

台本中の動作表現に基づく アニメーション原画検索システムの提案

藤川 雄翔^{1,a)} 杉江 直哉^{1,b)} 畑 玲音^{1,c)} 松下 光範^{1,d)} 山西 良典^{1,e)}

概要: アニメ制作にかかせないキャラクターやオブジェクトの動作を複数枚の絵で描く原画は、キャラクターの特徴やシチュエーションといったコンテキストを考慮した動作の描き方を参考にする上で有用であると考えられる。しかし現状では、描かれている動作が原画にメタデータとして付与されておらず、大量の原画の中から原画を描く制作者が企図する動作に類似した動作が描かれた原画を探し出すことは困難である。また、1作品に用いられる原画の枚数が膨大なことから、原画に人手でメタデータを付与することは莫大な時間と労力を要する。そこで本研究は、原画に対し動作を表す言語表現である動作表現をメタデータとして機械的に付与することで、制作者が企図する原画を効率的に探すことを支援することを目指す。本稿では、台本に記載される画面の内容を説明した文章である画面説明文に動作表現が含まれる点に着目した。画面説明文から抽出した動作表現をメタデータとして対応する原画に付与することで、動作表現で検索可能なアニメーション原画の検索システムの構築を試みる。

キーワード: アニメ, 原画, 動作表現, 検索

1. はじめに

日本では、アニメーション（以下、アニメ）が重要なエンタテインメントの形態の1つとして位置づけられている。アニメ産業レポート 2023 [8]によると、インターネットの普及とストーリーミングサービスの台頭により、日本で制作されたアニメは世界中で容易に視聴可能になった。その結果、現在では、アニメ産業は急速に拡大しており、海外を軸に 2012 年から 2022 年の 10 年間で市場規模が 1.3 兆円から 2.9 兆円と 2 倍以上になった。国内のテレビで放映されるアニメに限っても、2022 年には 317 本の作品が制作されており一大産業となっている。

アニメ制作は、大別すると (1) 企画やシナリオなどのプロダクション、(2) 絵コンテや動画などのプロダクション、(3) 編集や音響などのポストプロダクションの 3 工程で構成される [7]。これらの工程は、ベルトコンベア方式で進行し、原則として一度完了した工程への逆戻りはできない。制作過程では、自然言語、静止画、動画、音響など多くの

データを扱い、30 分のテレビアニメであっても 200 ~ 300 人と関わる人数が多い [13] が、予算と納期の双方の制約により、各工程において多くの時間を割くことは困難である。アニメは全て紙と鉛筆のみのアナログ環境で制作されていたが、情報技術が進化したことで、幅広い業界におけるデジタル化と同様に、アニメ産業においても予算と時間の確保の観点からデジタル技術の導入が始まっている [7]。これにより、作業効率の向上やエフェクトの表現の自由度が拡大したが、一方で、デジタル化に踏み込めない工程も複数存在する。そのため、プリプロダクションの段階においてアニメクリエイターを支援する研究がこれまでに行われている (e.g., [6], [9], [11])。特に、アニメの品質に重要な影響を及ぼす原画制作の過程では、繊細なタッチが求められることから、いまだに紙と鉛筆等の物理的媒体を利用することが多く [3]、原画制作には膨大な時間を要する。

アニメで表現される動作は、原画制作の段階で大きく決定される。原画はアニメにおける動きの要所を描いた素材であり、これをもとにアニメが制作される。例えば、人がジャンプするシーンを描く場合、(1) ジャンプする前の屈伸、(2) 上昇途中、(3) 最高点に達した瞬間、(4) 降下途中、(5) 着地の瞬間、の 5 つの段階が原画として用意される。その後、これらの原画の間に「中割り」と呼ばれる補助的な絵を挿入することで、より滑らかな動作となり、臨場感を

¹ 関西大学
Kansai University, 2-1-1 Ryozenji, Takatsuki, Osaka 569-1095, Japan
a) k920357@kansai-u.ac.jp
b) k831248@kansai-u.ac.jp
c) k223167@kansai-u.ac.jp
d) mat@res.kansai-u.ac.jp
e) ryama@kansai-u.ac.jp

与えることができる。アニメにおけるキャラクターの動作は同一の動作であっても、そのキャラクターの属性や性格、そのキャラクターの置かれた状況や感情に応じて描かれ方は異なる。例えば、キャラクターが走る場面であった場合、活発な少年は手足を大きく、素早く動かして胸を張った綺麗なフォームで走るが、運動が苦手な少年は息を切らしながらフラフラと走るといった描画の違いが考えられる。このように物語のコンテキストに応じてキャラクターの動作の描き方を変更することにより、物語を視覚的に語るが可能となる。

