

物語の構造抽出のための慣用的表現の処理

Processing Idiomatic Representations for Extracting Story Structures

村井 源^{*1}

Hajime Murai

^{*1} 東京工業大学

Tokyo Institute of Technology #1

In order to automatically extract narrative structure from story text, these are necessary to extract agents such as characters which acts subjectively and to identify their behaviors in described order. In the cases which one behavior corresponds to one verb, it can be processed by ordinary morphological analysis. In other cases, it is necessary to analyze relationships between contents of behaviors and expressions. In this research, focusing on idioms and pragmatic expressions, methods of extraction and identifications are proposed.

1. はじめに

近年発達してきた情報処理技術の活用により、計量書誌学をはじめとする文学の計量分析への試みは、様々な形で行われてきている。機械による物語理解の実現は現状では困難だが、物語分析に計量的指標を取り入れ、物語分析の客観性を高めうる可能性はありうると考えられる。このような方向性の元、人文的な物語分析で一般的な構造分析(バルト 1979)の中で古典的なプロット分析(プロップ 1987)を用い、分析結果に計量的解析を行い、物語構造の特徴、物語パターンの変化を人間の解釈と計量的指標の折衷により捕える試みも行われている(村井 2011)。また、単に物語を記述するのみではなく、登場人物の言動に基づいた構造を用いることで物語のオチをデータとして記述することも実現されており、人間が物語を読んで面白いと思えることの原因を構造の側面からデータ化し分析することも可能である(Murai 2014, 2015)。

しかしながら、これらの分析で用いる物語構造は人手によって物語を解釈して作られている物であり、コンピュータが物語を読んでその構造を抽出しているわけではない。現在の自然言語処理技術においては、コンピュータが自然言語で記されたテキストの内容までを意味的に把握することは困難性が高く、物語中でどの登場人物がどのような言動を行っているかを正確に理解した上で精度よく特定し抽出することは実現されていない。

そこで本研究においては、物語中の登場人物の言動とそれらによって生じる物語上の機能のみに焦点を絞って、どのような表現を分析できれば、登場人物の言動から構成される物語構造を実用的な精度で抽出可能であるか検討し、そのために必要な慣用的な表現の処理についてデータ構造と分析のアルゴリズムを提案する。

汎用性が高くかつ物語の本質に踏み込めるようなデータ構造の自動的な獲得が実現できれば、物語間の計量的な比較やコンピュータ等での物語構造の自動作成などへの応用分野にも道が開かれると期待される(松原 2013)。

2. 物語の登場人物の言動分析

2.1 言動分析の課題

物語中から登場人物の言動を抽出するために、物語中で意

思を持って主体的に行動する対象を指し示す語彙を集めた主体語彙辞書を作成し、主体語彙を含む命題を抽出したところ、物語中の重要な展開を示す表現の大部分を取得することができた(村井 2015)。しかし、これらの主体語彙を含む命題表現が登場人物のどのような言動を示し、物語中でどのような機能を持っているかは、例えば単純に動詞を抽出することによっては特定が困難である。

その理由の一つとして、日本語に限らず一般的に言語には様々な慣用的な表現があり、命題中の動詞と他の単語の組み合わせによって単体の同士の場合とは異なる意味を示す用法が多数あることが挙げられる。例えば、単に「立つ」の場合は直立するような状態への変化を示すが、「気が立つ」では興奮や怒りを表すが、「腕が立つ」ではすぐれている状態を表す。このような慣用的表現の収集とリストの作成は自然言語処理の研究領域において進められてきている(佐藤理史 2007)。

また一つの動詞が複数の意味を持ち、慣用的な表現程明確ではないがその動作対象や主体と結びつく助詞や助動詞のパターンによって異なる意味となる場合も多数存在する(荻野 2003)。本研究ではこれらも慣用的表現として一般的な慣用句と同様に扱った。

さらに、日本語の場合、会話中の発話文の文末に特定の助詞や助動詞の組み合わせ付けることで疑問や命令、批判などの機能を持たせることが可能である(国立国語研究所 1960)。例えば、「話す」という動詞の場合、あとに「なよ」(話すなよ)が来れば禁止の意味になるが、「なさい」(話しなさい)がつけば命令の意味になる。これらは主体の言動の機能としては「禁止+言う」と「命令+言う」で、同じ動詞を用いているにもかかわらずその後続く助詞等によって全く逆の意味になる。

物語構造の自動的な抽出においては、これらの動作的表現の特徴を踏まえた上で、物語テキスト中に出現する主体の動作を示す表現からどのような言動を行っているかを特定して抽出する必要がある。

2.2 分析の方針

登場人物の言動の分析のために、まず主体を含んだ命題を抽出し、それを対象としてまずデータベース中の慣用句表現と一致するかを判定し、その後慣用句として用いられていない部分に関して特定の文末表現と一致するかをさらに判定し、これら二つに含まれなかった動詞に関して物語の機能としての分類(Murai 2014)を行った。図 1 に分析の流れを示す。

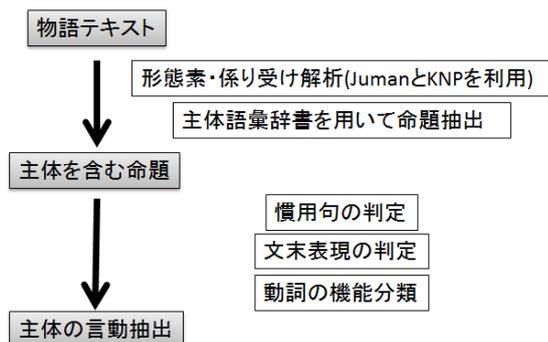


図1 分析の流れ

2.3 用いたデータと構造

慣用句のデータとしては、『基本慣用句五種対照表』(佐藤理史 2007)中の動詞で終わるが諺ではない慣用句をベースとし、一般的な慣用句リストには含まれないが名詞との組み合わせで慣用的に意味の変化するパターンなどを加えて 2185 のパターンをデータとして登録した。文末表現のデータとしては、『話しことばの文型 1 対話資料による研究』(国立国語研究所 1960)記載の文例をベースに、197 パターンを登録した。

