

知識グラフ構築に向けた物語文の構造分析

佐藤 浩輔¹ 頼 展 韜¹

¹ 株式会社バンダイナムコ研究所
{k18-sato,z-lai}@bandainamco-mirai.jp

概要

ゲームやアニメ、映画、漫画などといった近年大規模化し領域横断的に展開するコンテンツ産業において、物語およびその設定に関わる知識をどのように整備・運用するかが課題になっている。物語においては世界内の状態が動的に変化するため、通常の知識グラフでは十分に扱いきれないという問題がある。そこで本研究では、知識グラフの構築の下準備として、物語文（グリム童話）を対象に物語を構成する述語の観点から予備的な分析を行い、課題を抽出した。自然言語処理タスクとしての物語文からの知識グラフ構築について議論する。

1 はじめに

近年、ゲームやアニメ、映画、漫画といった物語に関わるコンテンツの大規模化が進んでいる。

単一の作品が長期継続することで規模が大きくなるケースがある一方で、シェアード・ユニバースとしてシリーズとして共通ないしは関連する世界設定のもとに複数の作品を展開するような場合もある。

並行して、知的財産としてのコンテンツを展開する IP (Intellectual Property) ビジネスが成長しており、グッズ展開、アニメ化やゲーム化といった従来の展開にとどまらず、コラボカフェなどといった幅広い業態・領域でコンテンツを活用し、経済圏を日夜拡大している [1]。

このようなコンテンツの大規模化・越境化にともなう、それらコンテンツに関わる知識をどのように管理するかが課題になっている。

たとえばコンテンツの世界観やキャラクター設定、あるいは既存のエピソードといかに整合的に新規コンテンツを作成・展開するかがしばしば問題になる。具体的には、キャラクターを他メディアへ展開する場合、原作の世界観やキャラクターなどの設定を踏襲することが求められ、キャラクターの性格や口調などの設定が矛盾する場合に作品イメージの

毀損であったり低評価につながることもある。

このような事態を避けるため、設定資料をまとめたり、企画・制作の段階で人手による監修を行うのが一般的であるが、一方でコンテンツが大規模化すればするほど前提・参照すべき知識が多くなり、コストが飛躍的に増大することになる。

大規模な知識を管理する手法のひとつとして、知識グラフ (knowledge graph) が着目されている。

知識グラフとは、様々な知識を体系的に連結し、グラフ構造によって表現したもの¹⁾で、典型的には RDF[3] のような主語・述語・目的語の三つ組みとして記述されたデータから、人や事物をノード、述語 (関係) をエッジとして構成される。

知識グラフを用いることで特定のエンティティ (人や事物) に関連する情報を取得したり、エンティティ間の関係性を調べることができるため、コンテンツのストーリーや世界観、設定資料などを知識グラフとして構築することで、コンテンツに関わる知識を効率的に管理できる可能性がある。

しかし、知識グラフのこの領域に特化した文献はこれまでなく、どのような固有の課題があるか必ずしも明確化されていないかった。

そこで本研究では物語文を対象に構造を分析し、物語に関わるコンテンツにおける知識グラフ構築に向けてどのような課題があるか明らかにした。

物語に関わるコンテンツは、コンテンツ本体に限らず脚本や設定資料など自然言語で記述された膨大な様々な文書群を背景に持つ。これら文書からの情報抽出および知識グラフ構築という観点から自然言語処理タスクを整理し、取り組むべき課題を明らかにした点が本研究の貢献となる。

1.1 物語と知識グラフ

広義の物語 (narrative) とは、語り手から受け手への事象 (event) の報告を指す [4]。物語研究の古

1) 「知識をグラフ構造で表現する」という点では共通しているものの、様々な定義が乱立している [2]。

典型的な分類では、物語は物語内容 (story) と物語言説 (discourse) からなる。物語内容が内容、すなわち「何を」語るかを指すのに対して、物語言説は物語の表現、すなわち「いかに」語るかを指す。ロシアフォルマリズムでは前者をファーブラ (fabula)、後者をシュジェート (sjuzět) と呼び [5]、英語ではそれぞれストーリーとプロット (plot) の別に対応する。ストーリーが物語世界内の事象を起きた順序で並べたものであるのに対しそれをどのような順序・方法で語るかがプロットである。物語内容がひとつの物語につきただひとつであるのに対し物語言説は多数存在できるため、知識グラフを構築する場合には物語内容に基づいて情報を抽出することになる。

