

ゲームマニュアルにおける没入的記述の構造化分析 ―大規模言語モデルを用いた疑似ラベリング手法の適用―

福田 一史（立命館大学 映像学部） fukudakz@fc.ritsumeikan.ac.jp

概要：本研究は、ゲームマニュアルが持つ「説明」と「没入」という二重の機能のうち、後者に着目する。これまで体系的な分析がされてこなかったマニュアルの物語的要素に対し、大規模言語モデル（LLM）を用いた疑似ラベリングにより、テキストを「物語」や「キャラクター」といったセクションへ自動で構造化する手法を提案する。事例研究として「ニンテンドークラシックミニ」収録の30タイトルを分析した結果、初期のマニュアルにも豊かな物語記述が存在し、ファンタジーやSFといったジャンル類型が特定できた。本アプローチは、マニュアルの大規模なデジタルアーカイブ化やメタデータ自動生成を可能にし、ゲーム史研究の一次資料としての価値を高めるものである。

キーワード：ゲーム説明書、デジタルアーカイブ、疑似ラベリング、ナラティブ分析、ゲームスタディーズ

A Structural Analysis of Immersive Descriptions in Game Manuals: Applying a Pseudo-Labeling Method with Large Language Models

Kazufumi Fukuda (Faculty of Image Arts and Sciences, Ritsumeikan University)

Abstract: This study focuses on the immersive function of game manuals, one of their dual roles alongside instruction. To address the lack of systematic analysis of the narrative elements in manuals, this research proposes a method for automatically structuring text into sections such as "narrative" and "character" using pseudo-labeling with Large Language Models (LLMs). A case study analyzing 30 titles from the "Nintendo Classic Mini" revealed that even early manuals contain rich narrative descriptions, and identified genre patterns such as fantasy and science fiction. This approach enables the large-scale digital archiving of manuals and automatic metadata generation, enhancing their value as primary sources for game history research.

Keywords: Game Manuals, Digital Archive, Pseudo Labeling, Narrative Analysis, Game Studies

1. はじめに：ゲームマニュアルのパラテキスト・二重性と資料化

ゲームの周辺資料であるマニュアルには、「説明（Instruction）」と「没入（Immersion）」という二重の機能が共存しているといわれているが[1][2]、単に操作説明書にとどまらず、ゲーム世界の物語や雰囲気伝える役割も果たしてきた。例えば、初期の家庭用ゲームコンソールであるマグナボックス・オデッセイ（1972）には、36 ページにも及ぶ詳細なマニュアルが付属しており、テレビとの接続方法からコントローラの使い方まで丁寧に解説されていた[3]。また、1980 年代のファミリーコンピュータなどに代表されるビデオゲームが家庭に普及する時期には、各ソフトのマニュアルの前半部に「ものがたり」や「キャラクター」などの短い物語説明が含まれるようになり、操作法を教える説明的要素に加え、プレイヤーをゲーム世界に惹きつける没入的要素が付与された。近年では、物理メディアの衰退とともにマニュアルは簡素化され、デジタル形式やゲーム内チュート

リアルが主流となっているが、一方でそれゆえにマニュアルの文化的価値や回帰的な関心も高まっている[4][5]。

これまでの研究を概観すると、ゲームマニュアルに関する議論は主に次の 3 つの観点に集約できる：

1. **説明と没入の二重機能：**マニュアルは操作方法やゲームルールを説明する実用的文書であり、他方でイラストや背景設定を通じてプレイヤーをゲーム世界へ引き込む物語ツールでもある。Townesley（2020）は、複数のマニュアルを分析し、この二重性を焦点化した。
2. **パラテキストとしての役割：**ゲーム本編に付随するマニュアルは、Genette（1997）が述べた「パラテキスト」[6]の一例と見なせる。Consalvo はこのパラテキストの概念をゲームに拡張し、ゲームの核となるコードやプログラムとそれによって規定される基本的なゲームプレイを「テキスト」と定義し、その付随的な

パラテキストとしてゲームマニュアルを例示した。そして、広告や動画など外部のパラテキストも含めゲームプレイをそのような多様なメディア経験から構成される全体的な活動（Assemblage）と再定義した。このような考え方はゲーム・スタディーズ分野では多くの文献で用いられる一つの重要な研究視点である[2]。

3. **アーカイブ資料としての価値：**ゲームプレイがプレイヤーごとに多様な表現を持つ一方、マニュアルは印刷物であるため安定的なテキスト形態を保つ。尾鼻らの研究では、マニュアルをデジタル化しデータベース化することがゲーム史研究に資するとの指摘がなされている[8][9][10]。Schmidt（2023）なども、マニュアルの消失はゲーム体験の一部を失うことに等しいと論じており、保存・再評価の重要性が強調される[4]。