慣用表現のデータの構造は図 2 のように、文節ごとに単位を区切っており(文末表現は形態素ごと)、各文節に対して表記揺れや別表現、敬語の場合の表現などを複数持つことができる。また動詞部分に関しては、動詞本体以外に、サ変(する)、受け身(られる/れる/せる)、否定(ない/ぬ)とそれ以外(よう/そ/やすい)などをデータとして分割記述するようにしている。

また慣用表現の一致の判定においては、まず最後の動詞部分に係り受けする文節が慣用句データにマッチするかをチェックし、マッチしない場合は次に最後の動詞から指定数の文節内で慣用句データにマッチするかをチェックする形にしている。このため、「足が地に着かない」のような基本的な形だけでなく、「あしが全く地につかず」のような間に修飾的な形態素を含んだり、動詞が活用していたりする場合にも適用が可能である。

足／あしガ 地ニ 着く／つく ない／ぬ

図2 慣用表現のデータ構造例

文末表現のデータは単語単位になっており、対象となる物語テキストの各文の末尾から数えて最長一致するパターンを出力する形になっている。文末表現に関しては、実際の形態素以外に活用する形態素の特定の活用形が出現する場合にマッチするという表現も許容している。例えば、図 3 のように「動詞基本形 + 「な」 + 「よ」というパターンは、「休むなよ」や「動くなよ」などの表現とマッチする形になっている。

動詞基本形 な よ

図3 文末表現のデータ構造例

慣用表現・文末表現のいずれともマッチしない動詞が命題中に含まれる場合、物語の展開に関連の深い動詞のカテゴリ表を用いて機能の分類を行う。動詞の機能分類は「対象」「自身」「状況」「意図」「評価」の 5 大カテゴリの下に 18 の小カテゴリがあり、それぞれのカテゴリが肯定・否定および人称の属性を持つ。

表1 情報カテゴリでの動詞分類の例

情報	一人称	二人称	三人称
肯定	分かる,知る	言う,教える	放送
否定	忘れる,誤解	騙す,隠す	検閲

例えば表1のように対象大カテゴリの下の情報小カテゴリの場合、肯定形は情報を得ること、否定形は情報を失うことや間違った情報を得ることに相当する動詞をカテゴリ化する。主体本人の言動は一人称、相手がいる場合は二人称、不特定多数が主体や対象の場合は三人称の分類している。

現在 831 の動詞がカテゴリ化されており、これらとマッチする場合にはそれを動詞の言動の機能として判定する。なお、各慣用表現及び文末表現もその機能が同じカテゴリで付与されている。

3. 結論と今後の課題

慣用表現と文末表現を通常の動詞と同様に物語上の機能で分類することが可能になったが、現在データとして登録されているパターンの数は決して十分とはいえず、今後他のリストを追加したり、特定のコーパス中に出現するパターンを人手でチェックしたりするなどして増やしていく必要がある。またある程度十分な量のパターンが蓄積できたら実際の物語テキストを用いてその精度の判定などを行い有効性を確認する必要がある。

参考文献

- [バルト 1979] ロラン・バルト(花輪光訳), 物語の構造分析, みすず書房, 1979.
- [ブロップ 1987] ウラジーミル・ブロップ(北岡誠司, 福田美智代訳), 「昔話の形態学」, 水声社, 1987.
- [村井 2011] 村井 源, 松本 斉子, 佐藤 知恵, 往住 彰文: 物語構造の計量分析に向けて-星新一のショートショート-の物語構造の特徴-, 情報知識学会誌, Vol. 21, No. 1, pp.6-17, 2011.
- [Murai 2014] Hajime Murai: Plot Analysis for Describing Punch Line Functions in Shinichi Hoshi's Microfiction, 2014 Workshop on Computational Models of Narrative, OpenAccess Series in Informatics, Dagstuhl Publishing, Vol. 41, pp. 121-129, 2014.
- [Murai 2015] Hajime Murai: Towards Automatic Extraction of Punch-lines: Plot Analysis of Shinichi Hoshi's Flash Fictions, JADH Annual Conference 2015, Proceedings of JADH Annual Conference 2015, JADH, pp. 17-18, Sep. 2015.
- [松原 2013] 松原 仁, 佐藤 理史, 赤石 美奈, 角 薫, 迎山 和司, 中島 秀之, 瀬名 秀明, 村井 源, 大塚 裕子: コンピュータに星新一のようなショートショートを創作させる試み, 人工知能学会全国大会, 2D1-1, 2013.
- [村井 2015] 村井源: 物語の構造抽出のための主体語彙辞書の作成, じんもんこん 2015, 情報処理学会シンポジウムシリーズ, Vol. 2015, No. 2, pp. 111-116, 2015.
- [佐藤理史 2007] 基本慣用句五種対照表の作成, 情報処理学会研究報告 自然言語処理 (NL), 2007-NL-178, pp. 1-6, 2007.
- [荻野 2003] 荻野孝野, 小林正博, 井佐原 均: 日本語動詞の結合価, 三省堂, 2003.
- [国立国語研究所 1960] 国立国語研究所, 話しことばの文型 1 対話資料による研究, 秀英出版, 1960.