

ラップバトルにおけるアンサーの類型化及びアンサーの有無と表現の自動分類*

三林 亮太^{†a)} 佃 洗撰^{††} 渡邊 研斗^{††} 中野 倫靖^{††}
後藤 真孝^{††} 山本 岳洋[†] 大島 裕明[†]

Taxonomization of Answers in Rap Battles and Automatic Classification of Answer Presence and Expression*

Ryota MIBAYASHI^{†a)}, Kosetsu TSUKUDA^{††}, Kento WATANABE^{††}, Tomoyasu NAKANO^{††},
Masataka GOTO^{††}, Takehiro YAMAMOTO[†], and Hiroaki OHSHIMA[†]

あらまし 本論文では、ラップバトルにおけるアンサーの類型化と、それに基づくアンサーの有無と表現の自動分類について述べる。アンサーはラップバトル固有の技法で、対戦相手のラップ文に対して、その内容に沿って返答したラップ文をアンサーと呼ぶ。アンサーには様々な表現があるが、一般的な対話等の類型化の先行研究ではその類型化が未達成だったため、本研究ではまず、アンサーの類型化に取り組んで3種類の返答類型を作成した。次に、これらの返答類型をより細かく分類するためのサブ返答類型を6種類作成した。更に返答類型とは別に、どのような表現がラップ文に用いられるかを示す表現類型を11種類作成した。以上の類型に基づいて、アンサーの有無と表現のそれぞれを自動的に分類するために、BERTを用いた分類手法とルールベースの分類手法を提案し、評価実験を実施した。対戦相手のラップ文に対する返答のラップ文においてアンサーの有無を2値分類するタスクでは、F値が0.7708と高い分類性能を得た。アンサーの表現類型を判別する分類タスクにも取り組み、F値が0.4154と中程度の分類性能を得た。

キーワード ラップバトル、アンサー、類型化、テキスト分類

1. ま え が き

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。それぞれの番で発言する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。ラップバトルの例を示した図1では、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2文で構成されるバース1に対して、ラッパーBは2文のバース2で返答している。こうした返答を複数回繰り返すとバトルは終了し、観客が多数決で勝敗を決める。

ラップバトルでは、バースの質が勝敗の評価基準となる。バースの質は主に、ライム、フロー、アンサーの三つを対象に評価される[1],[2]。ライムは母音列が一致または類似する単語のペアを指し、そうしたペアを含む返答は高く評価される。図1のバース2の「特別(“TOKUBETSU”)」と「後列(“KOURETSU”)」は、“OUEU”という母音列が一致するライムを含んでい

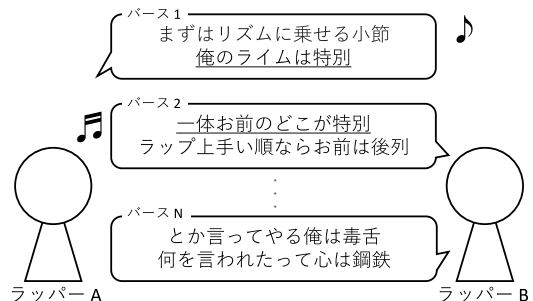


図1 ラップバトルの例

[†] 兵庫県立大学大学院情報科学研究科, 神戸市

Graduate School of Information Science, University of Hyogo, 8-2-1 Gakuen-nishimachi, Nishi-ku, Kobe-shi, 651-2197 Japan

^{††} 産業技術総合研究所, つくば市

National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba-shi, 305-8568 Japan

a) E-mail: af22h007@guh.u-hyogo.ac.jp

* 本論文は、データ工学研究専門委員会推薦論文である。

DOI:10.14923/transinfj.2024DET0001

る。次に、フローはリズムに合わせたラップ特有の歌い方を指し、良いリズムで歌った返答は高く評価される。最後に、アンサーは相手のバースの内容に沿った返答を指し、アンサーを含む返答は高く評価される。図1のバース1の「俺のライムは特別」に対して、バース2の「一体お前のどこが特別」は、相手の「特別」という内容に沿った返答なのでアンサーとなっている。

これら三つの評価基準のうち、ライムとフローは一般的なラップ楽曲でも用いられるのに対し、アンサーはラップバトル固有の技法である。相手のバースを理解しなければ、その内容に沿ったアンサーを含むバースで返答できないので、事前準備ができず即興で考えた証明とみなされる。そのためアンサーの有無やその質は、ラップバトルの重要な評価基準となっている。

しかし、バースの各ラップ文がアンサーとして成立するかどうかは評価が難しく、従来その分析は不十分で、アンサーとして成立するラップ文にどのような種類の表現が用いられるかは明らかになっていない。そのため、アンサーの質を適切に評価してラップバトルの勝敗を公平に決めることは観客にとっても難しかった。勝敗を公平に判定できない場合、誤審が起きたり延長戦になったりする場合もあり、ラッパーの負担が大きくなるため、勝敗を公平に判定することはラッパーにとっても重要である。また、もしアンサーの有無やその表現の種類を自動分類して提示できれば、観客は勝敗の判定の参考になる。著名なラップバトルの大会であるUMBでは、2007年に歓声の大きさを機械で判定し勝敗の判定基準に取り込むなど、勝敗の判定の参考になるようなデータの提供をおこなっていることから、そうした分類結果の提示は観客にとって有意義であると考えられる。

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。まず、ラップバトルの動画を文字起こしたコーパスから「対戦相手のラップ文」と「そのラップ文に対する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(3.1)。次に、作成したペアとラップバトルに関する書籍を参考に、ラップバトル視聴歴10年以上の筆頭筆者1名と3年以上の非著者1名の2名がアンサーの類型を作成した。その結果、対戦相手のラップ文に対して「単語」で返答するアンサー、「ライム」で返答するアンサー、「対話」で返答するアンサーの3種類の返答類型を作成した(3.2)。また、返答類型をより詳細に説明するためのサブ返答類

型を6種類作成した(3.3)。更に、これら返答類型とは別に、各ラップ文がどのような表現を用いているかを示す表現類型を11種類作成した(3.4)。以上の類型の妥当性を検証するため、作成したペアデータに対して類型ラベルのアノテーション作業をおこなった。その結果、アノテータ間の回答の合意(一致率)を表す指標であるKrippendorffの α 係数[3]を用いた分析では、「単語」で返答するアンサーの返答類型と表現類型ではそれぞれ0.5を超える一致率を、「ライム」及び「対話」で返答するアンサーの返答類型では0.3を超える一致率を示した(4.1)。

本研究では更に、作成した類型を基に、BERT[4]や母音列のマッチングを用いてアンサーを自動分類した(5.1~5.6)。その結果、「単語」「ライム」「対話」の3種類の返答類型のうちの少なくとも一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F値が0.7708となり、高い性能でアンサーの有無の自動分類ができた。更に、各アンサーがどのような表現類型を含むかを推定するマルチラベリング問題にも取り組み、F値が0.4154と中程度の分類性能を得た。

2. 先行研究

類型化の先行研究では、類型化の方法、対話に関する類型化、時間表現に関する類型化、有害表現に関する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[5]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、適切な類型を作成する。類似する手法として、PustejovskyらはMATTER サイクル[5]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、後述する類型化の先行研究においても、MATTER サイクルではなく、MAMA サイクルが採用されている[6],[7]。本研究においてもMAMA サイクルを用いることで、適切な類型を作成する。

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[8]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」の三つのカテゴリとそれ

