6	<p>写真をパッと見てどこの国の話かと思ったらまさか日本とは、何も考えないで騒ぐ奴らがこんなにいるんだね。コロナが落ち着くのを心から願ってる人たちにホント失礼。</p>
7	<p>「何を語るのか」ではなく「誰が語るか」だと思います！</p>
8	<p>白鵬関は十二分に貢献してこられたと思う。昔のままの世界を維持したいなら最低限外国人力士に頼ってはいけない。相手の常識は認めないのに自分達は気に入らないという根拠でいちいち難癖つけるのは凶々しいのでは？ それだけの見返りは与えてないでしょう。対価なく要求するなんてただのブラックですよ。</p>
9	<p>不倫してるわけでもなし、(そもそも不倫でもどうでもいい) そっとしてあげればいいのに。お子さんたちは立派に育っているんだし、突っ込むところないと思いますが</p>
10	<p>だったら GOTO 再開してほしい。ワクチンパスポート導入の上で。若年層に給付してもワクチン打ってない比率が高いのと、お金を使う場所がインターネット上の可能性も高いので、本気で経済回すことを考えたら一律給付にするのが妥当だと思う。</p>

表 5.25: ランダムに抜き出した”思う”を含む低評価コメント 20 件 (2)

11	<p>この 10 万の支給は大賛成です。 子育ての世代が一番大変なだから皆で応援しましょう！ 将来その子たちから年金という形で養ってもらいます。 この一律ってのがさすが公明党だと思う。 後は子供手当を 22 歳の学生までに延長して、 子供のいる家庭を皆で支援していきましょう</p>
12	<p>凄すぎる。でも、こんな歳まで働きたいとは思わないな。 この歳になったら考えも変わるのだろうか。</p>
13	<p>子育て世代からすると嬉しい話。 もちろん生活困窮者はいろんな世帯に いるわけですが、困窮の理由はそれぞれ。 大半は心身の問題や境遇などでしょうが、そうでない人もいます。 子供は最低限かかる教育費、生活費があり、もちろん稼げないので 親がその生活費を賄うわけで、出費は多くなってしまふ。 そして将来の国のためにいなければいけない存在。 子供がいないから税金払ってるのに 納得いかないというのはわかりますが、 それ言ったら病院行かないのに健康保険料払ってる人も沢山いるし、 雇用保険使わないのに払っている人もいる。 税金は必ず払った分が戻ってくるわけではなく、 むしろ自分に返ってくるものの方が少ないと 理解しないとイケないと思います。</p>
14	<p>神戸って知らない人はオシャレな感じ都会と思ってる人多いけど、 実際はただの田舎の狭い街だから、若者には物足りないでしょう。</p>
15	<p>そうそう SNS でも自身の NPD と ASD を こじらせた発達 (あとにしょうがいつけて) の 女が「自分さえ良ければいいとモラハラばかりして 責任取れないなら (殺したり 妻に全部育児を押し付ける等いろんな意味で) 子供を作るな」とかいう意見が 出るや否や「障害者は子供作るなど言っている！ 差別だ！」と騒ぎ発言者を叩いてる。 弱さを盾にしてくるから健常者は 当たり前頃すらこそそそ言わないとイケない。 マイノリティに優しい社会が加害されても黙ってろって社会と 同じってイケないと思うんですよね。</p>

表 5.26: ランダムに抜き出した”思う”を含む低評価コメント 20 件 (3)

16	<p>神戸は「阪神大震災が起きているから『大震災の震源地』になる可能性は低い」と私は思う。九州なら「福岡より熊本を選ぶ」。「一番危なそうな関東首都圏」(3つのプレートがあるから、状況が非常に予測しにくい)以外なら、私はこのルールで選びたい…</p>
17	<p>おめでとう！国民にも祝福幸せムードがこれまでにないくらい漂っています！このコロナ禍に小さな幸せをありがとう！お母さんの借金は私たち国民が頑張って働いて納税した税金で返してください！私たちは小室さんとかこ様の暮らしのために納税をしています。アメリカの暮らしは物価が高いと思うので、遠慮なく私たち国民の税金を使ってください！幸せになってください！小室さんが幸せなら ok です！</p>
18	<p>菅首相地味だが良いと思う。前任者と違い見苦しく弁解しないのが好感が持てる。非難の捌け口になって気の毒な面がある。無派閥だから当然政治手法不自由な点があろう。通信料が一変したのは有難い。この点だけでも評価に値する。</p>
19	<p>悪いけど 出所した人に基本的人権などいらないと思います。安心して暮らしたい。</p>
20	<p>まあ、誘導的な記事だけでも、実際そうなんだろうな。山口県は工業地帯が万年渋滞なワリには農道とかがめっちゃ整備されてて、一言で言えば「歪」だと思う。「や、整備順序おかしいでしょ?!」みたいな。某かの理由あるいは忖度があるのは確かだろうねえ。まあ僕は福岡県民なんですけどね。2号線だけは絶対使わない。</p>