物語を知識グラフ化する際の課題は、第一に物語世界においては現実世界との連続性がない可能性がある点がある。現実世界における諸法則が物語内世界でも成立するとは限らない上、現実世界には存在しない概念や事象が存在したり、現実には存在する語であっても違う意味付けで用いられることがある。そのような場合既存のオントロジー（概念体系）や汎用的な知識グラフをそのまま使うことができず、設定資料などをもとに作品固有の知識グラフを構築する必要がある。よって、与えられた物語の本文や周辺資料から情報を抽出し、知識グラフを構築できるようなシステムが必要とされる。

第二に、物語が動的な性質を持つ点である。物語がイベントの系列である以上、状態の変化が伴う。それゆえ参照する時点によって真理値が異なる命題が存在する。加えて物語においては「誤解」や「欺き」などといった登場人物の知識や認識が重要なケースがしばしばある。このような要素は古典論理の枠組みを超えており、適切に扱うためには例えば時相論理、認識論理といった様相論理 (modal logic) などを取り入れる必要がある。

物語の知識グラフ化に向け、本研究では既存の物語集を対象に予備的な分析を行った。

1.2 関連研究

1.2.1 物語のモデル化

物語の構造分析の先駆的な業績である Propp[6] はロシア民話における「魔法昔話」と呼ばれるカテゴリの物語について、登場人物の果たす機能の観点から分析し、その話の筋が本質的には一つであると主張した。Lévi-Strauss[7]、Barthes[8] など構造主義者の仕事を経て、Todorov[9] によって研究分野として

の物語論 (narratology) が提唱された。

Propp ら初期の仕事が物語内のできごと、すなわち物語内容に関わるものであったのに対し、後に成立した物語論では文学的表現としての物語言説の分析に重きを置いており、より抽象的かつ精緻な議論が積み重ねられていた一方で理論をプログラムとして実装することが困難になっている [10]。そのため、物語からの知識グラフ構築という本研究の目的に適うものではない。

本研究では、物語内容にのみ着目し、より基本的な単位である物語内のイベント、さらにはイベントを構成する命題を単位として分析する。

1.2.2 知識グラフ

静的でない知識を扱うために、イベントを扱うための知識グラフが提案されている (e.g., [11])。ただし、これらは基本的には汎用的な知識を想定しており、必ずしも物語に使えるとは限らない。たとえば [12] は Narrative Graph を提案しているがニュースのストーリーライン生成の文脈において現実世界の多数のニュースを対象にしたものであり、フィクションを対象にしたものではない。

また、ビデオゲームの文脈で、知識グラフと OpenAI GPT-2[13] のような大規模言語モデルと組み合わせることで、ゲーム内クエストを生成しようという試みがある [14]。ただしここで想定している知識グラフはあくまでゲームを通して変わらない静的な要素からなるものであり、進行状態によって変わる動的な要素を想定していない。

本研究では、物語という固有ドメインにおける知識グラフを構築するための方法を見出すことを目的とする。その際、物語内において世界が動的に変化するという前提のもとに物語を構成する命題に含まれる述語の分析を行う。

2 方法

2.1 データ

著名な物語集のひとつであるグリム童話 [15] を対象に分析を行った。この物語集は創作ではなく収集された民話を編纂したもので、版によって収録されている物語に異同がある。本研究ではグリム童話に含まれる物語のうち、日本語版 Wikipedia²⁾ 上に項目とあらすじの存在するものを使用した。2023 年 12 月の段階で、あらすじの存在する物語は 88 件で