らに対するサブカテゴリー（「タスク」のサブカテゴリーとしては「情報要求」や「検索支援」など）を作成した。類型は Cohen の κ 係数 [9] を用いて、サブカテゴリーレベルで評価され 0.71 の高い一致率を示した。

Higashinaka ら [10] は対話破綻の誤り分類に取り組み、作成した 1,146 件の対話データの中からランダムに選択した 100 件に対して 24 名がアノテーションをおこない、付与したコメントを基に分類した。「文法エラー」「解釈不能」「発話無視」などの 17 のメインカテゴリー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 クラスに分けて評価した。Fleiss の κ 係数 [11] が 3 クラスの分類では 0.276 の一致率を示し、2 クラスの分類では 0.396 の一致率を示した。

こうしたデータに基づく類型化とは異なるアプローチとして、Higashinaka ら [12] は対話理論から逸脱した対話を対話破綻とみなして類型を作成し、「発話」「応答」「文脈」などの 16 のメインカテゴリーと、それらに対するサブカテゴリー（「発話」のサブカテゴリーとしては「構文的誤り」「意味的誤り」など）を作成した。580 件の対話に対してそれらの類型をアノテーションして評価した結果、Fleiss の κ 係数がメインカテゴリーでは 0.400、サブカテゴリーでは 0.239 の一致率を示した。

更に、Higashinaka ら [13] はこれら二つの類型を統合した新たな類型を作成した。この類型では、Fleiss の κ 係数において、専門アノテータによるアノテーションでは 0.567、クラウドワーカーによるアノテーションでは 0.488 の一致率を示し、いずれも高かった。

2.3 時間表現に関する類型化

保田ら [6] 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 MAMA サイクルを用いて作成した。Cohen の κ 係数により評価し、0.653 の高い一致率を示した。

坂口ら [14] は、京都大学テキストコーパスを対象に時間表現の類型を作成してアノテーションした。評価した結果、Krippendorff の α [3] が厳しい条件では 0.526、緩和した条件では 0.867 の一致率を示した。作成した類型を基に、双方向 GRU モデル [15] を用いて自動分類した結果、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 の正解率を示した。

2.4 有害表現に関する類型化

小林ら [7] 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に対し、有害レベルと有害カテゴリーの 2 種類のアノテーションを付与した。有害レベルは 4 段階でラベル付けし、有害カテゴリーの類型は MAMA サイクル [5] を用いて作成した。結果として、「差別」「卑語」「迷惑行為」などの七つのメインカテゴリーと、それらに対する 41 のサブカテゴリー（「差別」のサブカテゴリーとしては「人種差別」や「職業差別」など）を作成した。これらを Krippendorff の α を用いて評価し、有害レベルの 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帰属問題として解くことで、高い性能で有害検知を実現した。

2.5 本研究の位置づけ

以上の先行研究では、対話破綻、時間表現、有害表現を対象に類型を作成していた。一方で、著者の知る限り、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、まだ十分に分析さ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 [2], [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作成することは重要な貢献となる。

また、ラップバトルにおけるパースの自動生成に関する研究も幾つか取り組まれている [17]~[19]。本研究でアンサーの自動分類が新たに可能になれば、パースの自動生成だけでなく、パースの自動評価やラップバトルの勝敗判定の補助などへの応用も切り拓ける。

3. アンサーの類型化

アンサーを類型化するために、まず「対戦相手のラップ文」と「そのラップ文に対するアンサーとして成立するラップ文」のペアデータを作成する。次にそのデータも踏まえながら、アンサーの返答類型と、返答類型をより詳細に説明するためのサブ返答類型を作成する。更に、ラップ文自体の表現類型も作成する。

3.1 アンサーとして成立するラップ文のペアデータの作成

我々は、Mibayashi ら [19] が作成した、ラップバトルの動画を文字起こしたコーパスを利用した。三つの大会 (UMB^(注1), 凱旋 MCbattle^(注2), 戦極 MCBAT-

(注1) : <https://www.youtube.com/user/umbofficial>

(注2) : https://www.youtube.com/channel/UCe_EvY8GrvYgx8PbwRBc75g

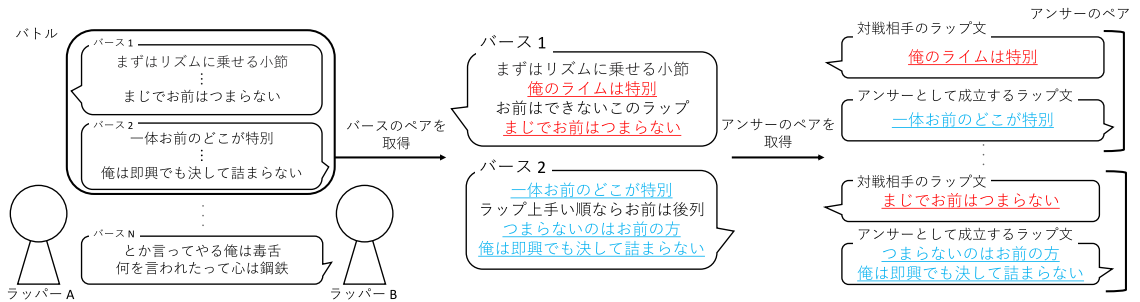


図2 対戦相手のラップ文とアンサーとして成立するラップ文のペアデータの作成方法

TLE^(注3)と一つのテレビ番組（フリースタイルダンジョン）について、それぞれ25件のラップバトルをコーパスからランダムにサンプリングして、計100件を取得した。

取得した100件のラップバトルから、「対戦相手のラップ文」と「そのラップ文に対するアンサーとして成立するラップ文」のペアを抽出する。まず、図2の左側のように、一つのラップバトルから隣接するバースを取得する。返答バースの各ラップ文がアンサーとは限らないため、返答バースから「アンサーとして成立するラップ文」を抽出し、対戦相手のバースからそれに対応してペアとなるラップ文を抽出する必要がある。例えば図2の右側のように、「一体お前のどこが特別」というアンサーとして成立するラップ文をバース2から抽出し、「俺のライムは特別」というラップ文をバース1から抽出してアンサーのペアを作る。その際、アンサーが複数文にまたがる場合は文同士を連結する。例えば、対戦相手のラップ文「まででお前はつまらない」に対するアンサーが「つまらないのはお前の方」「俺は即興でも決して詰まらない」のように2文で構成される場合、これら2文を「つまらないのはお前の方俺は即興でも決して詰まらない」のように区切り文字なしで連結する。こうして100件のバトル内の全ての隣接バースの組み合わせに対して、ラップバトル視聴歴10年以上の筆頭著者がアンサーとして成立するかを判断しながらペアを抽出し、ラップ文のペアデータ429件を作成した。

3.2 アンサーの返答類型の作成

429件のペアデータ内の事例に加えて、著名なプロのラッパーらがラップバトルの戦略やテクニック、どのような返答をアンサーと判別しているか等に関して

記したラップバトルの解説書[1], [20]を参考にして、アンサーの類型として以下の三つを作成した。

(1) 単語によるアンサー： 対戦相手のラップ文中の単語と関連した単語が含まれている。
具体例（「テンション」と関連した「情熱」が含まれている）：

相手のラップ文：俺のテンションは高いぞ
返答のラップ文：情熱は確かにいいね

(2) ライムによるアンサー： 対戦相手のラップ文中の単語に対してライムとなる単語が含まれている。
具体例（「テンション」に対するライムである「ファッション」が含まれている）：