表 5.27: 総合評価別の”思う”スコア値のトップ30

低評価コメント			中評価コメント			高評価コメント		
抽出語	品詞	スコア	抽出語	品詞	スコア	抽出語	品詞	スコア
ない	否定助動詞	980.95	ない	否定助動詞	5140.4	ない	否定助動詞	773.333
いい	形容詞 (非自立)	202.55	いい	形容詞 (非自立)	770.633	人	名詞 C	118.583
良い	形容詞	174.317	良い	形容詞	719.033	良い	形容詞	105.383
人	名詞 C	142.333	人	名詞 C	711.467	いい	形容詞 (非自立)	92.7
良い	形容詞 (非自立)	102.733	良い	形容詞 (非自立)	434.617	良い	形容詞 (非自立)	58.9
必要	形容動詞	75.683	ん	否定助動詞	331.15	必要	形容動詞	51.133
ん	否定助動詞	68.5	必要	形容動詞	301.15	今	副詞可能	45.767
今	副詞可能	60.617	無い	形容詞	284.9	国民	名詞	45.017
無い	形容詞	52.55	多い	形容詞	270.9	ん	否定助動詞	41.45
問題	ナイ形容	49.55	自分	名詞	262.533	無い	形容詞	40.983
多い	形容詞	49.15	今	副詞可能	244.25	多い	形容詞	40.967
自分	名詞	45.417	国民	名詞	236.417	自分	名詞	36.75
違う	動詞	44.8	問題	ナイ形容	202.717	問題	ナイ形容	34.733
見る	動詞	42.75	本当に	副詞	193.05	本当に	副詞	34.317
欲しい	形容詞 (非自立)	42.167	言う	動詞	192.383	ぬ	否定助動詞	30.4
国民	名詞	39.6	違う	動詞	180.767	ワクチン	名詞	29.55
悪い	形容詞	37.233	悪い	形容詞	170.367	言う	動詞	28.367
日本	地名	34.983	見る	動詞	168.15	日本	地名	27.383
本当に	副詞	33.783	日本	地名	160.717	打つ	動詞	26.05
言う	動詞	32.667	ぬ	否定助動詞	160.35	違う	動詞	25.833
ワクチン	名詞	28.95	欲しい	形容詞 (非自立)	147.05	悪い	形容詞	25.267
出来る	動詞	28.9	子供	名詞	146.267	見る	動詞	24.067
ぬ	否定助動詞	28.733	何とも	副詞	144.417	出来る	動詞	23.967
子供	名詞	26.933	出来る	動詞	135.017	ほしい	形容詞 (非自立)	23.85
ほしい	形容詞 (非自立)	23.117	思う	動詞	130.1	欲しい	形容詞 (非自立)	23.017
思う	動詞	21.8	ほしい	形容詞 (非自立)	123.35	結婚	サ変名詞	21.483
別に	副詞	21.55	出る	動詞	114.567	思う	動詞	19.933
接種	サ変名詞	20.683	変わる	動詞	101.433	疑問	形容動詞	19.917
難しい	形容詞	20.6	素晴らしい	形容詞	93.6	高い	形容詞	17.3
変わる	動詞	20.417	ワクチン	名詞	87.95	出る	動詞	16.633

