

# 物語テキストからのストーリー抽出について

## On Extraction of Story from Narrative Text

大石 顕祐<sup>\*1</sup>  
Kensuke Oishi

小方 孝<sup>\*2</sup>  
Takashi Ogata

<sup>\*1</sup> 岩手県立大学大学院  
Graduate School of Iwate Prefectural University

<sup>\*2</sup> 岩手県立大学  
Iwate Prefectural University

In this paper, we propose a method and system extracting sentences that compose a story (events occurred actually in a narrative) from narrative language text using grammatical knowledge such as modality and various rhetorical effects. The purpose of this research is related to the concept of “intertextuality” as a modern literary theory. Intertextuality is a thought that try to think a text in the diverse relations with other texts. We have a fundamental concept of narrative generation by de-composition and re-composition of existing narrative texts, the system proposed in this paper gives a tool for de-composition of text in intertextual mechanism.

### 1. まえがき

Kristeva[Kristeva 1970]のテキスト間相互関連性(intertextualité)(間テキスト性)の理論に基づけば、テキストの受容とはその当事者が連想的に他のテキストを記憶から引き出し、現在のテキストと関連付けながら行う経験を意味する。一方制作の側面においては、あるテキストの産出に当って当事者は様々なテキストの記憶や連想の網の中で制作行為を行う。

我々は、物語生成システムを間テキスト性の観点から、既存の物語やその部分を様々な形式の情報に変換する分解機構と、入力され蓄積された情報を加工し、システム内部に保存・蓄積する保存機構、蓄積された情報を基に新たな物語を再構成する再構成機構から成るものとして再構成する事を示した[大石 2007]。図 1 に間テキスト性を取り込んだ形での物語生成システムの枠組みを示す。

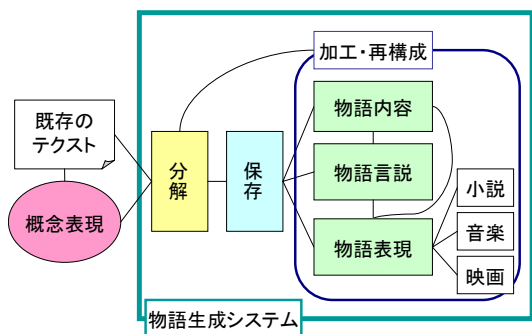


図 1 間テキスト的な物語生成システムの枠組み

これまでの研究では、中嶋ら[中嶋 2006]が保存・蓄積機構及び再構成機構の案を提示した。このシステムでは入力となる物語の文章を手で分解し、概念表現として記述したが、本研究の大きな目標は、この分解機構の自動化であり、ここではその最初のアプローチとして、物語文章からストーリーを構成する文のみを抽出するシステムの試作を示し、分解機構へ向けた一つの足掛りとする。

物語生成システムでは、図 1 に示すように物語の構成要素を物語内容(story)・物語言説(discourse)・物語表現に分類してい

る。物語内容は「物語は何を語っているか」を表す諸要素、すなわち登場人物や舞台などの設定・物語中で起こった出来事とその構成・主題等によって構成される。それに対して物語言説は視点・時間順序・距離・速度などの修辭的技法や叙法の技法によって物語内容を修飾したものであり、物語表現は言語・映像など直接的な媒体による表現である。

本研究では、抽出対象とするストーリーを「物語テキストから叙法や修辭的效果等を除去し、物語中に実際に起こった出来事のみを対象に、それを時間順に並べたもの」と定義する。

### 2. システムの方法

#### 2.1 全体像

ストーリー文抽出・構造化システムは、図 2 に示すように日本語で記述された物語テキストからストーリーを構成する文のみを抽出するストーリー抽出システムと、ストーリー抽出システムで得られたデータに基づき、行為(action)を基点に、行為者(agent)、被行為者(counter-agent)、行為対象(object)、行為が行われた場所(object)等を含む格フレームを抽出する事象概念自動抽出システムから構成されるものとする。

