言語モデルもワクワクする？ -言語モデルを用いたオノマトペの感情分析-

中林 祐太[1]　嶋田 和孝[2]

概要：オノマトペは，音・状態・情動・行為を象徴的に表現し，「ワクワク」「ドキドキ」などの形で感情を直感的に伝達する役割を持つ．本研究では，言語モデルによるオノマトペの感情推定への影響度の分析を行う．特に，言語モデルがオノマトペに内包された感情的な意味をどのように捉え，感情推定にどのように活用しているかを実験的に検証する．複数の学習データに基づく言語モデルの比較や表記の違い（カタカナとひらがな）についての分析を行う．感情推定モデルとしては BERT を利用し，Plutchik の基本 8 感情に分類するモデルを構築する．ファインチューニングの際に，訓練データにオノマトペを含むもしくは含まないなどの条件を加えて，3 つのモデルを構築する．この 3 つのモデルに対してオノマトペを含まない文，カタカナのオノマトペを含む文，ひらがなのオノマトペを含む文を入力し，出力分布を KL Divergence を用いて比較する．またトークン単位での感情ラベルへの影響度を調べるために SHAP 分析を用いる．実験結果より，ひらがなよりカタカナのほうが感情推定への影響が大きいことがわかった．一方で，それが訓練データの分布にも依存することが実験的に確かめられた．

キーワード：オノマトペ，感情分析，言語モデル，BERT

Emotion Analysis of Onomatopoeia Using Language Models

Abstract: Onomatopoeia symbolically represents sounds, states, emotions, and actions, playing a role in intuitively conveying emotions through expressions such as ”wakuwaku” and ”dokidoki.” This study analyzes the impact of onomatopoeia on emotion classification by language models. We experimentally investigate how a language model perceives the emotional meaning embedded in onomatopoeia and how the onomatopoeia is utilized it for emotion classification. We conduct comparisons among BERT models trained on different datasets and analyze the effect of different notations (katakana and hiragana). Each BERT model is fine-tuned, on the basis of Plutchik’s eight emotions. For these three models, we input sentences without onomatopoeia, sentences containing katakana onomatopoeia, and sentences containing hiragana onomatopoeia, and compare the output distributions using Kullback-Leibler (KL) divergence. Additionally, we employ SHAP analysis to examine the impact of each token on each emotion label. The experimental results show that katakana onomatopoeia has a greater impact on emotion than hiragana onomatopoeia. However, it was also experimentally confirmed that this effect depends on the distribution of the training data.

Keywords: Onomatopoeia, Emotion Analysis, Language Models, BERT

1. はじめに

オノマトペとは，音・状態・情動・行為などを象徴的に表現する語であり，「ワクワク」や「ドキドキ」などの例が挙げられる．これらは特定の感情を直感的に伝達し，コミュニケーションを円滑にする重要な役割を担っている [1]．オノマトペの使用は文脈や感情の種類に依存するものの，感情表現の一手段として広く用いられている．これまでの研

[1] 九州工業大学 情報工学部 知能情報工学科
Department of Artificial Intelligence, Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN
[2] 九州工業大学 大学院情報工学研究院 知能情報工学研究系
Department of Artificial Intelligence, Kyushu Institute of Technology 680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

究では，オノマトペと感情の関連性について多くの議論がなされ，様々な分析や実験が実施されてきた [1], [2], [3]．一方で，自然言語処理の分野では特に，言語モデルを活用した文章の感情推定の研究が進んでおり，その精度や応用範囲が広がっている [4], [5], [6]．これらの背景を踏まえ，本研究では，言語モデルを用いた感情推定においてオノマトペの影響を定量的に分析し，言語モデルがオノマトペに内包された感情をどのように捉え，感情推定にどのように活用しているのかを検証する．

本研究の主な目的は，オノマトペが感情推定に影響を与えるかどうかを明らかにすることである．具体的には，文章中にオノマトペを含む場合と含まない場合の感情推定結果を比較し，オノマトペの存在が感情予測の確率分布にどのような変化をもたらすのかを定量的に評価する．また，オノマトペが感情推定に与える影響の大きさを測るために，モデル内部の処理を解析し，オノマトペが感情推定の寄与度を算出する．これにより，感情推定におけるオノマトペの役割を明確化することを目指す．本研究の成果により，言語モデルがオノマトペをどの程度理解し，感情推定に活用しているのかを解明し，オノマトペの持つ情報が感情分析においてどのような意味を持つのかを示すことができる．

2. 関連研究

オノマトペは，感情の表現や伝達に重要な役割を果たすと考えられており，これまでにも多くの研究が行われてきた．井上ら [1] は，基本感情（喜び，怒り，悲しみ，恐れ，嫌悪，驚き）を表す顔面表情のイラストを用い，それらを表現するオノマトペを大学生に挙げてもらう調査を行った．その結果，特に喜びや悲しみは特定のオノマトペ（例：「にこにこ」「しくしく」）で表現される傾向が見られた．この研究は，特定の感情カテゴリーが特定のオノマトペ（例：「にこにこ」「しくしく」）と強く結びついていることを示しており，感情推定においてオノマトペが重要な手がかりとなる可能性を示唆している．増子ら [7] は，感情極性値に基づいた印象に曖昧さを含むオノマトペの可視化を行った．その結果，オノマトペの感情極性のバラツキを可視化することで，単純な低次元マッピングでは捉えにくい印象の曖昧さを明確にし，感覚的表現の理解を支援できることを示した．これは，オノマトペによって印象の曖昧さ（感情極性のバラツキ）が異なることを示した．