以上のように、マニュアルは「説明書以上の何か」であることが知られている。特に「没入」を担う部分は他の家電製品の取扱説明書と本質的に異なり、ゲーム世界の構築に寄与する点で特筆される。本研究では、これらマニュアルの豊かな内容を活用可能な形でデジタルアーカイブ化し、メタデータを体系化することを目指す。具体的には、大規模言語モデル（LLM）を用いた疑似ラベリングによってマニュアルを構造化し、そのうえで抽出した没入的記述とりわけ物語要素のテキスト分析を行うことで、ゲーム史・メディア史の一次資料としての再利用性向上を図る。

2. 研究の方法

2.1 研究方法の概要

本研究では、ゲームマニュアルにデジタル・アーカイブ化のための標準的な方法論を適用しつつ、セクション分類に LLM を導入して自動化を試みる。具体的には、マニュアルから抽出したテキストを意味のまとまりごとにセクション分割し、各セクションにタイプラベルを疑似付与する。この工程により、説明セクションと没入セクションを区分し、大規模な内容分析を可能とする。一方、分類の正確性や LLM の出力は確定的なものではないため「疑似ラベリング」と位置付け、必要に応じて後処理や統制を加えている。

こうして構造化したマニュアルデータから、

没入に関係する物語セクションのテキストを抽出し、自然言語処理技術を用いて分析する。具体的には、抽出テキストに対して固有表現抽出（NER）、トピックモデリング、ジャンル分類、クラスタリングなどを適用し、マニュアル物語がどのような特徴を持つか探索する。これらの分析結果は、収集・構造化データを活用した自動メタデータ生成の有効性を示すとともに、ゲームコレクションを横断的に検索・比較し、時代・ジャンル別の語彙傾向分析を行うなどといったアプローチへ応用可能である。

また、固有表現抽出で取得するエンティティは説明と没入のどちらのセクションに登場するかによって、多義性を解消できると考えられる。例えば、アイテムとして「ボタン」を認識した場合、説明セクションでは物理的なコントローラのボタンを指す一方、没入セクションではゲーム世界内のアイテムやゲームの仕掛けであるボタンを指す可能性が高い。同様に「市ヶ谷」、「フィンランド」などといった地名のエンティティでは、説明セクションでは開発会社所在地などの現実世界のメタ情報を示すのに対し、没入セクションでは物語の舞台そのものを示す。このようにセクションの文脈を明確に区別することでエンティティの多義性を解消し、ナラティブ研究と企業研究の両側面において、データベースの検索精度と分析の信頼性を大幅に向上させることができる。

2.2 分析対象

本研究ではウェブで公開されている「ニンテンドークラシックミニ ファミリーコンピュータ」の説明書をサンプルとして用いる[11]。ニンテンドークラシックミニ ファミリーコンピュータは 2016 年に任天堂により発売されたゲーム機であり、HDMI ケーブルで映像・音声出力が可能となっており、手軽に 30 本の復刻されたファミリーコンピュータタイトルを当時の内容に近い形でプレイすることができる。本機収録の各ソフトに対応し、ウェブで公開されているマニュアル PDF を収集し、上記手法を試行的に適用した。

2.3 資料化と分析のプロセス

本研究のパイプラインは図 1 のとおりである。



図 1. 本研究のパイプライン

このプロセスにより、OCR による文字抽出、LLM による構造化と分類、それに基づく分析対象のテキスト抽出と分析という一連の流れを実装・検証した。前半の3つの資料化のプロセスを3.において、最後の分析の詳細を4.において論じる。

3. ゲームマニュアルの資料化

ここでは、前述の通り「ニンテンドークラシックミニ ファミリーコンピュータ」のゲームマニュアルを対象に、デジタル人文学の標準的な方法論を用いて、資料化を実施する。

まず、ゲームマニュアルの PDF ファイルからページ単位の画像に分割した。そして、ゲームマニュアルに頻出する非定型で複雑なレイアウトに強い Cloud Vision OCR[14]を用いて全てのページ画像から文字情報を抽出した。