相手のラップ文：俺のテンションは高いぞ
返答のラップ文：でもファッションは残念だな

(3) 対話によるアンサー： 対戦相手のラップ文の内容に対して反応した内容が含まれている。

具体例（相手の内容を肯定する対話となっている）：

相手のラップ文：俺のテンションは高いぞ
返答のラップ文：そうだな確かにそう見える

これらの「対戦相手のラップ文に対するアンサーとして成立するラップ文」の類型を返答類型と呼ぶ。各返答のラップ文は、複数の返答類型に属することが可能である。どの類型にも当てはまらない返答のラップ文は、「アンサーでない」ラップ文とみなされる。

3.3 アンサーのサブ返答類型の作成

前節で作成した三つの返答類型のそれぞれを更に細かく分類したサブ返答類型を作成する。具体的には、ラップバトル視聴歴10年以上の筆頭著者1名と3年以上の非著者1名の2名が協議しながら以下の手順で実施した。まず3.1で作成したラップ文のペアデータ429件から300件をランダムに選択し、返答類型を参考にして、考えるサブ返答類型の候補を9種類列挙した。それらに対して、MAMAサイクル[5]の(1)

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

表1 作成したアンサーの返答類型（3 種類）及びサブ返答類型（6 種類）とその説明

返答類型	サブ返答類型	サブ返答類型の説明
単語によるアンサー	同じ語	相手のラップ文中の単語を含む
	関連語	相手のラップ文中の単語に関連する単語を含む
返答類型「単語によるアンサー」に該当しない		相手のラップ文中の単語に関連する単語を含まない
ライムによるアンサー	ライムを含んでいる	相手のラップ文中の単語に対するライムを含む
返答類型「ライムによるアンサー」に該当しない		相手のラップ文中の単語に対するライムを含まない
対話によるアンサー	肯定	相手のラップ文の内容に対して肯定している
	否定	相手のラップ文の内容に対して否定している
	肯定でも否定でもない	相手のラップ文の内容に対して肯定も否定もしていない
返答類型「対話によるアンサー」に該当しない		相手のラップ文の内容に対して反応した内容が含まれていない

表2 MAMA サイクルの過程で追加・統合・削除をしたサブ返答類型とその説明

返答類型	サブ返答類型	追加・統合・削除	サブ返答類型の説明
単語によるアンサー	類義語	「関連語」との区別が難しいため削除	相手のラップ文中の単語の類義語を含む
	その他関連語	「類義語」の削除に伴い、「関連語」のみをサブ返答類型として採用したため削除	相手のラップ文中の単語に関連する類義語以外の単語を含む
ライムによるアンサー	ライムを含んでいる可能性がある	追加したが、一致率の向上に寄与しなかったため「ライムを含んでいる」に統合	相手のラップ文中の単語に対するライムを含んでいる可能性がある
対話によるアンサー	肯定と否定の両方	該当する事例が少なかったため削除	相手のラップ文の内容に対して肯定と否定をしている

基準策定, (2) アノテーション作業, (3) 評価, (4) 修正作業の工程を3回繰り返した。その結果, 1種類が追加され, 4種類が統合または削除されて6種類となり, 「単語によるアンサー」に対して2種類, 「ライムによるアンサー」に対しては1種類, 「対話によるアンサー」に対しては3種類のサブ返答類型が得られた。作成したサブ返答類型の概要とその説明を表1に示す。また, MAMA サイクルの過程で追加・統合・削除をしたサブ返答類型を表2に示す。

サブ返答類型について, 具体例を用いて説明する。

(1) 単語によるアンサーのサブ返答類型 相手のラップ文中の単語に対して, どう関連する単語を含むかによって, 以下の2種類のサブ返答類型を作成した。

- 同じ語

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

- 関連語

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

以下のように, これら2種類に該当しない場合には, 返答類型「単語によるアンサー」に該当しない。

相手: 俺のテンションは高いぞ

返答: それなら良かったじゃねえか

(2) ライムによるアンサーのサブ返答類型 サブ返答類型を2種類にできるかをMAMA サイクルで検討したが, 最終的に以下の1種類のサブ返答類型となった。

- ライムを含んでいる

相手: 俺のテンションは高いぞ

返答1: でもファッショんは残念だな

返答2: つまりダンジョンでやろうセッション

以下のように, これに該当しない場合には, 返答類型「ライムによるアンサー」に該当しない。

相手: 俺のテンションは高いぞ

返答1: 楽しそうでなにより

返答2: テンションがどうこう俺は知らんな^(注4)

(3) 対話によるアンサーのサブ返答類型 相手のラップ文の内容に, どう反応するかによって, 以下の3種類のサブ返答類型を作成した。

- 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

- 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

以下のように, これら3種類に該当しない場合には, 返答類型「対話によるアンサー」に該当しない。

相手: 俺のテンションは高いぞ

返答: こいつの地元を知ってるか

(注4): 同じ語を含むだけではライムにはならない。

表3 作成したラップ文の表現類型（11 種類）とその使用例

表現類型	説明	返答の例
ライム	同じ母音列をもつ単語を返答に含むこと	つまりパッションでやろうセッション
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	俺はめちゃくちゃラップ上手い
リスペクト	相手の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	自分の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
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	そんなことで俺は動じない
レベゼン	自分の地元や所属をアピールすること	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話しかけること	お客さんどっちについていくか決めた
表現なし	いずれの表現類型にも属さない	俺はこうやってラップしてる

表4 MAMA サイクルの過程で統合・削除をした表現類型とその使用例

表現類型	説明	統合・削除	返答の例
優勢	自分が有利な立場にあることを主張すること	「自己賛美」に内包されるため 「自己賛美」に統合	会場を沸かした数は俺のほうが多い
開き直り	相手の発言が正しいうえでそれに対して開き直ること	「平気」と「否定」の二つの類型が含まれており 表現類型に相応しくないため削除	たしかに俺はこの前逃げたけどそれがなんだ

3.4 ラップ文の表現類型の作成

前節までの返答類型によって、アンサーが成立するか否かの分類は可能となったが、ラップ文がどのような表現となっているかを分類することはできない。ラップバトルでは、相手を褒める表現よりも不快にする表現の方が評価される傾向にあるなど、ラップ文の表現はアンサーの質を判断する上で重要な情報となる。そこで本節では、アンサーかどうかは問わずに、各ラップ文の表現を分類可能にするための**表現類型**を作成する。例えば

返答：お前のラップは下手くそだ

では、相手の嫌がる内容が含まれているため、このラップ文の表現類型は「ディス」と判別する。

表現類型でも、**3.3**と同様のデータに対して同様の手順で、ラップバトルに精通した2名が協議しながら13種類の表現類型を作成した。それらの表現類型に対して、MAMA サイクル [5] を計3回繰り返すことで、13種類の表現類型のうち2種類を統合または削除し、表3の11種類の表現類型を得た。MAMA サイクルの過程で統合・削除をした表現類型を表4に示す。

4. 返答類型と表現類型のアノテーション

作成した返答類型と表現類型の妥当性を検証するために、**3.1**で作成したラップ文のペアデータ429件に対して、類型ラベルのアノテーション作業をおこなった。アノテータは、クラウドソーシングサービス Lancers にてラップバトル視聴歴3年以上の3名を

