

## テレビゲームコーパスの構築とその利活用

著者	麻 子軒
雑誌名	言語資源ワークショップ発表論文集
巻	1
ページ	117-126
発行年	2023
URL	<a href="http://doi.org/10.15084/00003731">http://doi.org/10.15084/00003731</a>

## テレビゲームコーパスの構築とその利活用

麻 子軒 (関西大学国際教育センター) †

### Construction and Utilization of Video Game Corpus

Ma Tzuhsuan (Kansai University)

#### 要旨

コーパス言語学の普及に伴い、これまでは書籍・雑誌・新聞・小説・漫画など、様々な媒体によるコーパスが構築されてきたが、テレビゲームを対象に作成されたものはまだない。本稿では、キャラクターのセリフが多く含まれるロールプレイングゲームを例に取り、テレビゲームコーパスを構築する方法論を述べた上で、その利活用の例を示し、同コーパスに言語資源としての価値があることを主張する。具体的に、役割語や日本語教育といった分野における研究例が挙げられる。また、ロールプレイングゲームには、ナレーションによる状況描写が少ないことや、一部会話のやり取りがプレイヤーの想像で補完しなければならないことなどの特徴がある。このことから、本稿で作成したテレビゲームコーパスを用いれば、同じ「作られた話し言葉」として代表される小説・漫画では観察できない言語的特徴を見出すことができると考えられる。

#### 1. はじめに

コンピュータの計算能力が飛躍的に向上したことで、大規模な言語データを高速に処理することが可能となり、コーパス言語学が大いに発展を遂げた。2011年に、国立国語研究所によって日本語に関する初めての大规模コーパス『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)が公開された(国立国語研究所 2015)。それ以来、ウェブコーパス、歴史コーパス、学習者コーパスなど、異なる目的で開発されたコーパスが次々と世に出た。また、安価かつ高性能なコンピュータの発売により、研究者も各自の研究目的に応じてコーパスを自作することが可能となった。

こうした背景の下、これまで書籍・雑誌・新聞・小説・漫画<sup>1</sup>など、様々な媒体によるコーパスが構築されてきた。しかし、テレビゲーム(以下、ゲーム)を対象に作成されたコーパスはまだない。その理由として、ゲームの文字情報を表示させるのにプレイヤーの操作が必要であることや、テレビ画面上に表示された文字情報を電子化する作業が煩雑であることが考えられる。媒体が紙である小説や新聞の場合は、紙に印刷された文字を画像にスキャンし、OCRソフトウェアを用いれば電子化作業がある程度自動化できるが、ゲームはその媒体の性質上同じ処理ができない。また、ゲームは子どもの遊びである固定観念がまだ強く、研究の対象として扱われていないことが多い。

そこで、本稿ではゲームコーパスを構築する方法論を述べた上で、その利活用の例を示し、同コーパスに言語資源としての価値があることを主張する。今回は言語情報が比較的多く含むと思われるロールプレイングゲームを例に取るが、本稿で提案した方法論が今後ほかのジャンルのゲームコーパス作成の参考になればと思う。

† kenji.ma@kansai-u.ac.jp

<sup>1</sup> 公開されていないが、研究のために各研究者によって作成された漫画コーパスはある。自作の漫画コーパスで漫画の言語的特徴について考察したウンサーシュッツ (2013) がその一例である。

## 2. コーパスの構築方法

構築方法は、まず、ゲームを一つ選定し、そのゲーム内の言語情報をすべて電子化し、通し番号や文字区分などの情報を付与した後、プログラミング言語で処理しやすいように CSV 形式の構造化データとして保存する（以下、本文データ）。次に、本文データの言語情報のみ形態素解析器で処理を行い、語彙素や品詞などのアノテーションが付与されたデータを別個の CSV 形式で保存する（以下、形態素解析データ）。以下、手順を紹介する。