次に、抽出したページ単位のテキストファイルを用いて、ページ情報を付与した仮 XML を作成した。続いて、LLM (claude-4-sonnet)を用いて、テキストの構造化と疑似ラベリングとしてセクションタイプを付与した。その他の LLM として、ChatGPT、Gemini、DeepSeek も試行したが、ChatGPT (gpt-4o)と Gemini (Gemini-2.5-Pro)はトークン制限や簡潔化バイアスが原因かと想定されるが、セクションの一部省略による抜け落ちが複数回発見され、全文のデータ化が求められる本研究には適さないと評価した。DeepSeek (DeepSeek-v3.1)と Claude (claude-4-sonnet)はほぼ同程度の性能であった。ただし、DeepSeek はレスポンスが早い点は利点であったものの、Claude は複数ファイルの一貫性保持に強いといった性質があるため[13]、ここではその性質を優先し Claude の結果を使用する。

また、本研究では以下の XSD スキーマで、LLM で生成する XML の構造を定義した。一つのマニュアル全体を示す manual は必須の title 属性を持ち、それは必須の number 属性を持つ各ページ (page) で構成される。page は1件以上の section を持ち、それは必須の type 属性を持ち、head と p で構成されるほか、さらに subsection に区切る事もできる、というものである。TEI/XML を用いることも選択肢として考えられるが、ここでは求められる機能が限られ要件が少ないこと、TEI のゲームのテキストエンコーディングに関する機能は現時点では数少なくスキーマ拡張の柔軟性に難があること、といった前提条件があるため独自の構造を定義した。

さらにお手本となるゴールドデータを1件作成した上で前述の XSD を提示するプロンプトで変換したサンプル[14]は下記である。

```
<?xml version='1.0' encoding='utf-8'?>
<!-- Created by claude-4-sonnet -->
<manual title=" マリオブラザーズ 取扱説明書 "
lang="ja" game_id="CLV-P-HAAWJ">

  <!-- 表紙ページ -->
  <page number="1">
    <section type="cover">
      <head>ファミリーコンピュータ マリオブラザーズ
    </head>
    <p>HVC-MA</p>
    <p>取扱説明書</p>
    <p>Nintendo</p>
    </section>
  </page>

  <!-- 前書きと目次 -->
  <page number="2">
    <section type="preface">
      <p>このたびは任天堂ファミリーコンピュータカセット・マリオブラザーズ (HVC-MA) をお買いあげいただきまして、誠にありがとうございました。</p>
      <p>ご使用前に取扱い方、使用上の注意等、この「取扱説明書」をよくお読みいただき、正しい使用方法でご愛用ください。</p>
      <p>なお、この「取扱説明書」は大切に保管してください。</p>
    </section>
    <section type="toc">
      <head>目次</head>
      <p>1. 使用上の注意</p>
      <p>2. コントローラー各部の名称と操作の説明</p>
      <p>3. マリオブラザーズの遊び方</p>
    </section>
  </page>
  <!-- 以下続く -->
```

各セクションに適切な見出し (<head>) と本文 (<p>) が配置され、type 属性が cover, toc, warning などが付与されている。

最後に LLM 出力後の後処理として、得られた type 値を正規化・統合した。具体的には、類似ラベルの統合 (character と characters の統一など) や、詳細タイプから大分類へのマッピング (action や combat を gameplay に統合など) を行い、後述の19カテゴリを定義した。最終的に各説明書から600強のセクションが得られ、そのうち narrative ラベルが付与されたセクションは51件あった。

主要なタイプとしては、gameplay (105件)、controls (9件)、character (31件)、item (24件)、system (28件)、narrative (51件)、instruction (137件)、reference (53件)、warning (34件)、legal (42件)、enemy (20件) の11種類である。その他のタイプと

しては、cover (25 件)、preface (20 件)、greeting (3 件)、toc (17 件)、company_info (1 件)、blank (4 件)、empty (1 件)、memo (1 件) である。

そして、生成した XML ファイルの type 属性の値を上記のマッピングに沿った値に一括で編集した。このプロセスにより、ノイズの多い生データは、後続の定量的・定性的分析に耐えうる、クリーンで一貫性のある XML 形式の構造化データセットへと変換された。

4. 分析と考察

4.1 没入と説明のセクション分量比較

まず、前述の疑似ラベルを用いて、説明と没入を特定し、それぞれの分量を示したのが表 1 である。没入に分類されるセクションは gameplay, character, item, narrative, enemy として、その他は説明に分類される。

表 1. 説明と没入のセクションの分量比較

カテゴリ	セクション数	文字数	平均文字数
説明	495	49320	99.6
没入	323	23263	72.0

表 1 を見ると、説明系セクションの文字量が没入系セクションより大きい。これは選んだサンプルが 1980 年代中心のファミリーコンピュータ作品であるため、当時は説明が多めであったためと考えられる。近年ではゲーム内チュートリアルが発達等で説明量が減る傾向にあるとも考えられるため、将来的には年代別に比較行うことでその検証が可能である。

