

物語の構造抽出のための主体語彙辞書の作成

村井源 (東京工業大学 大学院社会理工学研究科)

物語のテキストから物語の構造を自動的に抽出する目的を達成するために必要となる手法のロードマップを作成した。そしてその第一段階である登場人物による動作表現の特定を実現するために必要となる主体語彙の辞書をその階層的構造と合わせて作成した。また主体語彙の辞書を用いて登場人物の動作に相当する命題のリストを物語のテキストから抽出した。今後の課題として、得られた候補の命題リストに対する重要度の評価や、主体語彙の含まれていない命題であっても動作表現などから登場人物の言動として判定して抽出することなどを検討する必要がある。

Creating a subject vocabulary dictionary for story structure extraction

Hajime Murai

(Graduate School of Decision Science and Technology, Tokyo Institute of Technology)

In order to achieve the automatic extraction of narrative structure from story text, a roadmap of required techniques was made. As a first step, a dictionary of subjective vocabulary was made combined with its hierarchical structure for enabling identification of behavior representation of story characters. Moreover, a list of propositions corresponding to the operation of the story character was extracted from the story text by utilizing the dictionary of the subjective vocabulary. Future works are those: to evaluate importance of obtained candidate propositions' list; to examine verbs and to extract propositions which don't include subjective vocabulary.

1. まえがき

物語に関する研究はナラトロジーなどと呼ばれ、登場人物の機能や役割[1]、物語の意味的な構造[2]、語りの順序や文体など[3]の様々な側面から物語とは何かを理解するための研究が進められてきている。一方で近年物語に関する研究においてもコンピュータを用いたものが増えてきているがそれらの多くは物語の自動生成に関する研究か、生成のために少数の例を分析した研究が主流である[4]。しかし、最終的な目標が物語の自動生成であるとしても、そのためにはコンピュータが物語の性質を適切に扱えることが必要となる。すなわち、物語を理解し解釈するメカニズムをコンピュータを用いて計量的・科学的に明らかにすることがコンピュータを用いた物語研究の基礎として重要であると考えられる。

筆者は物語を人間が読んだ場合の面白さの一つであるオチの構造に焦点を当て、物語の構造を解析したデータをコンピュータに分析させることでいくつかの種類の逆転形式のオチの自動的な抽出が実現できることを明らかにしてきた[5, 6, 7, 8]。しかし、現段階では分析の元となる物語の構造表現の作成は全て人手によるものであり、コンピュータによる完全な自動的解釈の実現に

は至っていない。

もし物語構造の自動抽出が行えるようになれば、オチの抽出以外にも様々な物語のパターンの分析を自動的に行うことが可能となり、しかも人手を介さずに行えるようになるため大規模で再現性のある物語研究が実現可能となる。

そのため本研究では、コンピュータによる自動的な物語構造の分析を実現するために、テキスト本文から物語構造自動的に抽出する方法について検討し、その実現のために必要であると考えられる主体語彙の構造を策定して基礎的な辞書を作成し、それを用いた物語構造抽出のケーススタディを行う。

本研究では検討用のケーススタディとして、物語構造の技巧に定評のある星新一のショートショート作品を対象データとし、その物語構造分析を行う前提でのデータの作成を行っている。

2. 物語構造の検討と自動化に向けて

物語の構造をコンピュータやデータベース等でも用いられる形式として記述する場合、物語の要素であるプロットをどのように記述するかが問題となる。古典的なプロット分析ではプロットを物語の各部分の機能の配列として記述することが一般的である[1]。ただし、特定のテキスト箇所における物語機能は一つに決定できるとは

限らず、複数の登場人物の言動から構築される物語機能が並行的に存在する場合がある。またテキストの小部分における機能がそれを含みより大きな部分での機能の一部となっているなどの複雑な階層的構造がありうることも考慮に入れておく必要がある[5]。

そこで、各場面の一つの物語機能（記号）を割り当てるという古典的な方法ではなく、その場面ではどのような言動が登場人物たち（悪魔や天使、言葉話すコンピュータやロボット等も含む）によってなされているかを並行的に記述するというアプローチを採用する。場面の分割は、登場人物の変化、場所の移動、時間の推移、何らかのイベントの発生が起こった箇所を切れ目として分析者が判定して行っている。また物語の構造分析に用いるため、登場人物によって行われる言動の表記も言動そのものだけではなく、言動の対象、その動機、結果がどのようになったかという情報を付加している[6, 7]。図1に星新一作品『契約者』の物語構造を上記の方針で分析した場合の例を示す。