### 2.1 対象ゲーム

ゲームにもアクションゲームやパズルゲームなどの様々なジャンルがある。本稿では言語学として研究する可能性を考慮し、ロールプレイングゲーム（以下、RPG）というジャンルをコーパス化の対象とする。RPG とはプレイヤーがゲームの主人公になり、架空の世界でほかのキャラクターに話しかける、情報を集めるなどして、物語を展開させていくものである。そのため、キャラクター間の会話が大量含まれており、言語情報がほかのゲームジャンルと比べて比較的多いと思われる。

具体的に選定したのは、1988年にエニックス社（現在はスクウェア・エニックス）によって発売された『ドラゴンクエスト3』である。同ゲームが日本で最も代表的なRPGの一つであり、後に複数のリメイク版が制作された。本稿ではコーパス化する際に、1996年のリメイク版のものを使用する。昔のゲームを選んだ理由は、画面構成とシステムが比較的単純であり、方法論を確立するために適しているからである。

### 2.2 電子化

コーパスは、機械可読性を持たせるために、電子化する必要がある。電子化の対象は、ゲーム内におけるすべての言語情報である。RPGのゲーム画面に現れる言語情報は、主にメニュー、メッセージ、セリフの3種類に分けられる。「メニュー」は図1の「ヒール」「HPを30～回復する」のようなゲーム内の道具・魔法名やそれに対する説明文（HPの数字などの変数は含まない）で、「メッセージ」は図2の「主人公はタンスを調べた！300ゴールドを手に入れた！」のようなゲーム内の進行状況を語るナレーション（小説での地の文に該当する）で、「セリフ」は図3の「こんにちは。ここは国立国語研究所ですわ。」のような物語の展開と関わるキャラクターによる発話（小説での会話文に該当する）である<sup>2</sup>。電子化する際、一つの枠に表示されたすべてのテキストを一行の文字列として記録する。

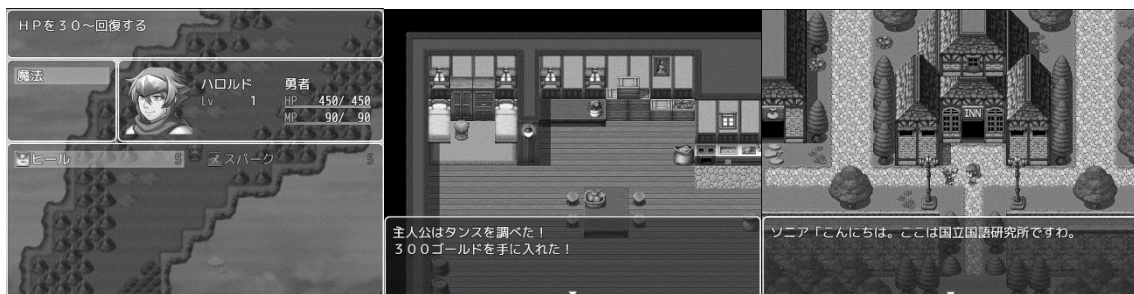


図1 メニュー

図2 メッセージ

図3 セリフ

<sup>2</sup> 著作権の関係で、実際に発売されたゲームの画面を掲載することができないため、図1～3は筆者が「RPG ツクールMV」というソフトウェアで作成したイメージである。

電子化する最も効率的な方法は、ゲーム媒体に記録された情報から直接文字を読み出すことであるが、昔のゲームの文字コードは PC 用の Shift-JIS や UTF-8 といった標準化されたものでなく、各ゲームメーカーの独自の方法でコーディングされたものが多かった。そのコード表が既知の場合は、理論的にそれを利用して文字情報を自動的に取り出せるが、直接ゲームの記録媒体にアクセスし、情報を抽出することは「解析」に該当する行為で、ゲームの利用規約に違反するおそれがある。

そこで、本稿では代わりに下記の方法を提案する。まず、実際にゲームを一通りプレイし、言語情報が画面に表示されるたびにスクリーンショット画像に保存した後、OCR ソフトウェアで文字を認識させる<sup>3</sup>。次に、一行の文字列ごとに番号や文字区分などの属性を付与し、後の整形や形態素解析に備え、プレーンテキスト（文字コード UTF-8）の CSV 形式で「本文データ」と名付けて保存する。設定した属性は表 1 の通りである<sup>4</sup>。