これまでの研究では，オノマトペが読者への印象や文章理解に及ぼす影響について広く注目されてきた．本研究では，従来の研究成果を踏まえ，感情推定への影響を分析する．具体的には，オノマトペの感情表現上の特性に加え，文章理解における認知的・情緒的効果に着目し，最新の言語モデルを用いた感情推定タスクにおいて，オノマトペが推定プロセスに与える影響を定量的に評価することを目的とする．

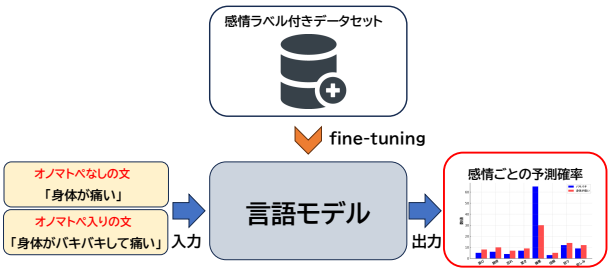


図 1 本論文での処理の概略．

3. オノマトペの影響の分析

本節では，オノマトペが言語モデルの感情推定に与える影響を分析し，その特徴を明らかにすることを目的とする．具体的な分析手法として，文章中にオノマトペを含む場合と含まない場合の感情推定結果を比較し，オノマトペの存在が感情予測の確率分布にどのような変化をもたらすのかを定量的に評価する．また，オノマトペが感情推定に与える影響の大きさを測るために，モデル内部の処理を解析し，オノマトペが感情推定の寄与度を算出する．

3.1 分析手法

分析における感情推定のプロセスを図 1 に示す．本手法では，まず感情ラベル付きデータセットを用いて言語モデルを学習し，感情推定モデルを構築する．構築した感情推定モデルは，入力された文章に対して感情推定を行い，Plutchik の基本 8 感情[*1]に対応する各感情の予測確率を算出し，出力する．本分析では，オノマトペを含む文，含まない文の大きく 2 パターンを入力文とする．入力により得られた 8 感情の予測確率を確率分布とみなし，オノマトペの有無による文章間の結果を比較する．

本分析では，オノマトペの影響度を測るために，以下の 3 つの観点に着目する．まず，3.1.1 節で，オノマトペの有無が感情推定結果にどのような違いをもたらすかを評価するための指標を定義する．次に，3.1.2 節で，感情推定モデルがオノマトペをどの程度重視しているかを分析し，単語レベルでの影響度を測る指標を提示する．最後に， 3.1.3 節で，感情ラベル付きデータセットの訓練データ内におけるオノマトペの有無が，感情推定モデルの出力にどのような違いをもたらすのかを評価するために，複数の訓練データの設定について述べる．

3.1.1 文章中のオノマトペの有無による感情推定への影響度指標

本分析では，オノマトペを含まない文章の感情推定結果を基準となる確率分布とし，オノマトペを含む文の感情推

[*1] Plutchik が提唱した，人間の感情は「喜び・悲しみ・期待・驚き・怒り・恐れ・嫌悪・信頼」の 8 つの基本感情に分類されるというもの

定結果を比較対象の分布とする．これら二つの確率分布の差異を測定することで，オノマトペの付与が感情推定に与える影響を KL Divergence (Kullback-Leibler Divergence) を用いて定量的に評価する．

KL Divergence は，基準となる分布 (オノマトペを含まない文) と比較対象の分布 (オノマトペを含む文) の違いを測る指標で，この値が大きいほど 2 つの確率分布の違いが大きいことを意味する．つまり，この値が大きいほどオノマトペの影響度は大きいといえる．KL Divergence は以下の式で計算される．

$$D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{1}$$

ここで，$P(i)$ は対象を感情推定モデルに入力した結果得られる感情予測値の確率分布であり，$Q(i)$ は基準となるオノマトペを含まない文の確率分布である．

3.1.2 トークン単位の影響度指標

オノマトペの感情推定モデルにおけるトークン単位の影響度を計る指標として，SHAP 分析（SHapley Additive exPlanations）[8] を用いる．SHAP 分析は機械学習モデルの予測結果に対して，各特徴（入力データの要素）がどの程度影響を与えたかをトークン単位で定量的に説明する手法である．本研究では，感情推定モデルに対して，SHAP 分析を用いることで，予測結果に対する各単語（トークン）の影響度を数値化する．

3.1.3 訓練データ中のオノマトペの有無による複数感情推定モデルの構築

感情ラベル付きデータセットの訓練データ内のオノマトペの有無が，感情推定モデルの出力にどのような違いをもたらすかを評価し，オノマトペにの影響度を測るために様々な訓練データの設定で分析を行う．そこで本研究では，以下の 3 つの感情推定モデルを構築する．

- **BaseModel（訓練データ中のオノマトペを除外してファインチューニング）**訓練データにオノマトペが存在しない場合の感情推定モデル[*2]である．このモデルを用いて 8 感情の感情値を推定する．これは基盤モデルであり，モデルの出力はオノマトペに関してファインチューニングしていない感情推定モデルが，どのようにオノマトペを捉えているかの指標になる．
- **FullModel（訓練データすべてを用いてファインチューニング）**オノマトペを含む訓練データによってファインチューニングしたモデルである．ファインチューニングによってオノマトペと 8 つの感情値との関係を学習していることが期待されている．
- **FilteredModel（分析対象のオノマトペを訓練データから除外してファインチューニング）**訓練データから後述の比較分析の際に利用するオノマトペを除外し，ファインチューニングしたモデルである．分析の際に利用するオノマトペについて直接学習していないモデル[*3]であり，FullModel との比較対象となる．

[*2] ただし，事前学習の際に学習している可能性は十分にある．

表 1 各モデルのデータセットの詳細.

