

## 物語創作理論書の計量テキスト分析

### —小説・映画脚本・演劇脚本の執筆における概念構造の比較—

高橋 椋一 村井 源 猪原 健弘  
東京工業大学大学院 社会理工学研究科

本研究の目的は、物語作品の創作におけるジャンル間の差を明らかにすることである。まず、物語作品の作り方を解説した書籍である物語創作理論書を収集し、小説・映画脚本・演劇脚本の創作理論書の電子コーパスを作成した。つづいて、物語創作に関連した名詞を抽出し、それらを32のカテゴリーに分類した。そして、各カテゴリーに含まれる語の出現頻度を算出し、カイ二乗検定を用いて3つのジャンル間で比較した。その結果、各ジャンルを特徴づける概念カテゴリーはそれぞれ異なり、小説は文章と物語展開に関わるカテゴリーに、映画脚本は登場人物に関わるカテゴリーに、演劇脚本は作品製作に関わるカテゴリーに特徴づけられることが明らかになった。

### Identifying the Conceptual Structure of Narrative Writing: A Quantitative Comparison of Screenwriting, Playwriting, and Fiction Writing

Ryoichi Takahashi Hajime Murai Takehiro Inohara  
Graduate School of Decision Science and Technology  
Tokyo Institute of Technology

This study aimed to investigate the differences between screenwriting, playwriting, and fiction writing. We first built a narrative writing books corpus, consisting of thirty books written by expert writers about how to write a screenplay, playscript, or fiction novel. We then extracted nouns related to storytelling from the corpus. The extracted words were classified into thirty-two categories, and through a chi-square test, the frequency of categorized words was compared across the three genres of writing. The results revealed that the categories that characterize screenwriting, playwriting, and fiction writing differ.

#### 1. 問題と目的

人が物語を産み出す際の心や認知の働きについては、主に心理学・認知科学の分野において研究が行われている。そうした研究は、文章産出研究とクリエイティブ・ライティング研究の2つに大別が可能である。文章産出研究は、心理学における文章理解研究の流れを汲んでおり、人間一般が物語を産み出す際の認知メカニズムやその発達過程の解明が、研究の主眼となっている(e.g., [1][2])。一方、クリエイティブ・ライティング研究は創造性研究の流れを汲んでいる。研究の対象は熟達した創造的な作り手であり、彼らの持つ特性や技能の分析が行われている[3]。

こうした物語の産出に関わる既存研究には、2つの問題点を指摘することが出来る。1点目は、熟達者を対象とした標本調査が不足していることである。物語産出研究では人間一般の物語る能力とその発達過程が対象となるため、学生や幼児を対象とした標本調査が行われることが多い。一方、クリエイティブ・ライティング研究において

は、熟達した作り手から得た発話プロトコルを分析するというアプローチや、伝記や日記といったデータを用いた人文学的・精神分析的アプローチが試みられているが、こうした研究は分析対象となる作り手を少数に絞った事例研究が多い。したがって、結果の一般化可能性が担保されないという欠点が存在する。熟達者を対象とした標本調査が少ない原因としては、芸術心理学的研究に特有の、データ収集における困難さが考えられる。量的データ・質的データを問わず、プロフェッショナルな芸術家から心理学的データを収集することは一般的に困難である。歴史に残る著名な芸術家にはそもそも存命の人物が多くなく、そうした芸術家を対象に実験・調査を直接的に行うことは困難である。また、存命の芸術家であっても、研究協力に対するインセンティブや研究倫理における諸問題が障壁となり、分析可能なデータを収集することには困難が伴うと考えられる。こうした理由から、熟達者を対象とした標本調査が不足していると推測される。

2点目は、表現媒体やジャンルによる執筆方略の差が考慮されておらず、領域固有性の問題が検討されていないことである。心理学研究において「物語文」は「説明文」と対比される概念であり、両者の差についての検討がなされることが多かった。そのため、「物語文」の中におけるジャンルの差違(e.g., 小説, 昔話, 脚本, ゲーム)は考慮されないことが多い。しかし、人間の認知や学習はその対象となる領域に依存し、他の領域から比較的独立しているという領域固有性[4]の理論が物語産出にも該当すると仮定した場合、産出する物語の媒体やジャンルの差違は、物語産出の様相に影響を及ぼすと考えられる。しかし、物語を産み出すプロセスをジャンル横断的に比較した研究は見当たらない。