表 1 「本文データ」の構成

属性	説明
番号	すべての行の文字列に対し、重複しないように一意に付与された通し番号
区分	言語情報の種類を表す属性で、「メニュー」「メッセージ」「セリフ」の3種類がある
画像	キャラクターの外見を表す画像を名付けた名称（区分がセリフの場合のみ）
性別	キャラクターの性別（区分がセリフの場合のみ。判断不能の場合は「*」と記す）
名前	キャラクターの名前（区分がセリフの場合のみ。設定されていない場合は「*」と記す）
場所	キャラクターの所在（判断不能の場合は「*」と記す）
内容	ゲーム画面に表示された言語情報 <sup>5</sup> （区分がセリフの場合は発話内容）

実際に作成した「本文データ」の例は表 2 である。

表 2 「本文データ」の作成例

番号	区分	画像	性別	名前	場所	内容
932	セリフ	王	男	*	アリアハン	もはや 魔王に いどめるような者は おらぬ。 われわれには もう 希望はないのか...
933	セリフ	主人公の母	女	*	アリアハン	いいえ 王さま。 この子が います。 オルテガの 血をひく [主人公の名前] が。
934	セリフ	主人公の母	女	*	アリアハン	夫の 遺志は きっと この子が ついでみせますわ！
935	メッセージ				オリビアの岬	どこからともなく 悲しげな 歌声が 聞こえる.....。
936	セリフ	オリビア	女	オリビア	オリビアの岬	ああ エリック！ 私の 愛しきひと。 あなたをずっと まっていたわ。
937	セリフ	エリック	男	エリック	オリビアの岬	オリビア ぼくのオリビア もう君を はなさない！
938	セリフ	オリビア	女	オリビア	オリビアの岬	エリックーっ！

<sup>3</sup> 同一対象に何度話しかけても同様の内容が表示される場合、電子化するの是最初の一回のみとする。ただ、ゲームの進行によってセリフが変わるキャラクターの場合は異なる発話をすべて収録する。

<sup>4</sup> 今回のような小規模コーパスは必要ないが、複数のゲームを含むコーパスを作成する場合は、ゲームタイトル、ゲームジャンル、制作会社などの属性も必要になってくると思われる。なお、画像や名前を入れたのは、後に述べる役割語研究において必要な情報だからである。

<sup>5</sup> 主人公の名前のような、プレイヤーの入力によって変わる変数は適当な文字列を付けて [ ] で示す。

### 2.3 形態素解析

検索を効率的に行うためには、文字列を形態素に分割し、品詞などのアノテーション情報を付与する必要がある。ただ、半角文字が混在すると、後のプログラミングによる処理でトラブルに繋がることが多いため、形態素解析を施す前に、前作業として半角を全角に置換する。また、ゲームに出現するキャラクター名や地名などの固有名詞を事前にユーザー辞書に登録しておくことで解析精度が上がる。

実際に形態素解析を行う対象は「本文データ」の「内容」欄の情報である。用いる解析器は MeCab0.996<sup>6</sup>、辞書は UniDic1.3.12<sup>7</sup>である。解析結果をプレーンテキスト（文字コード UTF-8）の CSV 形式で「形態素解析データ」と名付けて保存する。ここで、「本文データ」で付与された「番号」を残し、両データが互いに紐付けられるようにしておけば、区分や名前などの情報が必要な際に、プログラミング言語を用いて抽出することができる。MeCab の処理で得られるのは短単位データであるが、Comainu<sup>8</sup>を用いれば、長単位データを得ることもできる。表 3 は、番号 934 の内容に対応する形態素解析の結果である。