モデル名	訓練データ	検証データ
BaseModel	15,476 文	1,042 文
FullModel	16,919 文	1,121 文
FilteredModel	16,919 文	1,121 文

表 2 ファインチューニングの設定.

損失関数	CrossEntropy
最適化関数	AdamW
学習率	5e-5
バッチサイズ	8
エポック数	3

3.2 分析設定

本節では，本研究における分析の設定について説明する．具体的には利用するデータセット，使用するモデル，分析の方法について順に説明する．

本研究では，日本語の文章に対する感情がアノテーションされたデータセットとして，Suzuki ら [9] が公開している「WRIME: 主観と客観の感情分析データセット」を使用する．このデータセットには，日本語の SNS テキスト全 43,200 文に対して，Plutchik の基本 8 感情に基づく感情強度（0〜3）が 3 人のアノテータによって付与されている．本研究では，3 人のアノテータの感情強度の平均の値を使用する．本論文では，以下「WRIME データセット」と呼ぶ．

感情推定モデルには，東北大学が公開している BERT モデル[*4]を用いる．この BERT モデルを 3.1.3 節の指針に基づきファインチューニングして使用する．この指針に基づき訓練データを選別するため，3 つのモデルがファインチューニングに利用するデータの母数が異なる．なお，本論文では，WRIME データセットのうち，一文中の最大感情強度が 2 以上の文章を用いている．訓練データと検証データの統計値を表 1 に示す．また，本分析で構築する 3 つモデルの設定を表 2 に示す．

すべてのオノマトペを分析の対象にすることは困難であるため，本論文では，分析対象を選定する．まず，井上ら [1] の研究と同様に，XYXY 型[*5]のオノマトペを対象とする．井上ら [1] は，大学生 45 名を対象に感情ごとのオノマトペ使用についてアンケート調査を実施した．本論文では，その結果をもとに WRIME データセットに含まれるも

[*3] BaseModel と同様に，事前学習の際に学習している可能性は十分にある．
[*4] https://huggingface.co/tohoku-nlp/bert-base-japanese-v2
[*5] 同じ 2 文字（X と Y）の組み合わせを繰り返す形

表 3 使用するオノマトペと WRIME データセット中で使用されている文の数.

表現	合計数	カタカナ	ひらがな
にこにこ	25	22	3
いらいら	89	84	5
ぷんぷん	1	0	1
しくしく	1	0	1
むかむか	4	4	0
びくびく	8	8	0
どきどき	55	45	10
ぶるぶる	5	4	1
わくわく	72	49	23
がくがく	2	2	0
つんつん	2	1	1
にやにや	14	9	5
めそめそ	4	2	2
うきうき	16	13	3
そわそわ	16	6	10

ので 15 種の XYXY 型オノマトペを選定した．使用するオノマトペの一覧と，WRIME データセットにおける出現数を表 3 に示す．表中のカタカナとはそのオノマトペがカタカナ表記（たとえば，ニコニコ）で出現した回数を，ひらがなとはひらがな表記（にこにこ）で出現した回数を意味する．

本論文では，「帰ります。」という文をベースにし，それに表 3 のオノマトペを追加した文（「XYXY しながら帰ります。」）を用意し，ベース文との予測確率の分布の差を検証する．具体例を挙げると，「にこにこ」というオノマトペに対して，「帰ります。」「ニコニコしながら帰ります。」「にこにこしながら帰ります。」という 3 つの文が用意される．それぞれを 3.1.3 節で説明した 3 つのモデルに入力し，ベース文との KL Divergence や SHAP 値を計算することで，オノマトペの影響を調査する．

3.3 分析結果

前節で説明した設定に基づき，オノマトペの有無による差（3.3.1 節）や感情推定モデルがオノマトペをどの程度重視しているのかを検証（3.3.2）を行う．

3.3.1 オノマトペの有無による感情推定への影響分析

元の文「帰ります。」の感情予測の確率分布と前節で作成した文との感情予測の確率分布の違いを KL Divergence で測る．図 2 に KL Divergence の結果を示す．図 2 では色が濃いほど，KL Divergence の値が大きいことを意味する．たとえば，図 2 の最上部にある「にこにこ」を例に取る．一番左の 0.066 は BaseModel が「帰ります」という文について 8 感情の予測をした場合の確率分布と「にこにこしながら帰ります」という文の確率分布との KL Divergence の値である．KL Divergence が大きいということは，2 つの文に対する 8 つの感情の確率分布の傾向が異なることを意味しており，これはオノマトペの有無の影響が大きいことを意味する．このケースでは右の FilteredModel は「にこにこ」というオノマトペの存在が感情推定に強い影響を与えていることを意味している．