表 5.28: 中評価コメントの”思う”の左右に多く含まれている語の上位 10 単語

左 1		左 2		左 3		左 4		左 5			
抽出語	品詞	回数	抽出語	品詞	回数	抽出語	品詞	回数	抽出語	品詞	回数
何とも	副詞	141	ない	否定助動詞	4244	ない	否定助動詞	1244	ない	否定助動詞	1045
ない	否定助動詞	90	いい	形容詞 (非自立)	1350	必要	形容動詞	440	人	名詞 C	385
毎回	副詞可能	77	良い	形容詞	992	問題	ナイ形容	317	言う	動詞	318
今	副詞可能	50	良い	形容詞 (非自立)	697	良い	形容詞	307	国民	名詞	240
本当に	副詞	39	無い	形容詞	395	人	名詞 C	218	自分	名詞	182
改めて	副詞	38	多い	形容詞	339	変わる	動詞	172	見る	動詞	152
全く	副詞	33	違う	動詞	272	無理	形容動詞	157	良い	形容詞	151
強い	形容詞	27	欲しい	形容詞 (非自立)	225	良い	形容詞 (非自立)	144	思う	動詞	127
最近	副詞可能	27	ほしい	形容詞 (非自立)	184	ダメ	形容動詞	139	本当に	副詞	118
と	感動詞	24	悪い	形容詞	155	見る	動詞	137	必要	形容動詞	95
右 1		右 2		右 3		右 4		右 5			
抽出語	品詞	回数	抽出語	品詞	回数	抽出語	品詞	回数	抽出語	品詞	回数
ない	否定助動詞	1446	ない	否定助動詞	493	ない	否定助動詞	352	ない	否定助動詞	373
人	名詞 C	195	ん	否定助動詞	245	人	名詞 C	186	人	名詞 C	159
ぬ	否定助動詞	116	人	名詞 C	174	自分	名詞	165	日本	地名	145
ん	否定助動詞	102	自分	名詞	73	今	副詞可能	161	自分	名詞	130
浮かぶ	動詞	58	今	副詞可能	60	日本	地名	103	今	副詞可能	119
壺	名詞 C	57	日本	地名	52	子供	名詞	94	国民	名詞	93
気持ち	名詞	27	子供	名詞	51	国民	名詞	89	子供	名詞	92
ツボ	名詞	22	国民	名詞	49	多い	形容詞	71	仕方	ナイ形容	85
今日	副詞可能	21	本当に	副詞	47	ん	否定助動詞	52	言う	動詞	76
反面	副詞可能	20	見る	動詞	37	ワクチン	名詞	52	ん	否定助動詞	71

表 5.29: 各評価コメントの”思う”の左右に多く含まれている語の上位 10 単語

左 1			左 2			左 3			左 4			左 5		
低	中	高	低	中	高	低	中	高	低	中	高	低	中	高
何とも	何とも	何とも	ない	ない	ない	ない	ない	ない	ない	ない	ない	ない	ない	ない
ない	ない	今	いい	いい	いい	必要	必要	必要	必要	必要	必要	人	人	人
今	毎回	強い	良い	良い	良い	問題	問題	問題	問題	問題	問題	言う	言う	言う
全く	今	改めて	良い	良い	良い	良い	良い	良い	良い	良い	良い	言う	言う	言う
毎回	本常に	最近	無い	無い	無い	人	人	人	人	人	人	見る	見る	見る
嬉しい	改めて	ない	多い	多い	多い	大事	変わる	人	人	人	人	自分	見る	見る
心から	全く	良い	欲しい	違う	違う	見る	無理	出来る	出来る	出来る	出来る	思う	思う	思う
最近	強い	本常に	多い	欲しい	欲しい	良い	良い	大事	大事	大事	大事	必要	思う	見る
いい	最近	心から	ほしい	ほしい	ほしい	話	ダメ	変わる	変わる	変わる	変わる	国民	国民	国民
本常に	と	毎回	悪い	悪い	疑問	関係	見る	無理	無理	無理	無理	本常に	意味	問題
右 1			右 2			右 3			右 4			右 5		
低	中	高	低	中	高	低	中	高	低	中	高	低	中	高
ない	ない	ない	ない	ない	ない	ない	ない	人	人	人	人	ない	ない	ない
人	人	人	人	人	人	人	人	人	人	人	人	人	人	人
ぬ	ぬ	ぬ	人	人	人	今	自分	ワクチン	ワクチン	ワクチン	ワクチン	自分	日本	ワクチン
人	人	人	今	自分	今	今	今	日本	日本	日本	日本	ワクチン	自分	今
気持ち	浮かぶ	浮かぶ	自分	自分	自分	ワクチン	ワクチン	今	今	今	人	日本	今	国民
浮かぶ	壺	壺	見る	日本	国民	子供	自分	日本	日本	日本	日本	今	国民	接種
壺	気持ち	幸せ	日本	子供	日本	ワクチン	子供	国民	国民	国民	国民	子供	子供	小室
必要	ツボ	必要	ワクチン	国民	打つ	小室	言う	多い	子供	子供	小室	国民	言う	自分
記事	今日	気持ち	国民	本常に	小室	本常に	接種	人	個人	個人	個人	接種	思う	結婚
部分	反面	行動	W	見る	出来る	国民	ワクチン	ワクチン	本常に	本常に	本常に	自民党	ワクチン	思う