場面 人物	1	2	3	4	5	6
悪魔	放棄 魔王 仕事 怠惰 失敗	契約 男 命 仕事 成功		契約 男 命 仕事 成功	放棄 魔王 仕事 失敗	侮蔑 男 復讐 成功
男			破棄 悪魔 命 安全 失敗	契約 悪魔 命 安全 成功		
魔王	依頼 悪魔 仕事 開始					

図1 星新一作品『契約者』の物語構造表現例

図1では、横軸に場面、縦軸に登場人物を取り、各マスに上から言動、対象、動機、結果を記述している。このような表記を用いた場合、図中の網掛け部分のような対応する言動を特定することが可能であり、星新一のショートショート作品に特徴的な逆転形式のオチの機械的な抽出も可能である[8]。

このように、物語構造を特定後の分析において他の場面での登場人物の言動との関連性を分析したり比較を行ったりするためには、言動を一定

の形式に整理分類してデータベースへ格納することも必要である[6, 7]。

逆転形式のオチを分類して特定するため、物語構造中に現れた言動を5つの大領域(対象, 自身, 状況, 意図, 評価)とその下位の15の小領域に分けた。さらに動作主の主体性の強い動作(一人称)、対象となる相手を伴う動作(二人称)、動作主が必ずしも明確ではない動作(三人称)の三種類に分割し、肯定と否定の極性を付与した。表1に場人物の言動分類表と例を示す。

表1 登場人物の言動分類表と例

大領域	小領域	極性	一人称の例	二人称の例	三人称の例	
対象	情報	肯定	調査	説明	放送	
		否定	忘却	隠蔽	検閲	
	事物	肯定	作成	取得	発生	
		否定	破壊	贈与	崩壊	
	仕事遊び	肯定	労働	就職	勤務	
		否定	怠惰	解雇	無職	
自信	存在	肯定	誕生	出産	人口増加	
		否定	死亡	殺害	人口減少	
	身体	肯定	成長	育成	強化	
		否定	衰退	危害	弱化	
	正体	肯定	正体	類似	発覚	
		否定	変身	交代	変化	
	認知	肯定	覚醒	気付け	正気	
		否定	幻覚	洗脳	狂気	
	状況	移動拘束	肯定	移動	送迎	輸送
			否定	静止	拘束	施錠
		環境関係	肯定	自立	友好	平和
			否定	孤独	敵対	動乱
境遇		肯定	満足	共栄	繁栄	
		否定	不満	排他	没落	
意図	秩序約束	肯定	従順	契約	制定	
		否定	反抗	不履行	廃止	
	指示支配	肯定	依頼	許可	提案	
		否定	制止	禁止	警告	
	目標願望	肯定	努力	助力	祈り	
		否定	放棄	妨害	呪い	
評価	評価信頼	肯定	自信	称賛	好評	
		否定	後悔	非難	不評	
	感情	肯定	歓喜	好意	祝い	
		否定	悲嘆	憎悪	哀悼	

図1のような登場人物の言動を抽出して作られた物語構造のデータと、表1のような言動の分類表を用いれば、機械的な物語の構造分析が可能となる。ただし、現段階では物語構造のデータは

人手で作られたものであり、完全に自動化された物語分析を実現できているわけではなく、また、物語構造の特定における分析者の主観の影響を排除しきれない点で問題が残っている。

分析者の主観を可能な限り排除した客観的な物語分析を実現するためには、テキスト本文から物語構造を自動的に抽出するメカニズムを実現することが望ましい。

テキスト本文から物語構造を抽出するためには、テキスト本文からの登場人物の言動の抽出と物語における場面の分割を自動的に行う必要がある。また、登場人物の言動のある程度は物語の表現を彩るための文体的な装飾であり、必ずしも物語上の機能を有しているとは言えないことも合わせ考えると、抽出された登場人物の言動から物語構造上重要な言動を選別する必要もあろう。それらの処理が自動で可能になれば物語構造の抽出から特徴の分析まで一貫して機械的に行うことが可能である。