採用した^(注5)。アノテータは著者が作成したアノテーション用の Web ページにアクセスし、各ペアに対して、(1) ラップ文のペアに対するサブ返答類型のラベル付け作業と、(2) 返答ラップ文に対する表現類型のラベル付け作業を実施した。これを、アノテータごとに異なる表示順で429回繰り返した。サブ返答類型のラベル付けでは、「単語によるアンサー」(2種類)、「ライムによるアンサー」(1種類)、「対話によるアンサー」(3種類)のそれぞれに対し、該当するサブ返答類型があればそれを一つ付与し、なければ「類型に該当しない」を付与するように指示した。表現類型のラベル付けでは、11種類の中から該当する全ての表現類型を、複数付与するように指示した。ただし「表現なし」を付与する場合は、他は付与できない。

4.1 アノテータ間のアノテーション結果の一致率

3名以上の評価者の一致率が計算可能である Fleiss の κ 係数と Krippendorff の α 係数を用いて、アノテーション結果の一致率を評価した。Fleiss の κ 係数は-1から1までの値を取り、Krippendorff の α 係数は(理論上負の値を取ることが可能だが通常は)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。「類型に該当しない」ラベルを含まない場合(「いずれかの類型」の場合)の一致率

(注5)：ワークを採用する際には、我々が用意したラップバトルの知識に関する四つの質問に回答してもらい、その内容を基に、ラップバトル視聴歴10年以上の筆頭著者が視聴歴3年以上であるかを主観的に判断することで、信頼性の高いワークを採用するように努めた。

の計算では、「類型に該当しない」ラベルが一つでも付けられている文ペアはデータから除外して計算した。

評価結果を表5に示す。まず、「単語によるアンサー」のサブ返答類型ラベルと「類型に該当しない」ラベルを含んだアノテーション作業の一致率は $\kappa = 0.5565$, $\alpha = 0.5569$ と他の返答類型のアノテーション結果の一致率と比較して高い結果となった。これは、Higashinaka ら [12] の研究での専門業者らによる対話破綻ラベルのアノテーション一致率 $\kappa = 0.567$ に匹敵する高い一致率を示している。特に、サブ返答類型「同じ語」は $\kappa = 0.7028$, $\alpha = 0.7030$ と最も高い一致率だったが、これはラップ文のペアに同じ単語があるかどうかを確かめるだけでラベル付けが容易だったからである。しかし「同じ語」でも一致しない事例を調査した結果、

相手：麦の海で溺れてな

返答：いくらだって溺れて来たよ

の下線部のような、「同じ語」と「関連語」の判断が分かれる場合があることが判明した。このようにほぼ同じ語でも、細かな表記の違いでアノテータ間の一致率が低下することがあった。「関連語」の一致率も同様の理由で低下していた。

「ライムによるアンサー」の一致率は $\kappa = 0.3634$, $\alpha = 0.3639$ となった。サブ返答類型を一つしかもたないため、「類型に該当しない」ラベルの一致率は原理上同じ値となるので、表5には記載していない。この結果は、Higashinaka ら [10] による対話破綻の2クラスの分類における $\kappa = 0.396$ に近い一致率であるが、 $\kappa = 0.4$ をわずかに下回ることから、適度なレベルの

一致率と言える。単語によるアンサーほど一致率が高くない原因としては、アノテータによる漢字の読み間違いや、実際に音読をしないとライムであると気づきにくい場合の見落としなどが挙げられる。漢字の読み間違いの具体例としては、「お前は枚方を上げれず終わるぜ」と「イカサマ若しくは枚方」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」という母音列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。更に、見落としの具体例としては、「分かるかな見せつける俺たちだけのクロスオーバー」と「動くローカルクロスオーバー向こうがどうだこうだ」というアンサーペアに対しては、「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。これはテキスト上では、「クロスオーバー（“UOUOOAA”）」と「どうだこうだ（“OUAOUA”）」という母音列になるが、実際に読み上げる場合は、「ドーだこーだ（“OOA”）」のように読むことができる。「こーだ」と読む場合、母音列は“OOA”で一致するためライムと判断できるが、これは実際に音読をしないとライムの判断が難しいため、アノテータ間でラベルの不一致が起きた可能性がある。これらのような、判断が難しいアンサーペアに対して、アノテータの一部が正しく判断できなかったため、一致率が低下したと考えられる。

「対話によるアンサー」のサブ返答類型ラベルと「類型に該当しない」ラベルを含んだアノテーション作業の一致率は $\kappa = 0.3091$, $\alpha = 0.3096$ となった。Higashinaka ら [10] は、対話破綻の3クラスの分類における $\kappa = 0.276$ を比較的低い一致率であると述べており、それに近い $\kappa = 0.3091$ も比較的低い一致率であると言える。比較的低い一致率である原因としては、「否定」と「肯定でも否定でもない」の判別が難しかったことが影響していた。ラベルの組み合わせを調査した結果、「否定、否定、肯定でも否定でもない」が56件、「否定、肯定でも否定でもない、肯定でも否定でもない」が34件あり、アノテータ間でラベルが不一致した組み合わせとして多かった。これは、アノテータに提示した例が明瞭でなかったため、「否定」と「肯定でも否定でもない」の判別が難しくなった可能性がある。アノテータには、「否定」の例として「俺のテンションは高いぞ」という相手の発言に対して、「いや俺

表5 返答類型における3名のアノテータ間のアノテーション結果の一致率。 κ は Fleiss の κ 係数を、 α は Krippendorff の α 係数を表す。下線部は「類型に該当しない」ラベルを含む類型全体での一致率を示す。

返答類型	サブ返答類型	κ	α
単語によるアンサー	同じ語	0.7028	0.7030
	関連語	0.3545	0.3550
	(いずれかの類型)	0.5295	0.5300
	(「類型に該当しない」ラベルも含んだいずれかの類型)	<u>0.5565</u>	<u>0.5569</u>
ライムによるアンサー	ライムを含んでいる	0.3634	0.3639
対話によるアンサー	肯定	0.4343	0.4348
	否定	0.4423	0.4427
	肯定でも否定でもない	0.1578	0.1584
	(いずれかの類型)	0.3878	0.3885
	(「類型に該当しない」ラベルも含んだいずれかの類型)	<u>0.3091</u>	<u>0.3096</u>

のほうが上回ってる」と「いやお前全然テンション低いよ」という 2 種類の返答のアンサーペアを示していたが、「いや俺のほうが上回ってる」という返答が自己賛美をしているようにも受け取れるため、相手の発言を否定しているか不明瞭であったと考えられる。そのため、アノテータによっては「否定」の定義が不明瞭になり一致率が低下した可能性が考えられる。今後は、各類型における例を明瞭になるよう精査していく必要がある。

次に、表現類型のアノテーション作業の一致率を表 6 に示す。11 種類から複数選択をした組み合わせがアノテータ間で完全に同一となる一致率は、一つでも異なると不一致になるので $\kappa = 0.1249, \alpha = 0.1256$ と低かった。一方、どれか 1 種類でも表現類型が全アノテータ間で同じであれば一致とみなす条件だと、 $\kappa = 0.5661, \alpha = 0.5664$ と高い一致率を示している。

各表現類型を個別に検討すると、一致率は幅広く分布していることがわかる。まず、「ディス」と「レペゼン」は一致率が 0.5 台で他の類型よりも高く、判別しやすいラベルだったことがわかる。特に「レペゼン」は地域名や県名を含むことが多く、判別しやすい。次に、「自己賛美」と「リスペクト」の一致率が 0.4 台と中程度であり、どちらも自分か他人の能力を認める肯定的な表現で似ている。そうした似た表現でも、自分に向けた表現なのか、相手に向けた表現なのかが区別できていた。また、「ライム」の一致率は 0.3 弱で少し低く、上記の「ライムによるアンサー」で述べた理由と同様だと考えられる。「否定」と「指摘」はいずれも一致率が低かったが、これらは「ディス」と表現が似