2) <https://ja.wikipedia.org/>

あった。KHM38「奥様きつねの結婚」のようにひとつのタイトルに複数の物語が含まれるものは別個の物語として扱い、合計 90 件の物語を分析対象とした。利用した物語の一覧は付録の表 A.1 に示す。

2.2 手続き

2.2.1 イベントの抽出

著者が手動でコーディングし、あらすじの中から主要なイベントを抽出した。イベントには主要な名詞と動詞が少なくともひとつずつ含まれるようにした。さらにイベントから命題を抽出した。命題は主語、述語、目的語の三つ組みを基本として表現した。

簡便のため述語は基本的に日本語の動詞を用いたが、「仕立て屋には 3 人の息子がいた」などのような関係や所有の命題化には“has”、「竜は悪魔であった」のような性質や状態・属性を表す場合は“is”を述語として用いた。

データとして用いたあらすじには記述の粒度にばらつきがあった。描写の多寡の影響を減らすため、物語の進行上省略可能な命題については省略した。

例として KHM26「赤ずきん」からイベントと命題を抽出した場合の例を表 1 に示した。ここから状態の変化を図示したものを付録の図 A.1 に示す。

上記手続きにより抽出された命題の数は全部で 1,712 件であった。命題の個数の平均は 19.3、最大値は 47、最小値は 4 であった。

2.2.2 述語の集計

前処理を行ったうえで命題に含まれる述語を集計した。前処理および集計は Python(3.11.3)で行った。助ける/助けられるのような動詞の表層形の違いを吸収するため、MeCab (0.996) を用いて形態素解析を行い、主要な動詞を見出し語化したものを集計した。形態素解析の際、「連れていく」(連れて-行く)のような複数の語に分解されてしまう語はまとめて 1 語の述語として扱った。分析用に述語の表現の違い(意志、受動、否定)の情報を付した。

「行く」「来る」「着く」のような似た意味を持つ述語の違いを吸収するため、似た意味を持つ述語群を手動でカテゴリとしてまとめ、別途集計した。

3 結果

集計結果のうち、出現頻度が上位の術語を表 2 に示した。「行く」がもっとも頻度が多く、次いで関係や所有を表す“has”が多く、状態を表す“is”、変化

を示す「なる」も多かった。「結婚する」「食べる」「死ぬ」「生む」「暮らす」といった生活やライフイベントに関わる単語が多かった。また、(条件を)「提示する」、「要求する」などといった取引や約束にかかわる述語も多かった。

述語の表現の違いについて集計したところ、意志(「~しようとする」)が 52 件、受動が 55 件、否定が 52 件であった。各表現の出現のを物語ごとに集計したところ、90 件の物語のうち、意志が 33 件、受動が 32 件、否定が 33 件であった。

述語のカテゴリ化を行った後の集計結果の上位 10 件を表 3 に示した。その結果、「行く」「帰る」などの移動にかかわる単語カテゴリがもっとも多かった。次いで「手に入れる」「奪う」などといった所有の移転、“has”、“is”などといった(静的な)状態の記述、「提案する」「約束する」といった取引/約束、「死ぬ」「生む」といった生死、(ある状態に)「なる」「変わる」といった変化、の順で多かった。

4 考察

本研究では、物語文から物語を構成する主要な命題を抽出し、分析を行った。

その結果、移動・所有の移転・取引/約束・生死・変化・婚姻といった、状態の変更をとともなう述語が物語を構成する命題の中に多く含まれていた。この結果は物語文において状態の動的な変更が省略可能でなく、重要な位置を占めることを示している。

また「~しようとする」といった意志の表現が、1/3 程度の物語に含まれていた。これは特に登場人物が何か行動をしようとしたときに他者が妨害するといった物語上重要な場面で出てきやすく、物語において特別な役割を果たしていることを示唆している。このような表現を状態の動的な変化を考慮せずに記述することは難しいと考えられる。

本研究の結果は、イベント系列である物語の知識グラフ化を行う際に動的な側面を考慮することの必要性を例証するものである。

物語の知識グラフ化にあたって、情報抽出にあたっては少なくとも以下のタスクを解決する必要があると考えられる。

- ① **イベント系列の抽出:** 物語世界内のエンティティに関するできごとについての情報。
- ② **静的な情報の抽出:** 物語世界内のエンティティに関する、時間的に変化しない情報。
- ③ **状態の変化と時間的情報の抽出:** (1) 人や事物・

