

# ラップバトルにおけるアンサーの類型化および類型の自動分類

三林 亮太<sup>†</sup> 佃 洗撰<sup>††</sup> 渡邊 研斗<sup>††</sup> 中野 倫靖<sup>††</sup> 後藤 真孝<sup>††</sup>

山本 岳洋<sup>†</sup> 大島 裕明<sup>†</sup>

<sup>†</sup> 兵庫県立大学 情報科学研究科 〒651-2197 兵庫県神戸市西区学園西町 8-2-1

<sup>††</sup> 産業技術総合研究所 〒305-8568 茨城県つくば市梅園 1-1-1

E-mail: <sup>†</sup>af22h007@guh.u-hyogo.ac.jp, <sup>††</sup>{k.tsukuda,kento.watanabe,t.nakano,m.goto}@aist.go.jp,

<sup>†††</sup>t.yamamoto@gsis.u-hyogo.ac.jp, <sup>††††</sup>ohshima@ai.u-hyogo.ac.jp

**あらまし** ラップバトルとは、2名のラッパーが即興のラップを交互におこなう競技で、その勝敗を判定するための評価指標に、相手のラップへの返答であるアンサーの質がある。アンサーの質は判別が難しく、どのような場合にアンサーとして成立するのか、どのような種類のアンサーがあるのかは明らかでなく、勝敗を判定することは容易ではない。アンサーの類型化とその自動分類が可能になれば、どのようなアンサーが即興ラップに含まれているのかを明確にすることができ、勝敗の判定補助が可能となる。そこで本研究では、ラップバトルにおけるアンサーの類型化および類型の自動分類に取り組む。まず、ラップバトルの動画を文字起こししたコーパスから、アンサーとして成立するラップ文のペアを429件作成した。それらに対して、3年間以上ラップバトルの視聴経験があるアノテータ2名によるカテゴリの作成とアノテーションをおこなうことで、アンサーを類型化した。結果として、アンサーかアンサーでないかを判別する3種類のメインカテゴリと、アンサーがどのような表現かを判別する1種類のメインカテゴリの合計4種類のメインカテゴリから成る類型を作成した。さらに、アンサーの類型をより詳細に分類可能にするため、4種類のメインカテゴリに対して合計24種類のサブカテゴリの類型を作成した。作成した類型の内、2種類のメインカテゴリにおいて、アノテーションにおけるKrippendorffの $\alpha$ が0.5を超える一致率を示した。アンサーの自動分類では、BERTを用いた自動分類に取り組んだ。相手のラップ文に対する返答のラップ文が、作成した3つの類型の内の少なくとも1つに該当するか否かを判別する分類タスクに取り組み、正解率が0.8554と高い分類結果を示した。さらに、アンサーがどのような表現かを判別する分類タスクにも取り組み、F値が0.4641と中程度の分類性能を示した。

**キーワード** ラップバトル, アンサー, 類型化, 自動分類

## 1 はじめに

ラップバトルとは、2名のラッパーが即興のラップを交互におこなう競技である。ラッパーが自分の番で発言する即興のラップをバースと呼び、バースは複数のラップ文で構成される。図1にラップバトルの例を示す。ラッパーAの「まずはリズムに乗せる小節/俺のライムは特別」というバースに対して、ラッパーBは「一体お前のどこが特別/ラップ上手い順ならお前は後列」というバースを返答する。このようなやり取りを複数回繰り返すことでラップバトルは終了する。一般的に、ラップバトルの勝敗は、観客による多数決によって決められる。

ラップバトルでは、バースの質が勝敗の判断基準となる。バースの質は主に、ライム、アンサー、フローの3つを対象に評価される[1],[2]。ライムは母音列が一致する単語のペアを指し、そのような単語のペアをバースに含むことができていたかを評価する。たとえば、図1のバース2の「特別(TOKUBETSU)」と「後列(KOURETSU)」は「OUEU」という母音が一致するためライムである。次に、アンサーは相手のバースを理解し、その内容に沿った返答をする行為を指し、そのような返答をバースに含むことができていたかを評価する。たとえば、図1

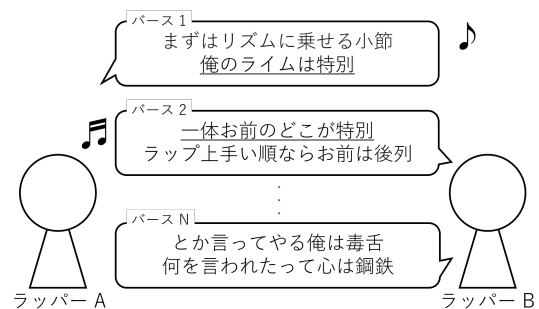


図1 ラップバトルの例

のバース1内の「俺のライムは特別」に対して、バース2内の「一体お前のどこが特別」は、相手の「特別」というバースの内容に沿ったアンサーである。最後に、フローはラップの歌い方のことを指し、バースをどれだけリズム良く歌うことができたかを評価する。

これら3つの中でも特に、アンサーの質はラップバトルにおいて重要な評価指標である。即興でおこなわれるラップバトルでは、その場で即興のバースを考えたことを証明する必要がある。つまり、事前にクオリティの高いバースを準備していないことを示す必要がある。アンサーは、相手の内容に沿ったバースを事前に準備できないため、即興で考えた証明としても利用

される。また、ライムとフローは一般的なラップの楽曲においても評価対象として用いられるのに対して、アンサーは対話形式のラップバトルに特化した評価対象である点でも重要であると言える。

しかし、アンサーの質は判別が難しい評価指標であり、どのようなパースの掛け合いがアンサーとして成立するかは、これまでに分析されていない。さらに、アンサーが成立した場合においても、そのアンサーがどのような種類の表現であるかは、基準が明らかになっていない。そのため、観客にとってアンサーを評価することは容易ではなく、ラップバトルの勝敗を公平に決めることは難しい。もし、パース内にアンサーが含まれているかを判別することができ、それがどのような表現であるかを自動的に提示できれば、観客は勝敗の判定が容易になる。そこで本研究では、ラップバトルにおけるアンサーの類型化をおこない、その類型を基にアンサーの自動分類に取り組む。

具体的には、まずラップバトルの動画を文字起こししたコーパスからアンサーのペアを429件作成し、3年間以上ラップバトルの視聴経験があるアノテータ2名によるカテゴリの作成とアノテーションによってアンサーを類型化する。本研究では、「単語」「ライム」「対話」の各観点でアンサーに該当するか否かを判別するための3種類のメインカテゴリと、アンサーの表現の種類を判別するための1種類のメインカテゴリの合計4種類のメインカテゴリから成る類型を作成した。

さらに、アンサーの類型をより詳細に分類可能にするため、4種類のメインカテゴリに対して合計24種類のサブカテゴリの類型を作成した。作成したメインカテゴリとサブカテゴリの妥当性を検証するため、クラウドソーシングを用いたラベル付けタスクをおこなった。その結果、4つのメインカテゴリの内2つのメインカテゴリでKrippendorffの $\alpha$ が0.5を超える一致率を、残りの2つのメインカテゴリで0.2を超える一致率を示した。また、サブカテゴリについても、Krippendorffの $\alpha$ を調べた結果、24種類のサブカテゴリの内4種類で0.5を超える一致率を、16種類で0.2を超える一致率を得られた。