原画を描く制作者にとって、原画制作の前工程である絵コンテやレイアウトの情報だけではコンテキストを考慮しながら動作を描くことは容易ではない。このような場合、制作者は他の作品の原画を参考にすることがあるが、大量の原画の中から制作者が企図する動作に類似した動作が描かれた原画を探し出すことは困難である。さらに現状の原画にはメタデータが付与されておらず、その作業を人手で行う場合は膨大な時間と労力を要することが想定される。このようなメタデータが付与されていないコンテンツからユーザの求めるシーンに類似したシーンの抽出や検索を試みた研究はこれまでも行われている (e.g., [1], [5], [12])。

本研究は、動作を表す言葉 (以下、動作表現) から原画の検索を可能にすることを目的とする。本稿では動作表現から目当ての原画と同一の画面切り替えの区切り (以下、カット) に属する他の原画を組み合わせた動画が検索可能か検証を行う。原画に対し動作表現をメタデータとして機械的に付与するために、絵コンテで描かれている内容を説明する台本に記載された文章 (以下、画面説明文) に動作表現が含まれる点に着目した。アニメの原画は、カットごとに分けたカット袋で管理されており、それぞれのカットには番号が与えられている。画面説明文から抽出した動作表現をメタデータとして同じカットの原画に付与することで、動作表現を検索キーワードとしたアニメ原画の検索システムの構築を試みる。

2. 動作表現の抽出と原画への付与

2.1 使用する素材

本研究では、トリガーデータセット [4] にある、原画と画面説明文が記述される台本を使用した。原画は動きのポイント (動き始め、重要な中間ポイント、動き終わり) が描かれた絵であり、カットごとにフォルダ化されている。台本は、声優がアフレコを行う時に使用する素材であり、カット番号、画面説明文、登場キャラクターの音声文の情報を有している。

本稿ではカット番号と画面説明文から動作表現を抽出し、メタデータとして対応する原画に機械的に付与する。

2.2 動作表現の抽出

画面説明文から動作表現の抽出を行う。形態素解析ツールである MeCab を用いて、各カットの画面説明文を形態素解析し、抽出する動作表現を機械的に選定した。加えて、不要な動作表現を排除するためにストップワードリストを独自に設定する。抽出条件は以下のように決定した。

- MeCab による形態素解析をした後、動詞を抽出
- 「れる」や「せる」等の助動詞は除外
- (名詞 or 副詞) + サ変動詞「する」の複合語は抽出

ユーザが入力する検索キーワードは日本語の口語文法における動詞を想定し、動詞の終止形で抽出する。そのため、受け身や可能等の意味がある助動詞「れる」や使役の意味がある助動詞「せる」等は抽出対象から除外した。また、サ変動詞「する」と名詞や副詞が組み合わさった複合語 (e.g., “破壊する”, “ゴソゴソする”) はユーザが検索キーワードとして使用する可能性が高いと判断し、これらも抽出対象とした。抽出する際には、後述する類似する動作表現を検索する際の処理を考慮し、名詞または副詞の部分のみ抽出する。MeCab は最小の意味単位に分解するツールであるため、複合語を抽出する際は検索キーワードとして想定外の単語に分解される可能性がある。このような場合は原画を確認し、原画で中心的に描かれる動作として適切な動作表現のみをメタデータとして扱う。例えば、「踊り出す」という動詞は「踊る」と「出す」に分解される。この場合、検索キーワードとして想定される「踊る」を抽出し、「出す」を除外する。また、MeCab は非標準的な言葉使いや口語表現に対しては誤解析をすることがあるため、必要な動詞のみを抽出し、残りをストップワードとして除外する。例えば、「防いでる」は「防ぐ」、「い」、「て」、「いる」に分解されることが望ましいが、「防ぐ」と「でる」と誤解析したため、「防ぐ」を抽出し、「でる」を除外する。動作表現を抽出する際は各カットに 1 つの動作表現とは限らず、必要に応じて複数の動作表現を抽出する。これらの処理により、ユーザの検索効率と精度の向上を図り、システムの有効性を高めることが可能となる。

2.3 データベースの作成

画面説明文から抽出した動作表現をメタデータとして対応する原画に付与するためにデータベースを作成する。データベースには、(1) カット番号、(2) 原画クリップ (3 章で後述)、(3) 画面説明文、(4) 画面説明文から抽出した動作表現、を集約する。原画には 17 カットの欠損が確認されたが、これらのカットには図 1 に示す画像を挿入することで代替処理を行なった。作成したデータベースの一部を表 1 に示す。