ただし、まず登場人物の言動を抽出するというタスク自体がそれほど容易ではない。特に日本語の物語文においては主語の省略が広く行われており、機械的に形態素解析や構文解析を行っても主体と言動の関係性をつかむことは難しい。また、文学的な文章においては比喩や婉曲表現などの間接的な表現を用いる修辞法も多用される。

その点も考慮に入れた上での物語の自動分析に向けてのロードマップを図2に示した。

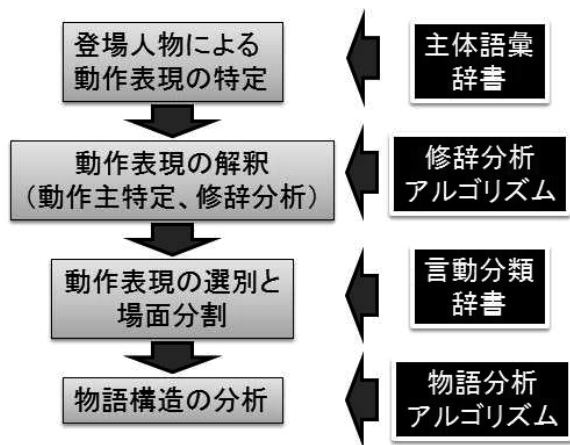


図2 物語の機械的分析実現に向けての方針

本稿では主体語彙辞書の整備とそれを用いた動作表現の抽出の方針について述べる。

3. 主体語彙と階層構造

コンピュータによって自動的に登場人物の言動をデータとして抽出し整理分類を行うためには、物語のテキストの本文から登場人物とその言動にあたるテキスト箇所を特定して抽出する必要がある。

まず、物語において登場人物となりうるのは、若干の例外はあるが、一般的には主体的に意志を持って行動する何らかの存在である。多くの場合には人間であるが、ジャンルによっては宇宙人や宇宙生物、悪魔や幽霊、ロボットや自律的に動くコンピュータなども、物語の主体として登場する。

また人間を指し示す場合にも固有名詞だけではなく代名詞(例:私、僕)、家族内での関係(例:父、母)、職業的關係(例:社長、社員)、一般的人間関係(例:敵、味方)など様々な種類の関係性を示す語彙が物語中では主体を指し示す語彙としてとして用いられている。他に物語においては個人が主体となるのみではなく、特定の集団や組織、地域や国家、種族や民族などが主体的な存在として扱われる場合も少なくない。

このように種々の物語の主体的な行動をとる存在を指し示す語彙は多岐に亘るが、これらの語彙を物語中からピックアップしてリスト化し、また語彙の種類や関係性をデータとして構造化し、機械可読な辞書を構築することは可能である。構築された主体語彙のリストを用いることで、物語における主体とその言動を記述しているテキスト箇所の候補を抽出することが容易となると考えられる。可能性の高い候補を抽出できればその後の物語の分析をより正確に進めるための基礎を構築できると期待される。

また主体語彙の種類を分類し、様々な付加的な属性情報をデータ化することで、主体語彙の種類ごとの用いられ方の相違の分析や、代名詞とその他の主体語彙の關係の分析なども可能になると考えられる。特に物語文の中においては、特定の同一の主体に対する呼称が場面と誰からの視点によるかで変化する場合が多く(例:「エス氏」は社会的には「会社員」であり社内では「部下」であり、家庭では「父」であるというような場合、さらに代名詞では「彼」や「その男」となったりもする)、これらの複数の呼称の結びつきや關係性を分析する上ではこれらの種別が活用可能ではないかと考えている。

そこで物語中に用いられる主体を示す語彙や一般的な辞書等に掲載されているような物語中でも主体を示す語彙として用いられうる語彙を分類し、物語中での主体を示しうる語彙の階層構造とそこに分類される語彙のリストを作成した(表1)。

種別の大分類としては、主体自体がどのような存在であるかを一般名詞で表現する語彙を分類した「存在」、主体が他の主体とどのような關係性にあるかによって与えられる呼称を分類した「關係」、主体に固有的に与えられている呼称である「固有名詞」、そして一般的に省略的な表現