「思う」の中評価コメントを抜き出した結果を表 5.30, 表 5.31, 表 5.32, 表 5.33 に示す。

まず, 高評価コメントを利用した表 5.20, 表 5.21, 表 5.22, 表 5.23 と比較してみた。一見すると高評価コメントと中評価コメントの違いがあまり見受けられない。しかし, 高評価コメントよりも問題提起の内容は少ない印象を受けた。高評価コメントではコメント 1, 5, 6, 9 など複数現れていたが, 一方中評価コメントでは表 5.30 のコメント 5 くらいしかなかった。そのため, 問題提起をすると高評価になりやすい構図が何となく見えた。

また, 中評価コメントでは公人などを擁護するコメントも散見され, コメント 7, 10, 15, 19 があった。

次に低評価コメントの表 5.24, 表 5.25, 表 5.26 と比較してみる。

低評価コメントと比較すると, 低評価コメントでの批判の少なさに驚いた。他のコメントに対する批判は, 低評価コメントではコメント 4, 9, 19 とあったが, 記事が書かれている対象に対する批判は少なく, むしろ擁護するようなコメントがコメント 8, 9, 17, 18 と多く見受けられた。中評価コメントでは, コメント 8, 13, 14, 18 と記事が書かれている対象に対する批判も多かった。

また中評価コメントでは, コメント 2, コメント 8, コメント 11 と「…」の使用が多かった。

まとめると, 中評価コメントの特徴は以下であると思う。

- 高評価, 低評価コメントと比べて使用された語の差異は少ない。
- 高評価コメントに比べると, 問題提起をするようなコメントは少なかった。
- 低評価コメントと比べると, 記事が書かれた対象に対する批判が多かった。
- 「…」の使用が多かった。
- 高評価コメントと低評価コメントのどちらの傾向もあり, 特に高評価コメントとよく似ていた。

表 5.30: ランダムに抜き出した”思う”を含む中評価コメント 20 件 (1)

1	<p>「詭弁だ」というコメントが多いですね。 文春のやり口を常日頃批判的に見ている私ですが、 この記事は「なるほど、そういう葛藤があるのか」と思い読みました。 同時に、でもやってることとズレてるだろ、 その影響力からして許されないズレだろ、とも思いました。 加えて、やはりそれでも意義があると論じる姿勢は 「上から目線」と感じざるを得ません。 それはさておき、要するにこれが「詭弁」、 すなわち悪意を伴う意図的な嘘であると 一方的に断じられてしまう社会だ、ということなんですよ。 そういう社会において文春のやり口は、 やはり功罪のうち「罪」が大きすぎると思います。</p>
2	<p>河野さんは、あんたにだけは言われたくないと思っている事でしょう…。</p>
3	<p>ポッポを批判する人がずいぶん多いが、 ポッポの方がブレないのは事実だと思うし、むしろ信念の人といってもよい。 ポッポの政治信条には全く同意できないけどね。 ブロックデマ太郎の最近の変節ぶりには開いた口が塞がらないよ。 さすがあの洋平の息子の二世政治家だね。</p>
4	<p>普段はマスゴミと言っているのに 何故この匿名の元婚約者については 100% 信用し善人扱いになるのか？ マスコミにした「日本人ってチョロい！」と、思っているでしょうね ただし金銭トラブルと皇室利用は別ですが！</p>
5	<p>新規感染者がワクチン打った人なのか、何回打ったのか、打ってないのか、 子供たちは何割打ったのか、という情報を添えるべきでは？ 増えてるぞ～だけではただ不安をあおるだけだと思う。</p>

表 5.31: ランダムに抜き出した”思う”を含む中評価コメント 20 件 (2)