表 3 「形態素解析データ」の作成例

番号	原文文字列	品詞(大)	品詞(中)	品詞(小)	活用型	活用形	語彙素	語彙素読み	語種
934	夫	名詞	普通名詞	一般	*	*	夫	オット	和
934	の	助詞	格助詞	*	*	*	の	ノ	和
934	遺志	名詞	普通名詞	一般	*	*	遺志	イシ	漢
934	は	助詞	係助詞	*	*	*	は	ワ	和
934	きつと	副詞	*	*	*	*	急度	キット	和
934	この	連体詞	*	*	*	*	此の	コノ	和
934	子	名詞	普通名詞	一般	*	*	子	コ	和
934	が	助詞	格助詞	*	*	*	が	ガ	和
934	つい	動詞	一般	*	五段-ガ行	連用形-イ音便	繼ぐ	ツグ	和
934	で	助詞	接続助詞	*	*	*	て	デ	和
934	みせ	動詞	非自立可能	*	下一段-サ行	連用形-一般	見せる	ミセル	和
934	ます	助動詞	*	*	助動詞-マス	終止形-一般	ます	マス	和
934	わ	助詞	終助詞	*	*	*	わ	ワ	和
934	!	補助記号	句点	*	*	*	!		記号

上記の手順を経て得られた「本文データ」の文字列は 5,900 行で、内訳はメニューが 72、メッセージが 1,162、セリフが 4,666 である。キャラクターは全部で 79 種類である。

「形態素解析データ」の総語数は延べ 67,715 語、異なり 3,164 語で、記号類を除けば延べ 58,407 語、異なり 3,150 語になる。今回用いたのは 1996 年発売のゲームであるが、記録媒体の大容量化が進み、近年制作されたゲームはさらに規模が増えると予想される。

<sup>6</sup> <https://taku910.github.io/mecab/>

<sup>7</sup> 最新版は 3.1.10 である。 <https://clrd.ninjal.ac.jp/unidic/>

<sup>8</sup> <http://comainu.org/>

## 2.4 注意点

ゲームは、書籍や新聞などの媒体とは異なる性質を持つため、コーパス化する際にいくつかの注意点がある。以下にそれを記しておく。

一つ目、ストーリーの時系列が一定でないゲームがある。書籍の場合、最初のページから最後のページまで読む順番が決まっており、ストーリーの進行が一本道である。それに対し、自由度が高いゲームの場合、先にAに行ってからBに行くこともできれば、Bに行ってからAに行くこともできる。文脈を問題にする研究の場合は要注意である。

二つ目、ストーリーの本筋とは別に、サブイベントが設定されているゲームがある。これらのイベントは見なくてもゲームをクリアできる。書籍の場合、書かれている内容は意図的に飛ばさない限り、すべて目にするのが原則であるが、ゲームの場合はプレイヤーのよって目に入る言語情報の量が異なる。また、いわゆる隠しイベントもあるため、熟練したプレイヤーでないと、すべての言語情報を抽出できない可能性がある。

三つ目、プレイヤーの行動によってストーリーが変わるゲームがある。マルチエンディングを採用しているゲームがその典型例である。具体的に、Aルートに進めば、Bルートのストーリーが見られないため、一周回のプレイですべての言語情報を収集できない。研究目的によって、作品総語数の認定が難しくなる場合がある。

四つ目、ゲームの歴史は書籍に比べてまだ浅く、著作権はほとんどまだゲーム会社が持っており、著作権処理は個人研究者として交渉することが困難であろう。著作権が切れた書籍のように自由に公開できるゲームがほぼない現状では、ゲームコーパスは個人研究用として構築することが可能であっても、公開が難しいことが予想される。

五つ目、ゲームは最初に発売されたオリジナル版と、後に異なるプラットフォームに移植するため新たに作成されたりリメイク版とが存在する場合がある。後者には内容の追加と削除が行われることがあり、収録された言語情報も変化する可能性がある。書籍に例えるならば異なる「版」に類似した概念である。コーパス化する際に、どのプラットフォームのバージョンを用いたかを明記する必要がある。

本節では、ゲームコーパスを構築する一例として、『ドラゴンクエスト3』というRPGを取り上げた。RPGに特化した方法ではあるが、ほかのゲームジャンルにも応用できると思われる。ただ、以上見てきたように、ゲームはその媒体の特殊性によって、書籍や新聞コーパスと同列に扱えない点にも注意が必要である。次節では、ゲームコーパスが実際にどの分野の研究に役立つかについて述べる。

