

---

## 報告書

---

### 1 今週の進捗

- 脚本の情報を用いた絵コンテの OCR

### 2 研究の目的と全体構成

本研究の目的は、アニメーション制作に用いられる絵コンテ画像から、シーン内容・台詞・演出情報を構造化データとして抽出し、脚本との整合性を保ちながら意味的に理解可能な表現へ変換することである。

本研究では処理を以下の段階 (Stage) に分割して設計している。

- Stage0：脚本の前処理（知識化）
- Stage1：絵コンテ画像からの OCR と CUT 単位への分割
- Stage2：脚本情報を用いた OCR 結果の整形・補正
- Stage3（今後）：意味理解およびシーン説明生成

本報告では、主に Stage1 および Stage2 の実装と進捗について述べる。

#### 2.1 Stage0：脚本の知識化

脚本 (docx 形式) を入力とし、以下の 2 種類のデータを生成する。

- 登場人物・用語辞書 (`lexicon_entities.json`)
- 段落・台詞単位に分割された脚本コーパス (`script_chunks.jsonl`)

脚本中の「登場人物」セクションから正規表記を抽出し、人物名の正解集合として保持する。また、物語本文はシーン番号や台詞・地の文の構造を保持したまま分割し、後段での検索 (RAG 的利用) を可能にしている。

#### 2.2 Stage1：絵コンテ画像からの OCR

Stage1 では、絵コンテ画像を CUT 単位で処理する。各ページを行 (row) に分割し、以下の列画像を入力とする。

- CUT 番号
- PICTURE
- アクションメモ
- ダイアログ
- TIM (時間)

各列画像に対して OCR を行い、行単位の結果を保存する。その後、CUT 番号の推定結果を用いて行を統合し、CUT 単位の JSON データ (`cutXXXX.stage1.json`) を生成する。

### 2.2.1 使用モデル

Stage1 の OCR には視覚言語モデルを用いる。

- 使用モデル : Qwen3-VL (Vision-Language Model)
- 用途 : 画像入力に対するテキスト生成 (OCR)

一般的な OCR エンジンではなく、視覚言語モデルを用いることで、日本語の手書き・装飾文字・制作現場特有の表記にも対応可能とした。

### 2.2.2 CUT 番号の系列補正

CUT 番号は OCR 誤認識が起こりやすいため、前後関係を用いた系列補正を導入している。

前の CUT 番号を  $c_{t-1}$ 、現在の OCR 推定値を  $\hat{c}_t$  とすると、期待値は

$$c_t^{exp} = c_{t-1} + 1$$

と定義できる。

$\hat{c}_t$  が  $c_t^{exp}$  から大きく乖離している場合、数字誤認（例：0 と 9）を考慮して補正を行う。これにより CUT 番号の飛びや誤結合を抑制している。

## 2.3 Stage2 : 脚本情報を用いた整形・補正

Stage2 では、Stage1 で得られた OCR 結果を、脚本由来の知識を用いて整形・補正する。

### 2.3.1 OCR 結果の正規化

OCR 結果には、

- 無意味なプレースホルダ（例： $\square$ ）
- ノイズ行

が多く含まれるため、連続するプレースホルダを 1 行にまとめるなどの正規化を行う。

### 2.3.2 脚本との対応付け

絵コンテ中の台詞は脚本中にも存在するという前提のもと、OCR で得られた台詞をクエリとして、脚本コパスから一致・類似する台詞を検索する。

本研究では、

- 台詞の完全一致（正規化後）
- 類似度スコア（しきい値付き）

を用いて候補を抽出している。

### 2.3.3 登場人物名の誤認識補正

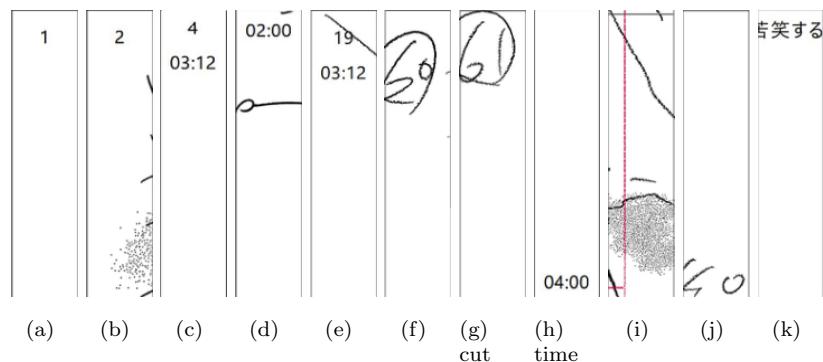
OCRにより誤認識された登場人物名（例：「ベルリ」→「ベリル」）に対し、脚本由来の人物辞書を用いた補正を行う。