次に、作成した類型を基に、BERT [3] を用いたアンサーの自動分類をおこなった。その結果、「単語」「ライム」「対話」の3種類のメインカテゴリの少なくとも1つでアンサーに該当するか否かを分類する2値分類タスクにおいて正解率は0.855となり、高い性能でアンサーの自動分類が可能であることを示した。さらに、アンサーがどのような表現かを判別する分類タスクにも取り組み、F値が0.4641と中程度の分類性能を得た。

本研究の貢献を以下に示す。

- ラップバトルコーパスから抽出した429件のアンサーペアに基づいて、4種類のメインカテゴリと24種類のサブカテゴリから成る、ラップバトルにおけるアンサーの類型を作成した。
- 作成したアンサー類型を基に、クラウドソーシングを用いたアンサータイプのラベル付けタスクを実施し、メインカテゴリとサブカテゴリのいずれにおいても作成した類型に一定の妥当性があることを示した。
- BERTを用いたアンサーの自動分類に取り組み、正解率が0.855という高い分類性能が得られることを示した。

## 2 関連研究

類型化に関する研究は多くおこなわれており、これまでに、類型化の方法、対話に関する類型化、コーパスに対する時間表現の類型化、有害表現の類型化などが取り組まれてきた。

### 2.1 類型化の方法

類型化の代表的な方法として、MAMA(Model-Annotate-Model-Annotate) サイクルがある。MAMA サイクル [4] は Pustejovsky らが提案したアノテーション方法で、基準策定、アノテーション作業、評価、修正作業を繰り返すことで、最適な類型を作成する。類似する手法として、Pustejovsky らは MATTER サイクル [4] を提案している。MATTER サイクルは、構造学習器を学習するステップを MAMA サイクルに追加した手法である。ただし、アノテーションの方針が決まっていない初期の状態では MAMA サイクルをおこなうことが多く、後述する類型化の関連研究においても、MATTER サイクルではなく、MAMA サイクルが採用されている [4] [5]。

### 2.2 対話に関する類型化

対話に関する類型化の研究は多く取り組まれている。Trippas ら [6] は、音声対話に関する類型化をおこない、「タスク」「会話」「その他」の3つのカテゴリとそれらに対するサブカテゴリ（「タスク」のサブカテゴリとしては「情報要求」や「検索支援」など）を定義している。類型は Cohen の  $k$  係数 [7] を用いて、サブカテゴリレベルで評価され0.71の高い一致率を示した。

Higashinaka ら [8] は、対話破綻の誤り分類に取り組んでいる。1,146件のデータを作成し、その中からランダムに選択した100件の対話データに対して24名がアノテーションをおこない、そこで付与したコメントを基に対話破綻の誤りを分類した。「文法エラー」「解釈不能」「発話無視」などの、17のメインカテゴリからなる類型を作成し、それらを「破綻」「破綻の可能性がある」「破綻でない」の3クラスと、「破綻」「破綻でない」の2クラスに分け、評価をおこなっている。評価では、Fleiss の  $k$  係数 [9] が3クラスの分類において0.276の一致率を示し、2クラスの分類では0.396の一致率を示している。

また、このようなデータに基づく類型化とは異なる類型化のアプローチとして、Higashinaka ら [10] は、対話理論から逸脱した対話を対話破綻とみなし、類型を作成している。結果として、「発話」「応答」「文脈」などの、16のメインカテゴリと、それらに対するサブカテゴリ（「発話」のサブカテゴリとしては「構文的誤り」や「意味的誤り」など）を定義している。作成した類型を基に、1,046件の対話に対して、アノテーションをおこなっている。評価では、Fleiss の  $k$  係数 [9] がメインカテゴリでは0.400、サブカテゴリでは0.239の一致率を示している。

さらに、Higashinaka らはこれら2つの類型を統合した新たな類型を提案している [11]。この類型では、Fleiss の  $k$  係数 [9] において、専門作業によるアノテーションでは一致率が0.567、クラウドワーカによるアノテーションでは一致率が0.488となって、いずれも高かった。

### 2.3 コーパスに対する時間表現の類型化

保田ら [5] は、現代日本語書き言葉均衡コーパスを対象に、時間的順序関係の解析を補助するための類型を提案している。類型の作成には、MAMA サイクルを用いている。類型は、Cohen の  $k$  係数 [7] により評価し、0.653 の高い一致率を示した。また、MAMA サイクルでは一般的に、アノテーション一致率という指標を良くする方向に最適化するが、言語学的には多様なラベルも必要であり、必ずしも一致率を最適化することが適しているわけではないと保田らは主張している。

坂口ら [12] は、京都大学テキストコーパスを対象に、時間表現のアノテーションをおこなっている。類型は、Krippendorff の  $\alpha$  [13] により評価され、厳しい条件において 0.526、緩和した条件において 0.867 の一致率を示している。次に、作成した類型を基に、双方向 GRU モデル [14] を用いた自動分類に取り組んでいる。結果として、時間性判定タスクでは、0.905、事象の時間的長さ分類タスクでは 0.606、事象の発生時期分類タスクでは 0.496 の正解率を示している。

### 2.4 有害表現に関する類型化

有害表現を類型化する研究もおこなわれている。小林ら [15] は日本語の有害表現の類型化に取り組み、Twitter, ガルちゃん, 5ch といったプラットフォームから取得したコーパスを使用している。アノテーションは 2 種類おこない、有害レベルと有害カテゴリの 2 種類に対してラベルを付与した。有害レベルは 4 段階のスコアをつけており、有害カテゴリは MAMA サイクル [4] を用いて類型を作成している。結果として、「差別」「卑語」「迷惑行為」などの 7 つのメインカテゴリと、それらに対する 41 つのサブカテゴリ（「差別」のサブカテゴリとしては「人種差別」や「職業差別」など）を定義している。これらを、Krippendorff の  $\alpha$  [13] を用いて評価し、有害レベルにおいて、4 値分類では 0.40、2 値分類では 0.78 の一致率を示した。また、作成した類型を基に、BERT を用いた有害検知器を作成しており、回帰問題として解いている。結果として、いくつかの評価尺度に対して高い性能で有害検知ができています。

### 2.5 本研究の位置づけ

これまでに紹介した関連研究では、対話破綻、時間表現、有害表現を対象に類型を作成している。著者らの知る限り、ラップバトルにおけるアンサーの表現を対象にした類型はまだ作成されていない。これはラップバトルが比較的新しい競技であり、まだ十分に分析がおこなわれていないためだと考えられる。しかし、近年では、ラップバトルの言語学的側面からの分析がおこなわれており [1], [16], ラップバトルという競技において、アンサーに関する類型を定義することは重要な貢献となる。

さらに、ラップバトルにおけるバースの自動生成に関する研究もいくつか提案されている [17], [18], [19]。本研究において、アンサーの自動分類に取り組むことで、バースの自動評価やラップバトルの勝敗判定の補助など、バースの自動生成に対する応用が可能になると考えられる。そのため、坂口ら [12] や小林ら [15] と同様に、作成した類型の自動分類にも取り組む。