3. 動作表現に基づくアニメ原画の検索手法

各カットに対して複数枚の原画が独立して検索結果とし

表 1 データベースの一部

Cut	Clip	Script	Verbs
1	1.gif	野外ライブステージ PD	□
~	~	~	~
309	309.gif	慌てて逃げるアッコ達 尖塔ドーム内に逃げ込む 猛スピードで外を横切るドラゴン	['慌てる', '逃げる', '逃げ込む', '横切る']
310	310.gif	ドーム内の 2 人なめ通過するドラゴン	['通過']
311	311.gif	床に叩きつけられるアッコ, ロッチ	['叩きつける']
312	312.gif	塔の上に降りるドラゴン	['降りる']
~	~	~	~
398	398.gif	一同から学校全景に TB T「おしまい」 FO 黒味	□

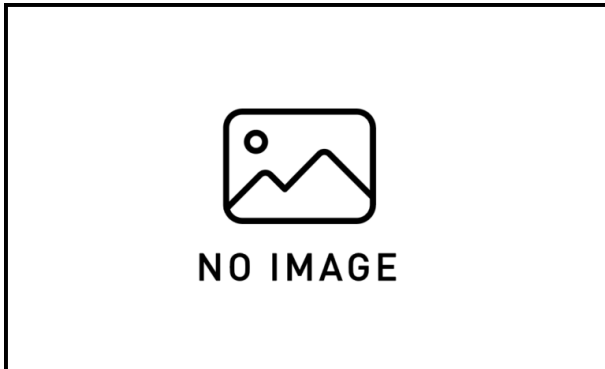


図 1 欠損カットに対して挿入する画像

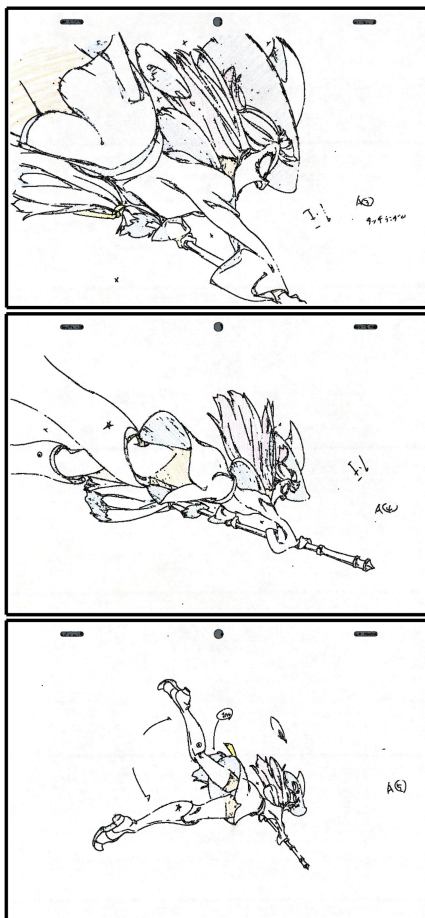


図 2 カット 92「真っ逆さまに落ちるアッコ」([4]より引用)

て表示された場合、それらをコマ送りで見たとときに、どのようなアニメ効果が得られるかを想像することは困難である。そのため、検索結果として提示された原画がどのような動作の一部となるのかを視覚的に理解できることが重要である。また、動作表現をキーワードとして原画を検索する場合、入力した動作表現とメタデータとして付与された動作表現が必ずしも一致するとは限らない。言語には同じ意味を持つ言葉や似た意味を持つ言葉が存在することから、検索キーワードと類似した動作表現であれば検索できることが望ましい。さらに、提案手法では画面説明文から抽出した動作表現をメタデータとして機械的に付与するため、検索結果として提示された原画がユーザの描きたいキャラクターの特徴やシチュエーションといったコンテキストに応じた動作と異なる可能性がある。そのため、検索結果として提示された原画のコンテキストを理解し、企図する動作に応じているかを判断できることが求められる。

これらの要件を踏まえて、制作者が企図するコンテキストに応じた動作が描かれた原画を動作表現により検索可能にするための本システムに必要な仕組みを (1) 原画クリップの作成, (2) 類似した動作表現の検索, (3) 画面説明文の提示, の 3 点に定めた。(1) では、それぞれのカットごとに原画を短い動画 (以下, 原画クリップ) にすることで原画で描かれる動作の視覚的な把握を可能にする。(2) では、動作表現を分散表現として扱うことで類似した動作表現による検索を可能にする。(3) では、原画と画面説明文を同時に提示することで原画のコンテキストの理解を可能にする。