6	大麻もピストルもそういうものが良ければ所持が認可されている国に移住すれば良いと思うよ。 それしか方法がないでしょ？
7	行政も問題なんだけど、一番問題なのはやはり、母親だよ。加害者と同罪だと思う。
8	<p>自己主張の強い独裁主義の国の民は飢餓に苦しむ …。</p> <p>その中でミサイル開発をし、多額の金を使う。</p> <p>相反する行為と思いませんか。</p> <p>あの国の独裁者は国民は皆働きバチで一部の上位者の為に働き、働けないものは不要と思っている。</p> <p>幾ら援助しても決して飢えに苦しむと</p> <p>言われている国民には行き届かないで搾取され、</p> <p>援助された収益はミサイル開発に使われる。</p> <p>今、北朝鮮の情報が何が正しく、何が正しくないか、判断するのは難しい。</p> <p>自国で飢餓に苦しむ人が本当に存在するのか。</p> <p>独裁者に並列して、富裕層の一部の特権者が優遇されて、</p> <p>正常な政治経済の中で飢餓が発生している状況とは思えない中で、</p> <p>援助を外部からするのは疑問視される。</p>
9	<p>オリックスファンなら怒るのは当たり前ですね。残念なのが、</p> <p>一部の人が千賀の SNS に文句を言い言ったり、</p> <p>死ね等の文言もありました。</p> <p>こう言う行為をしていると、オリックスファン全体が</p> <p>このような事をしていると勘違いされてしまう。</p> <p>来年以降にオリックスの山本由伸が同じ行為をしてしまった場合を</p> <p>想像出来るのであれば、千賀に怒る気持ちはあれど、</p> <p>誹謗中傷をしようとは思わない。</p> <p>怒る気持ちは分かるが、あまりにも酷いコメントが散見される。</p>
10	元婚約者の方が責任を感じる必要も謝る必要もないと思います。どちらかと言えば被害者です。

表 5.32: ランダムに抜き出した”思う”を含む中評価コメント 20 件 (3)

11	<p>元婚約者の方が、なぜ？申し訳なく思う必要があるのか？ 事の発端は小室親子！ 眞子さまの PTSD にしろ、今回の騒ぎにしろ… 全て小室親子が原因なのでは？ 眞子さまには申し訳ないが… 皇族の結婚って、もっと格式あるものだと思う。 もし…それをわかった上で… 一般人になり 1 人の女性として結婚されるなら… 正式に皇族（秋篠宮家）と離縁して… 税金や国の力に頼ること無く結婚されればいいと思う。 そうすれば…会見も必要ないと思う。 そして…結婚後（入籍後）は一般人なのだから… 護衛や生活費に税金を一切使わないで欲しいと 1 人の国民として思う。</p>
12	<p>一時金はいりませんと言えば、 国民が黙るとでも思っているのでしょうか？ だとしたら、国民を馬鹿にしすぎです。 そもそも、やましいことがなければ、 正々堂々と一時金をもらえばいいのに、 やっぱり受け取れないような理由があるんですよね？ この度の KK の態度や風貌、顔つきを見て、改めて思いました。 結婚しては駄目な人だって。</p>
13	<p>人の子を平気で殺すクズも許せないけど 母親の神経を疑う。 なぜ子供を守らなかったのか。 私も子供 2 人いますがとてもじゃないけど 虐待とか考えられません。 自分も親が再婚で手を出されたり 辛い思いもたくさんしてきたから 同じ事を していいとかしようとか思いません。 こんな悲しい事件がいつまでも減らない事も 悲しく思います。 つい腹が立ってコメントしてしまいました。 長々と失礼しました。</p>
14	<p>不倫ってある程度仕方ないと思うところあるんだけど、 小さい子供がいる親の不倫は最も許せない 自分で作った子供なのに子育てはどうしたの？ 女と合ってる時間があったら子育てしたら？ A 子だってインスタグラマーなら素人じゃないだろう、 名前だしなさいよねと思いました。</p>
15	<p>なんかね 負けが込んで出た言葉で一時的だと思うけど 巨人ファンではないけれど、頑張ってる貫きたい</p>

表 5.33: ランダムに抜き出した”思う”を含む中評価コメント 20 件 (4)