## 3 アンサーの類型化

本節では、アンサーの類型化の方法について述べる。まず、ラップバトルの動画を文字起こししたコーパスを用いて、アンサーのペアデータを作成する。作成したアンサーのペアデータに対して、3 年間以上ラップバトルの視聴経験があるアノテータ 2 名によるカテゴリの作成とアノテーションをおこない、アンサーを類型化する。作成した類型は Fleiss の  $k$  係数 [9] と Krippendorff の  $\alpha$  [13] によって、アノテータ間のラベルの一致率を評価する。

### 3.1 アンサーのペアデータの作成

我々は、Mibayashi ら [18] が作成した、ラップバトルの動画を文字起こししたコーパスを利用した。3 つの大会 (UMB<sup>1</sup>, 凱旋 MCBattle<sup>2</sup>, 戦極 MCBATTLE<sup>3</sup>) と 1 つのテレビ番組 (フリースタイルダンジョン) について、それぞれ 25 件のバトルをコーパスからランダムにサンプリングすることで、合計 100 件のバトルを取得した。

取得した 100 件のバトルに対して、アンサーのペアデータを作成する。その作成方法は、図 2 の左側に示すように、まず、1 つのバトルからバースのペアを取得する。ただし、これらのバースはすべてがアンサーで構成されているわけではなく、アンサーでない内容も含まれている。そこで、このバースのペアから、アンサーとそのアンサーをやるに至った相手のラップ内容 (アンサーの対象) をアンサーのペアとして文単位で人手で抽出する。バースのペアには、アンサーとアンサーの対象が複数件含まれている場合があるため、その場合は該当するものをすべてアンサーのペアとして抽出する。

たとえば、図 2 に示すように、「俺のライムは特別」というアンサーの対象の文と、「一体お前のどこが特別」というアンサーの文はアンサーのペアである。バースによっては、アンサーが複数文にまたがることもあるため、その場合は文同士を連結した。たとえば、「まじでお前はつまらない」に対して「つまらないのはお前の方/俺は即興でも決してつまらない」といったように、アンサーが 2 文にまたがる場合、「まじでお前はつまらない」に対して「つまらないのはお前の方俺は即興でも決してつまらない」と連結した文をペアにした。100 件のバトル内のバースのペアすべてに対して、10 年間以上ラップバトルの視聴経験がある筆頭著者が、アンサーと判断したものをアンサーのペアとして取得した。結果として、429 件のアンサーのペアデータを作成した。

### 3.2 ラベル付けの基準作成

本研究ではラベル付けの基準作成のために、ラップバトルの戦略やテクニックについて詳細な情報が書かれた著名なラッパーらの書籍 [2], [20] を参考にした。書籍内には、実例を基にしたラップバトルの解説があり、プロのラッパーがどのような

1 : <https://www.youtube.com/user/umboofficial>

2 : [https://www.youtube.com/channel/UCe\\_EvY8GrvYgx8PbwrBc75g](https://www.youtube.com/channel/UCe_EvY8GrvYgx8PbwrBc75g)

3 : <https://www.youtube.com/user/senritumc>

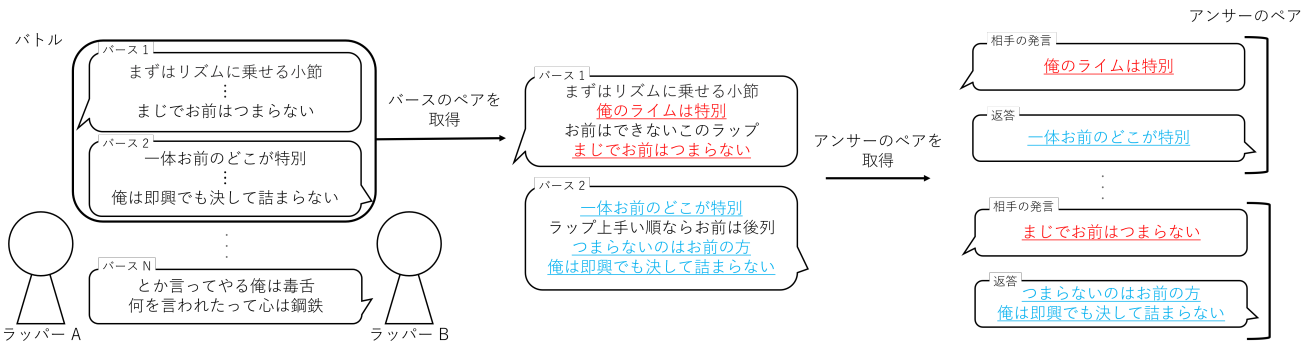


図2 アンサーペアデータの作成方法

返答をアンサーと判別しているかが書かれている。これらの内容を参考に、以下の3つの各観点においてアンサーに該当するか否かを判別するためのメインカテゴリを設計した。

- (1) 相手の発言を考慮した**単語**が返答に含まれるか
- (2) 相手の発言を考慮した**ライム**が返答に含まれるか
- (3) 相手の発言を考慮した**対話**が返答に含まれるか

以下、メインカテゴリ1を「**単語によるアンサー**」、メインカテゴリ2を「**ライムによるアンサー**」、メインカテゴリ3を「**対話によるアンサー**」と呼ぶ。

3つのアンサーの該当を判別するメインカテゴリに加えて、アンサーがどのような表現であるかを判別する以下のカテゴリも作成した。

- (4) アンサーの表現の種類

まず、1つ目の単語によるアンサーは、相手の発言を考慮した単語が返答に含まれるかを判別するカテゴリである。たとえば以下のような例が挙げられる。

相手：俺のテンションは高いぞ

返答：テンション確かにいいね

これはラッパーAの「テンション」という単語がラッパーBの発言に含まれているため、相手の発言を考慮しているというアンサーである。

2つ目のライムによるアンサーは、相手の発言を考慮したライムが返答に含まれるかを判別するカテゴリである。たとえば以下のような例が挙げられる。

相手：俺のテンションは高いぞ

返答：でもファッションは残念だな

これは、ラッパーAの「テンション」という単語に対して、ライムである「ファッション」という単語がラッパーBの発言に含まれているため、相手の発言をライムで考慮しているアンサーである。

3つ目の対話によるアンサーは、相手の発言を考慮した対話が返答に含まれるかを判別するカテゴリである。たとえば以下のような例が挙げられる。

相手：俺のテンションは高いぞ

返答：そうだな確かにそう見える

これは、直接的に相手の発言を含んでいるわけではないが、相手の発言に対して、肯定や否定といった、対話的な返答が含まれているアンサーである。

4つ目のアンサーの表現の種類は、アンサーが成立した際に、どのような表現の種類のアンサーであるかを返答の内容のみで判別するカテゴリである。たとえば以下のような例が挙げられる。

返答：お前のラップは下手くそだ

これは、返答自体に含まれる表現の種類のことであり、上記の例では、相手を否定する表現が含まれている。

### 3.3 メインカテゴリに対するサブカテゴリの作成

前節で作成したメインカテゴリをさらに細かく分類できるように、4つのメインカテゴリそれぞれに対するサブカテゴリを作成する。まず、サブカテゴリの作成のため、10年間以上ラップバトルの視聴経験がある筆頭著者1名と3年以上ラップバトルの視聴経験がある著者ではないアノテータ1名の合計2名で、アンサーペアに対するアノテーションをおこなった。アノテーションでは、同じアンサーペアに対して、2名のアノテータ間で協議をしながらサブカテゴリを付与した。3.1節で作成した429件のアンサーペアから300件をランダムに選択し、それらに対してサブカテゴリを付与した。300件のアノテーション後、さらに残りの129件のアンサーペアからランダムに選択した100件のアンサーペアに対してアノテーションをおこない、それらに対して新しいサブカテゴリが付けられなかったことを確認した上で、アノテーションを終了した。アノテーションしたサブカテゴリを精査し、各サブカテゴリは「単語によるアンサー」において4種類、「ライムによるアンサー」において2種類、「対話によるアンサー」において5種類、「アンサーの表現の種類」において13種類作成した。