そこで本研究では、(1)熟達した物語の作り手に関する一般化可能性の高い結果を得ることと、(2)物語産出におけるジャンル間の差違を明らかにすることを目的とする。そして上記の目的を達成するアプローチとして、「物語創作理論書」を対象とした「計量テキスト分析」を行うことを提案する。

## 2. 方法

### 2.1 分析対象：物語創作理論書

本研究では、一般に「創作論」や「シナリオ論」といった言葉で呼ばれる、作家が物語創作の技術・ノウハウを記述した「ハウツー」のテキストを対象として分析を行う。現在のところ、こうした物語創作におけるハウツー本を表す語は論者によって異なり、明確な定義もない。日本においては、「小説作法」「脚本術」「シナリオ論」などとジャンルによって呼称が異なり、物語創作における理論全般を表す名詞が存在しない。また、英語圏においても How-to Write a Narrative Story や Storytelling Theory など様々な名詞句が存在し、統一的な呼称も見当たらない。そこで本研究では、呼称を物語創作理論(theory of narrative writing)と統一した上で、物語創作理論書を(1)物語創作の熟達者が、(2)自身の物語創作の技術を解説した、(3)説明的文章と定義する。

物語創作理論書は、物語創作の初心者向けに出版されることが多く、いわば物語創作の「参考書」である。したがって、それが「初心者向け」に「要点を中心として」語られた文章であるという点において、作家の執筆過程における認知や思考のプロセスのすべてを直接的に反映したテキストであるとはいえない。しかし、熟達した作り手が物語の創作過程においてどのような点を重要視しているかが明示的に記述されているという点で、熟達した作り手の持つ技能や知識を分析するには適当なテキストであるということができる。また、物語創作理論書には、小説・文学・映画脚本・演劇脚本・テレビドラマ・テレビゲーム

といった複数のジャンルのものが存在する。したがって、物語作りにおいて表現媒体やジャンルの違いに関わらず共通している側面と、それぞれの媒体・ジャンルに固有の側面という、二つの側面を分析することが可能である。そして、数多くの物語創作理論書が出版されているため、サンプリングによる標本調査が行えるという大きな利点があるといえる。

### 2.2 分析手法：計量テキスト分析

本研究では、物語創作理論書を分析する手法として、計量テキスト分析を用いる。計量テキスト分析とは、テキスト型データに対して統計学的手法を用いた定量的解析を行うことで、再現可能性を担保した形でテキストから情報を得る分析方法である。コンピュータを用いた計量テキスト分析には、大量のテキストを一様の基準で分析することができるという大きなメリットが存在する。

計量テキスト分析は近年、心理学・認知科学研究にも適用されており、テキストの解析を通して人間のこころの働きを解明することで、一定の成果をあげている。従来では、質問紙や実験における自由記述回答・言語プロトコルといった言語データに対して量的分析が行われることは稀であった。しかし、最近ではこうした言語データに計量分析を行うことで、質的分析とは異なった角度から人間の心理メカニズムを明らかにした研究が存在している[5][6][7]。また、実験協力者から直接的に取得される言語データに留まらず、書籍や雑誌に対して解析を行うことで、文章の書き手の価値構造や思考パターンを明らかにした研究も存在する[8][9]。

こうした理由から、多数の物語創作理論書を客観的な基準のもと分析するには計量テキスト分析が最適であると考えられたため、分析手法として採用した。

## 3. コーパスの構築

分析対象となる「物語創作理論書コーパス」の構築を行った。まず、物語創作理論書コーパスに含める書籍は、米国で出版・流通している英語で書かれた原著の書籍から選定することとした。その理由は2つある。第1の理由は、現在全世界で流通している物語創作理論書の多くが、米国の原著を翻訳したものであるという点である。現在日本で流通している物語創作理論書のうち、純文学やポピュラー文学といったジャンルにおいては、日本語が原著のものも見られる。しかし、映画・演劇・テレビドラマなどの物語創作理論書は、米国で出版された書籍が日本語に翻訳されたものが多く、日本語が原著のものはほとんど存在しないと言ってよい。また、異なる言語で記述された文章を比較する場合、内容の差違だけでなく言語体系の差違も検討する必要が生じることから、分析対象となるテキストの言語は統一されている