## 3. ゲームコーパスの応用例

### 3.1 役割語に関する研究

金水 (2003) は、特定の人物像を思い浮かべるある特定の言葉づかいを役割語と定義した。また、金水 (2014) は、「役割語は、言葉のステレオタイプ」とした上で、「ステレオタイプが最も効力を発揮するのは、フィクションにおいて」と指摘した。そのフィクションには、具体的に漫画、アニメ、小説などが含まれ、これまで多くの研究が行われたが、ゲームに関しては麻 (2019a) 以外あまり見られず、以下にその例を挙げる。

麻 (2019a) は、『ドラゴンクエスト3』を対象に、役割語の調査を行った。具体的に、作中の登場キャラクターのセリフに出現した人称代名詞・助動詞・助詞・感動詞の4つの品詞を持つ表現の量的分布を統計量とし、クラスター分析を適用した。その結果、キャラクターをその話し方によって、「異人ことば」「女ことば」「中性ことば」「老人ことば」「男ことば」

「少年ことば」の6グループに分類することができた。さらに、この6グループの分類結果を利用して、特化係数で各グループのキャラクターの特徴的な言語表現を表4のように抽出した（麻 2019a より一部改変）。

表4 各グループのキャラクターの特徴語

	代名詞	助動詞	格助	副助	接助	終助	準体	感動詞
異人	わし, おまえ, だれ, わたし	ていく, てる, ない		か				おお
女	あたし	ぎます, てらっしゃる, てあげる, ちゃう, ておく		ったら, なんて, って	けれど, けど	かしら, もの, わ, の, ね (他1語)		きゃー, まあ, おほほほほ, あら, はあ (他8語)
中性	オイラ, あなた, 彼女	ます, ベシ						やあ
老人	おぬし, そなた, わし, おのれ	じゃ, とる, である, まい		ぞ, なん, なぞ		のう, ぞい, ぞ		はくしょん, ふーむ, ふあ, やれやれ, ふむ (他3語)
男	あたい, おめえ, オレ, あんた, われ (他1語)	やがる, ちまう, てやる, らしい, てる		なんか, って, しか	けれど, ちゃ	ぜ, い, さ, や, な	ん	へい, おい, へー, おっと, ひゃー (他4語)
少年	ボク	ちゃう, てる, ていく		って, なんて	けれど, から	よう, さ, や, なあ, の (他2語)	ん	わーい, ねえ, ん, え

麻 (2019a) は、ゲームコーパスにも言語資源としての価値があることを示唆した。また、上記の研究では言及されていないが、(1)～(3)のような侍ことば、幼児ことば、ロボットことばもゲームで観察できる。(3)は表記の手段で表現された役割語である。

- (1) わしもいく**でござる**！世界をこのままほうっておくわけにはい**か**ない**でござる**！  
 (『ファイナルファンタジー6』カイエン)
- (2) ええっ！？フェアリーちゃんが？た**た**、たいへ**ん**で**ち**！どうすればたすけられ**ま**ちか？めがみさま！ (『聖剣伝説3』シャルロット)
- (3) ケガのこうみょうというやつ**デショウカ**？ぐうぜんにも、飛行機能が手に入り**マシタネ**！ (『クロノ・トリガー』ロボ)

さらに、定延 (2007) などで述べられたような、人工的な表現または方言を用いるキャラクターもゲームに存在する。下記の3例がこれに該当する。(4)が人工的な表現で、(5)(6)が関西弁と名古屋弁をベースにした表現だと思われる。

- (4) みんな！みんな死んじ**ゃ**ったと思**っ**た**クポ**！生きてた**クポ**ー！よ**か**った**クポ**ー！  
 ボクもい**っ**しよに戦**うクポ**！！ (『ファイナルファンタジー6』モーグリ)

- (5) ちゃうねん！ええ話やわあ。あんさん達の話に、めっちゃ感動しとったんよ！マナが少なくなったせいか、最近どうも涙もろうて... (『聖剣伝説3』 ウンディーネ)
- (6) でも、度重なるじしんで、洞窟の入り口がふさがってしまったぎゃー。山の行き止まりの大岩の近くだぎゃー。(『聖剣伝説3』 ダークプリースト)