人物名  $x$  (OCR結果) と正解候補  $y$  に対し、編集距離  $d(x, y)$  を計算する。隣接文字の転置も考慮した距離を用い、距離が十分小さい場合に補正候補とする。

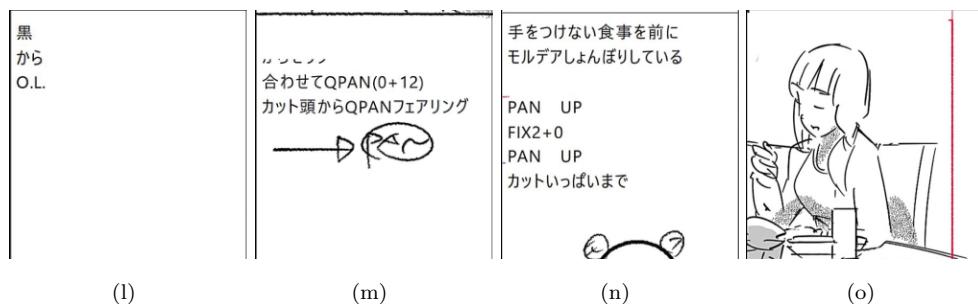
補正是元のOCR結果を上書きせず、正規化結果と補正ログを別フィールドとして保持する。

## 2.4 データ

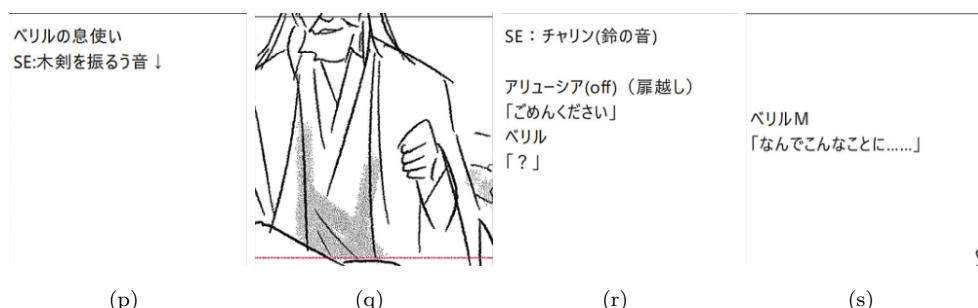
### 2.4.1 CUT, TIM



### 2.4.2 アクションメモ



### 2.4.3 ダイアログ



## 2.5 結果

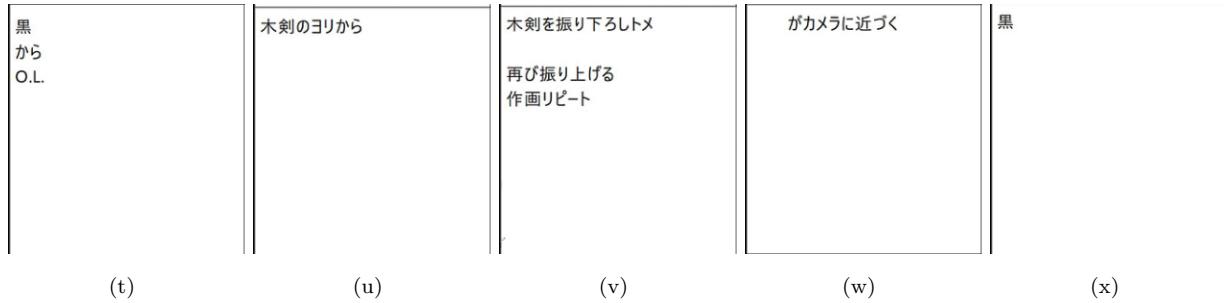
### 2.5.1 Stage0 (脚本の前処理)

脚本から登場人物の名前を取得.

```
{
  "episode_id": "episode01",
  "source_docx": "data/script_phase1/episode01.docx",
  "entities": [
    {
      "canonical": "ベリル・ガーデナント",
      "type": "character",
      "aliases": [],
      "evidence": [
        {
          "paragraph_index": 9,
          "text_snippet": "ベリル・ガーデナント"
        }
      ]
    },
    {
      "canonical": "ファナリー（セリフなし）",
      "type": "character",
      "aliases": [],
      "evidence": [
        {
          "paragraph_index": 22,
          "text_snippet": "ファナリー（セリフなし）"
        }
      ]
    },
    {
      "canonical": "モルデア・ガーデナント",
      "type": "character",
      "aliases": [],
      "evidence": [
        {
          "paragraph_index": 25,
          "text_snippet": "モルデア・ガーデナント"
        }
      ]
    },
    :
  ]}
```

脚本からセリフや地の文についても取得.