### 3.4 アンサーの種類の詳細

作成したアンサーの種類の概要を表1に示す。以下では、各メインカテゴリに対するサブカテゴリについて、具体例を用いて説明する。

#### 3.4.1 単語によるアンサー

単語によるアンサーでは、相手のバース内の単語に関連する単語を含むことで、本メインカテゴリのアンサーに該当すると判別する。単語は同じものとは限らず、類義語やその他の関連語を含む。

表 1 作成したアンサーの類型とアノテータ間のカテゴリの一致率

メインカテゴリ	サブカテゴリ	サブカテゴリの説明	Krippendorff の $\alpha$ [13]
(1) 単語によるアンサー	同じ語	相手のバース内の単語を含む	0.6994
	類義語	相手のバース内の単語に類似する単語を含む	0.1846
	その他関連語	相手のバース内の単語に関連する単語を含む	0.4153
	関連語なし	相手のバース内の単語に関連する単語を含まない	0.6497
(2) ライムによるアンサー	ライムあり	相手のバース内の単語に対するライムを含む	0.2372
	ライムなし	相手のバース内の単語に対するライムを含まない	0.2372
(3) 対話によるアンサー	肯定	相手のバースに対して肯定している	0.3629
	否定	相手のバースに対して否定している	0.3865
	肯定と否定の両方	相手のバースに対して肯定と否定をしている	0.2642
	肯定でも否定でもない	相手のバースに対して肯定も否定もしていない	0.0335
	対話になっていない	相手のバースに対する返答になっていない	0.2273
(4) アンサーの表現の種類	ライム	同じ母音列を持つ単語を返答に含むこと	0.3669
	ディス	暴言や相手が嫌がる内容を含むこと	0.5028
	否定	相手の主張や発言を認めないこと	0.1851
	指摘	相手の発言に対して指摘をすること	0.2952
	自己賛美	自分の能力をアピールすること	0.3557
	賛同	発言に対して称賛したり認めたりすること	0.4725
	文意不明	発言の意味が不明であること	0.2061
	開き直り	相手の発言は正しいがそれに対して開き直ること	0.2680
	優勢	自分が有利な立場にあることを主張すること	0.1644
	平気	相手の発言を気にしていないことをアピールすること	0.0721
	レペゼン	自分の地元や所属をアピールすること	0.6065
	観客	観客に対してアピールすることや観客に喋りかけること	0.3060
	表現なし	これまでのカテゴリのどれにも属さないもの	0.0927

表 2 アンサーの表現の種類の類型

サブカテゴリ	説明	返答の例
ライム	同じ母音列を持つ単語を返答に含むこと	ラップにおいてスキルは重要な内容
ディス	暴言や相手が嫌がる内容を含むこと	お前のラップは下手くそだ
否定	相手の主張や発言を認めないこと	理解できないその思想
指摘	相手の発言に対して指摘をすること	こないだもそれ言ってたけど
自己賛美	自分の能力をアピールすること	俺はめっちゃくちゃラップ上手い
賛同	発言に対して称賛したり認めたりすること	お前のアンサーは最高だな
文意不明	発言の意味が不明であること	俺はラップで崩壊する火事だ
開き直り	相手の発言は正しいがそれに対して開き直ること	たしかに俺はこの前逃げたけどそれがなんだ
優勢	自分が有利な立場にあることを主張すること	会場を沸かした数は俺のほうが多い
平気	相手の発言を気にしていないことをアピールすること	別に構わないぜ俺はそんなことでは動じない
レペゼン	自分の地元や所属をアピールすること	俺は地元大阪からはるばる来たぜ
観客	観客に対してアピールすることや観客に喋りかけること	お客さんどっちについていくか決めな
表現なし	これまでのカテゴリのどれにも属さないもの	俺はそうやってやってる

- **同じ語**：相手のバース内の単語を含む
  - － 相手：俺のテンションは高いぞ
  - － 返答：テンション確かにいいね
- **類義語**：相手のバース内の単語に類似する単語を含む
  - － 相手：俺のテンションは高いぞ
  - － 返答：そうだな気持ちは大事
- **その他関連語**：相手のバース内の単語に関連する単語を含む
  - － 相手：俺のテンションは高いぞ
  - － 返答：じゃあ体育会の飲み会みたいに盛り上げる
- **関連語なし**：相手のバース内の単語に関する単語を含まない（本メインカテゴリのアンサーに該当しない）
  - － 相手：俺のテンションは高いぞ
  - － 返答：俺は今日もラップがうまい

### 3.4.2 ライムによるアンサー

ライムによるアンサーでは、相手のバース内のライムに関連する単語を返答のバース内に含むことで、本メインカテゴリのアンサーに該当すると判別する。

#### ● ライムあり

- － 相手：俺のテンションは高いぞ
- － 返答 1：でもファッションは残念だな
- － 返答 2：つまりダンジョンでやろうセッション

#### ● ライムなし（本メインカテゴリのアンサーに該当しない）

- － 相手：俺のテンションは高いぞ
- － 返答 1：楽しそうだなにより
- － 返答 2：テンションがどうこう俺は知らんな<sup>4</sup>

### 3.4.3 対話によるアンサー

対話によるアンサーでは、相手のバース内の単語やライムを

4：同じ語を含むだけではライムにはならない。

返答のバース内を含むのではなく、対話として成立する返答がバース内に含まれていれば、本メインカテゴリのアンサーに該当すると判別する。対話には、肯定や否定といった基本的なサブカテゴリに加え、肯定と否定の両方や肯定でも否定でもない、といったサブカテゴリも考慮した。

- **肯定**

- 相手：俺のテンションは高いぞ
- 返答：そうだな確かにお前は高い

- **否定**

- 相手：俺のテンションは高いぞ
- 返答：いや俺のほうが上回ってるテンション

- **肯定と否定の両方**

- 相手：俺のテンションは高いぞ
- 返答：確かにお前は高いでも低いときもあるだろう

- **肯定でも否定でもない**

- 相手：俺のテンションは高いぞ
- 返答：テンションをアピールしているお前

- **対話になってない** (本メインカテゴリのアンサーに該当しない)

- 相手：俺のテンションは高いぞ
- 返答：高い低いでも俺の価値は高い

### 3.4.4 アンサーの表現の種類

アンサーの表現の種類では返答のみを対象とし、アンサーがどのような表現であるかを判別する。その種類は13種類と多いため、サブカテゴリの説明と具体例は表2に示す。

### 3.5 ラップバトルに精通したアノテータによるアノテーション

前節で作成したアンサーのタイプを評価するために、ラップバトルに精通したアノテータによるアノテーションをおこなった。アノテータは、クラウドソーシングサービス Lancers にて、ラップバトルの視聴歴が3年以上の者を3名採用した。アノテータは、著者らが作成したアノテーション作業用のサイトにアクセスし、3.1節で作成した429件のアンサーペアデータに対してアノテーションをおこなった。