以上見てきたように、ゲームコーパスは役割語の分析に役立つことが分かった。この点は次に述べる日本語教育にも関連している。

### 3.2 日本語教育に関する研究

国際交流基金が2018年に実施した「海外日本語教育機関調査」によると、世界の日本語教育機関に在籍する学習者の学習目的・理由として挙げた項目のうち最も回答が多かったのは「マンガ・アニメ・J-POP・ファッション等への興味」(66.0%)であるという(国際交流基金2020)。このなかにゲームも含まれており、具体的な内訳の割合が提示されていないが、ゲームも日本語学習者にとって重要な動機づけになっていると思われる。

ゲームと日本語教育の関係に焦点を当てた研究に麻(2008, 2019b)がある。麻(2008)は3種のゲームからそれぞれ延べ1,000語まで抽出する方法で語彙調査を行い、ゲームに出現した語彙の旧日本語能力試験における難易度を明らかにしようとしたが、調査単位が統一されていないなどの問題点がある。麻(2019b)はその改良版として『ドラゴンクエスト3』を対象に行った調査である。その結果の一部を表5に示す(麻2019bより引用)。

表5 旧日本語能力試験レベルの分布状況

	級外	1級	2級	3級	4級
異なり語数	1,003 (33.19%)	358 (11.85%)	889 (29.42%)	355 (11.75%)	417 (13.80%)
延べ語数	3,333 (11.23%)	1,205 (4.06%)	5,526 (18.62%)	5,485 (18.48%)	14,131 (47.61%)

異なり語数で見れば最も多く占めたのは級外の33.19%で、その次は2級の29.42%である。一方、延べ語数で見れば最も多いのは4級の47.61%であることが分かる。なお、最も難しい1級では異なり語数でも延べ語数でも低い割合を占めていることも読み取れた。また、名詞に関しては、「魔王」「勇者」「魔物」「魔法」「剣」のような、ゲームの世界観を反映させたものが多く、形容詞(イ形容詞)には「愛しい」「麗しい」「劳しい」のような文語の形容詞が多いという結論も同調査で得られた。

このように、ゲームには初中級の言葉が多い反面、出現した名詞と形容詞の頻出語が日常生活のそれとは異なる側面をもっており、特に教師のいない独学の学習者の場合は、目にした表現をそのまま覚えることがあるため、注意が必要である。この点は前述した役割語に関しても同じことが言える。これもゲームコーパスを調査してはじめて分かったことである。

### 3.3 その他の分析例

ほかにゲームコーパスに見られる言語表現に関する分析例として、二例挙げておく。

一つ目の例は「って」という表現の使用である。この表現は一般的に「引用する内容」を表す格助詞として用いられているが、ゲームの中には(7)(8)のような、日常生活であまり耳にしない使い方が存在する。



- (7) 戦争が、始まるかも知れないんだ**って**!? ホントウなのかい? あんた傭兵なんだから、何とかしておくれよ! (『聖剣伝説3』老婆)
- (8) ちょっと待って...いっしょに、ここから出よう。どうやってだ**って**? ふん、オイラに、まかせろ! (『聖剣伝説3』ケヴィン)

上記の例はいずれも主人公が別のキャラクターに話しかけた後に返ってきた返事である。「って」で表す引用の内容は、主人公が話しかけた内容の繰り返しだと思われるが、日常生活でわざわざ相手の発話内容を「って」で繰り返す人はあまりいない。それにもかかわらず、ゲームの世界ではこのような発話が多く見られる理由は何だろう。一つ考えられるのは、ゲーム内の発話は登場キャラクター間のコミュニケーションだけではなく、ゲーム制作者が登場キャラクターを通してプレイヤーに伝えるメッセージも含まれる、という解釈である。一般的に、プレイヤーに何らかの状況を説明するのに最も直接的な手段はナレーションだが、RPGは「プレイヤー自身が主人公である」という一人称視点の雰囲気を作るため、ナレーションと相性が悪い。残された手段は発話になるが、RPGは上記の一人称視点という理由で、主人公の発話を表示せず、代わりにプレイヤー自身にその内容を脳内補完させることが多い。つまり、プレイヤーは「話しかける」という行為を実行した後に、相手の返事内容を見て主人公の発話を想像するわけである。制作者がこのゲーム内特有の仕組みを利用して、プレイヤーに伝達したい情報を登場キャラクター経由で言わせている可能性がある。