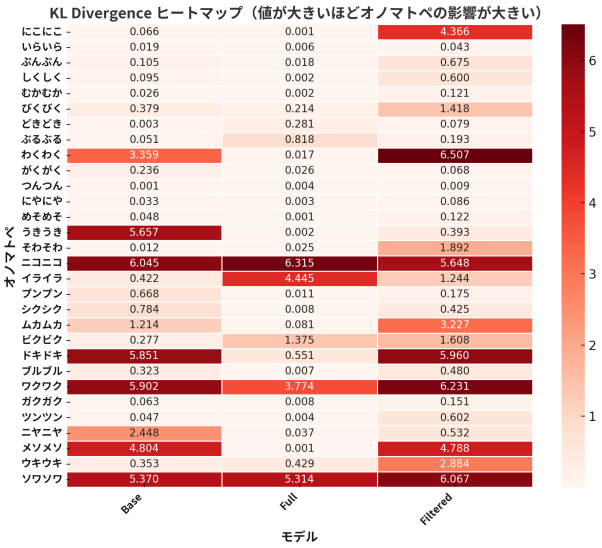


図 2 元の文章との確率分布の違い.

図 2 の結果を分析する．まず，BaseModel の結果を見ると，オノマトペを学習していない BaseModel においても，「うきうき」，「ニコニコ」，「ドキドキ」，「ワクワク」，「メソメソ」，「ソワソワ」といった一部のオノマトペで KL Divergence が大きい値を示しており，感情予測が変化することが確認された．一方，FullModel の分析結果では，「ニコニコ」「イライラ」「ワクワク」「ソワソワ」などの一部のカタカナ表記のオノマトペを除き，KL Divergence の値は全体的に小さく，オノマトペの影響度が小さいことを表す．FilteredModel では，BaseModel や FullModel と比較して，KL Divergence の値が大きくなり（右側の列が濃い色が多い），複数のオノマトペで影響度が大きいことがわかる．全体的な傾向として，ひらがな表記のオノマトペよりもカタカナ表記のオノマトペの方が，確率分布の変化が大きい傾向がある（上半分よりも下半分に濃い色が多い）こともわかる．図 2 を見ると，訓練データ中のすべてのオノマトペを学習した FullModel の影響度合いが全体的に低い（濃い部分が少ない）が，「ニコニコ」，「イライラ」，「ワクワク」といったオノマトペについては影響が特に顕著であった．これは，訓練データ中の事例数に大きく影響を受けたと考えられる．表 3 に示したようにこれらのオノマトペは訓練データ中の出現頻度が高い．すなわち，訓練データの分布がモデルの予測に大きな影響を与えていることがわかる．

次に，事例分析として，KL Divergence の値が大きい「わくわく」と「ワクワク」に絞って，詳しい分析を行う．まず，比較のために，ベース文である「帰ります。」の確率分布を図 3 に示す．図からわかるように，このベース文は

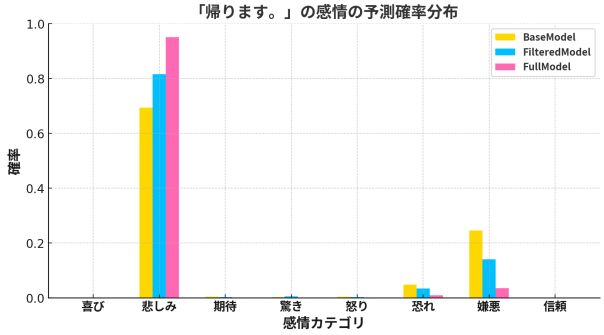
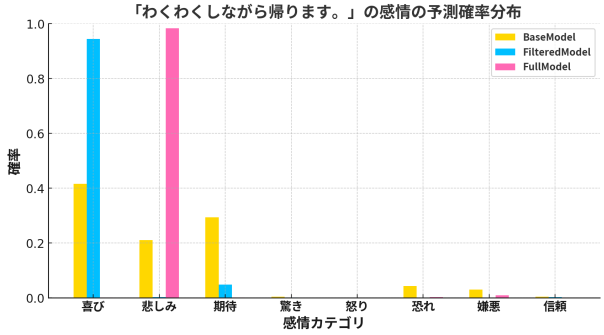
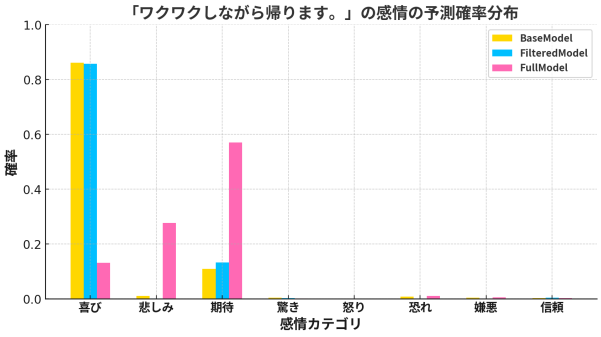


図 3 「帰ります。」の感情の予測確率分布.



(a) わくわく（ひらがな）の確率分布



(b) ワクワク（カタカナ）の確率分布

図 4 「わくわく/ワクワク」の感情予測の確率分布.