アノテータには、4種類のメインカテゴリのそれぞれに対してサブカテゴリを付与するように指示した。まず、「単語によるアンサー」「ライムによるアンサー」「対話によるアンサー」の3種類のメインカテゴリに対しては、サブカテゴリを1つだけ付与するように指示した。これら3種類のメインカテゴリ内のサブカテゴリは排反であるため、複数付与することはできない。次に、「アンサーの表現の種類」に対しては、サブカテゴリの複数付与を可能とし、どれにも当てはまらない場合は「表現なし」を付与するように指示した。アノテータに提示するアンサーペアデータの順序はアノテータごとにランダムに決め、作業の順番がアノテータ間で同じにならないようにした。

結果を図3に示す。図中の $n$ の値は、アノテータ3名が1つのアンサーペアに対して付与したサブカテゴリの一致数を示している。たとえば、(1)「単語によるアンサー」において、アノテータ3名が付与したサブカテゴリがそれぞれ「同じ語」「同じ語」「類義語」だった場合、「同じ語」に $n=2$ が1つカウ

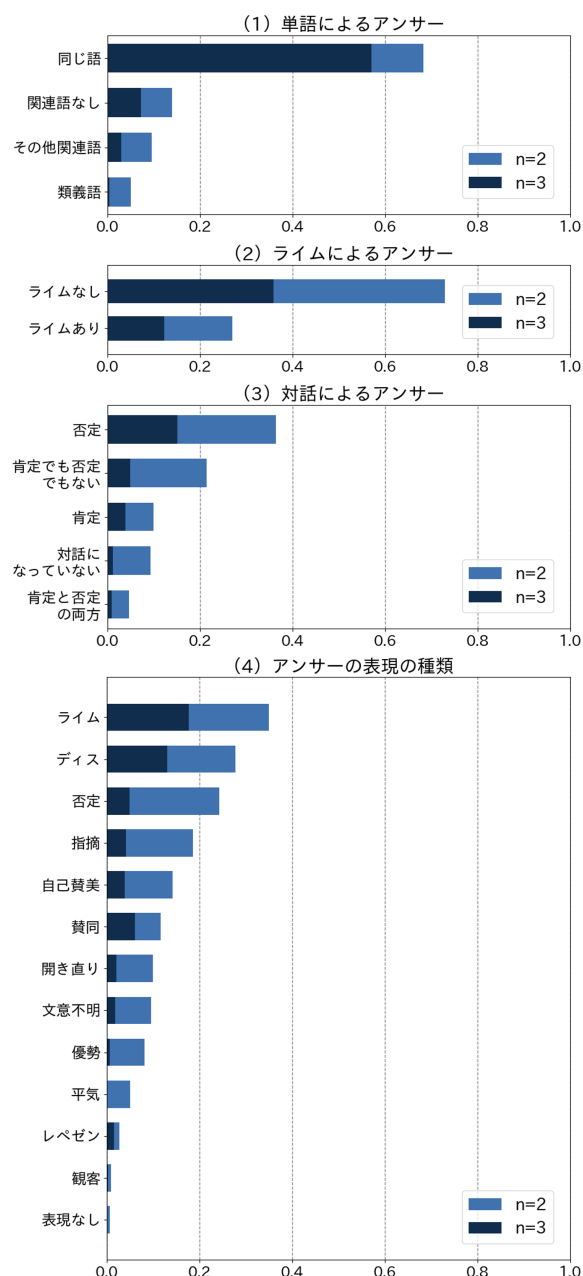


図3 4つのメインカテゴリに対するアノテーションによるサブカテゴリの割合 (横軸: 割合)

トされる。全員が「同じ語」なら、それに $n=3$ が1つカウントされる。アノテーションの信頼性の観点から $n=1$ は扱わず、アノテータ2名以上が同じサブカテゴリを付与した事を示す $n=2$ と $n=3$ を合算した割合 ( $n$ が2以上の割合) に基づいて考察する。そこで $n=2$ と $n=3$ のそれぞれをカウントした後、アンサーペアデータの総数429で割ることで、各サブカテゴリの割合を計算した。図3の各グラフでは、合算した割合の降順にサブカテゴリを表示している。

まず、(1)「単語によるアンサー」については、「同じ語」が68.29%と最も高い割合で、「関連語なし」は13.98%と低い割合であった。このことから、相手のバースに対して何らかの関連する単語がアンサーに含まれる傾向にあり、中でも「同じ語」が最も含まれていることがわかる。次に、(2)「ライムによるアンサー」については、「ライムあり」が27.03%と「ライムなし」

表3 メインカテゴリ別の評価

メインカテゴリ	Fleiss の $k$ 係数 [9]	Krippendorff の $\alpha$ [13]
(1) 単語によるアンサー	0.5582	0.5585
(2) ライムによるアンサー	0.2366	0.2372
(3) 対話によるアンサー	0.2450	0.2456
(4) アンサーの表現の種類 (複数ラベルの完全一致)	0.0955	0.0962
(4) アンサーの表現の種類 (複数ラベルの内どれか一つでも一致)	0.5533	0.5536

より低い割合で、ライムは望ましいもののアンサーにはライムが含まれない場合が多いことがわかる。そして、(3)「対話によるアンサー」については、「対話になっていない」は9.32%と低い割合なので、多くは対話になっていたことがわかる。対話になっている中では、「否定」が36.36%と最も高い割合な一方で「肯定」は10.02%と低い割合であり、アンサーでは否定的になりがちなのことがわかる。

最後に、(4)「アンサーの表現の種類」については、「ライム」が34.96%と最も割合が多かった。これはラップバトルでライムが多用されているために、アノテータが付与したサブカテゴリ(ラベル)が多かったと考えられる。次に多い「ディス」「否定」「指摘」は相手に対して否定的な表現や攻撃的な表現を含む類型である。これらの割合が18%を超えていることから、ラップバトルではそうした表現が多いことがわかる。「自己賛美」「開き直り」「優勢」「平気」は自身が優位な状態にあることを表現する類型であり、「賛同」は相手を褒めることを表現する類型である。これらのタイプの割合が5%から15%の間にあることから、自身の優位な状態を示すことや相手を褒めることより、上記の相手に対して攻撃的な表現を含む場合のほうが多いことがわかる。「文意不明」も割合が5%から10%の間にあるが、これは即興的で確かな表現を考えてアンサーを返すことがいかに難しいかを示している。最後に、「表現なし」は0.69%とかなり低い割合であることから、これらのタイプにより、アンサーの表現は十分に網羅できていると考えられる。

### 3.6 類型に対するアノテーションの評価

前節でアノテーションしたアンサーのタイプを評価する。今回、アノテータが3名であるため、3名以上の評価に対応している Fleiss の  $k$  係数 [9] と Krippendorff の  $\alpha$  [13] を用いて評価した。Fleiss の  $k$  係数は  $-1$  から  $1$  までの値を取り、Krippendorff の  $\alpha$  は理論上負の値を取ることが可能であるが、一般的には  $0$  から  $1$  までの値を取る。どちらの評価尺度も  $1$  に近づくほど一致率が高いことを示す。本実験では、Krippendorff の  $\alpha$  については名義尺度で評価をおこなった。

