

昔話の自動シナリオ化システムの構築

今 誠一[†], 吉田 文彦[‡], 内田 理^{††}, 菊池 浩明^{††}, 中西 祥八郎^{††}
[†] 東海大学大学院工学研究科, [‡] 東海大学文学部, ^{††} 東海大学電子情報学部
e-mail: now@ep.u-tokai.ac.jp

1. はじめに

近年、多くのものが放映されている映画やドラマ、劇等のシナリオは、シナリオライターが題材となる文章を読み、決められた形式に書き換えることによって作られている。シナリオは省略の芸術とも言われ、必要最低限の文で内容を的確に表している。そのため、シナリオ作成者は、元となる物語の内容を理解していなければならないと考えられる。

シナリオを自動的に作成する研究は、著者が知る限り見当たらない。シナリオを自動作成できれば、シナリオライターの負担を軽減できると考えられる。

そこで本稿では、自動シナリオ化システムを構築する。そのために、場面（シーン）決定と、場面それぞれのタイトル決定手法を提案する。さらに、シナリオの根幹となる柱、ト書き、セリフの抽出方法について提案する。

2. 昔話

本稿は昔話を題材とし、システムの構築を行った。昔話の特徴として、以下のことが挙げられる。

- I. 感情的表現や抽象的表現が極めて少なく客観的事実によって語られている。
- II. 主人公も敵も道道具・小道具も孤立的に語られている。
- III. 同じ場面をほとんど同じ言葉で繰り返す。
- IV. その様子をリアルには語らない。
- V. 「移動」によって「場所」の変化を見分けることが、一般的な文章に比べて判断しやすい。

昔話は語り継がれていたものであるため、その内容は簡潔になっており、文章構造を把握することが容易である。

今回はドコモ電子図書館[1]内の民話・昔話コレクションに所蔵されている昔話 108 話を題材として、54 話をシステム作成用に、残りの 54 話を評価用に用いた。

3. シナリオ化システム

3.1 概要

本システムは、大きく 2 つに分かれている。まず昔話を場面ごとに分ける。これにより、物語の内容を大まかに把握することができる。その後シナリオ化を行い、

シナリオを作成する。図 1 にその概要を示す。

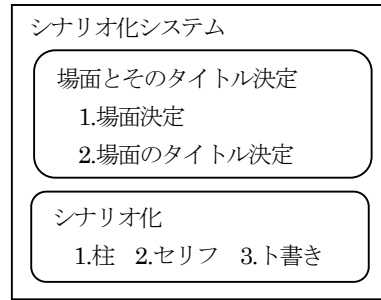


図 1. シナリオ化システムの概要

場面とは、昔話を内容ごとに分けたものであり、シナリオを作る際の柱を立てる部分として扱うことができる。

3.2 場面とそのタイトル決定

3.2.1 場面決定

場面は 4 つの行程により決定する。その流れを図 2 に表し、以下に手順を示す。

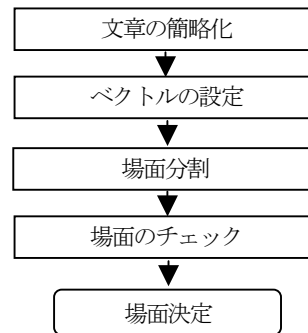


図 2. 場面決定システムの流れ

手順

- (1) 主語が複数ある文に対し、文章を簡略化し内容を把握しやすくするために、係り受けを調べる。そして、主語と述語が一对一の文になるように、一文を複数文にわけ、空行のある部分は、確実に場面変化しているものとし、予め区別しておく。
- (2) 元の文が何文目かを表す文番号(x)、何段落目に所属していたかを表す段落番号(y)、また、場所を示す単語に物語の進行順に順番に番号をつけた場所番号(z)の 3 次元ベクトル(x, y, z)を各文に対応させる。

- (3) 対応させた3次元ベクトルに基づきクラスタ分析を行い、分類されたクラスタ^(注1)をそれぞれ一場面とする。
- (4) (3)で作られた場面をチェックし、隣り合う場面の最後の3行と最初の3行を比較し、登場人物と時間について調べる。この2点において変化がない場合、2つの場面を1つに統合する。全ての場面で登場人物と時間についてチェックを行い、場面を統合して、最終的に場面を決定する。

これで場面分けが行えるのは、昔話が2章で述べた特徴I・Vを持っているからである。

3.2.2 チャート作成

登場人物の動作を簡易的に表したチャートを作成する。チャートの一部を図3に示す。チャートによって表示されるのは、登場人物の場所での動作と、他の登場人物に対する動作である。