これは、同じ「作られた話し言葉」である漫画や小説では観察しにくい現象である。漫画では、長方形の吹き出しで「ナレーション」、角を泡にした吹き出しで「心的描写」を表現することができる。小説でも「地の文」で類似した効果が得られる。漫画や小説では、作者がこれらの手段を用いて読者にメッセージを伝えることができる一方、ゲームの場合はそれがしにくいいため、「って」という表現を用いて同様の効果を狙うのである。

二つ目の例は、命令形の多用である。命令形は「命令」や「指示」を表す一種の待遇表現である。人間関係などの制限により、現実世界では命令形は限られた場面でしか用いられない。しかし、ゲームでは(9)～(11)のように、比較的多く現れている。

- (9) さあ! この水筒をもってとつと**いきやがれっ!** うぐっ! ..... (『ドラゴンクエスト3』あらくれ)
- (10) よし、じゃあふたてに分かれよう。おまえらは、港に**急げ**。この町から脱出するんだ! (『聖剣伝説3』デュラン)
- (11) セリス将軍!! さあ、もう芝居はよい。そいつらの魔石を持ってこっちへ**来い** (『ファイナルファンタジー6』ケフカ)

ゲームはほとんど架空の世界観を持っており、シチュエーションや登場キャラクター間の人間関係も現実世界では実在しにくいものが多い。それが結果的に「命令形」を使いやすい環境に仕立て上げたのだと思われる。

このように、ゲームというフィクションの世界での言語表現は、リアルの世界のそれとは異なることがある。本節で取り上げた「って」、命令形、及び役割語は現実世界で多用すると、場合によって幼稚に見られ、不利益を被るおそれがある。特に日本語を独学している学習者は、教師によるフィードバックがないため、自分で判断することが困難である。こうした注意が必要な表現も、ゲームコーパスの調査を通して発見することが可能である。

#### 4. ゲームコーパスの位置づけ

ここまで、ゲームコーパスが言語資源としての活用例を見てきたが、本節では、ゲームコーパスがコーパス言語学での位置づけを考える。結論を先に述べると、均衡コーパスの一部、対訳コーパス、マルチメディア・コーパスとして活用する可能性があると思われる。以下、このことについて少し触れておく。

##### 4.1 均衡コーパスの一部として活用する可能性

最も均衡性が確保されている BCCWJ であっても、作業上などの理由でゲームが収録されていない。すでに1節で述べたように、確かにゲームの言語情報を電子化するのに手間がかかるが、ゲームに現れた言語表現は日本語の中に実在している。また、3節で見たように、ゲームはその特殊な性質で、ほかの言語使用域とは異なる用法も複数存在する。これらの理由を考慮し、今後、ゲームを均衡コーパスの母集団の一部として取り入れる検討も必要になってくると考えられる。無論、どのジャンルのどのゲームをどれほど対象に入れるかは、日本全国の制作本数を総合的に考慮し、慎重にサンプリングする必要があるが、日本語における言語使用実態を解明する一環としてやる価値はあると思われる。

##### 4.2 対訳コーパスとして活用する可能性

ゲームコーパスは均衡コーパスとしてではなく、特別の目的を果たすために構築された簡易コーパスとしての役割もある。対訳コーパスがその一例である。ゲームは、ほかの言語作品と比べて、対訳コーパスの構築が比較的容易である。その理由は、人気なゲームは多言語版も発売されていることが多く、そして、ゲームは一度画面の枠に表示できる文字情報の量が限られているため、ページごとに保存されることが一般的である。そのページの情報を利用すれば、両言語のほぼ文単位での対応関係を得ることができる。