3.1 原画クリップの作成

1つのカットは複数枚の原画で構成されており、それぞれの原画で描かれる動作の一部分を並べて見るだけでは、それらがどのような動作を創り上げるのかを想像することは困難である (図 2 参照)。そのため、動作が描かれた複数枚の原画を実際に放映された完成作品と近い形である動画にして提示すると、ユーザとしては原画に描かれる動作を容易に理解することができる。提案手法では、それぞれのカットにおいて、カットを構成する複数枚の原画を一定間

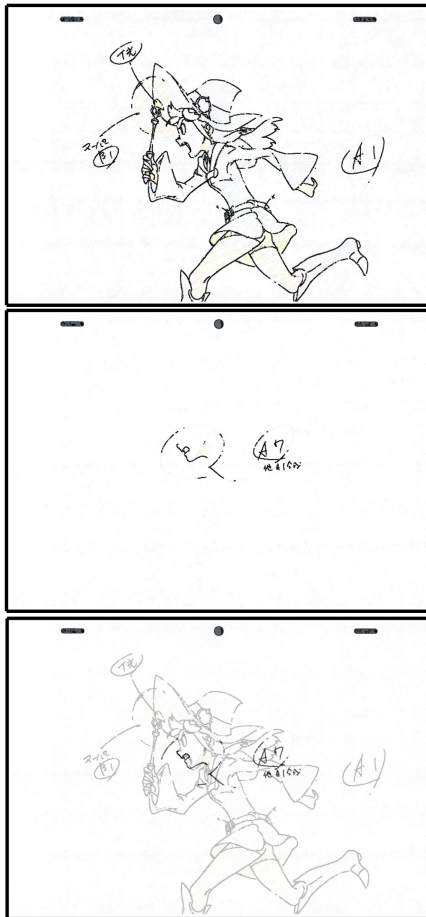


図 3 カット 151 の原画合成 ([4] より引用)

隔で切り替え、原画クリップにすることで、原画で描かれる動作の可視化を図る。実際に放映される完成作品では、原画と原画の間に中割りが入り、一般的に 1 秒間に 24 枚の原画と中割りを切り替えるため、中割りの存在も考慮し、1 秒間に 6 枚の原画を表示させる原画クリップを作成した。

3.1.1 原画合成の検討

原画制作において、前後の原画との微細なズレを排除するために、直前の原画の上に新規に描く原画を重ね合わせ、下方から照明を当てながら作画を行う手法がある。これにより、上の原画を透視しながら描くことが可能となり、動作の連続性と一貫性を維持することができる。このプロセスでは、同じ動作を繰り返すシーンにおいては異なる部分のみを作画するため、特定の部分のみが原画に描かれていることがある。例えば、12 枚の原画で構成された走るカット（カット 151）において、1～6 枚目の原画と 7～12 枚目の原画は同じ動作の繰り返しであるが、1～6 枚目は口を動かしながら走っており、7～12 枚目は歯を食いしばって走っているという違いがあり、実際にこのカットでは 7～12 枚目の原画はキャラクターの口元だけが描かれていた。このように、同じ動作を繰り返す場合は差分のみが作画されるため、データセットの原画をそのまま原画クリッ

プにするだけではどのような動作を表しているのかを理解することが難しくなる。これを踏まえ、原画の合成も試みた。手順は以下の通りである。

- (1) 合成する 2 つの原画の透過度を下げる
- (2) 2 つの原画を合成する
- (3) コントラストと彩度を上げる

原画合成は、全体が描かれている原画の上に差分のみが描かれている原画を重ね合わせて一枚の画像とする。重ね合わせる原画の組み合わせは原画に記されたメモ（e.g., 他 A1 合成）を参考に決定した。しかし、線の濃度を変更せずにそのまま貼り付けた場合、線と線の重なりが目立つことにより見ることが困難である。そのため、合成前に原画の透過度を下げることで比較的に見やすくすることができる。今回は透過度を全体が描かれている原画を 0.3、差分のみが描かれている原画を 0.7 に設定した。さらに、合成後の原画のコントラストと彩度を上げることで合成前の原画の濃度に近づけることを目指す。これらの操作をすることにより、原画クリップでの動作の連続性の維持を可能にした（図 3 参照）。