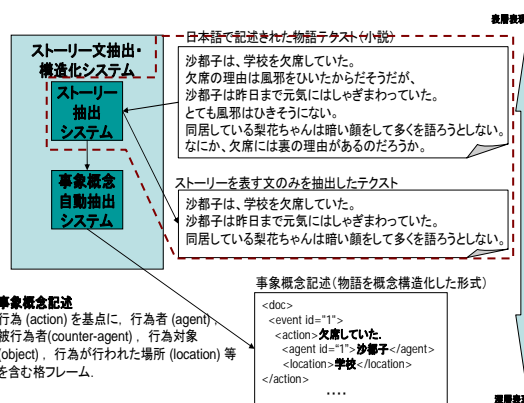


図 2 ストーリー文抽出・構造化システムの全体像

前節では本研究で抽出するストーリーを定義したが、今回は叙法や修辭的效果の除去と物語中に実際に起こった出来事の設定に焦点を絞り、時間順の並べ替えについては考慮しない。また、同義語の同定や道具格の特定のようなオントロジー知識が必要な部分、出来事の意味的分類のような文間関係の解析が

連絡先: 大石顕祐, 岩手県立大学大学院ソフトウェア情報学  
研究科, 岩手県岩手郡滝沢村滝沢字菓子 152-52,  
g231g010@edu. soft. iwate-pu. ac. jp

求められる部分については自動化の範囲外とする。時間順の並べ替えを除くストーリー抽出のプロセスのうち、ストーリーを自動抽出するシステム全体を、ストーリー文抽出・概念構造化システムと呼んでいる。

今回は、このうちストーリー抽出システムを試作する。ストーリー抽出システムの構成と入出力は次節以降で述べる。

## 2.2 ストーリーを構成する文の抽出

ここでは、ストーリーを構成する文の抽出方法を詳しく説明する。発話の階層のトップレベル、すなわち地の文の視点で客観的事実とされている行為について記述している文を、ストーリーを構成する文として主に文法知識を用いて自動抽出した。以下に、抽出の手順を説明する。

1. 行為の記述を同定する。行為の記述には、動詞・動詞句の述語、及び複合述語が含まれる。(詳細を 2.3 節で説明)
2. 行為実行の客観性を判断する。判断材料として、モダリティ(発話者の心的態度)の表現、及び発話の階層(地の文と台詞・注釈)を用いる。(詳細を 2.4 節で説明)
3. 客観的に実行された行為が述語になっている文をストーリーを構成する文として出力する。(詳細を 2.5 節で説明)

なお、実際に物語内容を自然言語で記述する場合、例えば任意の物語から粗筋を要約する場合には、先に述べた手順に加え重要文の抽出、換言すれば大幅な文の省略が行われると考えられるが、このプロセスは時間順の並べ替えと同様、物語内容の時間的順序に関わるものであることから、今回は時間順同様考慮しないこととした。

## 2.3 行為の記述同定

名詞文・形容詞文が事象の性質を表すのに対し、動詞文は事象の変化、すなわち行為を表す場合が多い。ここでは、動詞文の述語になる語句として、動詞及び動詞句(前項動詞・後項動詞から成る複合動詞、名詞に動詞スルが接続するスル動詞、動詞テ型に補助動詞が接続したもの、及び形容詞に動詞化する接辞が接続したものが挙げられる)を同定する。また単独の動詞と動詞に形容詞化する接辞が接続することで成立する形容詞句を区別するため、形容詞句についても併せて同定を行う。

なお、動詞・動詞句には様々な要素が付随し、複合述語を形成する場合がある。動詞文の場合、ボイス(行為の視点)、テンポラリティ(行為の時間的位置づけ)アスペクチュアリティ(行為の時間的展開の様態)、否定の表現、敬語・文体(待遇表現)、モダリティ(話し手の心的態度)の 6 つの機能・意味のカテゴリについて考慮する必要がある。以下に、各カテゴリの概要と、同定するについて説明する。