また本分類では、gameplay を没入として分類したが、これは action や technique など境界的な分類をマッピングされた概念である。厳密に説明/没入を分ける際にはさらなる細分化が必要かもしれない。

4.2 テキストアナリティクスによるナラティブ分析

3. において作成したゲームマニュアルの XML ではセクション疑似ラベルが付与されたことで、説明書から特定のタイプのテキストを抽出可能となった。そのためここでは没入的内容の代表といえる、narrative セクションのテキストを抽出し、自然言語処理、とりわけテキストアナリティクスの方法を用いて、ナラティブ分析を行う。構造化データセットから抽出したテキストについて改行や空白の正規化と長さしきい値による簡易前処理を行ったファイルを作成し、分析にはこれを用いた。

4.2.1 分析：固有表現抽出

テキストアナリティクスとしてはまず、固有表現抽出 (Named Entity Recognition: NER) を行った。使用した NER モデルは、単語とエンティティの知識拡張型訓練済み Transformer モデルである studio-ousia/luke-japanese-base-lite[15]であり、これを基盤に採用した BIO スキーマによるトークンレベル分類を行う。抽出は、独自辞書によるパターン照合・正規表現パターン照合・フィルタリングによるパターン抽出と、深層学習による抽出を組み合わせたハイブリッド方式を採用している。エンティティのタイプとして定義したものは CHARACTER, LOCATION, GAME_TITLE, ITEM, ENEMY の 5 種類とした。ファインチューニング機能を実装しており、ゴールドデータとして 5 作品のゲームマニュアルの narrative セクションのラベル付きデータを作成し、それを学習して実行した結果、65 件のエンティティが抽出された。代表的な抽出例は以下の通りである (種別数/出現回数)：

- CHARACTER (31 種類)：リンク (17)、マリオ (13)、ゼルダ姫 (8)、ベルモンド (3)
- LOCATION (13 種類)：キノコ王国 (10)、ハイラル (10)、プブブランド (16)、アトランチス (4)
- ITEM (12 種類)：トライフォース (17)、スターロッド (4)、海賊の本 (1)
- GAME_TITLE (7 種類)：スーパーマリオブラザーズ (2)、スーパーマリオブラザーズ 3 (2)
- ENEMY (2 種類)：十字架 (1)、ドラゴン (1)

大半の固有表現は適切に抽出できたが、CHARACTER として分類された「ベンチャークエスト」や「バッテリーバックア」のように、いくつかの誤って抽出されたエンティティ例が散見された。誤抽出の原因としては、BIO スキーマに基づくトークン列の途中でエンティティが途切れてしまう構造的な課題が挙げられる。たとえば「バッテリーバックアップ」という語が、サブワードトークン化により「バッテリー」「バック」「アッ」「プ」といった単位に分割された際、末尾の「プ」が O (非エンティティ) として誤認識されると、ラベル系列全体が崩れ、「バッテリーバックア」のような不完全な表現がエンティティとして抽出されてしまう。このような現象は、サブワード分割手法 (BPE や WordPiece など) と BIO スキーマの整合性に起因する典型的な問題の一つである。

ゴールドデータの増強や禁止語リストの指定による後処理などといった統合的な対応が必要になるものと思われる。

「スーパーマリオブラザーズ」シリーズや「ゼルダの伝説」シリーズのような任天堂の有名タイトルに関わるエンティティの抽出の割合が多いように思われる。これは本調査のサンプルとなるゲームの 30 件のうち 14 件が任天堂タイトルであり偏りがあることが影響しているものだと想定される。

ここでは、エンティティ抽出の手法としてパターン抽出と機械学習抽出 (LUKE 抽出) を行ったが、機械学習抽出の件数は全体の 5%ほどであり限定的であり、抽出の基準を落とすことで件数を増やすことはできるが、その場合、適切なエンティティの抽出はほぼなされなかった。一方、独自辞書や正規表現パターン及びフィルタリングからなるパターン照合で抽出した固有表現は概ね高精度で、特にゲーム固有の名詞に強みを示した。この結果から、学習データ量が少ない段階ではルールベースの補完が有効であり、今後は Human-in-the-loop 型で辞書リストと学習データを拡充していくのが望ましい。

4.2.2 分析：トピックモデリング

トピックモデリングとしては narrative セクション全体を対象に Latent Dirichlet Allocation (LDA) と BERTopic の 2 手法を試した。結果として、得られたトピック数は LDA では 2 件に留まったが、BERTopic では 3 件であり、解釈可能性についても後者が有用な結果であった。そのため BERTopic の結果をここで示す。このアプローチでは、テキストを SentenceTransformer で埋め込み、HDBSCAN によるクラスタリングを経て UMAP 次元削減している。その結果得られたトピック例を示す (カウントは所属セクション数)：