3.2 類似した動作表現の検索

メタデータとして付与した動作表現は台本中の画面説明文から抽出したものである。そのため、台本の作家が執筆した動作表現とユーザが検索時に入力する動作表現は必ずしも一致するわけではなく、相違が生じる可能性がある。この相違は、同義語や類義語の存在に起因する。これらの同義語と類義語を検索対象に含めることで、台本の作家とユーザ間の動作表現の相違を縮小することを試みる。提案手法では、動作表現を分散表現として扱うことで類似した動作表現による検索を可能にする。動作表現の分散表現は日本語 Wikipedia エンティティベクトル [2] を用いることで獲得し、検索キーワードと類似度の高い上位 10 単語を検索対象に含めることで類似した動作表現の検索を実現する。

3.3 画面説明文の提示

検索結果で提示された原画から制作者が企図する動作と類似した動作が描かれている原画を選ぶには、それぞれの原画のコンテキストを把握する必要がある。しかし、提案手法では画面説明文（図 4）から抽出した動作表現をメタデータとして機械的に付与するため、検索結果として提示される原画はユーザの企図するコンテキストに応じた動作と異なる可能性があり、検索結果として原画のみを提示する場合は選別が困難である。このことから提案手法では、検索結果として得られた原画の前後の画面説明文を読むことが原画のコンテキストの理解に有用であるという仮定の下、各カットに対応した画面説明文とその前後 2 カットず

つの画面説明文を原画クリップと併せて提示した。これにより、原画が描かれたコンテキストの理解を容易にし、原画に描かれている動作がユーザの企図するコンテキストに応じた動作であるかを判断できるようにすることを目指す。これらの方針に基づいてシステムを実装する。

4. 実装

本稿では、動作表現でアニメ原画を参照できる検索システムを Web アプリケーションとして構築する。本システムは Python3.8.18 を用いて開発し、主要なフレームワークには Flask を採用した。実装したシステムのインターフェースは検索画面、検索結果画面、原画クリップ詳細画面の3つの画面で構成される。

4.1 検索画面

検索画面はユーザからの言葉の入力を受ける最初の画面(図5)である。中央に配置したテキストボックス内に日本語で動作表現を入力し、検索ボタンをクリックすると検索結果画面(図6参照)に遷移する。

ユーザが入力した動作表現以外にも、類似した動作表現の検索も同時に行う。本稿で使用した Word2Vec 学習済みモデルである日本語 Wikipedia エンティティベクトルでは、一般的に漢字で表記があるものは漢字で学習されているため、このような単語をひらがなやカタカナを用いた場合、分散表現を獲得することができず、類似度の計算ができない可能性がある。そのため、ユーザに対しては基本的

8	7	6	5	4	3	2	1	
客席の上をかすめて飛ぶ星々ステージに集まってい	顔を手で覆う	光に照り返されるアッコ	夜空の星が落ちてくる	見上げたまま息を飲むアッコ	夜空に輝く星	客席の幼いアッコ キョロキョロ見回す 何かに気付く	客席の観客達	屋外ライブステージ P D

図4 台本中の画面説明文の一部



図5 検索画面



図6 検索結果画面

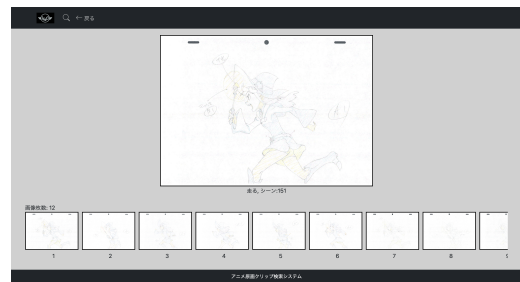


図7 原画クリップ詳細画面

に漢字が存在するものは漢字で書いてもらうよう指示する。

4.2 検索結果画面

検索結果画面は検索結果の一覧を表示する画面である。検索結果として、それぞれのカットごとに、(1) 原画クリップ、(2) シークバー、(3) 画面説明文、をセットとして提示する。

複数枚の原画から作成した原画クリップの下にはメタデータとして原画に付与した動作表現を提示する。原画クリップをクリックすると、クリップ詳細画面(図7参照)に遷移する。作品の最初のカット番号である1を左端に、最後のカット番号(本研究で使用したデータは398カット/話)を右端に配置し、線上における原画クリップのカット番号の位置にドットを配置した。これにより、ユーザはストーリー全体におけるカットの挿入位置をイメージすることができる。画面説明文は、各カットに対応した画面説明文とその前後2カットずつの画面説明文を提示する。また、最上部には、検索キーワードと類似した動作表現のヒット件数を提示し、瞬時に件数を把握できるようにした(図8参照)。