方が望ましい。したがって本研究では、様々なジャンルにおける物語創作理論書が原著で存在する、米国の書籍のみを対象とした。第2の理由は、量の観点である。米国においては物語創作理論書が大学や専門学校のテキストとして用いられていることが多く、数多くの書籍が出版されている。そのため、分析を行うのに十分な数の書籍が存在している。再版も多いため入手可能性が高いこともメリットとしてあげられることから、分析対象に適していると判断した。

また、本研究では小説・映画・演劇の3つのジャンルにおける物語創作理論書を分析対象とする。これは、上記の3ジャンルにおいては、テレビドラマやアニメーション、純文学といった他ジャンルと比較し、数多くの物語創作理論書が出版されているためである。

分析対象となる書籍の選定にあたっては、米国のオンライン書店 Amazon.com を参考にした。具体的には、小説(fiction writing)カテゴリー、映画脚本(screenwriting)カテゴリー、演劇脚本(playwriting)カテゴリーの3つのカテゴリーに含まれる書籍の中から選定を行った。選定においてはまず、映画脚本および演劇脚本カテゴリーには物語創作理論書だけでなく、作品自体(実際の映画または演劇の脚本)が含まれているため、それらを取り除いた。つづいて、3つのカテゴリーから10冊ずつ書籍のサンプリングを行った。その際には以下の2つのルールを設け、2つの基準全てを満たすように書籍をサンプリングした。

(1)現代における高い社会的価値: Amazon.com における3つの各カテゴリーにおいて、それぞれ上位100位以内にランクインしている書籍の中から分析対象を選定した。

(2)執筆者の多彩さ: 特定の著者の物語創作理論に結果が偏ることを避けるため、同一の著者による書籍は2冊以上サンプリングしないこととした。

つづいて、上記の手順でサンプリングした書籍をテキストデータに変換し、物語創作理論書コーパスを構築した。物語創作理論書コーパスは、小説創作理論書コーパス、映画脚本創作理論書コーパス、演劇脚本創作理論書コーパスの3つのサブ

コーパスからなる。表1は、3つのサブコーパスの概要および基本統計量を示したものである。なお、基本統計量の算出において句読点は除いている。

その後、コーパスに含まれる全テキストに対して TreeTagger[10]を用いた形態素解析を行い、品詞情報および見出し語情報を付与してタグ付きコーパスを構築した。

#### 4. 頻出名詞の抽出とカテゴリー分類

本研究では、物語創作に関わる基本的な概念を分析するため、テキストに出現する一般名詞に着目して分析を行う。分析ではまず、出現頻度の高い一般名詞の中から、物語創作に関連すると考えられる語彙を筆者が抽出し、分類を行った。分類の手順は以下の通りである。

はじめに、小説創作理論書コーパス、映画脚本創作理論書コーパス、演劇脚本創作理論書コーパスの各サブコーパスにおいて、出現頻度が100回以上の一般名詞を抽出した。その結果、小説創作理論書コーパスから306語、映画脚本創作理論書コーパスから353語、演劇脚本創作理論書コーパスから331語の、計990語が抽出された。これらの語から重複したものを除いた490語からなる語彙群を分類に用いることとした。

つづいて、得られた語彙群から以下の2つの条件に合致する語彙を抽出した。

(1)物語創作と関連性の高い語彙である: 物語創作と強い関連を持つ語彙とそうでない語彙とを分け、関連の弱い語は取り除いた。たとえば、lot は a lot of というイディオムとして使用されることが多く、物語創作と直接的な関連は持たない。また、week や day といった語は「1日・1週間で数ページを目安に執筆しましょう」といった形で、創作にわずかながら関連した文脈に登場する場合もあるが、物語創作理論書を強く特徴付ける語であるとは考えづらい。このような、ストップワード(大半の文章において一定程度出現するため、文章を特徴付ける役割を持たない語)であると考えられる語は、分類からは除外している。