「悲しみ」の確率が極端に大きい．図 4(a) と図 4(b) にオノマトペを含む文に対する 3 つのモデルの出力結果を示す．どの結果も元の文「帰ります。」（図 3）と異なる分布であることがわかる．

次に，図 4(a) をみると，訓練データに直接的にそのオノマトペが含まれていない BaseModel（黄）と FilteredModel（青）において「喜び」の予測確率が高くなっていることが確認された．図 4(b) においても同じ傾向が見られる．一方，「わくわく」と「ワクワク」の両方を学習している FullModel（ピンク）では，「わくわく」の感情推定結果が「帰ります.」の文と大きく変わらない傾向を示した．

3.3.2 オノマトペの感情推定モデルにおけるトークン単位の影響度の分析

オノマトペが感情推定に与えるトークン単位での影響を定量的に評価するために，SHAP 分析を用いた解析を行う．本分析では，3.3.1 節の結果に基づき，オノマトペの影響が顕著であった「わくわく」および「ワクワク」に着目し，その影響の詳細を検討する．具体的には，感情予測の確率が高くなる傾向を示した「喜び」および「期待」の感情カテゴリを対象に，SHAP 値を用いた分析を行う．なお，ある感情ラベルの予測において各トークンの SHAP 値が正の大きな値になることは，そのトークンがその感情ラベルを予測するのに大きな役割を担っていること意味する．負の値はその逆である．以降の図では青が正の値，赤が負の値を意味する．

図 5 と図 6 にそれぞれひらがなとカタカナの場合の SHAP 値を示す．まず，「わくわく」の予測結果について分析する．どちらにおいても，「喜び」の予測確率が最も高かった BaseModel，FilteredModel ではオノマトペ部分の SHAP 値が大きくなっており，予測確率を大幅に上昇させる要因となっていることがわかる．また，「期待」の予測確率が最も高かった FullModel についても，「期待」の結果を確認すると，特にカタカナの場合に予測確率の上昇に寄与していることが分かる．これらの結果から，オノマトペが感情推定において直接影響を与えていることがわかる．

最後に，モデルの観点でまとめると FullModel は訓練データの分布に大きな影響を受けてしまう．一方で，BaseModel は 8 感情へのファインチューニングの差異にオノマトペを訓練データに含んでいないため，オノマトペの影響は最小限である．FilteredModel は分析対象に関するファインチューニングはしておらず．FullModel のような訓練データからの影響を受けず，オノマトペの情報を学習している．BaseModel と FilteredModel によるさらなる深い分析が，オノマトペの解釈や感情との関係性を調べる上で重要だと考えられる．

また，表記の違いによる影響も顕著であった．図 2 において，カタカナ表記のオノマトペとひらがな表記のオノマトペで比較すると，感情推定モデルに与える影響が大きく，確率分布の変化がより顕著であることが確認された．この結果は，カタカナ表記がより強調された印象を与え，感情の喚起に強く寄与する可能性を示唆している．一方で，この違いがモデルの訓練データに起因する可能性も考えられ，表記の差異が感情推定に及ぼす影響について，より詳細な検証が必要である．

4. カタカナ表記の影響要因の検証

前節の図 2 とそれに関連する考察で，表記の違いに着目すると，カタカナ表記のほうが感情変化に大きな影響を与えていることがわかる（下半分の方が色が濃い）．本節では，これが，何に起因するのかを実験的に検証する．具体的にはカタカナとひらがなの予測分布の差を比較する．また，WRIME データセットでのカタカナの使用傾向を調査し，新たなモデルを作成することで，カタカナの影響度が