4.3 原画クリップ詳細画面

原画クリップ詳細画面では、検索結果画面で選択した原画クリップとそれを構成する各原画を提示する。原画ク

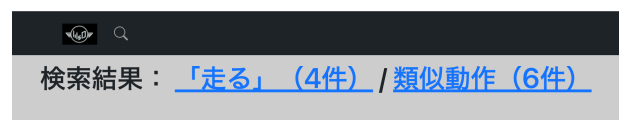


図8 ヒット件数の提示



図 9 原画の拡大表示

リップの下には原画に付与した動作表現を提示する。各カットを構成する原画の枚数が一目で分かるよう、原画の合計枚数を提示するとともに、各原画にはカットの始まりから順に付与した番号を提示した。また、原画枚数が多い場合の対応方法として、横へのスクロール機能を加えた。さらに、原画をクリックすることによる原画の拡大表示機能を加えるとともに、キーボードの左右のボタンで前後の原画も拡大した状態で見ることを可能にした(図9参照)。

5. 実験

提案システムを用いて実験を行うことは下記の2つである。

実験 1 類似動作検索の実験

実験 2 原画のコンテキスト理解の実験

実験 1 では、提案システムを用いて、類似した動作表現を検索対象に含めたことの有効性を検証する。実験協力者が動作表現を用いて原画を検索することが可能かどうか、またどのような動作表現を用いて原画を検索するかを測る。実験 2 では、原画クリップと併せて画面説明文を提示することの有効性を検証する。前後のカットの画面説明文の提示が原画のコンテキストの理解を促進するかを測る。

5.1 類似動作検索の実験

20歳以上の男女14名の実験協力者を対象にユーザ実験を行った。実験では、まず、実験協力者にシステムの操作方法について説明した。システムの操作方法については、動作表現を検索キーワードとすること、検索結果画面の画面構成、類似した動作表現も検索対象であること、原画クリップ詳細画面での原画の拡大表示方法、画面遷移の方法を説明した後、自由にシステムを触ってもらった。次に、実験協力者に「実験協力者はアニメーターの職業を目指してアニメの原画を描く練習している」という内容のシナリオを聞いてもらった。これは、実験タスクを説明する際、実験者の表情や声のトーンの変化などの非言語情報により、実験協力者ごとの環境が異なることを防ぐため、説明文を事前に音声で録音して行った。最後に、実際に放映された完成作品の中の異なる3つのカットの映像を提示し、シ

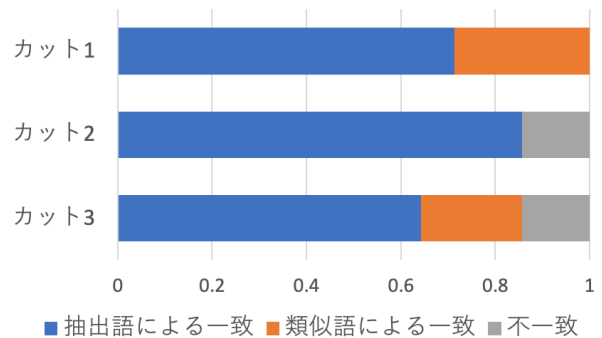


図 10 類似した動作表現の有効性

ステムを用いてそれらが描かれている原画を探索してもらった。

カットごとの検索結果を図10に示す。画面説明文から抽出した語彙によって正しいカットを選択した場合を抽出語による一致、類似した動作表現によって正しいカットを選択した場合を類似語による一致、それ以外を不一致として評価した。その結果、抽出語による一致が73.8%、類似語による一致が16.7%、不一致が9.5%であった。

5.2 原画のコンテキスト理解の実験

類似動作検索の実験とは別の20歳以上の男女14名を実験協力者として実施した。実験協力者には、3つのカットにおいて、原画クリップのみ、または原画クリップとそのカットの前後それぞれ2カットずつの画面説明文を見て、原画クリップで描かれる動作を具体的に記述させた。原画クリップのみを見て動作を記述する群(以下、単一資料提示群)と、原画クリップと画面説明文を見て動作を記述する群(以下、複合資料提示群)に無作為に7名ずつ振り分けていった。回答後には、複合資料提示群の実験協力者を対象に、画面説明文の提示は原画のコンテキストの理解に有効的であったかを問うインタビューを行った。その後、得られた回答が原画のコンテキストを妥当に反映した表現であるかを検証するため、各カットに対して5名の評価者に独立で5段階評価をしてもらった。