として用いられている「代名詞」の四種類とした。
 またこれらの主体語彙に合わせて、語頭や語尾に付随することで他の名詞に主体語彙的な機能を付加する主体語頭・主体語尾も分類し、物語のテキスト中で表れた場合にはその前後の単語と

合わせて主体語彙と同様に抽出することとした。
 主体語頭・主体語尾のリストは表2に示した。
 現在主体語彙に約3500種類、主体語頭・主体語尾に76種類の単語が分類されているが、今後必要に応じて適宜拡張していく予定である。

表1 主体語彙の分類と例

大分類	中分類	小分類	例
存在	人		人,人間,人物,小人,大男,宇宙人,異星人
	生物的		知的生物,宇宙生物,怪物,巨獣,生物,犬
	人工		ロボット,ロケット,円盤,自動機械,装置
	超自然		幽霊,悪魔,魔神,魔王,死神,鬼,仏
関係	生物的	血縁・性	兄,弟,姉,妹,息子,娘,親,親戚,夫婦
		成長	中年,婦人,長幼,老若,老少,初生児
		状態	病人,患者,経産婦,不美人,男前,佳人
	職業・資格的		医者,看護婦,弁護士,社員,社長,探検隊
	集団・社会・思想的		先輩,同僚,後輩,代表,責任者,所長,主人
	時空間		現代人,古代人,先人,未来人,東男,京女
固有名詞	人名		エヌ氏,エフ博士,佐藤,鈴木,太郎,花子
	種族名		地球人,インディアン,源氏,アジア人,マオリ族
代名詞	一人称		私,わたし,ぼく,僕,おれ,俺,不肖,吾輩,自分
	二人称		あなた,お前,君,君たち,あなた方,あんた
	三人称		彼,彼ら,彼女,彼女たち,あいつ,あいつら
	疑問		何物,誰か,だれ,だれだれ,どいつ,どなた

表2 主体語頭・主体語尾の分類と例

位置	大分類	小分類	例
主体語頭	呼称	性別	ミスター,ミス,ミセス,ムッシュー
主体語尾	集団		屋,たち,国,市,団,隊
	職種		係,官,士,工,王
		性別	氏,女,夫,婦人,婦
		その他	様,ちゃん,さん,くん
	敬称		師,先生,卿,公,御前,殿
死後		仏,信士,居士,大姉,信女	

4. 物語構造の抽出に向けて

主体の言動を物語の構造の要素とするため、物語の構造を分析するためには、主体とその動作を示す単語の関係をセットで物語の本文テキストから抽出する必要がある。このような動作の主と動作などからなる意味の最小単位を命題と呼ぶことが多い。登場人物の言動から成っている命題をテキストから抽出ができれば、その中から物語

中で中心的な機能を果たす箇所を選択することで基本的な物語構造を抽出できると考えられる。

一般的に、テキストの本文から命題を作成するためには、文を形態素に分解して、それらの係り受け関係を最初に解析する必要がある。ただし、係り受け関係中には、例えば名詞と名詞から成る複合名詞や、形容詞と名詞の組等から成る修飾的な状態を示す係り受けの関係も含まれる。そこで一般的な全命題を抽出するのではなく、登場人物の言動等に関連する可能性が高いと考えられる

動詞を含む係り受け関係を抽出する。また、動詞を含まないが登場人物のアイデンティティ情報に相当する体言止め等の表現(例:「私は悪魔」)も、物語構造上は重要な役割を果たす場合が多い(登場人物の隠された正体の暴露がオチとなる場合など)ため、命題として抽出する。

そこでテキストに対してKNP[9]を用いて形態素解析と係り受け解析を行い、結果として得られる係り受け関係から命題単位で主語と述語の組を抽出するプログラムを作成した。