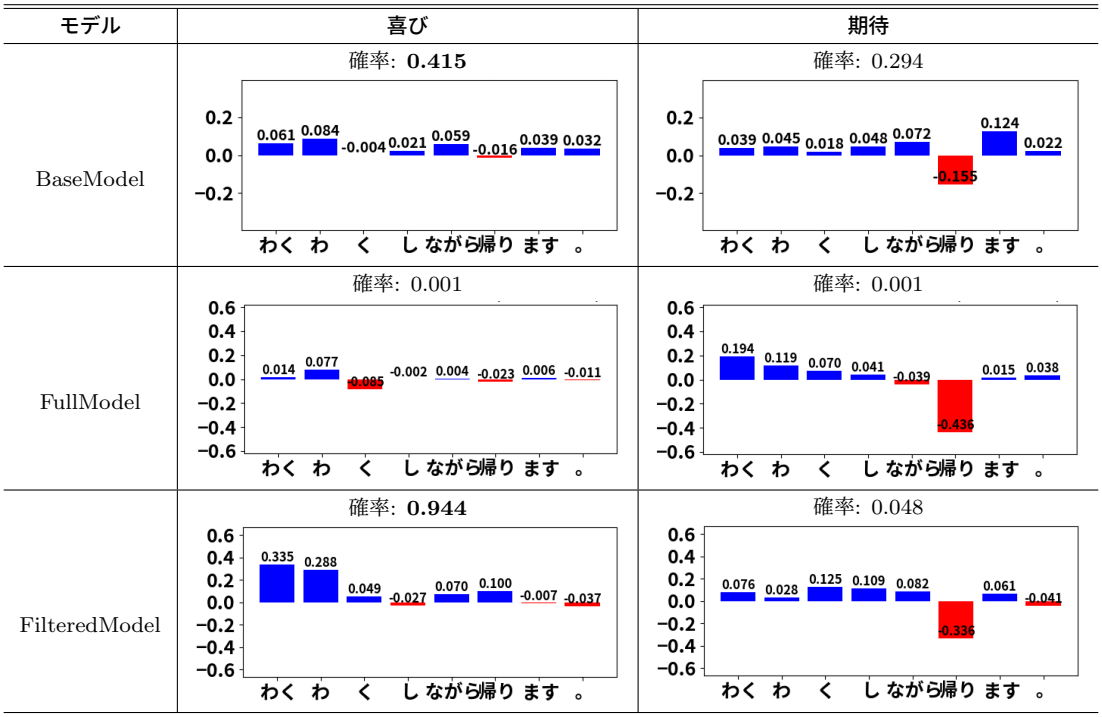


図 5 「わくわく」の SHAP 値の比較: 表内の確率はそれぞれの感情の予測確率．太字は予測確率が高いものを示す．

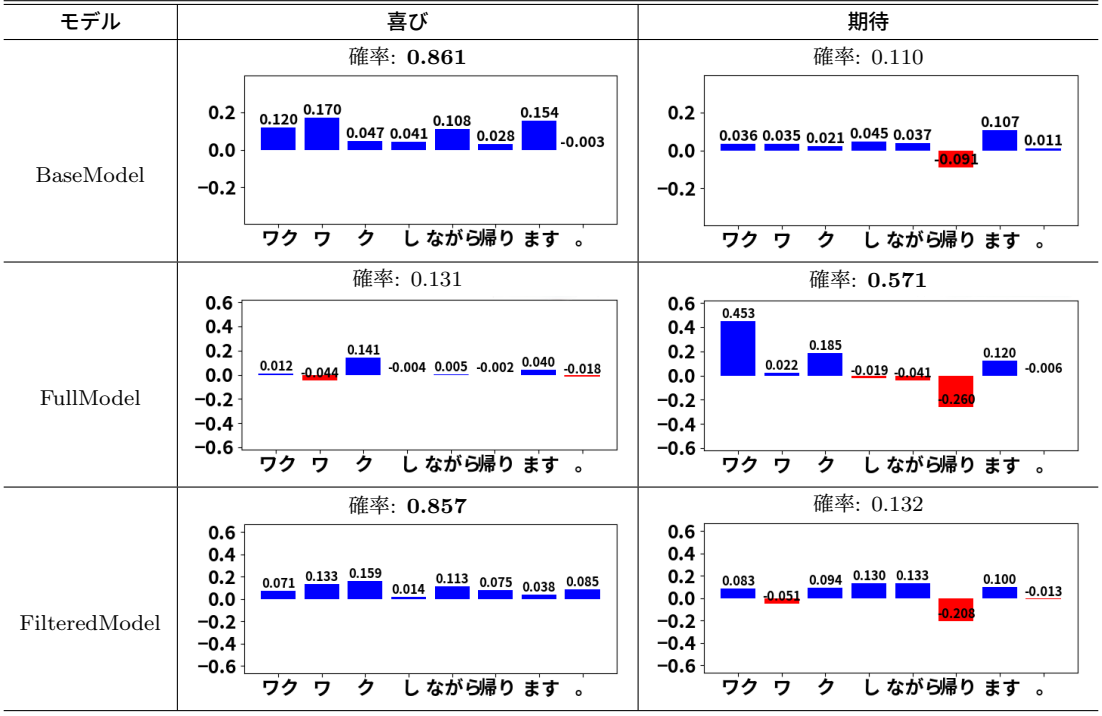


図 6 「ワクワク」の SHAP 値の比較: 表内の確率はそれぞれの感情の予測確率．太字は予測確率が高いものを示す．

大きくなる原因が訓練データの分布に由来するものであるのかを実験的に考察する．

カタカナ表記のオノマトペとひらがな表記のオノマトペによる感情予測の差異を定量的に評価するため，両者の予測確率の差分を算出しその平均を求め，比較する．図 7 にその結果を示す．この図は，赤色が濃くなるほどカタカナ表記のオノマトペの予測確率が，ひらがな表記よりも高いことを示している．青色はその逆を意味する．全体的な傾向として，「喜び」の感情がカタカナ表記のオノマトペで高くなる傾向が見られる．次に，「期待」の感情もカタカナ表記の方が高くなる傾向があることが分かる．一方で，「悲しみ」はひらがな表記の方が影響力を持っており，感情と字種に一定の関係があることが見てとれる．