16	<p>子供が一番可哀想。 浮気してバレたら離婚するかもな～と思う男は 子供を作るなど言いたい。</p>
17	<p>選手達は今の FC 東京をどう捉えているのか、興味あるな。 J1 のリーグ戦ではまだ一度も優勝をしていないが、 近年はそこまで悪い成績でもない。 カップ戦では複数回の優勝を経験している一方で、 ACL は常にベスト 16 止まり。 この調子で続けていけばいいと思う人もいるだろうし、 このままじゃいけないと思う人もいるだろう。個人的には、 日本の首都・東京をホームタウンにしているチームと考えると 少し頼りない。</p>
18	<p>少し長く権力を握りすぎましたと思いますが？</p>
19	<p>ここは場所柄、政府批判をしたいが為の 無責任コメントが大半だろうけど、 本当に舵取りが大変だと思います。判断基準も無いに等しい。 何かを優先する判断すれば何か必ず犠牲になり批判される、 自民党でほんとうに良かったと痛感する。</p>
20	<p>お金をむしり取られた元婚約者さんの心の傷の方が よっぽど大きいと思います。 しつこいじじいと親子で言ってるかも。 元婚約者さんは自分を騙していたのか騙していなかったかを はっきりしたいのでしょうか。</p>

第6章 SNSにおいて評価を得るための 考察

表5.1の単語から記事依存が少ない「本当に」を含む高評価コメントと「笑」を含む低評価コメントを比較し考察を述べたい。

「本当に」を含む高評価コメントの中では、「見直しが必要ですよね」や「責任は重くないですか？」など問題提起をするコメントや思っただけでも普段は言いづらい、言葉にしにくいコメントなどが多かった。また中には多くの人同意可能と思われるコメント、例えば感謝や感動などを伝えるコメントも高評価コメントに多かった。

「笑」を含む低評価コメントの中では、「めんどくせーな」などの伸ばし棒を使用したコメントが多く全体的に軽い印象を受けた。また一方、中には周囲が簡単には納得しがたいようなコメント(例えば予言など)や、他者への同意を求めたり他のコメントに言及し不快感を与えるコメントも低評価コメントに多かった。

以上より、評価を得るコメントには他人の視線を考え不快感を与えないコメントである必要があると思われる。その中で「本当に」や「今」などといった適切に自分の意見や考えを強調したコメントが高評価を得やすかったと思われる。自分の考えがしっかりと示されているか、他者に理解が得られやすいものなのかを投稿する前にしっかり考えることが重要と思われる。

第7章 今後の課題

今後の課題として、まずは高中低の選び方をもう少し厳密にやる必要があると思う。高中低の選ぶ方法として、本実験では式7.1を使用した。

$$\text{数値評価} = \frac{\text{GOODが押された数} - \text{BADが押された数}}{\text{時間評価}} \quad (7.1)$$

式7.1で時間評価とは、「何時間前にコメントが投稿されたのか」を示し、「 n 時間前」の時に $n + 1$ として処理をしていた。

ここで例えば n 時間で「GOODが押された数 - BADが押された数」が $6n$ であるモデルを考える。「1時間前」の時は「GOODが押された数 - BADが押された数」が6であるため数値評価は $6 \div (1 + 1) = 3$ で、この時総合評価は「中」である。一方、「5時間前」の時には数値評価は $(6 \cdot 5) \div (5 + 1) = 30 \div 6 = 5$ で、この時の総合評価は「高」である。

この例は本来ならば、ともに1時間当たり6ずつ評価があがっているので総合評価が統一されていなければならない。しかし、現在の方法ではコメントが投稿された時間が遅いほど総合評価が高になりやすいのでその点を注意する必要があると思う。

また、本研究では入力データのサンプル数が少なく、「皇室」や「オリンピック」などのような記事に固有の語が取り出されてしまうことが多かった。そのため、よりサンプル数を増やしてより厳密な結果を求める必要があると思う。

また、本研究では主に動詞や形容詞などの自立語を中心として分析をしていたため、「が」や「と」、「ます」といった付属語が排除されていた。そのため付属語を含めた分析についても検討する必要があると思う。

第8章 おわりに

本研究では、Yahoo ニュースのコメントを KH Coder で分析することによって評価を受けやすいコメントの特徴を分析した。

5.2.1 節の対応分析の結果では、高評価コメントでは「皇室」や「結婚」などの単語が、低評価コメントでは「コメント」や「批判」といった単語が多く出現していることが分かった。しかし、多くの単語は図の原点に集中しており分かりづらい結果となった。

5.2.2 節のクロス集計のソートの結果では、高評価コメントと低評価コメントにどのくらい偏って出現しているのかを明らかにするため「高の比率」を用いて数値的に明らかにした。結果、高評価コメントに多い語として「本当に」や「今回」など、低評価コメントに多い語として「笑」や「コメント」などが取り出せた。