表1 KHM26「赤ずきん」におけるイベント系列

番号	内容	命題による表現
①	赤ずきんはお使いを頼まれておばあさんの家へと向かう	(赤ずきん, 頼まれる, お使い) (赤ずきん, 向かう, おばあさんの家)
②	赤ずきんはお使いの途中で狼に遭う	(赤ずきん, 出会う, 狼)
③	赤ずきんは狼に唆されて道草をする	(赤ずきん, 道草する)
④	狼は赤ずきんの先回りをしておばあさんの家へ行く	(狼, 行く, おばあさんの家)
⑤	狼は家にいたおばあさんを食べる	(狼, 食べる, おばあさん)
⑥	狼はおばあさんの姿に変装して赤ずきんが来るのを待つ	(狼, 変装する, おばあさん)
⑦	赤ずきんがおばあさんの家に到着する	(赤ずきん, 到着する, おばあさんの家)
⑧	狼に赤ずきんは食べられてしまう	(狼, 食べる, 赤ずきん)
⑨	満腹になった狼は寝入る	(狼, 眠る)
⑩	猟師が通りかかる	(猟師, 通りかかる)
⑪	猟師が狼の腹の中から二人を助け出す	(猟師, 助ける, 赤ずきんとおばあさん)

表2 上位出現頻度の述語

述語	n	述語	n	述語	n
行く	52	変わる	22	受け入れる	15
has	48	提示する	22	変える	14
結婚する	35	やって来る	21	与える	13
is	34	死ぬ	20	出会う	13
帰る	31	見付ける	19	逃げる	12
食べる	29	暮らす	19	尋ねる	12
なる	27	要求する	17	言う	12
殺す	27	渡す	16	辿り着く	12
助ける	24	入る	16	命令する	12
手に入れる	22	生む	15	去る	12

表3 上位出現頻度の述語カテゴリ

述語/カテゴリ	述語の例	n
移動	行く, 帰る, 渡る	270
所有の移転	与える, 手に入れる, 奪う	100
状態	has, is, 持つ, 住む	100
取引/約束	提案する, 拒む, 約束する	92
変化	なる, 変わる, 変身する	78
生死	生む, 死ぬ	67
婚姻	結婚する, 求婚する	54
飲食	食べる, 飲む, 食事を与える	46
遭遇	出会う, 会う, 現れる	37
欺瞞	嘘をつく, 隠す, だます	28

所有権の移動、(2)物理的な状態の変化(生死・肉体・変身など)、(3)社会的な状態の変化(名称・称号・婚姻など)といった状態の変化・継続についての情報。

①と②についてはこれまで知識グラフ構築研究において扱われてきたものであるが、③についてはこれまで必ずしも十分に取り扱われてこなかった。自然文においては状態の開始・終了については不明瞭であったり暗に示されることも多く、談話レベルでイベント系列を抽出し、様々な時点におけるエンティティの状態を推定することは自然言語処理のタスクとしても容易なものではないと考えられる。

物語の知識グラフの設計・構築に向けては、加えて以下の事項を考慮する必要がある。

情報の粒度: 物語世界内のイベント・状態を、記述的・時間的にどの程度詳細に扱うか。

信念および知識: 登場人物の信念や知識をどう(煩雑にはならない形で)扱うか。

線形でない時系列: 物語がタイムループや分岐といった構造を持つ場合にどう扱うか。

物語言説の要素: 物語が時系列順に提示されていない場合や、受け手の知識・信念をどう扱うか。

これら課題をクリアすることで、「物語の知識グラフ」の構築につながると考えられる。

本研究の限界として、扱った物語の代表性や、物語における主要なイベントの抽出・命題化など本研究の分析が質的かつ著者の主観的判断によるところが大きいがあげられる。今後様々な種類の物語に対して妥当かつ客観的に評価できる方法を検討していく必要がある。