ているため、アノテータが判断に迷った際に「ディス」のみを付与した可能性がある。最後に、「表現なし」は一致率が最も低かったが、これは他の類型と排他的でアノテータ 3 名の中でも計 10 件と極めて少ない付与だったので一致しなかった。

4.2 返答類型と表現類型の割合

アノテーションされた返答類型と表現類型の割合及び後述するカウント数を図 3 に示す。図中の n の値は、一つのラップ文のペアに対してアノテータ 3 名が付与した類型の一致数を示している。返答類型「単語によるアンサー」において、アノテータ 3 名が付与し

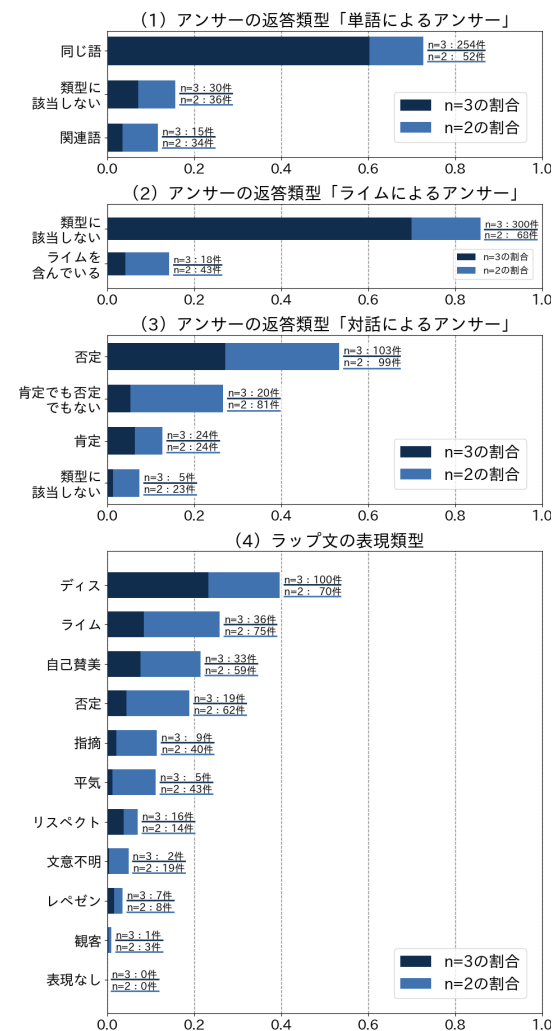


図 3 返答類型（サブ返答類型）と表現類型に対するアノテーション結果の割合（横軸は割合、図中に示す $n = 3$ と $n = 2$ はそれぞれカウントしたラベルの一致数）

表 6 表現類型における 3 名のアノテータ間のアノテーション結果の一致率。 κ は Fleiss の κ 係数を、 α は Krippendorff の α 係数を表す。下線部は「類型に該当しない」ラベルを含む類型全体での一致率を示す。

表現類型	κ	α
ライム	0.2906	0.2912
ディス	0.5013	0.5017
否定	0.2007	0.2013
指摘	0.1116	0.1123
自己賛美	0.4400	0.4404
リスペクト	0.4995	0.4999
文意不明	0.1434	0.1440
平気	0.2064	0.2070
レペゼン	0.5390	0.5394
観客	0.4222	0.4227
表現なし	-0.0078	-0.0070
複数ラベルの完全一致	<u>0.1249</u>	<u>0.1256</u>
どれか一つのラベルでも一致	<u>0.5661</u>	<u>0.5664</u>

たサブ返答類型が「同じ語」「同じ語」「関連語」だった場合、「同じ語」に $n = 2$ が一つカウントされる。全員が「同じ語」なら、それに $n = 3$ が一つカウントされる。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 名以上が同じサブ返答類型を付与した n が 2 以上の割合 ($n = 2$ と $n = 3$ を合算した割合) に基づいて、どのサブ返答類型が多く付与されていたかを考察する。そこで、各返答類型に対して、そのサブ返答タイプのそれぞれの n が 2 以上のカウント数をまず求め、次にその総数で割ることで、各サブ返答タイプの割合を計算した。図 3 の各グラフでは、その割合の降順にサブ返答類型を表示している。表現類型においても同様の計算をおこなった。

まず、返答類型「単語によるアンサー」では、「同じ語」が 72.68% と最も高く、「関連語」の 11.63% と大きな差がある結果となった。これは、対戦相手のラップで使われた単語と同じ語を繰り返す方が、関連語を考えて返答するよりも容易であるためだと考えられる。

次に、返答類型「ライムによるアンサー」では、「ライムを含んでいる」が 14.22% と低く、多くのアンサーにはライムが含まれていないことがわかる。ライムはラップ文化における代表的な特徴の一つであるため、これは予想外な結果だった。しかし、「ライムによるアンサー」は、「単語によるアンサー」と「対話によるアンサー」よりも難易度が高いから出現頻度が低くなっていると考えられ、稀少だからこそラップバトルでライムが高く評価されていることを示唆している。

そして、返答類型「対話によるアンサー」では、「否定」が 53.29% と高い一方で「肯定」は 12.66% と低かった。これは、ラップバトルで自身の優位性を確立する戦略として、対戦相手に対する否定的なラップを用いることが多いことに起因する。また、「対話によるアンサー」に該当しない割合は 0.73% と低く、アンサーとして成立するときには、相手の内容に反応できていることがわかる。

最後に、表現類型については、「デイス」が 39.62% と最も高かった。ラップバトルでは相手に対する否定的な表現や攻撃的な表現が多い傾向にあることが、この結果からも裏付けられた。2 番目に高いのは「ライム」の 25.87% で、これは返答類型「ライムによるアンサー」の 14.22% よりも大幅に高い結果だった。これは、対戦相手のラップ文に対するライムを含むよりも、自身のラップ文中でライムを含むことの方が容易だからである。「自己賛美」も 21.44% と高く、これも

「ライム」と同様に、対戦相手のラップにかかわらず、自身のラップ文中で能力をアピールするのは容易なことを示唆している。次に、「否定」「指摘」などの否定的な表現や攻撃的な表現を含む類型が 10% を超えており、「デイス」と同様にラップバトルではそうした表現が多いことがわかる。「平気」は自身が優位な状態にあること、「リスペクト」は相手を褒めることを表現する類型で、これらは約 5% から約 10% の間となっていて、肯定的な表現より否定的な表現の方が優勢なことがわかる。「文意不明」の約 4% は、即興で的確な表現を考えてアンサーを返すことの難しさを示している。「表現なし」は 0% で、作成した表現類型は多様な表現を十分に網羅できていたと結論づけられる。

5. アンサーの自動分類

対戦相手のラップ文に対する返答のラップ文が与えられたときに、作成した類型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評価実験をおこなった。まず、3 種類の返答タイプの少なくとも一つに該当するか否か（アンサーの有無）を 2 値分類する実験をした。次に、3 種類のそれぞれを対象に、返答のラップ文が、その返答類型に該当するか否かを 2 値分類する実験をした。本章では簡単のため、各返答類型に「該当する」「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。最後に、表現類型に対しても、返答のラップ文を 11 種類の表現類型に分類する実験をした。

5.1 BERT によるアンサーの返答タイプの自動分類

5.2, 5.3, 5.5 の評価実験では、共通の枠組みを用いて自動分類をするため、まずその枠組みを説明する。実験では、前章のアノテーション済みのラップ文のペアデータ 429 件を用いて、それぞれの返答タイプのサブ返答類型（「類型に該当しない」も含む）ごとに、アノテータ 3 名が付与したラベルの多数決を取って正解ラベルとした。3 名全員が別々のラベルを付与したペアデータは各実験において除外した^(注6)。

返答類型を自動分類するモデルとして、BERT [4] 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^(注7)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 64、

(注6)：サブ返答類型（「類型に該当しない」も含む）ごとに多数決を取った結果は異なるため（図 3 の n が 2 以上のカウント数に一致する）、返答類型ごとのペアデータ件数もそれぞれ異なる。

(注7)：<https://huggingface.co/cl-tohoku/bert-base-japanese-v3>

Optimizer は AdamW [21], ロス関数は Cross Entropy Loss, 学習率は $2e-6$, 最大文長は 256 とした. 10-fold 交差検証で学習・評価し, 正解ラベルは可能な限り各 fold で同数になるように分割した.