- **Topic 0 (39 件):** 「さあ」「しかし」「リンク」などの接続詞やキャッチフレーズ。
- **Topic 1 (10 件):** 「おどろいたことに」「魔法」「星」など、異世界設定や使命の説明など。
- **Topic 2 (6 件):** 「スポーツ」「ラウンド」「規定オーバー」など、スポーツゲームやルール説明など。

Topic 1 は「おどろいたことに」「地球から遠く離れた小さな星の小さな国」といった物語の

導入や世界観の説明等で構成されている。この結果は没入のための玄関を体現したものであり、パラテキストとしての説明書に求められる機能が強く表出したものだと解釈できる。

Topic 2 はゲームプレイやルールの説明を兼ねる内容であり、スポーツゲームジャンルに偏って生じているものと思われる。同ジャンルでは、世界観がゲーム経験に強く影響しないため、物語が没入と説明を兼ねる場合がある、という特徴が観察されたものと思われる。

Topic 0 は固有名詞や「さあ」「しかし」「する」とどうでしょう」「シャキ」などの多くの惹句フレーズで構成されている。物語に関するセクションであると同時に、読者に語りかけるような表現が多い。つまりこれらは、いわゆる第四の壁を超えたメタフィクションの手法の現れであると言える。このトピックの出現回数は多く、初期のゲームマニュアルの物語の語り口として広く用いられていたものであると考えられる。これは時代が進むにつれて変化していく可能性が考えられる。ただしこのトピックには、接続詞に類するフレーズがいくつか含まれており、弱いクラスターとも評価できる。ストップワードのチューニングにより、これらトピックはより明確化する可能性がある。

いずれにせよ、トピックモデリングでは数多くのトピックが得られたわけではなかったものの、当時のゲーム説明書の物語の単一的でないあり方が見て取れたと言える。

4.2.3 分析：物語ジャンル分類

抽出されたトピックを元に、キーワードベースのルールで物語ジャンル分類を行った。具体的には、Fantasy (魔法、呪い等) や SF (宇宙、ロボット等)、スポーツ、ホラーなど 7 ジャンルを定義し、各トピックの代表テキストからジャンルを割り当てた。結果を図 2 に示す。

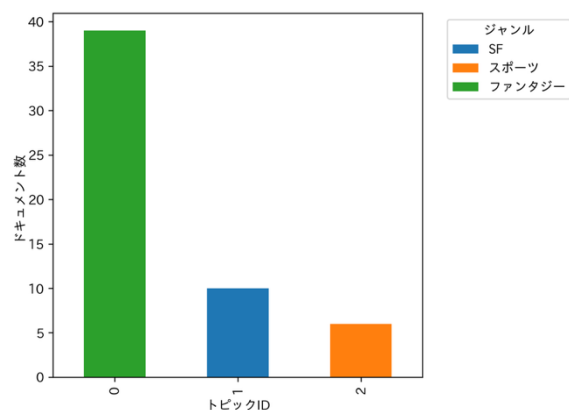


図 2. トピック別ジャンル分布

トピック 0 は「ファンタジー」を主要ジャンルとして、サブジャンルが「アドベンチャー」である。対象となるゲームの事例は「ゼルダの伝説」や「ソロモンの鍵」である。典型的なファンタジー冒険譚に該当する。

トピック 1 は、主要ジャンルが「SF」で、サブジャンルに「ファンタジー」「アドベンチャー」「ホラー」の 3 件が付与されている。代表例が「スーパーマリオ USA」「星のカービィ」「悪魔城ドラキュラ」である。SF を軸として様々な要素が含まれており、事例も多様であり、混在的なトピックである。

トピック 2 は主要ジャンルが「スポーツ」でサブジャンルは付与されていない。「マリオオープンゴルフ」「つっぱり大相撲」が代表例である。物語自体が主題となり難しいスポーツゲームは物語要素が控えめでジャンル特定が容易であるため、クラスタが明確に分かれた。

これらは、初期ゲームに共通する物語類型を反映している。とくに Topic 1 (SF 系) は多様な要素を含む混成的傾向が見え、初期ゲームでは SF を主題に様々な拡張が行われていた様子が示唆される。

4.2.4 分析：クラスタリング

最後に、全ゲームの narrative セクションについて、文書レベルのクラスタリングを行った。60 件のテキストを、`cl-tohoku/bert-base-japanese` でベクトル化し、UMAP による次元削減後に K-means クラスタリング (5 クラスタ) を適用した。図 3 に可視化を示す。