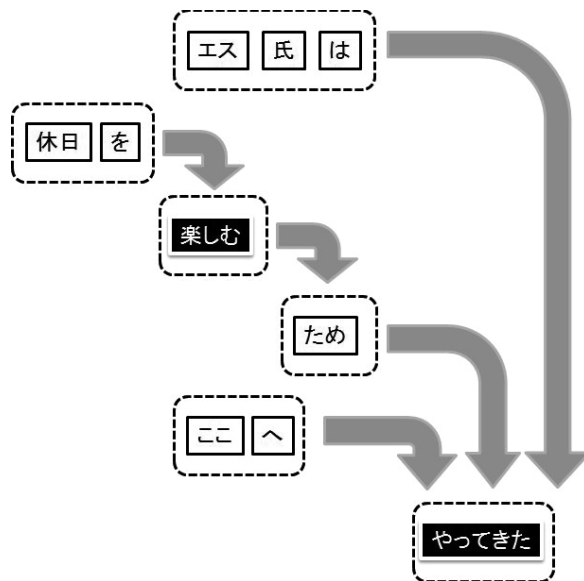


図3 形態素解析および係り受け解析結果の例

図3には「エス氏は休日を楽しむため、ここへやってきた」という例文をKNPを用いて形態素解析および係り受け解析した場合の結果を示す。図中では四角枠は形態素を、点線の丸枠は文節の単位を示している。また矢印は係り受けの関係を示している。このような分析結果の出力に対し、動詞箇所(「楽しむ」と「やってきた」)に係る文節をKNPの述語項構造解析の結果に基づいて分割抽出すると、「エス氏は 休日を楽しむため」と、「エス氏は ここへ やって来た」の二つの命題を取り出すことができる。

このようにして得られた命題から、主体語彙と主体語頭・主体語尾を含む物を選択することで、物語テキスト中に明示的に登場人物が行う言動が記されている箇所をピックアップすることが可能である。参考のため、Appendixに星新一のショートショート作品『悪魔』から抽出された主体語彙を含んだ命題のリストを示した。この作品の場合抽出された命題は44であった。得られたリストには物語の展開上重要と考えられる登場人物の言動も含まれているように見えるが(例:悪魔の出現, エス氏の悪魔への質問, エス氏の悪魔への金貨贈与の依頼, 悪魔のエス氏への金貨の

贈与, 悪魔の消失), 必ずしも重要とは言えない無関係な言動(たとえば「道具屋へ持っていても」など)や係り受け関係の解析に失敗した結果に見える物(たとえば「エス氏がない」など)も含まれている。また、逆に物語の展開上重要と考えられる金貨の消失は抽出された命題中に含まれていない。

命題の抽出結果より考察できることとしては、第一に登場人物が行う言動の全てが物語構造の分析上重要であるとは限らず、有用な物語構造を得るためには全ての命題に対して重要度の判定やその結果に基づく取舍選択を行う必要があるという点である。

また、日本語の物語テキスト中では動作言動の主体を省略する表現技法が多数用いられており、明示的に登場人物の言動である物以外の様々な言動も物語構造上重要な機能を果たす登場人物の言動候補として抽出する必要がある。このため、主体語彙が含まれていない命題においても直前に他の主体語彙を含む場合などには抽出対象に含めるような条件付けや、主体的語彙とセットで用いられることの多い主体的な動作を示す動詞や慣用句等をリストアップして辞書として整備し、主体的な動作を示す語彙が含まれている命題は主体的な語彙が含まれなくとも登場人物の言動の候補として抽出するなどの条件設定の検討が必要であると考えられる。

さらに、物語の構造分析上重要な対象物(例の『悪魔』の場合には金貨など)に関しては、主体語彙ではなくともそれに関連する命題を本文中から抽出することも考慮する必要がある。

また、物語中では主体を指し示す単語や表現が次々と変化していく状況が一般的にあり、同じ主体を指し示している複数の主体的語彙を含む表現の関係性を分析する必要もある。主体を示す表現の変化は冗長的表現の省略や主体に関する視点の変化に対応しているため、固有名詞表現と代名詞表現の対応関係を特定することや、すでに出ていた表現とは別の種類の関係性を示す主体語彙を特定することなどの同定手法を検討する必要があると考えられる。

5. 結論と今後の課題

本研究においては、物語のテキストから物語の構造を自動的に抽出する目的を達成するために必要となる手法のロードマップを作成した。そしてその第一段階である登場人物による動作表現の特定を実現するために必要となる主体語彙の辞書をその階層的構造と合わせて作成した。

また、KNPでの形態素・係り受け解析結果を用いて、係り受け関係に基づく命題を動詞の係り受け関係を中心に構成して主体語彙の含まれるものだけを抽出するプログラムを作成した。