(2)テキスト中での単語の用法が一貫している: 単語が複数の意味を持つ多義語である場合、その多義性を考慮しない場合には一般に、計量テキスト分析において分析精度は低下すると考えられる。そこで、カテゴリー分類を用いた分析の精度低下を避けるため、テキスト中での用法が一貫していない多義語の除外を行った。具体的には、語彙が複数のカテゴリーに分類されうると判断された場合に、コーパスから当該単語を含む用例をランダムに10例抽出し、単語の用法の判定を行った。10用例中7以上の用例が同様のカテゴリーに属する用法であると判断された単語のみを分類し、基準を満たさない単語は除外した。たとえば、characterization は「特徴の描写・説明」という意味で用いられる場合が多く見られるが、

表1 使用テキストデータの概要

	小説	映画脚本	演劇脚本
テキスト数(冊)	10	10	10
テキストサイズ(MB)	5.65	5.36	5.37
平均テキストサイズ(MB/冊)	0.57	0.54	0.54
平均単語数(語/冊)	96516.00	99157.20	98005.60
単語数の標準偏差(語)	42511.19	33010.61	33252.98
平均文数(文/冊)	5180.90	5389.10	5181.70
文数の標準偏差(文)	2830.82	1863.43	1877.89
平均文長(語)	19.49	18.53	19.40
文長の標準偏差(語)	2.85	2.22	3.19

表 2 一般名詞のカテゴリ分類における単語抽出例

分類対象語彙	分類から除外された語彙	
	用法が一貫していない	物語創作と関連性が低い
hero, protagonist, antagonist, character, phrase, description, exposition, narrator, dialogue, monologue, screenwriting, script, director, publisher	characterization, drama, style, reading, setting, situation, idea, development, curiosity, storytelling, tragedy, critic, adventure, fiction	lot, minute, week, day, month, past, future, clock, father, mother, man, woman, school, information, job, difference, material, issue,

表 3 カテゴリ分類された一般名詞の一覧

上位カテゴリー	下位カテゴリー	分類された一般名詞
登場人物	登場人物一般	character
	主人公	hero, protagonist
	協力者	ally
	敵対者	antagonist, adversary, villain, opponent
物語	構成	structure
	背景	backstory, background
	あらすじ	outline, summary, logline, synopsis, storyline, plot
	わきすじ	subplot
場面	物語	story, storey, tale, narrative
	場面一般	scene
	序盤	opening, setup, beginning
	中盤	middle, midpoint
	終盤	end, ending, climax, conclusion, resolution, solution
言語・記述	対立・争い	confrontation, opposition, battle, fight, struggle
	文章	phrase, subtext, context, sentence, passage, metaphor, paragraph, chapter, section, page, format,
	描写	description, exposition
	会話	dialogue, conversation
	語り	narrator, monologue
制作	視点	viewpoint
	作家	writer, author, novelist, screenwriter, storyteller, playwright
	執筆行為	writing, screenwriting, playwriting, storytelling
	制作関係者・組織	editor, producer, artist, manager, stuff, publisher, director, agent, studio, designer, partner, executive
	俳優	actor, cast
	演技	performance, act, rehearsal
	劇場	theatre, theater, stage
作品	セールス	money, sale, market, company, publishing
	小説・脚本	screenplay, novel, manuscript, script
	媒体	medium, film, movie, TV, screen, television, book
	主題	theme
鑑賞	ジャンル	genre
	鑑賞者	reader, audience, reading
	感情	emotion, fear, surprise, suspense, desire, passion, love

「登場人物の役割を演じること」を指す用法も多く見られる。また、body は俳優や演技に関する文脈では「身体・肉体」という一般的な用法が多いが、場合によっては「物語の主要な構造・部分」という意味でも用いられている。このような、同一の単語ではあるが文脈によって多義的に用いられている単語は、物語創作に関連する語であっても、分類から除外している。

上記の条件による語彙の抽出例を表 2 に示す。条件を満たすと判断されたのは 115 語であり、

109 語が用法の一貫しない多義語と、266 語が物語創作に関連の低い語と判断された。そして、条件を満たす 115 語は 32 のカテゴリーに分類された。また、得られた 32 のカテゴリーのうち類似のものをまとめた結果、7 の上位カテゴリーが得られた。分類した名詞の一覧は表 3 に示す。