- ボイス: 動詞に表現されている出来事をどの視点に立って見るかを区別する形式。今回は視点の位置が動作の主体になる能動態(無標)、動作の客体になる受動態(～レル、～ラレル)に加えて、出来事を実現させようとする人物になる使役態(～サセル)と、その受動態である(～サセラレル)までを同定する。
- テンポラリティ: 文レベルの出来事の時間的位置付けを表し分ける、機能・意味のカテゴリで、「今日、翌日、水曜日」などの語彙的なものを含む。このうち、単語の語形変化によるスル・シタの対立関係をテンスという。動詞・動詞句との関係から、テンスのみが同定範囲となる。

- アスペクチュアリティ: 動的な出来事の時間的展開の様態を表し分ける機能・意味のカテゴリ。今回は中核となるアスペクト(スル・シタの対立関係)に加えて、「シテアル、シテイル、シテシマウ」などの準アスペクトまでを同定範囲とする。
- 否定の表現: 否定の表現には、文法的否定形式と語彙的否定形式がある。前者が動詞の活用として表されるのに対し、後者は動詞自身が否定的な意味を持つ。今回は文法的否定形式のうち、文の命題の側面に否定形式(～ナイ)のみを同定した。
- 敬語・文体: 敬語・文体をまとめて待遇表現と呼ぶ。待遇表現は、発話者が話者自身と聞き手の社会的勢力の大小関係、あるいは話者自身と話題中の人物との社会的勢力の大小関係を認識していることを、聞き手に対してアピールするために用いる言語表現である。待遇表現のうち、文法的なものを文体、語彙的な表現が加わるものを敬語と呼ぶが、今回は文体のみを同定範囲とする。
- モダリティについては行為実行の客観性判断に用いるため、次節で説明する。

## 2.4 行為実行の客観性判定

行為実行の客観性判断には、モダリティによる判定と発話の階層による判定を行う。図 3 にモダリティの分類と用例、及び本システムで同定する範囲について図示する。

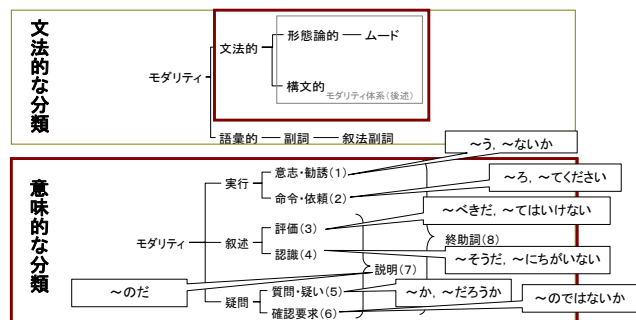


図 3 モダリティの表現の分類と用例

ここでいうモダリティとは、言語活動の基本単位としての文の述べ方についての話し手の態度を表し分ける、文レベルの機能・意味のカテゴリである[宮崎 2006]。従って、ここでいうモダリティには形態論的カテゴリであるムードの他、構文的な組み立て形式で表されるものや、叙法副詞による表現を含む。今回は、叙法副詞を除く各表現を同定し、客観性判断に用いた。モダリティには大きく分けて意思の表明や行為の要求など文の発話・伝達機能を表すものと、話し手の命題に対する把握の捉え方を表すものがある[守山 2000]。ここでは前者を実行のモダリティ、後者を認識のモダリティと呼ぶ。また、このいずれにも属さないものとして、聞き手に対する先行文脈などの説明を表す説明のモダリティが挙げられる。このうち、実行のモダリティの表現が付随する行為は、発話された時点では実行されていないと考えられる。また、確認のモダリティでは、表現によって行為が客観的に実行されたかどうかの確かさに差がある。ここでは認識のモダリティのうち、命題内容として書き出された事態の成立・存在を疑わないものとする確認のモダリティのみを抽出対象とする。

従ってモダリティによる判定では、実行のモダリティ、および確認以外の認識のモダリティの表現が含まれる行為は抽出対象から除外する。

なお、今回は発話の階層による判定では問題を単純化し、台詞や注釈などの括弧で括られた部分を抽出対象から除外し、地の文のみを抽出範囲とするに留める。

## 2.5 ストーリーを構成する文の出力

前処理として、複文(二つ以上の述語で構成された文)の分節化を行う。今回は接続助詞と読点で区切られた部分で分節化し、それぞれの文の述語は文末に書かれていると判断する。最後に、区切られた各文の述語から行為実行の客観性があると判断された文のみを抽出する。