評価結果を表3に示す。まず、単語によるアンサーでは、Fleiss の  $k$  係数は  $0.5582$  であり、Krippendorff の  $\alpha$  は  $0.5585$  であった。これは Higashinaka ら [10] の研究における、専門業者らによるアノテーション結果である Fleiss の  $k$  係数  $0.567$  に匹敵する高い一致率を示している。このような一致率になったのは、単語によるアンサーのサブカテゴリは比較的容易に判断できるものが多いためであると考えられる。ただし、表1に示すように、「類義語」と「その他関連語」の一致率が低い。これは両者の違いが曖昧であるため、アノテータ間でラベルが一致

しなかったと考えられる。

ライムによるアンサーについては、Fleiss の  $k$  係数は  $0.2366$  であり、Krippendorff の  $\alpha$  は  $0.2372$  であった。これは Higashinaka ら [8] による、対話破綻の3クラスの分類における Fleiss の  $k$  係数  $0.276$  に近い一致率であり、少し低い一致率と言える。原因としては、ライムが含まれているか否かをテキストで判別することは難しく、アノテータ間でラベルが一致しなかったと考えられる。具体的には、「お前は枚方を上げれば終わるぜ」と「イカサマもしくは枚方」というアンサーペアに対して、「ライムあり」のラベルが2件、「ライムなし」のラベルが1件付けられている。これは「枚方 (HIRAKATA)」と「イカサマ (IKASAMA)」が「IAAA」という母音で一致するライムである。しかし枚方という大阪の地名が正しく読めなかったため、「ライムなし」が付けられた可能性が高い。このように本来音声として聞いていけばライムと判別できるものが、テキストだけでは判別が難しくなり、不一致が起きたと考えられる。

対話によるアンサーについては、Fleiss の  $k$  係数は  $0.2450$  であり、Krippendorff の  $\alpha$  は  $0.2456$  であった。これも先ほどと同じく、少し低い一致率である。これは、「肯定でも否定でもない」が判別しにくく、一致率が低いことが影響していると考えられる。実際に3名のアノテータによるラベルの組み合わせを調べると、「否定、否定、肯定でも否定でもない」が78件、「否定、肯定でも否定でもない、肯定でも否定でもない」が23件あり、「否定」と「肯定でも否定でもない」が判別しにくいことがわかる。さらに、「対話になっていない、対話になっていない、肯定でも否定でもない」が25件、「対話になっていない、肯定でも否定でもない、肯定でも否定でもない」が16件あり、「対話になっていない」と「肯定でも否定でもない」も判別しにくいことがわかる。

アンサーの表現の種類については、複数ラベルの完全一致において、Fleiss の  $k$  係数は  $0.0955$  であり、Krippendorff の  $\alpha$  は  $0.0962$  であった。アンサーの表現の種類は複数ラベルであるため、完全一致することは非常に難しく、 $0.0962$  という低い一致率は自然であると考えられる。しかし、アノテータが付与した複数のサブカテゴリの中から、どれか1つでもサブカテゴリが全アノテータ間で一致していれば完全一致とみなすという緩和した条件においては、Fleiss の  $k$  係数は  $0.5533$  であり、Krippendorff の  $\alpha$  は  $0.5536$  と高い一致率を示している。

## 4 アンサーの自動分類

ラップバトルにおいて、相手の発言に対する返答のラップ文が与えられたときに、作成したアンサーのタイプに基づくアンサーの自動分類がどの程度できるかを検証するため、評価実験

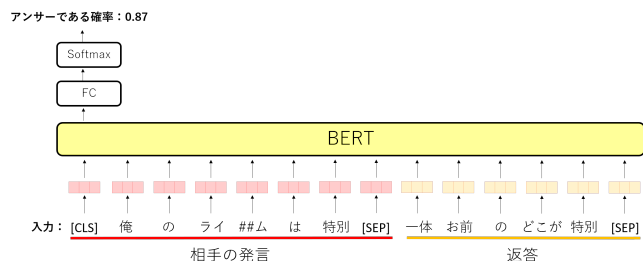


図4 BERTを用いたアンサーの分類方法

をおこなった。まず、「単語によるアンサー」「ライムによるアンサー」「対話によるアンサー」の3つのメインカテゴリを対象として、返答のラップ文が、各メインカテゴリのアンサーに該当するか否かの2値分類をおこなった。本章では簡単のため、「本メインカテゴリのアンサーに該当する」「本メインカテゴリのアンサーに該当しない」という表記を、「アンサーである」「アンサーでない」と記述する。次に、これら3つのメインカテゴリの内、少なくとも1つに該当するか否かを判別する2値分類にも取り組んだ。最後に、「アンサーの表現の種類」のメインカテゴリを対象として、返答のラップ文を13種類のサブカテゴリに分類する評価実験をおこなった。

#### 4.1 BERTによるアンサーの自動分類

4.2節から4.5節の評価実験では、共通の枠組みを用いてアンサーの分類をおこなっているため、まず以下でその共通の枠組みについて説明する。本実験では、アノテーション済みの429件のアンサーペアデータを基に分類タスクに取り組んだ。それぞれのメインカテゴリに対して、アノテーションしてもらったラベルを正解として学習用のデータを作成した。本実験では、3名のアノテータによって付与してもらったラベルの多数決を取って正解のラベルとした。多数決が取れない場合は学習用データから削除した。

アンサーの自動分類をするためのモデルとして、BERT [3]を用いた。BERTの事前学習モデルは東北大学が公開しているbert-base-japanese-v3<sup>5</sup>を用いた。BERTの学習に用いたハイパーパラメータは、バッチサイズは32、OptimizerはAdamW [21]、ロス関数はCross Entropy Loss、学習率は $2e-7$ 、Dropoutは0.1、最大文長は256とした。

BERTを用いたアンサー分類は図4に示すように、相手の発言を1文目、返答文を2文目とすることで入力とした。前処理として、1文目の先頭に[CLS]トークンを付与し、1文目と2文目の末尾には[SEP]トークンを付与した。本実験では、BERTから出力される、[CLS]トークンに対応する768次元のベクトルを用いて分類問題を解いた。

#### 4.2 単語によるアンサーの分類

単語によるアンサーの分類については、多数決によって正解ラベルを定義し、416件のデータセットを取得した。本実験では、返答文が単語カテゴリにおけるアンサーに該当するか否かを判別する2値分類に取り組むため、「関連語なし」以外の「同

表4 3つのメインカテゴリに対する分類結果

メインカテゴリ	正解率
(1) 単語によるアンサーの分類	0.7222
(2) ライムによるアンサーの分類	0.8139
(3) 対話によるアンサーの分類	0.8253
(1-3) 単語、ライム、対話によるアンサーの分類	0.8554