##### 4.3 マルチメディア・コーパスとして活用する可能性

石井・孫 (2013) はテレビ放送のデータを対象に、マルチメディア・コーパスを作成し、指示詞と指差しのような言語表現と非言語表現との関係を調べた。これが現実世界を反映させたマルチメディア・コーパスであれば、ゲームがフィクションの世界を反映させたマルチメディア・コーパスという位置づけになる。前述したように、フィクションでは言語表現としてのステレオタイプが観察されやすいが、同じことは非言語表現に対しても言える。近年のゲームは、キャラクターの身振りや表情なども工夫を凝らして作られることが多く、非言語表現のステレオタイプを観察するのに利用することもできよう。

#### 5. おわりに

本稿では、まず『ドラゴンクエスト3』という RPG を例にして、ゲームコーパスの構築方法を説明した。次に、ゲームコーパスを言語資源として活用する事例をいくつか紹介した。具体的に、役割語研究や日本語教育研究などの領域での利用例を見た。また、「って」のようなゲーム特有の言語的特徴も確認した。最後に、ゲームコーパスがコーパス言語学における位置づけを述べた。本稿が今後ゲームコーパス作成の参考、またはゲーム関連研究のきっかけになればと思う。

ただし、ゲームジャンルは RPG のみでなく、アクションゲームやパズルゲームなども存在する。本稿で提案した方法論で異なるジャンルのゲームコーパスを作成する際に、さらな

る調整が必要となるだろう。また、本格的にゲームコーパスを作成することになると、先行研究、例えば石川 (2012) でまとめられた、(1) 書き言葉や話し言葉などの現実の言語、(2) 大規模、(3) 基準に沿って網羅的・代表的に収集するもの、(4) コンピュータ上で処理できるデータとして保存するもの、(5) 言語研究に使用するもの、といったコーパスの観点を満たすように構築しなければならない。いずれも今後の課題とする。

## 文 献

- 石井正彦・孫栄爽 (2013) 『マルチメディア・コーパス言語学—テレビ放送の計量的表現行動研究—』大阪大学出版会
- 石川慎一郎 (2012) 『ベーシック コーパス言語学』ひつじ書房
- ウンサーシュッツ ジャンカーラ (2013) 『マンガにおける言語の役割—構造・語彙・登場人物という三つの観点から』一橋大学博士論文
- 金水敏 (2003) 『ヴァーチャル日本語 役割語の謎』岩波書店
- 金水敏 (編) (2014) 『<役割語>小辞典』研究社
- 国際交流基金 (2020) 『海外の日本語教育の現状 2018 年度日本語教育機関調査より』 <<https://www.jpff.go.jp/j/project/japanese/survey/result/dl/survey2018/all.pdf>>最終アクセス：2022 年 8 月 17 日
- 国立国語研究所 (2015) 『現代日本語書き言葉均衡コーパス利用の手引 第 1.1 版』 <<https://clrd.ninjal.ac.jp/bccwj/doc.html>>最終アクセス：2022 年 8 月 16 日
- 定延利之 (2007) 「キャラ助詞が現れる環境」金水敏 (編) 『役割語研究の地平』くろしお出版
- 麻子軒 (2008) 「テレビゲームが日本語教育の課外教材としての位置づけ—RPG を中心とした語彙調査の中間発表—」『大学院教育改革支援プログラム 平成 19 年度活動報告書 海外研修事業編』お茶の水女子大学, pp.245-249.
- 麻子軒 (2019a) 「計量的アプローチによる役割語の分類と抽出の試み—テレビゲーム『ドラゴンクエスト 3』を例に—」『計量国語学』32 巻 2 号, pp.103-116.
- 麻子軒 (2019b) 「テレビゲーム『ドラゴンクエスト 3』の語彙調査—日本語教育への応用可能性に向けて—」『日本語教育学会予稿集』2019 年度秋季大会, pp.131-136.

## 使用ゲーム

- スクウェア (開発元) (1994) 『ファイナルファンタジー6』スクウェア
- スクウェア (開発元) (1995) 『聖剣伝説3』スクウェア
- ハートビート・アルテピアッツァ (開発元) (1996) 『ドラゴンクエスト3』エニックス