## 3. ストーリー抽出システムの試作

前節では、ストーリー文抽出・概念構造化システムの構想と、ストーリー抽出システムにおけるストーリーの抽出手順について説明した。本節では、図4に示したストーリー抽出システムの構成に沿って、具体的な処理手順について説明する。

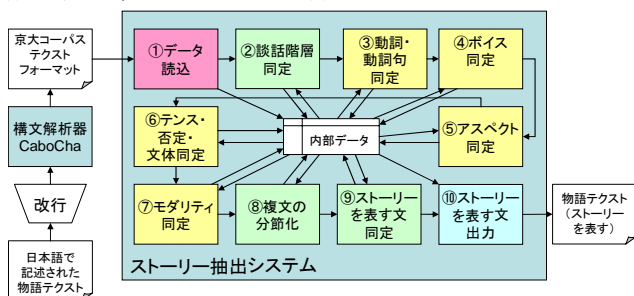


図4 ストーリー抽出システムの構成

前処理として入力する日本語で記述された物語テキストを文単位で改行、CaboChaで構文解析して、解析結果を京大テキストコーパスフォーマットにしたものを用意しておく。これは、文中での表現に対応する読み仮名、基本形、品詞、活用等が解析された形態素が分節ごとに分ち書きされている。係り受け構造は各文節の先頭に割り振られた番号と係り先の分節番号によって表現されている。

次に、ストーリー抽出システムの各モジュールについて説明する。①データ読み込みモジュールでは、入力データを内部データとして保存する。この時、用いるフォーマットでは、各形態素の構文解析の結果に、データ分類用の項目が付加される。以降の処理において、語句の種類、分節化後の文のid、ストーリーを構成する文かどうかの判定、談話の階層が書き加えられていく。②談話階層モジュールでは括弧記号の開閉から、談話の階層を同定し、内部データに書き加える。③～⑦の各モジュールでは、行為の記述における、動詞・動詞句と、動詞文の複合述語における6つの機能・意味のカテゴリーを同定し、それぞれの語句の種類を内部データに書き加える。このうち動詞・動詞句、ボイス、テンス、否定、文体、アスペクトについては、対象となる分類項目と一対一の関係にある。ただし、モダリティについては、説明のモダリティを他の有標のモダリティと区別する。⑧では、⑦までの処理において未判定の接続助詞から、複文を分節化して、分節化後の文のidを内部データに書き加える。⑨で、ここまでの処理の結果として書き加えられたデータ分類から、行為実行の客観性を判断し、ストーリーを表す文かどうかの判定を内部データに書き加える。最後に⑩の出力モジュールから、ストーリーを構成する文と判定された文のみを出力する。

## 4. 入出力と考察

入力データと出力データを比較し、入力データと出力データで意味の異なる文、及び意味不明な文を集計することで試作システムの評価を行った。ただし、構文の分節位置・分節された文節の意味的關係については評価を見送った。テストデータとして、有島武郎の小説『生まれいづる悩み』[有島 2000]を使用した。入力データの例を図5、出力データの例を図6にそれぞれ示す。

その横町を一町も行かない所に一軒の薬種店があって、それにつづいて小さな調剤所がしつらえてあった。君はそのガラス窓から中をのぞいて見る。ずらっとならべた薬種びんの下の調剤卓の前に、もたれのない袂り抜きの事務椅子に腰かけて、黒い事務マントを羽織った悒鬱そうな小柄な若い男が、一心に小形の書物に読みふけている。それはKと言って、君が岩内の町に持っているただ一人の心の友だ。君はくすんだガラス板に指先を持って行ってほとほとたたたく。Kは機敏に書物から目をあげてこちらを振りかえる。そして驚いたように座を立てて来てガラス障子をあける。