つぎに，WRIME データセットおけるカタカナと各感情

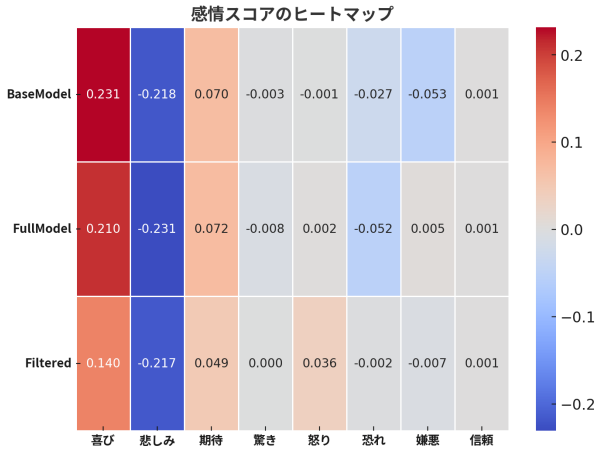


図 7 各モデルにおける「カタカナオノマトペの予測確率 − ひらがなオノマトペの予測確率」の平均値を示すヒートマップ．

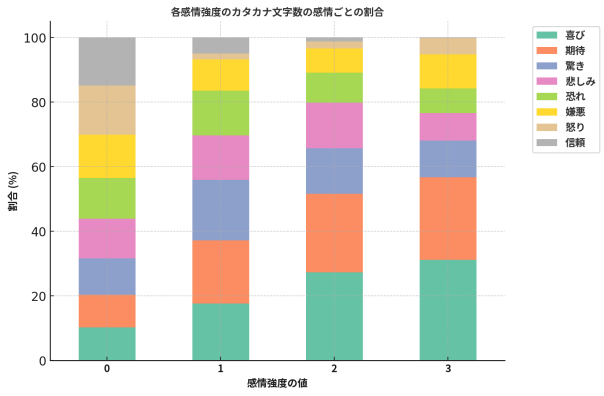


図 8 各感情強度のカタカナ文字数の割合．

および感情強度値の関係を見る．図 8 は，各感情強度におけるカタカナの文字の割合を示す．各感情強度は，WRIME データセットにアノテーションされた値であり，数値が大きい（最大値 3）ほどその感情の強度が高いことを示す．いずれの指標においても，感情強度が強くなるにつれて「喜び（緑）」と「期待（オレンジ）」の割合が大きくなる傾向が確認できる．一方で，「悲しみ（ピンク）」の感情は，感情強度が強くなるにつれて割合が小さくなる傾向が確認できる．これが，「カタカナ表記のオノマトペの方がひらがな表記のオノマトペよりも「喜び」や「期待」の感情が強くなり，「悲しみ」が小さくなる」という傾向の要因の一つであると考えられる．

これまでの分析により，感情推定モデルが学習データ中のカタカナ表記の分布に影響を受けている可能性が示唆される．しかしながら，訓練データの分布以外にも，カタカナ表記が持つ言語的な特徴に起因する可能性もある．この点を検証するために，オノマトペの表記を統一した訓練データによるファインチューニングモデルを考える．つまり，オノマトペの表記を統一することで，感情推定の確率分布が変化するかを検証する．

訓練データのオノマトペをすべてひらがな表記に統一

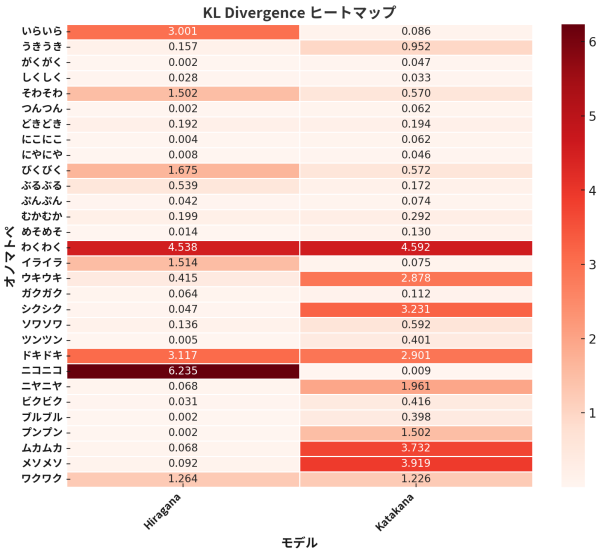


図 9 FullModel との統一表記モデルとの確率分布の違い．

したモデル（HiraganaModel）と，すべてカタカナ表記に統一したモデル（KatakanaModel）を作成する．作成した HiraganaModel および KatakanaModel の訓練データ数および検証データ数の設定は，FullModel と同一である．すなわち，FullModel の訓練データに含まれるオノマトペをすべてひらがな表記またはカタカナ表記に統一し，それぞれのデータを用いてモデルを訓練した．モデルの影響を比較するため，FullModel の感情推定結果を基準とし，HiraganaModel および KatakanaModel の感情推定確率分布との違いを定量的に評価する．具体的には，それぞれの確率分布との差分を測る指標として，KL Divergence を算出し，表記の違いが感情推定に与える影響を分析する．

図 9 に KL Divergence の結果をヒートマップで示す．結果より，KatakanaModel（右側）では，カタカナ表記のオノマトペに対して KL Divergence の値が大きくなっている．つまり，カタカナ表記に表記を統一することで，カタカナのオノマトペが感情推定に強く作用していることが確認された．一方，HiraganaModel においては，ひらがなへの影響は限定的であった．これらの結果から，感情推定モデルの出力に対するカタカナ表記の影響は，訓練データの分布に由来する側面が大きいことが示された．特に，WRIME データセットにおけるカタカナ表記の割合と感情強度との間にも関連性があり（図 8），訓練データ中の表記の偏りがモデルの予測に影響を与えていると考えられる．