物語の情報抽出にむけた自然言語処理分野におけるさらなる研究を期待する。

参考文献

- [1] PwC コンサルティング合同会社. コンテンツ ip を中心とした我が国のコンテンツ産業の競争力強化に向けた提言：諸外国のコンテンツ IP を中心とした競争力強化に関する調査事業 成果報告書. Technical report, 経済産業省, 2022.
- [2] Lisa Ehrlinger and Wolfram Wöb. Towards a definition of knowledge graphs. **International Conference on Semantic Systems**, 2016.
- [3] W3C. Rdf 1.1 concepts and abstract syntax. <https://www.w3.org/TR/rdf11-concepts/>, 2014. Accessed: 2024/01/11.
- [4] Gerald Prince. **A Dictionary of Narratology**. the University of Nebraska Press, 1987.
- [5] Boris Tomashevsky. Thematics. **Russian formalist criticism: Four essays**, pp. 61–95, 1925/1965.
- [6] Vladimir Iakovlevich Propp. **Morphology of the folktale**. University of Texas Press, 1928/1968.
- [7] Claude Lévi-Strauss. The structural study of myth. **The journal of American folklore**, Vol. 68, No. 270, pp. 428–444, 1955.
- [8] Roland Barthes. An introduction to the structural analysis of narrative. **New Literary History**, Vol. 6, No. 2, pp. 237–272, 1966.
- [9] T. Todorov. **Grammaire du Décaméron**. Approaches to semiotics. Mouton, 1969.
- [10] Pablo Gervás, Birte Lönneker-Rodman, Jan Christoph Meister, and Federico Peinado. Narrative models : Narratology meets artificial intelligence. In **International Conference on Language Resources and Evaluation. Satellite Workshop: Toward Computational Models of Literary Analysis**, pp. 44–51, 2006.
- [11] Willem Robert van Hage and Davide Ceolin. The simple event model. In **Situation awareness with systems of systems**, pp. 149–169. Springer, 2013.
- [12] Zhihua Yan and Xijin Tang. Narrative graph: Telling evolving stories based on event-centric temporal knowledge graph. **Journal of Systems Science and Systems Engineering**, Vol. 32, No. 2, pp. 206–221, April 2023.
- [13] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. **OpenAI blog**, Vol. 1, No. 8, p. 9, 2019.
- [14] Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. Personalized quest and dialogue generation in Role-Playing games: A knowledge graph-and language model-based approach. In **Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems**, No. Article 290 in CHI '23, pp. 1–20, New York, NY, USA, April 2023. Association for Computing Machinery.
- [15] Jacob Grimm and Wilhelm Grimm. **Kinder- und Hausmärchen**. 1812-1858.