「どこに」

君は黙ったまま懐中からスケッチ帳を取り出して見せる。そして二人は互いに理解するようにほほえみかわす。

「君はきょうは出られまい」

君は東京の遊学時代を記念するために、だいじにとっておいた書生の言葉を使えるのが、この友だちに会う時の一つの楽しみだった。

「だめだ。このごろは漁夫で岩内の人数が急にふえたせいかな忙しい。しかし今はまだ寒いだろう。手が自由に動かない」

「なに、絵はかけずとも山を見ていればそれでいいだ。久しく出て見ないから」

「僕は今これを読んでいたが(と言ってKはミケランジェロの書簡集を君の目の前にさし出して見せた)すばらしいもんだ。こうしてはいけなような気がするよ。だけどもとも及びもつかない。いいかげんな芸術家というものになって納まっているより、この薄暗い薬局で、黙りこくって一生を送るほうがやはり僕には似合っようなうだ」

そう言って君の友は、悒鬱な小柄な顔をひとときわ悒鬱にした。君は励ます言葉も慰める言葉も知らなかった。そして心がめぐるもののようにスケッチ帳をふところに納めてしまった。

図5 入力データの例

その横町を一町も行かない所に一軒の薬種店がある。それにつづく。小さな調剤所がしつらえてあった。君はそのガラス窓から中をのぞく。見る。ずらっとならべた薬種びんの下の調剤卓の前に、もたれのない袂り抜きの事務椅子に腰かける。黒い事務マントを羽織った悒鬱そうな小柄な若い男が、一心に小形の書物に読みふけている。それはKと言う。君はくすんだガラス板に指先を持って行く。ほとほとたたたく。Kは機敏に書物から目をあげる。こちらを振りかえる。そして驚いたように座を立てて来る。



ガラス障子をあける。  
君は黙ったまま懐中からスケッチ帳を取り出す。  
見せる。  
そして二人は互いに理解するようにほほえみかわす。  
そう言う。  
君の友は、悒鬱な小柄な顔をひときわ悒鬱にした。  
君は励ます言葉も慰める言葉も知らなかった。  
そして心がめするもののようにスケッチ帳をふところに納めてしまった。

図6 出力データの例

評価結果は、出力された文 1686 行のうち、誤回答が 216 行、正答率は約 87.2%。誤回答の主な原因としては、モダリティ表現の登録漏れと、分節化ルールの未整備が挙げられる。以下に誤回答の例を示す。

- モダリティ表現の登録漏れの例:「私は画面から目を放してもう一度君をみなおさないではいられなくなった。」⇒「私は画面から目を放す。もう一度君を見直さない。はいられなくなった。」この場合、「V-ではいられない」という形式がモダリティの表現として登録されていないため、不適切な位置で文章が分節化される。
- 分節化ルールの未整備の例:「君を木戸のところまで送り出してから、私は一人で手広いりんご畑の中を歩きまわった。」⇒「君を木戸のところまで送り出す。から、私は一人で手広いりんご畑の中を歩きまわった。」この場合、接続助詞の一部が除去されずに残る。

それ以外では構文解析のエラーを反映したものが散見された。

課題として、エラーの除去と、今回考慮しなかった、構文の文節位置・分節された文節の意味的關係を処理するルールの検討が挙げられる。

## 5. 物語生成における利用方法例

冒頭に述べたように、本研究の目的は、筆者らが研究・開発している物語生成システムのための間テキスト性の機構に適用することにある。[中嶋 2006]は物語文章からストーリーの部分を手作業で概念構造化し、その中の登場人物等の要素を置き換えたり、ストーリーの構成を変えたりして別の物語を作る試みをしているが、本研究の一つの目標は、この概念構造化の部分の手間をできるだけ省くようにすることである。

しかし現状のままでも間テキスト的な使用のイメージはいくつかある。例えば、概念表現化されていないとはいえ物語のストーリーの部分だけが抽出できれば、これをストーリーのパターンないしはスクリプトとして保存・利用することが可能性として考えられる。また、抽出された各文に対応するある出来事と他の出来事との外的時間関係(タクシス)[工藤 1995]を、それぞれの文におけるテンポラリティ、アスペクチュアリティの表現から推定することで、物語言説における時間順序・持続・頻度の構造を自動抽出するシステムも考えられる。