BERT を用いた分類では図 4 のように, 相手のラップ文を 1 文目, 返答のラップ文を 2 文目にして入力した. 相手のラップ文が 2 文以上で構成される場合は, 文同士を区切り文字なしで連結して 1 文とした. BERT に入力する前処理として, 相手のラップ文の先頭に [CLS] トークンを付与し, 相手のラップ文と返答のラップ文の末尾には [SEP] トークンを付与した. BERT から出力される [CLS] トークンに対応する 768 次元ベクトルを, FC 層で 2 次元に変換し, その後バッチ正規化 [22] をおこない Softmax 関数を適用することで分類問題を解いた.

5.2 アンサーの有無の分類

相手のラップ文と返答のラップ文のペアデータ 429 件を用いてアンサーの有無の 2 値分類を実験する. その準備として, まず, 各ペアデータの正解ラベルとして, 三つの返答類型ごとに多数決を取った結果, 三つの返答類型のうちの少なくとも一つの返答類型においてアンサーに該当すれば正例とした. これによりアンサーに該当する正例のペアデータ 412 件を得た. 次に, 2 種類の方法 (「相手のラップ文」「返答のラップ文」それぞれの差し替え) で, それらに対応する負例のペアデータ 412 件を作成した. 「相手のラップ文」を差し替える方法 1 では, 412 件からランダムに選んだ 206 件のペアデータに対して, 「相手のラップ文」が元々含まれていたバースから, 異なる「相手のラップ文」を新たにランダムに取得し, 元の「返答のラップ文」と負例ペアにした. 「返答のラップ文」を差し替える方法 2 では, 方法 1 で使用しなかった 206 件のペアデータに対して, 「返答のラップ文」が元々含まれていたバースから, 異なる「返答のラップ文」を新たにランダムに取得し, 元の「相手のラップ文」と負例ペア

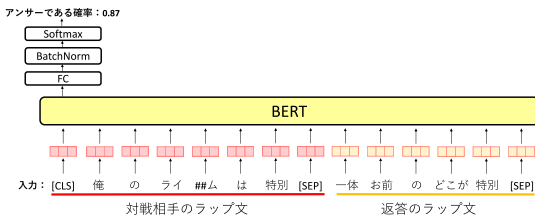


図 4 BERT を用いたアンサーの分類方法

にした. 作成した負例ペアは, ラップバトル視聴歴 10 年以上の筆頭著者が, 三つの返答類型のいずれの観点においても負例になっているかを目視で確認し, 負例になっていない場合は当該バースから負例になるラップ文を取得し直した. 以上により, 正例 412 件と負例 412 件の計 824 件のペアデータを準備して, 10-fold 交差検証で学習・評価した.

分類結果を表 7 の 1 行目に示す. 正解率が 0.7864, F 値が 0.7708 となり高い分類性能を示した. 図 5 左上の混同行列から, 正解ラベルが「アンサーである」と「アンサーでない」の両方に対して分類できたことがわかる. 正解ラベルが「アンサーである」のときに, 予測ラベルが「アンサーでない」と誤分類した 116 件を調査した結果, 最も多かったのが「単語によるアンサー」と「対話によるアンサー」の両方に該当する正例 69 件の誤分類だった. 両方に該当する正例は 298 件なので, そのうち 229 件は正しく分類できていた.

5.3 「単語によるアンサー」の分類

単語によるアンサーの正解ラベルを前述の多数決で付与し, ペアデータ 421 件を取得した. 返答のラップ文がこの返答類型に該当するか否かを 2 値分類するため, 「類型に該当しない」(66 件) 以外の, 「同じ語」「関連語」が付与された正例のペアデータ 355 件を抽出した. 同数の負例のペアデータを準備するために, 「類型に該当しない」の 66 件に加え, 5.2 で準備したアンサーに該当しないことが保証されている負例 412 件からランダムに 289 件を抽出し, 負例のペアデータ

表 7 三つの返答類型に対する分類結果とアンサーの有無に対する分類結果

実験設定	正解率	適合率	再現率	F 値
アンサーの有無	0.7864	0.8315	0.7184	0.7708
単語によるアンサー	0.8549	0.8865	0.8141	0.8488
ライムによるアンサー	0.7377	0.7736	0.6721	0.7193
対話によるアンサー	0.7835	0.8024	0.7521	0.7765

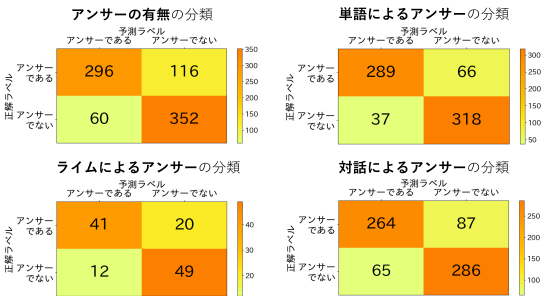


図 5 各分類タスクにおける 2 値分類の混同行列

355 件とした。

分類結果は表 7 に示すように、正解率が 0.8549、F 値は 0.8487 となり、特に高い性能で分類ができていることがわかる。これは、4.2 で述べたように「同じ語」の割合が 72.68% と高いため、ペアデータ内に「同じ語」が含まれているかを比較的容易に判別できていたためだと考えられる。実際に図 5 右上の混同行列において、正解ラベルが「アンサーである」のときに、予測ラベルも「アンサーである」と正しく分類した 289 件を調査した結果、「同じ語」が 280 件と多かった。

5.4 「ライムによるアンサー」の分類

ライムによるアンサーの正解ラベルを前述の多数決で付与し、ペアデータ 429 件を取得した。返答のラップ文がこの返答類型に該当するか否かを 2 値分類するため、「ライムを含んでいる」が付与された正例のペアデータ 61 件を抽出した。同数の負例のペアデータを準備するために、その残りの「類型に該当しない」が付与された 368 件からランダムに 61 件を抽出し、負例のペアデータとした。