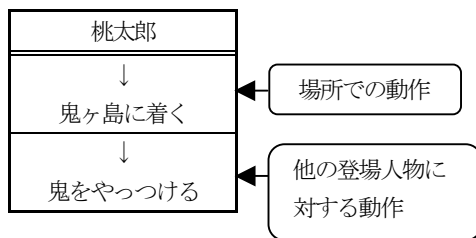


図3. チャートの一部

場所での動作は、場面決定の際に簡略化された文から名詞を抽出する。そして、名詞に対して、日本語語彙体系[2]より概念を調べ、概念が「場所」である名詞の含まれている文を抽出し、そこからさらに場所と動詞を抜き出し「場所での動作」とする。

同様に、簡略化された文に主語となる登場人物以外に登場人物がいた場合、「他の登場人物に対する動作」となる。昔話は2章のIIの特徴があるので、このように決定することができる。

この2つの動作を登場人物ごと、昔話の流れに沿って書き、チャートを完成させる。

3.2.3 場面のタイトル決定

タイトルは、

「(主語) が (述部)」

とし、その決定方法を以下に示す。

(A) 主語決定

- (1) チャートより、その場面内で最も登場回数の多い登場人物をタイトルの主語とする。
- (2) 主語の登場人物と同じ動作をしている人物は、主語としてタイトルに加える。

(B) 述部決定

- (1) 主語となると登場人物の場所での動作をタイトルの述部とする。図3において、動作は「鬼ヶ島に着く」、「鬼をやっつける」の部分である。
- (2) 場所の動作が複数ある場合は、以下に示す優先順位によって述部となる動作を決定する。
 - a. 主語の登場人物と同じ動作をしている人物が多い動作
 - b. 概念の深い名詞が含まれている動作
 - c. 出現頻度が多い名詞が含まれている動作
- (3) 上記によって決定できない場合、他の登場人物に対する動作をタイトルの述部とする。
- (4) 他の登場人物に対する動作が複数ある場合、(2)と同様の手法によりタイトルの述部を決定する。

以上により決定された主語と述部を「が」で接続し、タイトルを決定する。

チャートで表せない場合、具体的なタイトルをつけるため、名詞の概念の深さを調べ、その合計が一番大きい一文を抜き出し、構文解析を行う。そして、述語にかかっている枝で、最も近い部分をタイトルとする。

この例を図4で示す。抽出文として「桃太郎はどんどん食べて大きくなって、力持ちになりましたが、桃太郎は怠け者でした。」を考える。述部は「怠け者でした」なので、一番近い枝が「桃太郎は」となり、タイトルは「桃太郎は怠け者でした」となる。

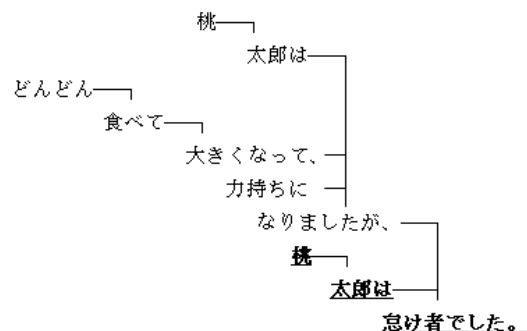


図4. 構文解析結果

3.3 シナリオ化

3.3.1 シナリオ

シナリオ本文は、「柱」、「セリフ」、「ト書き」の3要素から構成されている。

注1. クラスタ数を与える必要がある場合は、セリフを除いた行数の5分の1とする

柱とは、1つのシーンの始まりを示すもので、行の1文字目に「○」を書き、その下に場所を書く。似た場所がある場合は、「鈴木家・一郎の部屋」の様に、「・」で区切り、明確に場所を指定する。同じ場所であるなら全体で書き方を統一する。その下に括弧で朝・夕・夜などの時間帯を書く。指定がなければ、昼のシーンとみなす。

柱は場所を指定するものであるため、場所の変化が少しでもあれば柱を立ててもよい。ただし、同じ場所が続いたときは柱は立てず、同じ場所でも時間変化があれば柱を立てるものとする。

セリフは、登場人物の喋る内容を鍵括弧で書く。その人物の言った言葉のみを書くのが基本だが、細かな動作はト書きにするのではなく、セリフの中に括弧をつけていれることもできる。