カットごとの単一資料提示群と複合資料提示群の評価の平均値を図11に示す。実験の結果、カット1とカット2では、複合資料提示群の平均値が単一資料提示群の平均値を上回ったが、カット3では複合資料提示群の平均値が単一資料提示群の平均値を下回った。そのため、平均値の単純比較では大きな違いは確認されなかった。次に、単一資料提示群の平均値が複合資料提示群の平均値を上回っているという帰無仮説の下、2群の平均値に有意差があるかをWilcoxonの符号順位検定により調べた。その結果、 $p = 0.900$ であり帰無仮説は棄却できず、2群の平均値に有意差は認められなかった。

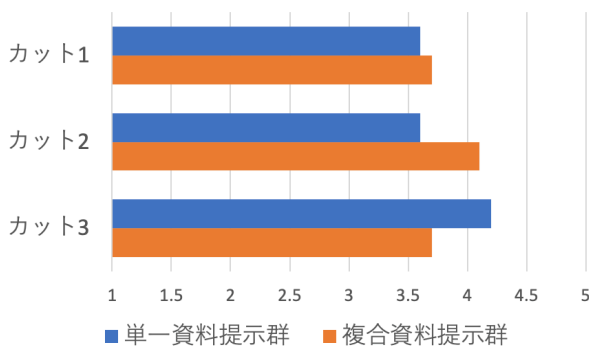


図 11 コンテキストの理解度

5.3 考察

類似動作検索の実験では、類似した動作表現から正しいカットを選択した実験協力者が確認されたことから、類似した動作表現を検索対象に含めることの有効性が示唆された。不一致と評価された検索キーワードには「びっくりする」という動作表現があった。この語は抽出語である「驚く」と同じ動作であるが単語分散表現の中に存在していなかったために検索できなかった。これは単語分散表現に含める語彙の拡充により解決できると考えられる。

原画のコンテキスト理解の実験では、単一資料提示群の平均値と複合資料提示群の平均値に有意差は認められなかった。一方で、回答後のインタビューから「画面説明文にある“PAN”や“なめ”といった専門用語がよく分からなかった」、「断片的な文章で読みづらかった」といった意見があったことから、画面説明文の提示が原画のコンテキストの理解において十分ではない可能性が示唆された。画面説明文には、主にキャラクターの動作は記述されるが、キャラクターの心情や動作の経緯の記述が不十分であるため、原画のコンテキストの理解に寄与しなかったことが推察される。

6. 展望

本稿で使用した台本に記載される画面説明文は、声優が各カットの内容を大まかに把握するためのものであるが、

縮まったドラゴンが放出した魔力を再吸収して石化していく。
落下していく石を見ながら。(すでに帯の上)
アッコ「ドラゴンが石に？」
ダイアナ「なるほど、魔竜石とはよくいったものですね」
アッコ「そうかあれが」
ダイアナ、アッコの持つ弓が気になる様子。コホンと咳して。
ダイアナ「でも最後のシャイニーアルク、あなたには見事でしたね」
アッコ「え？ダイアナ、シャイニーアルク知ってるの？」
ハッとして急に顔が真っ赤になるダイアナ。
ダイアナ「ま、まあ、それくらいは有名なので・・・」
アッコ「そっか、(無邪気に笑い) そりゃそうだよね」
冷や冷やのダイアナ。
アッコ「でもダイアナって意外といい奴だね」
ダイアナ「(すまして) 意外は不要です」
思わず笑うアッコ、釣られて笑うダイアナ。

図 12 脚本の一部

画面説明文を読むことによる原画のコンテキストの理解を促進するには不十分である。今後は、キャラクターの心情や動作の経緯を含んだ文章であり、断片的でなく全体的に構成された脚本(図 12 参照)の利用を検討する。画面説明文にはカットごとに管理されているためカット番号で原画と対応づけることができたが、脚本に記述されている文章はカットごとに管理されていないためカット番号が付与されていない。そのため、脚本をカットごとに区切り、それぞれにカット番号を付与する必要がある。また、本稿では動作表現を検索キーワードとして原画を検索し、原画と併せて提示した画面説明文を読むことで企図するコンテキストに応じた動作かどうかをユーザに判断してもらう手法を試みたが、データ数が増えれば画面説明文を読む量も増えることが考えられるため、原画探索に時間がかかる可能性がある。より効率的な検索を可能にするには、動作表現の他にもキャラクターの属性、性格、心情等もメタデータとして原画に付与し、検索条件を増やすことが必要である。