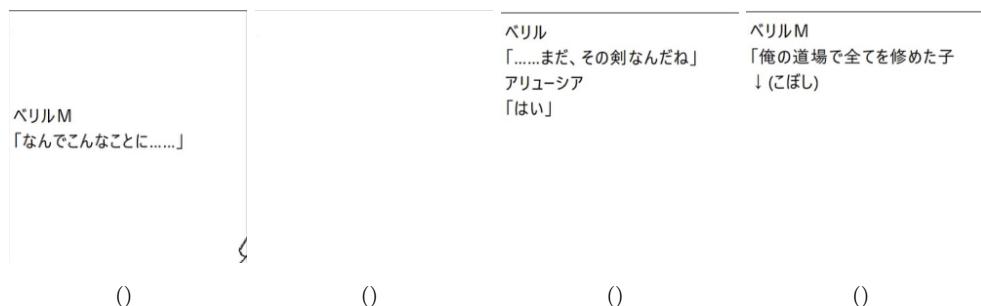
## 2.5.2 Stage1 (絵コンテ画像からの OCR)



(t) 黒\n から\n O.L. (u) 木剣のヨリから (v) 木剣を振り下ろすトメ\n 再び振り上げる\n 作画リピート  
(w) がカメラに近づく (x) 黒



(y) ベルルの息使い\n SE: 木剣を振る音 ↓ (z) □ () SE: 木剣を振るう音 !



(1) ベルリ M\n 「なんでこんなことに……」 (2) □ (3) ベルル\n 「……まだ、その剣なんだね」\n アリューシア  
\n 「はい」 (4) ベルリ M\n 「俺の道場で全てを修めた子\n ! (ごし)

## 2.5.3 Stage2 (脚本情報を用いた OCR 結果の整形・補正)

表 1 に脚本情報を用いた OCR 結果の整形・補正を示す。現状、セリフの前の話者のみ補正するようにしている。話者については脚本の登場人物の一覧から修正可能であることがわかった。

表 1: 修正前と修正後

カット番号	修正前	修正後
60	ベルリM「なんでこんなことに…」	ベリルM「なんでこんなことに…」
62	ベルル「……まだ、その剣なんだね」	ベリル「……まだ、その剣なんだね」
130	ベルルM「ぶっくりさせないでほしいな…」	ベリルM「ぶっくりさせないでほしいな…」

#### 2.5.4 以前の GPT 系の結果

比較項目	gpt-4o-mini gpt-4.1-mini gpt-4.1-nano	gpt-4o	gpt-4.1
page002_row5	Se: 木剣を振る音↓	SE: 木剣を振るう音 □	SE: 木剣を振るう音↓
page007_row2	ハンセン@合わせて QPAN(0+12)@ カット頭から QPAN フェアリング@ →@PAC	・・・ンジ@合わせて QPAN(0+12) @ カット頭から QPAN フェアリング	パンニング@合わせて QPAN(0+12)@ カット頭から QPAN フェアリング
page053_row3	ベリル、地図を見て	ベルリ、地図を見て	ベリル、地図を見て
page053_row5	ベリル立ち止まって アリュージアへ向き@ follow stop	ベルル立ち止まって アリュージアへ@向き@ follow stop	ベル立ち止まって アリューシアへ@向き@ follow stop
page060_row4	木剣を差し出す。	木剣を差し出す。	木判を差し出す。
page061_row1	ナビカ PANUP!@ ピ送り（土）→（木）へ@ じっと木鉢を見る@スレ 64	トウカ PANUP!@ ピン送り—@ じっと木鉢を見つめる@ スレ 64	↑ ICH PANUP!@ ピ送り 上→下へ@ じっと木釘を見つめる スレイ 64

### 3 今後の予定

- Stage2 における補正ルールの拡充
- Stage3 として、絵コンテ内容の意味理解およびシーン説明生成
- 定量評価指標の設計と比較実験

### 参考文献