この分類では、他の返答類型で用いた BERT に基づく方法ではなく、「相手のラップ文」と「返答のラップ文」のペアにおける母音の最長共通部分文字列（連続して一致する母音列）の長さを基準にルールベースで分類する方法を用いた。これは、BERT が文脈や意味を捉えることはできても、読みや母音などの音韻情報に基づいてライムの有無を判断することが難しく、予備実験で性能が低かったためである。具体的には、まず、同じ単語をライムと誤判定しないように、MeCab^(注8)で形態素解析をおこない、「相手のラップ文」と「返答のラップ文」の両方に同じ単語があれば、「返答のラップ文」から削除した。次に、各文を pykakasi^(注9)を用いてひらがな列とその母音列に変換した。撥音「ん」と促音「っ」はそれぞれ「N」と「Q」という特殊記号として扱い、長音「ー」は 1 文字前の母音と同じにした。母音列は、元の文を MeCab で形態素解析した際の分割単位に分け、文ペア間のそれぞれの分割単位の組み合わせに対して母音の最長共通部分文字列の長さを計算した。その長さが 3 文字以上であれば「ライムを含んでいる」とし、2 文字以下であれば「ライムを含んでいない」とした。

分類結果は表 7 に示すように、正解率が 0.7377、F

値は 0.7192 となり、高い性能で分類ができていることがわかる。図 5 左下の混同行列から、正解ラベルが「アンサーである」と「アンサーでない」の両方に対して分類できており、母音の最長共通部分文字列に基づくシンプルな分類方法でもある程度有効であることがわかる。

誤分類された 32 件のうち、False Positive (FP) が 12 件、False Negative (FN) が 20 件であることを確認し、その原因を分析した結果、最も多かったのは「音が類似する母音列をもつライムの誤分類」であり、これは 32 件中 10 件あった。具体的な例を示すと、アンサーペア中に「冥福 (“EIUU”）」と「迷惑 (“EIAU”）」といった単語のペアがあり、これに対してアノテータは「ライムあり」の正解ラベルを付けており、提案手法は「ライムなし」と予測している。この単語のペアは末尾から 2 番目の母音“U”と“A”が異なるが、単語を読み上げた際の音が類似しているため、アノテータがその類似を認識できて「ライムあり」と判断した可能性がある。その結果、母音が 3 文字以上一致するか否かで判別するシンプルな提案手法において誤分類が起きている。今後はこのような母音同士の音の類似を考慮した分類手法の提案を検討する必要がある。

次に多かったのは「単語中の部分母音列が一致するライムの誤分類」であり、これは 32 件中 9 件あった。具体的な例を示すと、アンサーペア中に「豆腐メンタル (“OUUENAU”）」と「ぶつけたら (“UUEAA”）」という単語のペアがあり、この単語のペアは“UUE”という部分母音列が 3 文字一致している。そのため、提案手法は「ライムあり」と予測したが、アノテータによる正解ラベルは「ライムなし」が付けられていたため誤分類となった。ラップバトル視聴歴 3 年以上のアノテータであっても、これらのペアデータに対して、ライムと判断できていないことから、“UUE”のような単語の途中までで一致する母音列をライムと判断するのが難しかったことがわかる。

次に、32 件中の 6 件は「語尾等の似た表現で母音列が一致するライムの誤分類」であった。具体的な例を示すと、「フリースタイル仕掛けんな」と「順序を間違えんな？」というアンサーのペアでは、相手の発言の「けんな (“ENA”）」と、返答の「えんな (“ENA”）」で 3 文字一致している。しかし、このように語尾を似た表現にするライムは簡易的であり、また、偶然発生する可能性が多いため、ラップバトルにおいては、著者の経験上ライムと判断しないことが多い。実際にアノ

(注8) : <https://taku910.github.io/mecab/>

(注9) : <https://github.com/miura/hrpykakasi>

テータも「ライムなし」の正解ラベルを付けていたのに対し、提案手法は「ライムあり」と予測してしまったため誤分類してしまっている。

次に、32 件中の 5 件は、「MeCab の分割単位による誤分類」であった。具体的な例を示すと、「音の上ではほどほど (“OOOO”) っぽいのが難しい」と「だけど今日はガッツリとほこほこ (“OOOO”) にしてきてくれ」といった「ライムあり」の正解ラベルが付いたアンサーペアに対して、MeCab による文の分割では「ほど (“OO”)」, ほど (“OO”)」と「ほこほこ (“OOOO”)」に分割されてしまい、これらの分割単位で母音一致を計算した際、母音が 3 文字以上一致せず、「ライムなし」と誤分類されてしまっている。

最後に、32 件中の 2 件は、「アノテータのラベル付けミスによる誤分類」であった。これは、アノテータのミスにより、「ライムあり」の正解ラベルを付けるべきアンサーペアに対して、「ライムなし」の正解ラベルが付いているものである。例えば、「お前枕元に千羽鶴」と「ゲン担ぐよりも千羽鶴」というアンサーペアは、相手の発言の「千羽鶴 (“ENAUU”)」と、返答の「ゲン担ぐ (“ENAUU”)」で母音が 5 文字一致しているライムであるが、正解ラベルには「ライムなし」が付けられている。これは、相手の発言と返答の両方に「千羽鶴」が含まれていることから、アンサーペア間でのライムなのか、返答内のみでのライムなのかを判断できなかった可能性がある。もう 1 件の事例についても同様である。

5.5 「対話によるアンサー」の分類

対話によるアンサーの正解ラベルを前述の多数決で付与し、ペアデータ 379 件を取得した。返答のラップ文がこの返答類型に該当するか否かを 2 値分類するため、「類型に該当しない」(28 件) 以外の、「肯定」「否定」「肯定でも否定でもない」が付与された正例のペアデータ 351 件を抽出した。同数の負例のペアデータを準備するために、「類型に該当しない」の 28 件に加え、5.2 で準備したアンサーに該当しないことが保証されている負例 412 件からランダムに 323 件を抽出し、負例のペアデータ 351 件とした。

分類結果は表 7 に示すように、正解率が 0.7834、F 値は 0.7764 となり、高い性能で分類ができていることがわかる。図 5 右下の混同行列から、正解ラベルが「アンサーである」と「アンサーでない」の両方に対してバランス良く分類できたことがわかる。これは、BERT の事前学習モデルが文同士の繋がりを的確に考

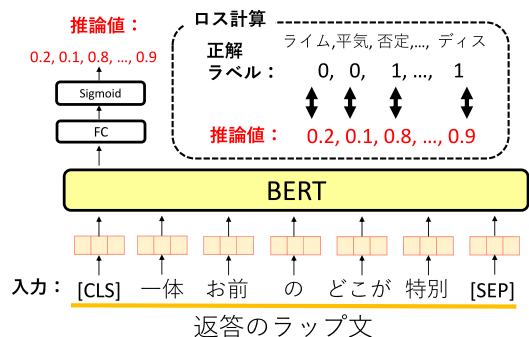
慮できるので、「相手のラップ文」の内容に対して反応した「返答のラップ文」になっているかどうかを、対話という文脈的な特徴を捉えて分類できたためだと考えられる。

5.6 BERT による表現類型の自動分類

ラップ文の表現類型の自動分類では、ある返答のラップ文に対して「表現なし」以外の 10 種類の表現類型が正解として該当するので、10 種類の表現類型を同時に予測するマルチラベリング問題として解いた。BERT 自体は 5.1 と同一だが、図 6 に示すように、返答のラップ文のみを入力とし、その 10 種類の表現類型を表す 10 次元のベクトルを出力する。前処理として、入力文の先頭に [CLS] トークンを付与し、末尾には [SEP] トークンを付与した。BERT から出力される [CLS] トークンに対応する 768 次元ベクトルを、FC 層で 10 次元に変換して出力する。過学習を防ぐために Dropout を 10% の確率でおこない、Binary Cross Entropy Loss によって正解の 10 種類の表現類型との誤差を計算した。ハイパーパラメータは 5.1 と同様にした。