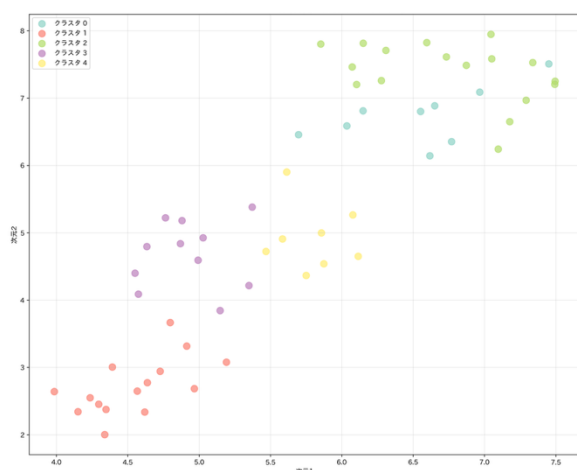


図 3. ゲームマニュアルの物語記述のクラスタ分布

各クラスタの特徴と代表例は以下の通りで

ある：

- **Cluster 0 (シリアス事件/復讐・対立)：** 9 件。代表作：『ダブルドラゴン II』(復讐劇)、『熱血物語』(抗争) など。事件発生から対立・戦闘へ。
- **Cluster 1 (説明・口上・短文)：** 15 件。代表作：『スーパーマリオ 3』の「マリオからのメッセージ」、『つっぱり大相撲』の口上など。説明的・案内的・キャッチフレーズ的な短文が多く含まれる。
- **Cluster 2 (SF・大規模戦闘)：** 17 件。代表作：『メトロイド』(宇宙・宇宙海賊)、『悪魔城ドラキュラ』(儀式・復活) など。宇宙・科学・儀式・大戦などシリアスでスケールの大きな設定が特徴。
- **Cluster 3 (夢・旅立ち・王国ファンタジー)：** 11 件。代表作：『スーパーマリオ USA』(夢/異世界 Subcon)、『星のカービィ 夢の泉』(夢・スターロッド) など。夢・異世界・王国・旅立ち等の叙述。
- **Cluster 4 (ゼルダ神話・伝承)：** 8 件。代表作：ゼルダシリーズなど。神話・伝承的設定が中心。

全体を俯瞰すると、直線状にプロットされており「ライト/メタメッセージ・明るい異世界冒険」と「ダーク/SF・シリアス展開」の二極化傾向が見られた。特に『ゼルダ』シリーズでは専用クラスタが形成され、独自性の高い世界観が際立っている。このように、クラスタリングにより各タイトルの物語近似性や分類軸が明らかとなったことで、学術的分析やコレクション検索に有用な構造化データとなる。

5. 考察

本研究により、LLM を用いたマニュアル構造化と疑似ラベリングが、説明系・没入系セクションの定量分析を可能にした。結果からは、少なくとも 1980 年代ファミコン期のマニュアルでは説明部分が多くを占める一方で、物語系セクションにも豊かな表現が含まれていることが定量的に裏付けられた。これは、Genette の言うパラテキストとしてのマニュアルが玄関の役割を果たし、多様なかたちでゲームと世界をつないでいる様相を示している。

とりわけ技術的な制約により映像表現に限界があったファミリーコンピュータにおいては、ゲームマニュアルは、このような多様な物語をもってして「世界観構築エンジン」として

の役割を担っていたといえる。また、これらのテキストがイラストレーションとセットで提供されていたため、視覚的な表現も物語への導入や世界観の構築に、強く影響したことが想像されるが、これは別途の研究が必要であろう。

5.1 方法論的意義

LLM を中核とした疑似ラベリングは、物理的に保管されているマニュアル群のデジタルアーカイブ化において、セクションレベルのメタデータ自動生成を可能にし、スケーラビリティの観点から有用である。

NER やトピック・クラスタリングと組み合わせることで、コレクション横断的な検索や比較、時代・ジャンルごとの語彙傾向分析に資する構造化データが得られる。

また、固有表現抽出やトピック・クラスタ分析により、1980 年代マニュアルに共通する物語要素や語り口パターンが浮かび上がった。例えば、ファンタジー冒険型（王国・伝承モチーフ）や SF/戦闘型、スポーツ型などの大まかなジャンルが確認され、マニュアルが提供する「世界観」「物語ジャンル」といったメタデータとして機能する可能性が示唆される。