プログラム実行の結果として得られた命題のリストには物語構造の分析上重要な命題が複数

含まれていたが、不要な命題が含まれていたり、重要と考えられる命題が抜けていたりするため、得られた命題のリストがそのまま物語構造の要素となるわけではない。

今後の課題として、得られた候補の命題リストに対する重要度の評価や取捨選択を行うことや、主体語彙の含まれていない命題であっても前後の関係や、用いられている動作表現などから登場人物の言動として判定して抽出することなどを検討する必要がある。また主体の表現が変化する場合の対象の同定を行う必要もある。

参考文献

- 1) Vladimir Propp: *Morphology of the Folk Tale*, University of Texas Press, (1968). 北岡誠司(訳), 福田美智代(訳): 昔話の形態学, 水声社, (1987).
- 2) Algirdas Julien Greimas: *Sémantique structurale : recherche de méthode*, Larousse, (1966). 田島宏(訳), 鳥居正文(訳), 構造意味論—方法の探求, 紀伊國屋書店, (1988).
- 3) Gerard Genette: *Discours du recit in Figures III, Seuil*, (1972). 花輪光(訳), 和泉涼一(訳), 物語のディスクール:方法論の試み, 水声社, (1985).
- 4) D. Fox Harrell, Dominic Kao, Chong-U Lim: *Computationally Modeling Narratives of Social Group Membership with the Chimeria System*, OpenAccess Series in Informatics, 2013 Workshop on Computational Models of Narrative, Vol. 32, pp. 123-128 (2013).
- 5) 村井 源: 物語プロットデータベースのためのデータ構造の検討, 情報知識学会誌, Vol. 23, No. 2, pp. 308-315, (2013).
- 6) 村井 源: 登場人物の言動に着目した物語プロットデータベースのためのデータ構造試案, 人文科学とコンピュータシンポジウム論文集, Vol. 2013, No. 4, pp. 255-260, (2013).
- 7) Hajime Murai: *Plot Analysis for Describing Punch Line Functions in Shinichi Hoshi's Microfiction*, OpenAccess Series in Informatics, Vol. 41, pp. 121-129, (2014).
- 8) Hajime Murai: *Towards Automatic Extraction of Punch-lines: Plot Analysis of Shinichi Hoshi's Flash fictions*, Proceedings of Japanese Association of Digital Humanities Annual Conference, pp. 17-18, 2015.
- 9) 黒橋・河原研究室, 日本語構文・格・照応解析システム KNP, <http://nlp.ist.i.kyoto-u.ac.jp/>? KNP, 2015/9/15 参照.

Appendix

星新一のショートショート作品『悪魔』から抽出された主体語彙を含んだ命題のリスト

湖は、国にあった。
エス氏は休日を楽しむため、
エス氏はここへやってきた。
道具屋へ持っていっても、
目を閉じ、相手が
みなれぬ相手が
そばに、相手が立っている。
色の黒い男で、
男で、しっぽがあった。
エス氏がふしぎそうに聞くと、
相手は、顔で答えた。
「わたしは悪魔」
悪魔も、絵にある
悪魔も、かっこうをしていたようだ。
「信じたくない人は、
人は、信じないでいればいい。
わたしは、だが、ここにいる」
目をエス氏は何度もこすり、
エス氏は気持ちを落ち着け、
エス氏はおそろおそろ質問した。
エス氏はしばらく考え、
エス氏はこう申し出た。
わたしに、お金をお与え下さいませんか」
悪魔は穴に、ちょっと手をつこんだかと
悪魔はつつこんだかと思うと、
悪魔は金貨をさし出した。
エス氏が首をかしげながら、
エス氏が手にとってみると、
エス氏が金貨にまちがい
エス氏がない。
「よくばりなやつだ」
エス氏は何回もねだり、
悪魔は金貨をたびに出してくれた。
悪魔はと言ったが、
エス氏は熱心にたのんだ。
悪魔はうなずき、また
悪魔は金貨をつかみ出し、
悪魔はそばに置いた。
エス氏はそうと気づいて、
エス氏は大急ぎで
エス氏は岸へとかけだした。
悪魔も、氷は割れ、
悪魔も、笑い声をあげている
悪魔も、みな底へと消えていった。