「表現なし」以外の 10 種類の表現類型のそれぞれで、返答のラップ文 429 件に対して、アノテータ 3 名のうち 2 名以上が同じ表現類型を付与している場合に、その表現類型を正解ラベルとした。これで正解ラベルが付かなかった 37 件を除いたラップ文 392 件を用いて、10-fold 交差検証で学習・評価した。

こうして 10 クラスのマルチラベリング問題として分類し、各表現ラベルにおける評価指標値を計算した。評価指標値の計算では、BERT から得られた予測結果である 10 次元のベクトル内の要素に対して、0.5 以上を 1 とし、0.5 未満を 0 として扱った。予測結果と正解ラベルを用いて、各次元ごとに True Positive (TP), False Positive (FP), True Negative (TN), False



Negative (FN) を計算し、マクロ平均を取った結果、F 値が 0.4154、適合率が 0.2999、再現率が 0.6826 となった。正解率については、10 種類全ての予測結果が完全に一致する場合のみを正解としてカウントし、その数をデータの総数 392 件で割って計算した結果、正解率が 0.1532 となった。候補のラベルが 10 種類と多い中で、10 種類全ての予測結果が完全に一致することは難しいため、正解率は想定通り低かった。しかし F 値は 0.4154 を示しており、本分類モデルは中程度の性能で分類が可能であることから、表現類型においてもある程度の自動分類がおこなえる可能性を示せた。

6. む す び

本論文では、ラップバトルで重要なアンサーの類型化に取り組み、それに基づく自動分類をした結果を述べた。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「対戦相手のラップ文」と「そのラップ文に対するアンサーとして成立するラップ文」のペアに基づいて、3 種類の返答類型、6 種類のサブ返答類型、11 種類の表現類型からなる、ラップバトルにおけるアンサーの類型を世界で初めて提案した。

- 提案したアンサーの類型を基に、クラウドソーシングを用いたアンサーのタイプのラベル付けタスクを実施し、3 種類のサブ返答類型と 5 種類の表現類型において中程度の一致率があることを示した。

- 提案したアンサーのタイプの自動分類が可能なことを示した。

今後は、本研究で作成した返答類型と表現類型に基づいて、その出現割合を基に類型ごとのアンサーの難易度を求める手法や、それに基づいて、ラップバトルの勝敗の評価基準の参考になるようなアンサーの評価値を自動算出できる手法の研究開発にも取り組む予定である。

謝 辞 本研究は JSPS 科研費 JP21H03775, JP21H03774, JP21H03554, JP22H03905 の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] DARTREIDER, MC バトル史から読み解く日本語ラップ入門, KADOKAWA, 2017.
- [2] V. Sykari, "Interactive oral composition: Resources, strategies, and the construction of improvised utterances in a Finnish freestyle rap battle," *J. American Folklore*, vol.132, no.523, pp.3–35, 2019.
- [3] K. Krippendorff, "Computing Krippendorff's alpha-reliability," *Departmental Papers (ASC)*, 43, pp.1–10, 2011.
- [4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional Transformers for language understanding," *Proc. NAACL'19*, pp.4171–4186, 2019.
- [5] J. Pustejovsky, H. Bunt, and A. Zaen, "Designing annotation schemes: From theory to model," *Handbook of Linguistic Annotation*, pp.21–72, 2017.
- [6] 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.
- [7] 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.
- [8] J.R. 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.
- [9] V. Susana, K. Uzay, and S. João, "Cohen's kappa coefficient as a performance measure for feature selection," *Proc. WCCI'10*, pp.1–8, 2010.
- [10] 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," *Proc. EMNLP'15*, pp.2243–2248, 2015.
- [11] J.L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol.76, no.5, pp.378–382, 1971.
- [12] R. Higashinaka, K. Funakoshi, M. Araki, H. Tsukahara, Y. Kobayashi, and M. Mizukami, "Towards taxonomy of errors in chat-oriented dialogue systems," *Proc. SIGDIAL'15*, pp.87–95, 2015.
- [13] 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.
- [14] 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.
- [15] R. Dey and F.M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," *Proc. MWSCAS'17*, pp.1597–1600, 2017.
- [16] H.S. Alim, J. Lee, L.M. Carris, and Q.E. 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," *Proc. IJCAI'15*, pp.2524–2530, 2015.
- [18] R. Savery, L. Zahray, and G. Weinberg, "Shimon the rapper: A real-time system for human-robot interactive rap battles," *Proc. ICCV'20*, pp.212–219, 2020.
- [19] 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," *Proc. CMMR'23*, pp.30–41, 2023.
- [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,” Proc. ICLR’19, pp.1–8, 2019.
- [22] V. Kocaman, O.M. Shir, and T. Bäck, “Improving model accuracy for imbalanced image classification by adding a final batch normalization layer,” Proc. ICPR’21, pp.10404–10411, 2021.

(2024 年 5 月 16 日受付, 10 月 21 日再受付,
2025 年 1 月 20 日早期公開)



三林 亮太

2022 兵庫県立大学大学院応用情報科学研究科博士前期課程了。修士（応用情報科学）。現在、兵庫県立大学大学院情報科学研究科博士後期課程在学中。日本データベース学会学生会員。



佃 洸摂

2014 京都大学大学院情報科学研究科博士後期課程了。博士（情報学）。現在、産業技術総合研究所主任研究員。主に音楽コンテンツ及びユーザ生成コンテンツを対象とした情報推薦・情報探索の研究に従事。情報処理学会、日本データベース学会各会員。



渡邊 研斗

2018 東北大学大学院情報科学研究科博士後期課程了。博士（情報科学）。現在、産業技術総合研究所人間情報インタラクション研究部門主任研究員。自然言語処理、音楽情報処理、歌詞情報処理の研究に従事。情報処理学会、言語処理学会各会員。



中野 倫靖

2008 筑波大学大学院図書館情報メディア研究科博士後期課程了。博士（情報学）。現在、産業技術総合研究所研究グループ長。情報処理学会、日本音響学会各会員。2009 情報処理学会山下記念研究賞（音楽情報科学研究会）、2013 Sound and Music Computing Conference (SMC 2013) The Best Paper Award, 2018 IEEE Pacific Visualization Symposium (PacificVis 2018) Honorable Mention Poster Award 等各受賞。



後藤 真孝（正員：フェロー）

1998 早稲田大学大学院理工学研究科博士後期課程了。博士（工学）。現在、産業技術総合研究所首席研究員。2009～2017 に IPA 未踏 IT 人材発掘・育成事業 PM, 2016～2022 に JST ACT-I「情報と未来」研究総括を兼任。現在、JST 創発 PO, 日本学術会議連携会員、統計数理研究所客員教授、筑波大学大学院教授（連携大学院）等を兼任。日本学士院学術奨励賞、日本学術振興会賞、市村学術賞等、69 件受賞。



山本 岳洋（正員）

2011 京都大学大学院情報科学研究科博士後期課程了。博士（情報学）。現在、兵庫県立大学大学院情報科学研究科/社会情報科学部准教授。主に情報検索におけるユーザインタラクションやユーザ理解に関する研究に従事。情報処理学会、日本データベース学会、ACM 各会員。



大島 裕明（正員）

2007 京都大学大学院情報科学研究科博士後期課程了。博士（情報学）。現在、兵庫県立大学大学院情報科学研究科/社会情報科学部准教授。主に情報検索、ソーシャルコンピューティング、デザイン学の研究に従事。情報処理学会、ACM、日本データベース学会各会員。