提案システムの想定するユーザは原画を描くアニメーターであるが、アニメーターを目指す人たちへの学習支援としての使用方法も考えられる。アニメ制作の初学者教育という観点から見ると、一般社団法人日本動画協会が原画の描き方の基礎を学習できる教材[10]を公開しているように、自発的な学びの素材に対するニーズは高い。既存作品の原画の蓄積と再利用がより柔軟に行えるようになれば、多様で「生きた」教材としてその用途に活用でき、より制作現場に沿った実践的な学びが可能になると期待される。

7. おわりに

本稿では、アニメの台本中の画面説明文から抽出した動作表現を対応する原画にメタデータとして機械的に付与することで、動作表現を検索キーワードとして原画を検索できるシステムを試作した。実験結果から、類似した動作表現を検索対象に含めたことの有効性が示唆された。一方で、原画クリップと併せて画面説明文を提示することは、原画のコンテキストの理解の促進に十分ではないことが示唆された。今後は、原画のコンテキストの理解を促進させるため脚本の利用を検討する。

謝辞

本研究は、科研費 22K12338 および 20K12130 の支援のもと行われた。国立情報学研究所の IDR データセット提供サービスにより株式会社トリガーから提供を受けた「トリガーデータセット」を利用した。記して謝意を表す。

参考文献

- [1] Haruyama, T., Takahashi, S., Ogawa, T. and Haseyama, M.: Similar scene retrieval in soccer videos with weak annotations by multimodal use of bidirectional LSTM,

- MMAasia '20: Proceedings of the 2nd ACM International Conference on Multimedia in Asia*, No. 27, pp. 1–8 (2021).
- [2] Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N. and Inui, K.: A Joint Neural Model for Fine-Grained Named Entity Classification of Wikipedia Articles, *IE-ICE TRANSACTIONS on Information and Systems*, Vol. E101-D, No. 1, pp. 73–81 (2018).
- [3] 渡辺英雄: 日本の商業アニメーション制作に於けるデジタル化に関する研究, 湘南工科大学紀要, Vol. 47, No. 1, pp. 91–104 (2013).
- [4] 株式会社トリガー: トリガーデータセット, 国立情報学研究所情報学研究データリポジトリ (データセット), <https://doi.org/10.32130/idr.15.1> (2022).
- [5] 三浦健仁, 中川裕志: シナリオを用いたドラマのシーン検索システム, 情報処理学会論文誌, Vol. 40, No. SIG03(TOD1), pp. 144–151 (1999).
- [6] 永田 幸, 茂木龍太, 兼松祥央, 鶴田直也, 三上浩司, 近藤邦雄: 身体に外見的特徴をもつキャラクターの原案制作のためのシミュレーションシステム, 映像情報メディア学会技術報告, Vol. 40, No. 11, pp. 249–252 (2016).
- [7] 神村幸子: アニメーションの基礎知識大百科, グラフィック社 (2020).
- [8] 増田弘道, 数土直志, 高達俊之, 松本 淳, 陸川和男, 石田具隆, 亀山泰夫, 森 祐治, 長谷川雅弘: アニメ産業レポート 2023, 一般社団法人日本動画協会 (2022).
- [9] 兼松祥央, 竹本祐太, 茂木龍太, 鶴田直也, 三上浩司, 近藤邦雄: ロボットアニメーションにおけるポーズ制作支援システムの開発, 画像電子学会誌, Vol. 46, No. 1, pp. 165–169 (2017).
- [10] 一般社団法人日本動画協会人材育成委員会: アニメーターの課題集 —動きの法則を理解するための第一歩—, <https://aja.gr.jp/jigyuu/ikusei/workbook> (2021).
- [11] 鈴木裕章, 兼松祥央, 茂木龍太, 三上浩司, 近藤邦雄: パーツ分類に基づく亜人類キャラクターの3Dモデリング手法, 映像情報メディア学会技術報告, Vol. 39, No. 14, pp. 119–122 (2015).
- [12] 祖父江亮, 山下隆義, 平川 翼, 藤吉弘巨, 中澤 満, Yeongnam, C., Björn, S.: レシピ画像と説明文を活用した類似シーン検索による調理動画の要約, 精密工学会誌, Vol. 86, No. 12, pp. 1026–1033 (2020).
- [13] 舛本和也: アニメを仕事に! トリガー流アニメ制作進行読本, 星海社 (2014).