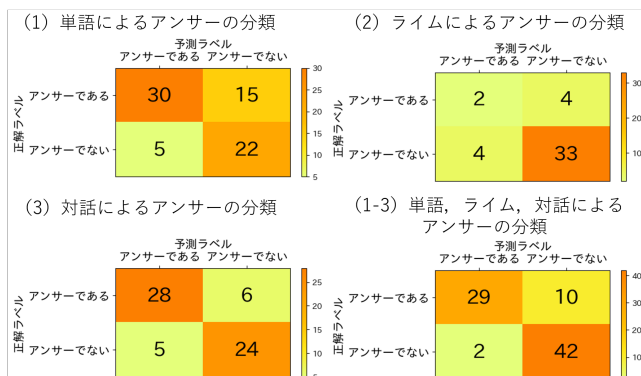


図5 各分類タスクにおける2値分類の混同行列

じ語」「類義語」「その他関連語」を正例とした、相手の発言と返答文のペアデータを356件取得した。それに対して、データセット内の別の返答文を付与したデータを負例として356件作成し学習をおこなった。結果として、データセットは712件となり、これを訓練用569件、検証用71件、テスト用72件に分けて学習をおこなった。

分類結果は、表4に示すように、正解率が0.7222となり、ある程度高い性能で分類ができていることがわかる。図5に示す混同行列では、本カテゴリに該当するアンサーは正しく分類できているが、アンサーであるものを、アンサーでないと間違えて分類していることがわかる。

#### 4.3 ライムによるアンサーの分類

ライムによるアンサーの分類についても、多数決によってラベルを定義し、429件のデータセットを取得した。本実験では、返答文がライムカテゴリにおけるアンサーに該当するか否かを判別する2値分類に取り組むため、「ライムあり」のラベルを正例、「ライムなし」のラベルを負例とした、相手の発言と返答文のペアデータを429件（「ライムあり」が116件、「ライムなし」が313件）取得した。これを訓練用343件、検証用43件、テスト用43件に分けて学習をおこなった。

分類結果は、表4に示すように、正解率が0.8139となり、高い分類性能となった。しかし、図5に示す混同行列では、予測結果のほとんどが「アンサーでない」となっていることがわかる。これは、データセット全体において、「ライムなし」のラベルが313件と多いことにより、このような偏った分類結果になっていると予想される。「ライムあり」の正例データのように、返答文中に相手の発言を考慮した自然なライムを含むことは、自動生成では困難であり、手作業でも時間がかかるため、正例と負例を同数にした学習は本実験ではおこなえなかった。今後はクラスウェイトを使用するなど、データセットのラベルを考慮した学習の改善が必要である。

<sup>5</sup> : <https://huggingface.co/cl-tohoku/bert-base-japanese-v3>



#### 4.4 対話によるアンサーの分類

対話によるアンサーの分類についても、多数決によってラベルを定義し、351件のデータセットを取得した。本実験では、返答文が対話カテゴリにおけるアンサーに該当するか否かを判別する2値分類に取り組むため、「対話になっていない」(40件)以外の「肯定」「否定」「肯定と否定の両方」「肯定でも否定でもない」を正例とした、相手の発言と返答文のペアデータを311件取得した。それに対して、データセット内の別の返答文を付与したデータを負例として311件作成した。結果として、データセットは622件となり、これを訓練用497件、検証用62件、テスト用63件に分けて学習をおこなった。

分類結果は、表4に示すように、正解率が0.8253となり、高い分類性能を示した。図5に示す混同行列では、本メインカテゴリのアンサーに該当するか否かを適切に分類できていることがわかる。これは、BERTの文同士の繋がりを考慮した事前学習により、少量のデータセットでもうまく分類ができたためだと考えられる。

#### 4.5 単語、ライム、対話によるアンサーの分類

単語、ライム、対話によるアンサーのカテゴリに対して、分類をおこなう。学習に使用するデータは429件の相手の発言と返答文のペアデータの内、3つのメインカテゴリのうち少なくとも1つのメインカテゴリにおいてアンサーに該当すれば正例とした。今回、412件のデータがアンサーと判断されたため、それらのデータを正例とした。それに対して、データセット内の別の返答文を付与したデータを負例として412件作成した。結果として、データセットは824件となり、これを訓練用659件、検証用82件、テスト用83件に分けて学習をおこなった。

分類結果は、表4に示すように、正解率が0.8554となり、高い分類性能を示した。図5に示す混同行列を見ると、本カテゴリでは「アンサーである」と「アンサーでない」の両方に対して適切に分類できていることがわかる。このことから、アンサーを判別する3つのメインカテゴリを同時に考慮することで、メインカテゴリを個別に分類する場合と比べて、高い性能で分類ができることを示した。

#### 4.6 アンサーの表現の種類の分類

最後に、アンサーの表現の種類を分類する。アンサーの表現の種類では、ある返答文に対して複数のサブカテゴリを同時に持つため、複数のサブカテゴリを同時に予測するマルチラベル問題として解いた。本実験では、図6に示すように、返答文のみを入力とし、返答文のサブカテゴリを表すベクトルを出力する。入力の前処理として、返答文の先頭に[CLS]トークンを付与し、末尾には[SEP]トークンを付与した。アンサーの表現の種類は13種類であるため、[CLS]トークンに対応する768次元のベクトルをFC層によって、13次元のベクトルとして出力し、Binary Cross Entropy Lossによって正解のサブカテゴリとの誤差を計算した。

学習に使用するデータは、アノテータ3人の内、2人以上が同じサブカテゴリを付与している場合、そのサブカテゴリを正

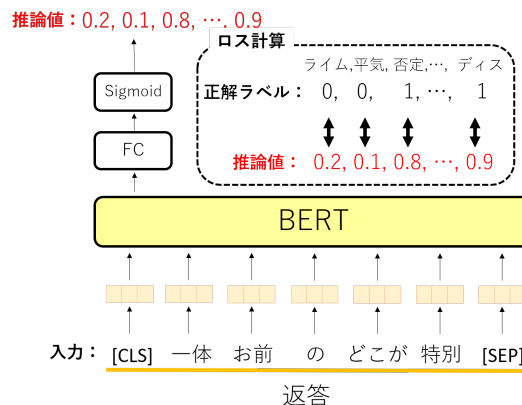


図6 BERTを用いたアンサーの表現の分類

解のサブカテゴリとして加えた。この操作によってラベルが付けられなかった23件のデータを除いた、406件をデータセットとした。これを訓練用324件、検証用41件、テスト用41件として使用した。

また、マルチラベル問題を正しく学習するためにクラスウェイトを使用した。マルチラベル問題では、図6に示すように、各サブカテゴリの有無を0と1で表したマルチホットベクトルを正解ラベルとして使用する。本実験で使用する正解ラベルのマルチホットベクトルには0が多く含まれているため、予測値をすべて0にするとといった過学習が起きる可能性がある。そのため、本実験ではサブカテゴリが含まれていることを示す1に対して、5.0のクラスウェイトをかけて学習をおこなった。学習後の推論時では、予測値に対して、0.5以上を1、0.5未満を0として扱った。

アンサーの表現の種類の分類結果を表5に示す。正解率は、予測値と正解ラベル内の13個の要素がすべて一致する際に正解とする場合(完全一致)と、予測値と正解ラベル内のそれぞれ対応する要素が一致する際に正解とする場合(要素ごとの一致)の2種類を計算した。まず、完全一致においては、クラスウェイトの有無に関わらず0.0243と非常に低い結果となった。これは、13種類のサブカテゴリの有無を完全に一致させる必要があるため、このような非常に低い正解率になったと考えられる。次に、要素ごとの一致においては、クラスウェイトなしが0.8592であり、クラスウェイトありが0.7617となった。これは、クラスウェイトなしの方が良い正解率に見えるが、実際はクラスウェイトなしの予測値がほとんど0であったため、0が多い正解ラベルにおいて正解率が高くなっている。