## 5. カテゴリーごとの頻度分析

つづいて、各下位カテゴリーに含まれる語彙の

表 4 残差分析による出現頻度のカテゴリー間比較

上位カテゴリー	下位カテゴリー	小説	映画脚本	演劇脚本
登場人物	登場人物一般	4792 ▽▽▽	5554 ▽▽▽	7711 ▲▲▲
	主人公	1581 ▲▲	2811 ▲▲▲	624 ▽▽▽
	協力者	14 ▽▽▽	113 ▲▲▲	84 ▲
	敵対者	253 ▽▽▽	910 ▲▲▲	263 ▽▽▽
物語	構成	529 ▲▲▲	505	348 ▽▽▽
	背景	242 ▲▲▲	154 ▽▽▽	204
	あらすじ	1223 ▲▲▲	1196 ▽▽▽	1037 ▽▽
	わきすじ	92 ▽▽▽	382 ▲▲▲	56 ▽▽▽
	物語	5710 ▲▲▲	6573 ▲▲▲	3120 ▽▽▽
場面	場面一般	2513 ▲▲▲	2931 ▽	2508
	序盤	511 ▽	776 ▲▲	598
	中盤	269 ▲▲▲	298	176 ▽▽▽
	終盤	1079 ▽▽▽	1584 ▽▽	1744 ▲▲▲
	対立・争い	383	664 ▲▲▲	278 ▽▽▽
言語・記述	文章	2711 ▲▲▲	2122 ▽▽▽	1973 ▽▽▽
	描写	253 ▽▽	424 ▲▲▲	297
	会話	453 ▽▽▽	1145 ▲▲▲	985 ▲▲▲
	語り	576 ▲▲▲	88 ▽▽▽	240 ▽▽▽
	視点	395 ▲▲▲	17 ▽▽▽	18 ▽▽▽
制作	作家	2893 ▲▲▲	2666 ▽▽▽	3350 ▲▲▲
	執筆行為	814 ▲	841 ▽▽▽	891 ▲▲
	制作関係者・組織	1686	2012 ▽▽▽	2111 ▲▲▲
	俳優	92 ▽▽▽	329 ▽▽▽	1386 ▲▲▲
	演技	159 ▽▽▽	693 ▽▽▽	1179 ▲▲▲
	劇場	185 ▽▽▽	296 ▽▽▽	2817 ▲▲▲
作品	セールス	623 ▲	764	570 ▽▽
	小説・脚本	2085	3394 ▲▲▲	1520 ▽▽▽
	媒体	2573 ▽▽	5246 ▲▲▲	1256 ▽▽▽
	主題	430 ▲▲	451	389
	ジャンル	119 ▽▽▽	375 ▲▲▲	178 ▽▽
鑑賞	鑑賞者	2881 ▲▲▲	2722 ▽▽▽	2829 ▲▲
	感情	1039 ▽▽▽	2164 ▲▲▲	1498

合計出現頻度をサブコーパスごとに算出した。その後、各サブコーパスにおける語彙の合計出現頻度を、カイ二乗検定を用いて比較した。カイ二乗検定および残差分析の結果を表 4 に示す。表 4 において▲は 5%の有意水準で、▲▲は 1%の有意水準で、▲▲▲は 0.1%の有意水準で多いことを示す。同様に、▽・▽▽・▽▽▽はそれぞれ 5%・1%・0.1%の有意水準で少ないことを示している。

残差分析の結果、小説の創作理論書では「文章」「語り」「視点」と、言語・記述に関わるカテゴリーにおいて言及が有意に多かった。これは、文章構成や人称の問題など、言語芸術である小説に特有の問題に、小説の書き手が強く意識を向けていることを示す結果である。また、「構成」「背景」「あらすじ」「物語」といったストーリー展開に関わるカテゴリーにおいても、小説では言及が有意に多く見られた。ストーリーに対する関心の高さも、小説の特徴のひとつだといえる。