5.2.3 節のコロケーション分析では、高評価コメントに偏っていた「本当に」と低評価コメントに偏っていた「笑」を元にその周囲にどのような語が出現しているかを見ることで、高評価コメントと低評価コメントの性質を明らかにしようとした。高評価コメントでは「ほしい」といった要望や意見を表した語や「残念」といった自分の感想を表すような語が多く見られた。低評価コメントでは「コメント」など他の人が出したコメントに対して言及をする語や「よね」などの自分のコメントに対して強く同意を求めるような語が多く見られた。

5.2.4 節の grep 分析や追加実験では、高評価コメントでは表 5.11 のコメント 10「望み通りもう無くしてしまっても良いのでは？」や表 5.12 のコメント 13「責任は重くないですか？」、5.20 のコメント 5「世界で助けるべき問題だと思う。」など問題提起をするようなコメントが多かったことが分かった。低評価コメントでは、表 5.16 のコメント 16「貰えない世帯がギャーギャー騒いでる」や表 5.24 のコメント 1「それは民主主義だからでしょ？」など記事が書かれた対象に対してはそれを擁護するようなコメントが多かった。

以上より、「見直しが必要ですよね」や「責任は重くないですか？」のような自分の考えをしっかりと示しながら他者に不快感を与えないコメントが高評価を得やすいのではないかという考察を得ることができた。

謝辞

最後に、この一年間研究を進めるにあたり、本研究のご指導をいただいた鳥取大学工学部電気情報系学科自然言語処理研究室の村田真樹教授，村上仁一准教授そして自然言語処理研究室の皆様へ深く感謝するとともに、心から御礼申し上げます。また、ご多忙の中被験者実験にご協力いただいた皆様にも深く感謝申し上げます。そして、参考にさせていただいた論文の著者の方々に対して深く感謝申し上げます。

参考文献

- [1] 末吉美喜. テキストマイニング入門 ExcelとKH Coderで分かるデータ分析. オーム社, 2019.
- [2] 岩森三千代, 百瀬健, 大久保みちる. KH Coderを活用した学生満足度調査における自由記述の解析. 新潟青陵大学短期大学部研究報告 第51号, pp. 33–41, 2021.
- [3] 小野拓哉, 石川直元, 安井裕, 佐倉宏. 訪問診療実習を通じて得た、医学生の学びの解析-KH Coderによるテキストマイニングから-. 東女医大誌 91(3), pp. 184–190, 2021.
- [4] 仁科光一. 保護者アンケートにおける「自由記述」の活用-学校評価におけるKH Coderによる分析の試み-. 日本教育学会大会研究発表要項 79巻, pp. 52–53, 2020.
- [5] 吉田伊武貴, 倉田陽平. 失敗談ツイートのテキスト分析による北海道における観光失敗リスクの把握. 観光情報学会第22回研究発表会講演論文集, pp. 25–28, 2021.
- [6] 安原有紗, 劉銘. 口コミの言語解析から見た日本人と外国人の浜離宮恩賜庭園に対する観光体験の相違. 環境情報科学論文集 35, pp. 227–232, 2021.
- [7] 渡邊憲二, 箕輪弘嗣. COVID-19におけるTwitterの利用傾向に関する探索的研究. 情報知識学会誌 31(2), pp. 204–210, 2021.
- [8] 樋口紘一. 言語研究の分野におけるKH Coder活用の可能性. 計量国語学 31巻号, pp. 36–45, 2017.
- [9] 久田祥平, 村山太一, 矢田俊太郎, 岩宮翔子, 荒牧英治. SNSコメントを用いたニュースメディアバイアスの分析. 人工知能学会全国大会論文集第35回, 1D2-OS-3a-04, pp. 1–4, 2021.
- [10] 内山香, 鈴木海渡, 田上翼, 塙一晃, 乾健太郎, 小宮篤史, 藤村厚夫, 町野明德, 楊井人文, 山下亮. ファクトチェックのための要検証記事探索の支援. 人工知能学会全国大会論文集第32回, 4Pin1-26, pp. 1–4, 2018.
- [11] 山下晃弘, 中村拓哉, 川村秀憲, 鈴木恵二. SNSにおける炎上リスク分析と対策システムの開発. 人工知能学会研究会資料第103回, SIG-KBS-B402-05, pp. 19–24, 2014.