A Appendix

表 A.1 分析に使用した物語の一覧

KHM	タイトル	KHM	タイトル	KHM	タイトル	KHM	タイトル
1	かえるの王さま、あるいは鉄のハインリヒ	37	おやゆびこぞう	125	悪魔とのおばあさん	196	リンクランクじいさん
2	猫とねずみとお友だち	38	奥様きつねの結婚*	126	誠実なフェレナントと不誠実なフェレナント	197	水晶玉
3	マリアの子ども	40	強盗のおむこさん	127	鉄のストーブ	198	マレーン姫
4	こわがることをおぼえるために旅にでかけた男	41	コルベスさま	129	ブラーケルの小娘	200	金の鍵
5	狼と七匹の子山羊	43	トゥルーデおばさん	139	旅にでる	6a	夜うぐいすとめくらとかげの話
6	忠臣ヨハネス	44	死神の名付け親	143	親不孝なむすこ	8a	ほうちょうをもった手
9	十二人兄弟	46	フィッチャーの鳥	145	神様の動物と悪魔の動物	22a	子どもたちが屠殺ごっこをした話*
10	ならずもの	47	ねずの木の話	148	梁の木	27a	死神とがちょうの番人
11	兄と妹	50	いばら姫	151	三人の怠け者	33a	長靴をはいた猫
12	ラプンツェル	52	つぐみの髭の王さま	153	星の銀貨	43a	奇妙なおよばれ
14	糸くり三人女	55	ルンベルシュティルツ	154	くすねた銅貨	54a	馬鹿のハンス
15	ヘンゼルとグレーテル	56	恋人ローランド	156	ぬらぬらの亜麻のかたまり	62a	青ひげ
18	わらと炭とそら豆	57	黄金の鳥	158	のらくら国のお話	68a	夏の庭と冬の庭
21	灰かぶり (シンデレラ)	64	黄金のがちょう	159	ディトマルツェンのほらばなし	73a	人殺し城
22	なぞなぞ	65	千匹皮	172	雪白と薔薇紅	81a	鍛冶屋と悪魔
23	はつかねずみと小鳥と腸づめの話	69	ヨリンデとヨリンゲル	181	池にすむ水の妖精	107a	からす
24	ホレのおばさん	73	狼と狐	182	小人の使い	143a	飢死にしそうな子どもたち
25	七羽のからす	76	なでしこ	184	釘	152a	哀しみの聖女
26	赤ずきん	89	がちょう番の女	190	机の上のパンくず		
27	プレーメンの音楽隊	103	おいしいお粥	195	土まんじゅう		
28	歌う骨	104	賢い人々				
31	手なしむすめ	108	ハンスぼっちゃん				
36	おぜんやご飯のしたくと金貨を生む驃馬と棍棒袋から出る	117	わがままな子供				
		122	キャベツろば				
		123	森のなかのばあさん				

* 2 件以上の話を含む物語

Note: 日本語版 Wikipedia 上の「グリム童話の一覧」ページに記載されているタイトル表記を利用した。

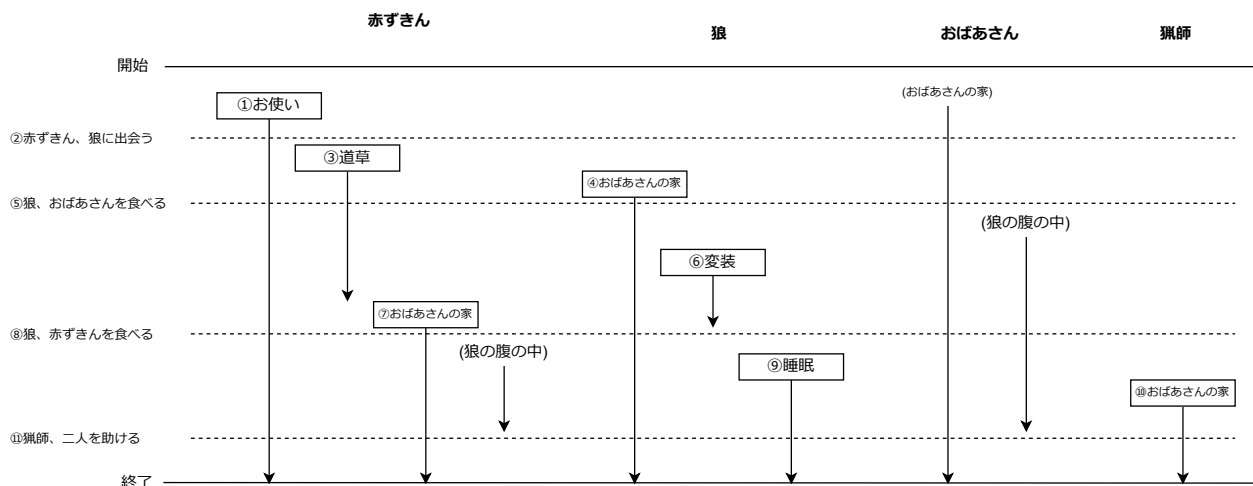


図 A.1 KHM26 「赤ずきん」におけるイベントと登場人物の動的な状態の変化

イベント番号は本文表 1 に対応。状態をともしないイベントを破線、状態の継続を矢印で表す。物語上必ずしも明示的に記述されていない状態を括弧で示す。