これらの分析は、マニュアルから得られる物語情報や登場人物・地名などの固有表現を自動抽出し、コーパス横断検索や類似物語探索に活用できる点で意義深い。とりわけキャラクターや地名といったエンティティは、検索のみならず、ゲームシリーズや世界観の共有、クロスオーバー、など作品をグルーピングする上での判定ツールとして極めて有用である。特に、固有表現抽出においては、ルールベースで高精度に抽出した語彙リストを用いてデータを拡充し、それを機械学習モデルの教師データに組み込む Human-in-the-Loop 戦略が今後の課題となる。深層学習型 NER の性能向上には、ゲーム研究に特化したアノテーション済データの整備が不可欠である。

本研究の手法は自動化に重点を置いており、膨大の数のゲームマニュアルコレクションにも適用可能である。より大規模なコーパスに拡張することで、分類や分析の精度はさらに高められるだろう。実際のアーカイブ構築では、登場キャラや世界名などの固有表現、さらにはストップワードや禁止語を統制したデータ処理も重要である。今後の改善点として、トークン制限・長文カットオフ対策や、LLM 間比較・プロンプト精錬による出力安定性向上などが挙

げられる。また、本サンプルは任天堂タイトル中心かつ 1980 年代作品偏重であったため、年代・メーカー・ジャンルを多様化したコーパスでの再評価も急務である。

5.2 技術的課題と改善点

本研究で提案した手法には、いくつかの技術的課題とそれに伴う改善点が存在する。

第一に、表記揺れや誤抽出の問題が挙げられる。OCR 処理時の認識誤りや、Transformer モデルにおけるサブワード分割が原因で、前述のような不完全なエンティティが抽出されるケースが見られた。この問題に対処するためには、処理後の正規化ルールの導入や、誤抽出されやすい語をあらかじめ指定する禁止語リストの整備が有効である。さらに、人手で修正したゴールドデータを増強し、モデルを再学習させることで、抽出精度を段階的に向上させることが可能である。

第二に、教師データの不足という課題である。特に、深層学習を用いた固有表現抽出（NER）の性能は、対象ドメインに特化した教師データの量と質に大きく依存する。現状ではゲーム研究に特化したアノテーション済みコーパスが乏しいため、今後は doccano[16]のようなアノテーションツールを活用し、研究者が効率的に教師データを整備する Human-in-the-loop のアプローチが鍵になるだろう。

第三に、大規模言語モデル（LLM）の出力の評価と安定性確保も考慮すべきである。本研究の試行において、モデルによりテキストの省略が発生するなど、出力特性に差異が見られた。そのため、継続的なプロンプトエンジニアリング手法と出力結果の評価と形式制御とモデル間性能比較を通じて、目的に応じて最適な手法を選択するプロセスが不可欠である。

そして、本研究のサンプリングの偏りも今後の課題である。分析対象が「ニンテンドークラシックミニ」収録タイトルに限定されたため、1980 年代の任天堂作品に偏りが生じている。本手法の汎用性を証明するためには、年代、メーカー、ジャンルをより広範にカバーする多様なマニュアルコーパスを用いた追加検証が急務である。

5.3 応用可能性

本研究で提案した手法と構築したデータセットは、いくつかの応用可能性を秘めている。

まず、ゲーム史・メディア史研究における一次資料の再評価に貢献できる。構造化されたデ

ータを用いることで、特定の時代やジャンルにおける物語パターンの変容や、使用される語彙の定量的変動を追跡する通時的研究が可能となる。これにより、マニュアルというパラテキストがゲームの世界観構築に果たしてきた役割を、よりマクロなデータに基づいて論じることが可能となる。

次に、本手法はゲームマニュアルに限らず、ゲームのその他のパラテキスト、例えばゲーム雑誌や攻略本、攻略サイトなどの分析へも応用可能である。「説明」と「没入」の二重機能を持つ周辺テキストの構造化・分析にも本アプローチを適用できるだろう。

さらに、デジタルアーカイブとしての検索機能にも直接的に寄与する。固有表現抽出によって得られた登場人物や地名、アイテム名などをメタデータとして付与することで、コレクションを横断した検索（例：「すべてのマリオシリーズのマニュアルから『クッパ』が登場する物語セクションを検索」）や、類似した物語を持つゲームを探索することが可能になり、資料の利便性が飛躍的に向上する。

最後に、本研究の枠組みを拡張し、新たな分析軸を導入することも考えられる。例えば、提案として、語り口の定量分析が挙げられる。「説明」と「没入」の各セクションで感情語や形容詞の使用頻度を比較することで、各々のセクションが持つ文体的特徴や、プレイヤーへの働きかけ方の違いを明らかにできる可能性がある。