5. おわりに

本研究では，言語モデルがオノマトペに内包された感情をどのように扱っているかを実験的に検証した．複数のモデルにより，オノマトペの有無によって Plutchik の基本 8 感情の分布がどのように変わるかを分析した．分布の差異については KL Divergence を用いた．また，トークン単位

の感情への影響度を確認するために SHAP 分析を用いた．

まず，オノマトペの有無が感情推定に与える影響について，BERT ベースの感情推定モデルを用いて分析を行った．その結果，オノマトペを含む文では，オノマトペを含んでいないと比較して，感情推定の確率分布が大きく変化することが確認された．特に，「わくわく」「ワクワク」などのオノマトペは，「喜び」や「期待」の感情の予測確率を顕著に上昇させる傾向を示した．さらに，SHAP 分析により，感情予測確率が高いカテゴリにおいて，オノマトペがトークン単位で，感情予測の予測確率を上昇させる方向に寄与していることが確認された．

次に，オノマトペの表記の違いが感情推定に与える影響を分析した．その結果，カタカナ表記のオノマトペの方が，ひらがな表記と比較して感情推定への影響が大きいことが明らかとなった．この現象について，訓練データの分布が感情推定に影響を与えていると仮定し，訓練データ内のオノマトペの表記を統一したモデル（HiraganaModel および KatakanaModel）を作成し，感情推定結果を比較した．その結果，KatakanaModel ではカタカナ表記のオノマトペに対する影響が顕著に強まり，カタカナ表記の訓練データがモデルの感情推定に強く作用していることが確認された．一方，HiraganaModel では，ひらがな表記の統一が感情推定結果に与える影響は限定的であり，モデルは本質的にカタカナ表記の影響を強く受けて学習している可能性が示された．

本研究の結果を総括すると，オノマトペは感情推定モデルの出力に大きな影響を与える要素であり，特にカタカナ表記のオノマトペが「喜び」や「期待」の感情をより強く喚起する傾向が確認された．また，訓練データ中の表記の偏りが，感情推定結果に影響を及ぼす要因の一つであることが示された．一方で，オノマトペ単独の影響をより厳密に評価するためには，文脈や共起する単語の影響を考慮したさらなる分析が必要である．

本研究の題名である「言語モデルもワクワクする？」は，言語モデルがオノマトペをどのように処理し，感情推定に活用しているのかという疑問に基づくものである．本研究の結果より，言語モデルは「わくわく」「ワクワク」といったオノマトペを認識し，特定の感情（「喜び」「期待」）の予測確率を上昇させる傾向があることが示された．これにより，言語モデルはオノマトペに内在する感情的意味を踏まえて，感情推定を行っていることが明らかになった．以上より，言語モデルはある意味で『ワクワクする』と言えるだろう．しかし，その影響はモデルの訓練データに大きく依存しており，特にカタカナ表記のオノマトペがより強く影響することが分かった．

オノマトペの表記の違いや文脈依存性に関するさらなる分析を進めることが今後の課題である．その分析を通じて言語モデルにおけるオノマトペの処理メカニズムをより詳細に解明することを目指す．また，本研究では XYXY 型のオノマトペを中心に分析を行ったが，他の形態のオノマトペについても同様の検証を行い，より包括的な知見を得ることが求められる．さらに，本研究では BERT を基盤とした感情推定モデルを用いたが，GPT 系モデルや RoBERTa などの異なる言語モデルを用いた場合の比較を行うことで，オノマトペの処理方法の一般性を評価することも今後の課題として挙げられる．

参考文献

[1] 井上弥, 野中陽一朗. オノマトペは基本感情を表現することばとして有効か ―顔面表情刺激を用いた探索的検討 ―. 学習開発学研究, Vol. 8, pp. 37–42, 2015.

[2] 清水祐一郎, 土斐崎龍一, 坂本真樹. オノマトペごとの微細な印象を推定するシステム. 人工知能学会論文誌, Vol. 29, No. 1, pp. 41–52, 2014.

[3] Tatsuki Kagitani, Mao Goto, Junji Watanbe, and Maki Sakamoto. Sound symbolic relationship between onomatopoeia and emotional evaluations in taste. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 36, 2014.

[4] Luna De Bruyne, Pranaydeep Singh, Orphée De Clercq, Els Lefever, and Véronique Hoste. How language-dependent is emotion detection? evidence from multilingual bert. In *Proceedings of the 2nd Workshop on Multilingual Representation Learning (MRL)*, pp. 76–85, 2022.

[5] Alexandra Ciobotaru and Liviu P. Dinu. Red: A novel dataset for romanian emotion detection from tweets. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, pp. 291–300, 2021.

[6] Wenbiao Yin and Lin Shang. Efficient nearest neighbor emotion classification with bert-whitening. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4738–4745, 2022.

[7] 増子由起, 齊藤史哲, 石津昌平. 感情極性値に基づいた印象に曖昧さを含むオノマトペの可視化: 自己組織化マップによる文書データの分析. 知能と情報, Vol. 28, No. 3, pp. 685–691, 2016.

[8] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, Vol. 30, pp. 4768–4777, 2017.

[9] Haruya Suzuki, Yuto Miyauchi, Kazuki Akiyama, Tomoyuki Kajiwara, Takashi Ninomiya, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. A japanese dataset for subjective and objective sentiment polarity classification in micro blog domain. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 7022–7028, 2022.