映画脚本の創作理論書では「主人公」「敵対者」および「協力者」に関する言及が、演劇・小説に比べて有意に多かった。ここから、映画脚本の書き手が主演・悪役・味方というはっきりした区別のもとに登場人物を位置づけており、登場人物間の役割分けを意識して物語を作り上げていることが読み取れる。一方、小説・演劇脚本では、そうした区別が映画ほど強くは意識されないことが示唆された。

演劇脚本の創作理論書では、「制作関係者」「俳優」「演技」といった、製作に関わるカテゴリーに対する言及が有意に多かった。これは、演劇脚本の書き手は脚本自体だけでなく、自身の書いた脚本がどのような人々によってどのように演技され、どのような場所で舞台化されるかといった事柄に高い関心を持っていることを示唆する結果である。一方、同じ総合芸術である映画において、製作に関わるカテゴリーの出現頻度は、有意に少ないものも多く見られた。これは、映画脚本の作り手は、自身の脚本がどのような人々によってどのように映像化されるかといった事柄に、他のジャンルよりも関心が低いことを示唆する結果である。

## 6. 結論と今後の課題

本研究では、小説・映画脚本・演劇脚本の物語創作理論書コーパスを構築し、コーパスを用いて分析の基礎となる語彙のカテゴリー分類を行った。そして、分類されたカテゴリーを用いて 3 ジャンルの物語創作理論書における一般名詞の出現頻度を算出し、カイ二乗検定を用いて比較を行った。その結果、各ジャンルを特徴づけるカテゴリーが明らかになった。

今後の課題としては、本研究で行った分類の妥当性を担保するため、複数人で分類を行った上で分類の一致度を測定することがあげられる。

また、本研究において、類似した下位カテゴリーをグループ化して上位カテゴリーを得る作業は、分析者の手作業によって行われた。しかし、下位カテゴリーに含まれる単語の出現頻度などから特徴ベクトルを作成し、因子分析やクラスター分析を行えば、コーパスにおける語彙の出現パターンを反映した上で上位カテゴリーを得ることが可能であると考えられる。今後はこうした方法を用いてカテゴリーのグルーピングを行うことで、より客観的な手順で概念間の関係性や概念構造を分析することができると考えられる。

そして、本研究では一般名詞が分析の対象となったが、固有名詞も分析することで、物語の作り手が参照している作家や作品群を抽出することが可能であると考えられる。

## 謝辞

本研究は科研費「修辞・物語構造のデータベース化による人文テキストの計量的分析基盤の構築」(26730168)の助成を受けた。

## 参考文献

- 1) 内田伸子：子どものディスコースの発達—物語 産出の基礎過程—, 風間書房 (1996).
- 2) 小林夏子：物語産出における人称表現の役割, 教育心理学研究, Vol. 38, No. 4, pp.379-388, (1990).
- 3) Kaufman, S. B. and Kaufman, J. C. (Eds.): *The Psychology of Creative Writing*, Cambridge University Press (2009).
- 4) Hirschfeld, L. A. and Gelman, S. A. (Eds.): *Mapping the mind: Domain specificity in cognition and culture*. Cambridge University Press (1994).
- 5) 玉利祐樹, 竹村和久：言語プロトコルの潜在意味解析モデルによる消費者の選好分析, 心理学研究, Vol. 82, No.6, pp.497-504 (2012).
- 6) 川島大輔, 小山達也, 川野健治, 伊藤弘人：希死念慮者へのメッセージにみる、自殺予防に対する医師の説明モデル—テキストマイニングによる分析, パーソナリティ研究, Vol.17, No.2, pp.121-132 (2009).
- 7) 真船浩介, 鈴木綾子, 大塚泰正：大学生におけるストレスの特徴-認知的評定, 及び心理的ストレス反応との関連の検討, 学校メンタルヘルス, Vol.9, pp.57-63 (2006).
- 8) 村井源, 川島隆徳：総合芸術の批評における批評対象の特徴：映画と演劇の批評の計量的比較, 情報知識学会誌, Vol.22, No.3, pp.203-222 (2012).
- 9) 工藤彰, 村井源, 往住彰文：計量分析による村上春樹長篇の関係性と歴史の変遷, 情報知識学会誌, Vol.21, No.1, pp.18-36 (2011).
- 10) Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees, *Proc. Proceedings of international conference on New Methods in Language Processing*, pp.44-49 (1994).