6. 課題と展望

本研究はまず、「ニンテンドークラシックミニ ファミリーコンピュータ」収録の 30 タイトルで方法論の有効性を確認したに過ぎない。今後はより広範な年代・メーカー・ジャンルをカバーするマニュアルコーパスを構築し、手法の汎用性を検証する必要がある。それを通じて、本研究では物語に焦点化した但、操作方法・ルール・メカニクスといったプレイ実装の側面やそれらと物語との融合関係についても特徴抽出や分類・クラスタリングが可能になる。また分析対象をゲーム雑誌や攻略本さらには攻略ウェブサイトなどに広げること想定しうる。

そしてスケールを拡大するうえで、人手により効率的にアノテーションを可能とする環境を構築し、ゲーム特有の固有表現リストを充実させ、それを基に NER モデルを強化する作業が重要である。また、LLM による構造化プロセスも、今後のモデルアップデートに合わせて継続的に評価・比較を行うべきであろう。さら

に、語りかけ文や物語構造を自動抽出する研究も、パラテキスト研究の発展に資するであろう。

将来的には、ゲームマニュアルの構造化・注釈付きコーパスを公開し、「ゲーム分析のためのシソーラス」や TEI 対応 XML など、デジタル人文的な資産として体系化することを目指す。本研究の手法はあくまで第一歩であり、さらなる検証と拡張により、ゲーム史研究における一次資料の豊かな利用へとつなげていきたい。

参考資料

- 1) Townsley, Geoffrey Kincaid. *Game Manuals: A Genre Analysis*. University of Arkansas at Little Rock, 2020, Master's Thesis.
- 2) Švelch, Jan. 2020. "Paratextuality in game studies: A theoretical review and citation analysis". *Game Studies* 20 (2).
- 3) Govia, Mario de. 2021. "Game Manuals". *Encyclopedia of Video Games: The Culture, Technology, and Art of Gaming*, Mark J. P. Wolf, vol. 2. Bloomsbury Publishing USA.
- 4) BrainiacBri. 2023. Please Save the Booklet for Future Reference - First Person Scholar. <https://www.firstpersonscholar.com/please-save-the-booklet-for-future-reference/>.
- 5) Birch, Nathan. 2013. "An Ode To Instructions: Why I' ll Miss Video Game Manuals". *UPROXX*, 2025 年 8 月 16 日. <https://uproxx.com/viral/an-ode-to-instructions-why-ill-miss-video-game-manuals/>.
- 6) Genette, Gérard. 1997. *Paratexts: Thresholds of interpretation*. Vol. 20. Cambridge University Press.
- 7) Consalvo, Mia. 2009. *Cheating: Gaining advantage in videogames*. Mit Press.
- 8) 尾鼻崇. ゲームマニュアルを対象としたビデオゲーム研究の可能性--デジタル保存とデータベース構築の意義と課題. *アート・リサーチ = Art research*. 2010, vol. 10, p. 101-110.
- 9) 尾鼻崇. ビデオゲームソフトウェア付属マニュアルの内容分析的研究 - 「物語設定」を対象とした調査と考察. *コア・エシックス*. 2011, vol. 7, p. 35-49.
- 10) 尾鼻崇, 毛利仁美. ゲームソフト付属取扱説明書の資料的価値に関する一考察. *日本デジタルゲーム学会 夏季研究発表大会 予稿集*. 2024, vol. 2024, no. 0, p. 31-36.
- 11) 「ニンテンドークラシックミニ ファミリーコンピュータ 説明書 | 任天堂」. <https://www.nintendo.co.jp/clv/manuals/ja/index.html>, (参照 2025-08-07).
- 12) 「画像内のテキストを検出する | Cloud Vision API | Google Cloud」. n.d. 参照 2025 年 8 月 8 日. <https://cloud.google.com/vision/docs/ocr?hl=ja>.
- 13) "Best AI chatbots of 2024 tried and tested- here's how they compare | Tom's Guide". n.d. 参照 2025 年 8 月 8 日. <https://www.tomsguide.com/ai/best-ai-chatbots-of-2024-tried-and-tested-heres-how-they-compare>.
- 14) 「マリオブラザーズ - CLV-P-HAAWJ.pdf」. <https://www.nintendo.co.jp/clv/manuals/ja/pdf/CLV-P-HAAWJ.pdf>, (参照 2025-08-16).
- 15) "studio-ousia/luke-japanese-base-lite · Hugging Face". n.d. 参照 2025 年 8 月 8 日. <https://huggingface.co/studio-ousia/luke-japanese-base-lite>.
- 16) "doccano/doccano: Open source annotation tool for machine learning practitioners.". n.d. 参照 2025 年 8 月 7 日. <https://github.com/doccano/doccano>.