このように、マルチラベル問題において、正解率だけで分類の性能を評価することは難しいため、正解率以外にも、適合率、再現率、F値を計算した。まず、適合率はクラスウェイトなしでは0.0512、クラスウェイトありでは0.7051であった。クラスウェイトありにおいては、サブカテゴリを含むと予測した際には、予測が正しくできていることがわかる。次に、再現率はクラスウェイトなしでは0.8000、クラスウェイトありでは0.3459であった。クラスウェイトなしはモデルの予測値がほとんど0であるため、再現率が0.8000と高くなっていることがわかる。クラスウェイトありは再現率が0.3459と低く、本来サブカテ

表5 アンサーの表現の種類の分類結果

実験設定	正解率 (完全一致)	正解率 (要素ごとの一致)	適合率	再現率	F 値
クラスウェイトなし	0.0243	0.8592	0.0512	0.8000	0.0963
クラスウェイトあり	0.0243	0.7617	0.7051	0.3459	0.4641

ゴリを含むと予測すべきものを予測できていないことがわかる。最後に、F 値はクラスウェイトなしでは 0.0963、クラスウェイトありでは 0.4641 であった。クラスウェイトなしでは適合率が非常に低いため、F 値も 0.0963 と低くなっている。クラスウェイトありでは、F 値が 0.4641 と中程度の値を示していることがわかる。これらのことから、アンサーの表現の種類の分類においては、クラスウェイトを使用した場合、サブカテゴリを含むと予測した場合には高い割合で正解できるが、本来予測すべきサブカテゴリを見逃してしまうことが多いことがわかる。しかし、F 値は中程度の値を示しているため、本分類モデルは中程度の性能で分類が可能であると言える。

## 5 まとめと今後の課題

本研究では、アンサーの自動分類に向けた類型化に取り組んだ。まず、ラップバトルコーパスからアンサーのペアデータを作成し、ラップバトルに精通した 2 名のアノテータによるラベルの作成とアノテーションにより、アンサーを類型化した。類型化の結果として、作成した類型の内、2 種類のメインカテゴリにおいて、Krippendorff の  $\alpha$  が 0.5 を超える一致率を示した。また、作成したアンサー類型を基に、BERT を用いた自動分類に取り組んだ。結果として、作成した 3 つのいずれかの類型のアンサーに該当するか否かを判別する分類タスクの正解率は 0.8554 となり、高い性能で分類ができていたことがわかった。さらに、アンサーがどのような表現かを判別する分類タスクにも取り組み、F 値が 0.4641 と中程度の分類性能を示した。今後はアンサーをバースの中から自動的に検知し、バース単位でのペアに対して、自動的にアンサーを分類できる手法に取り組む予定である。

謝辞 本研究は JSPS 科研費 JP21H03775, JP21H03774, JP21H03554, JP22H03905 の助成を受けた。

## 文 献

- [1] V. Sykari. Interactive oral composition: Resources, strategies, and the construction of improvised utterances in a Finnish freestyle rap battle. *The Journal of American Folklore*, Vol. 132, No. 523, pp. 3–35, 2019.
- [2] DARTHREIDER. MC バトル史から読み解く日本語ラップ入門. KADOKAWA, 2017.
- [3] J. Devlin, M. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL 2019, pp. 4171–4186, 2019.
- [4] P. James, B. Harry, and Z. Annie. Designing annotation schemes: From theory to model. *Handbook of Linguistic Annotation*, pp. 21–72, 2017.
- [5] 保田祥, 小西光, 浅原正幸, 今田水穂, 前川喜久雄. 『現代日本語書き言葉均衡コーパス』に対する時間情報表現・事象表現間の時間的順序関係アノテーション. *自然言語処理*, Vol. 20, No. 5, pp. 657–681, 2013.
- [6] JR. Trippas, D. Spina, P. Thomas, M. Sanderson, H. Joho, and L. Cavedon. Towards a model for spoken conversational search. *Information Processing & Management*, Vol. 57, No. 2, pp. 1–19, 2020.
- [7] V. Susana, K. Uzay, and S. João. Cohen’s kappa coefficient as a performance measure for feature selection. In *The 2010 IEEE World Congress on Computational Intelligence*, WCCI 2010, pp. 1–8, 2010.
- [8] R. Higashinaka, M. Mizukami, K. Funakoshi, M. Araki, H. Tsukahara, and Y. Kobayashi. Fatal or not? Finding errors that lead to dialogue breakdowns in chat-oriented dialogue systems. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2015, pp. 2243–2248, 2015.
- [9] JL. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.
- [10] R. Higashinaka, K. Funakoshi, M. Araki, H. Tsukahara, Y. Kobayashi, and M. Mizukami. Towards taxonomy of errors in chat-oriented dialogue systems. In *The 16th Annual SIGdial Meeting on Discourse and Dialogue*, SIGDIAL 2015, pp. 87–95, 2015.
- [11] 東中竜一郎, 荒木雅弘, 塚原裕史, 水上雅博. 雑談対話システムにおける対話破綻を生じさせる発話の類型化. *自然言語処理*, Vol. 29, No. 2, pp. 443–466, 2022.
- [12] 坂口智洋, 河原大輔, 黒橋禎夫. 事象に対する網羅的な時間情報アノテーションとその分析. *自然言語処理*, Vol. 26, No. 1, pp. 179–206, 2019.
- [13] K. Krippendorff. Computing Krippendorff’s alpha-reliability. *Departmental Papers (ASC)*. 43, pp. 1–10, 2011.
- [14] R. Dey and FM. Salem. Gate-variants of gated recurrent unit (GRU) neural networks. In *IEEE 60th International Midwest Symposium on Circuits and Systems*, MWSCAS 2017, pp. 1597–1600, 2017.
- [15] 小林滉河, 山崎天, 吉川克正, 牧田光晴, 中町礼文, 佐藤京也, 浅原正幸, 佐藤敏紀. 日本語有害表現スキーマの提案と評価. *言語処理学会第 29 回年次大会*, NLP 2023, pp. 933–938, 2023.
- [16] HS. Alim, J. Lee, LM. Carris, and QE. Williams. Linguistic creativity and the production of cisheteropatriarchy: A comparative analysis of improvised rap battles in Los Angeles and Cape Town. *Language Sciences*, Vol. 65, pp. 58–69, 2018.
- [17] D. Wu and K. Addanki. Learning to rap battle with bilingual recursive neural networks. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, IJCAI 2015, pp. 2524–2530, 2015.
- [18] R. Mibayashi, T. Yamamoto, K. Tsukuda, K. Watanabe, T. Nakano, M. Goto, and H. Ohshima. Verse generation by reverse generation considering rhyme and answer in Japanese rap battles. In *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research*, CMMR 2023, pp. 30–41, 2023.
- [19] R. Savery, L. Zahray, and G. Weinberg. Shimon the rapper: A real-time system for human-robot interactive rap battles. In *Proceedings of the 11th International Conference on Computational Creativity*, ICCO 2020, pp. 212–219, 2020.
- [20] 晋平太. フリースタイル・ラップの教科書. イースト・プレス, 2016.
- [21] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations*, ICLR 2019, pp. 1–8, 2019.