また、[佐藤 2008]は実際の多量の小説の本文からキーワードに合致する文を検索し、これをプロップによる物語の 31 の機能[Propp 1969]の順番に並べ、要素を一貫したものに置換することで、全て既存の文の利用と書き換えに基づいて物語文章を作成するシステムを提案しているが、その際ストーリーの構成要素としてはふさわしくない描写文や説明文なども検索にかかるという問題点があった。本稿で提案したシステムにより、ストーリーを構成する文のみを検索することで、この問題を解決することができる。

また、図 7 に例を示すように、抽出された各文を手作業で抜粋・並べ替えて新しい話を作るような利用も考えられる。

君は黙ったまま懐中からスケッチ帳を取り出す。  
そして二人は互いに理解するようにほほえみかわす。  
君はくすんだガラス板に指先を持って行く。  
君はそのガラス窓から中をのぞく。  
黒い事務マントを羽織った悒鬱そうな小柄な若い男が、一心に小形の書物に読みふけている。  
君の友は、悒鬱な小柄な顔をひときわ悒鬱にした。  
君は励ます言葉も慰める言葉も知らなかった。  
そして心がめするもののようにスケッチ帳をふところに納めてしまった。

図7 出力データを手作業で抜粋・並べ替えたもの

## 6. あとがき

本稿では、物語のテキストから叙法や修辭的效果等を除去し、物語中に実際に起こった出来事のみを対象に、それを時間順に並べたもの、すなわちストーリーを抽出するシステムについて検討した。今回は、その一手法として、ストーリーを表す文を抽出するシステムを試作した。また、試作プログラムを用いて、実際に物語テキストからストーリーを表す文を抽出した。

試作システムにおける課題として、上述したようなエラーの除去と、今回考慮しなかった、構文の文節位置・分節された文節の意味的關係を処理するルールの検討が挙げられる。また今後は、ストーリー抽出システムで得られたデータに基づき、物語テキストから物語内容の深層構造を記述する格フレームを抽出する事象概念自動抽出システムの実装を予定している。

## 参考文献

- [有島 2000]有島 武郎:生まれいづる悩み, 青空文庫, 2000, <http://www.aozora.gr.jp/cards/000025/card1111.html>, 最終アクセス 04/15/2008.
- [Kristeva 1970]Kristeva: *J.le texte du roman*, Mouton Publishers, The Hague, 1970. (谷口 勇 訳, テキストとしての小説, 国文社, 1985.)
- [工藤 1995]工藤 真由美:アスペクト・テンス体系とテキスト — 現代日本語の時間の表現—, ひつじ書房, 1995.
- [佐藤 2008]佐藤 秀樹・小林 厚太・安田 孝道・小方 孝:物語生成における文章表現に向けて, 日本認知科学会 文学と認知・コンピュータ研究分科会 II (LCCII) 第 14 回定例研究会予稿集, 14G-03, 2008.
- [宮崎 2006]宮崎 和人・安達 太郎・野田 春美・高梨 信乃:モダリティ(新日本語文法選書 4), くろしお出版, 2002.
- [守山 2000]守山 卓郎・仁田 義雄・工藤 浩:モダリティ(日本語の文法 3), 岩波書店, 2000.
- [中嶋 2006]中嶋 美由紀・小方 孝:物語生成システムと intertextuality-概念の整理と試作の考察-, 人工知能学会全国大会(第 20 回)論文集, 2E2-2, 2006.
- [大石 2007]大石 顕祐・小方 孝・中嶋美由紀・秋元 泰介:物語生成と間テキスト性の考察, 人工知能学会第二種研究会ことば工学研究会(第 26 回)資料, 55-62, 2007.
- [Propp 1969]Propp: V. (Пр о п п, В. Я.): Мор ф о л о г и я с к а э к и, Иэ, 2е. Москва:Н а у к а, 1969. (北岡 誠司, 福田 美智代訳:昔話の形態学, 白馬書房, 1987.)