ト書きは、登場人物の特徴と動作の説明、場所の説明および場面の状況を具体的に書く。必ず各行に、3文字分空白を頭につける。簡潔に、必要十分な情報だけを書き、過剰な形容詞や心理描写はしないことが重要である。

3.3.2 柱

(A) 場所の特定

柱にとって必要なのは場所と時間であり、まず、場所を抽出する方法について述べる。

- (1) 日本語彙体系より名詞の概念を調べ、場所の概念に含まれる名詞を含む文節を抽出する。
- (2) (1)の文節の助詞を調べる。助詞が「へ」であるなら次の文の前に柱を立て、「に・から」であるならその文の前に柱を立てる。
- (3) (1)の助詞が「の」であり、その後「場」の概念に含まれる名詞がきた場合、まとめて柱とする。違う場合は柱とはしない。
- (4) 一文の中に複数の場所の概念を持つ名詞が表れた場合、最も後ろにあるものを柱として抽出する。

(B) 時間の特定

次に、時間の抽出方法について述べる。時相名詞を抽出し、柱の一部とする。例外として、柱とする際に言葉を変化させなければならぬ語もある。そのような語に関しては、データベースを作成する。

例：夜が明けて→朝

一年がたった→一年後

3.3.3 セリフ

セリフで重要となるのは、そのセリフの話者の特定である。セリフの前後の文の書き方にはいくつかのパターンがあり、話者はそのパターンから特定できる。よって、パターンからいくつかのルールを作成して、話者を判定する。

まず、セリフは鍵括弧でくくられている部分であるの

で抽出する。次に、作成したルールより話者を特定する。以下、作成したルールについて、適用される順に述べる。

- (1) 「言う」や「話す」のようなセリフに関連する動詞を集め、データベースを作成する。そして、セリフと動詞から話者を特定する。「言う」を用いた例として以下を示す

- a. 「・・・」と言って、Aは～
- b. 「・・・」Aが言う～
- c. Aは言った。「・・・」

この様なパターンの場合、セリフはAの言ったものだと特定できる。

- (2) セリフ中に「Aです」「Aだ」「Aじゃ」とある場合、Aが話者だと特定できる。また、セリフ中に「A」「Aや」「Aよ」とあり、次にセリフがある場合、次のセリフはAのセリフだと特定できる。
- (3) セリフが連続している場合、登場人物は順番に話しており、一つのセリフの話者を特定できれば残りのセリフの話者を特定できる。
- (4) 登場人物には話し方に特徴がある場合もあり、特徴を抽出し話者を特定することができる。例えば、「俺、おれ、わし、あっし」がセリフの中にあれば、そのセリフは男性のセリフと特定できる。語尾が「～わ。」で終われば女性のセリフである。昔話は男性と女性の区別を明確にしているため、男性か女性かを特定することができる。
- (5) 以上で特定できない場合には、セリフの前後の文の主語が同じ登場人物であるなら、話者はその登場人物とする。違う時は、セリフの前の文の主語が登場人物なら話者とする。登場人物でないなら後の文の主語を話者とする。

また、ドコモ電子図書館所蔵の昔話の特徴として歌が入っている。歌は、文の始めに「♪」記号がついており、句読点がない。よって、「♪」のついた文から、句読点の入った文がくるまでが歌と特定できる。よって、歌はセリフとして抽出する。

3.3.4 ト書き

昔話は、客観的事実によって書かれているので、ト書きに似た書き方になっていることが多い。よって、ト書きへ変換するのが不要な部分を省くだけでよく、現代的な物語と比較して容易である。

以下、昔話をト書きへ変換するための方法を示す。

- (1) ドコモ電子図書館所蔵の全ての昔話は、物語の始まりが「とんと昔あったと。」であり、終わりは「昔こっぼり、てんぼろりん。」となっている。この2文はト書きには不要なので削除する。

- (2) 文末が「動詞の過去形+と。」または「動詞の過去形。」となっている文を残し、それ以外を削除する。その上、「動詞の過去形+と。」となっている文の「と」を削除し、文末の動詞を原型に直す。
- (3) 文末の動詞が思考動作の概念に含まれるのであれば、その文は削除する。
- (4) 思考動作の概念に含まれる動詞が文の途中で出てきた場合、その文の木構造を調べ、その動詞を含む塊を削除する。ただし、主語は削除しない。木構造[3]の例を図5に示す。図5において、「乞食は、不思議なことにでくわしたなあ、と」が塊である。よって、「不思議なことにでくわしたなあ、と」を削除する。
- (5) 繰り返されている語、『カタカナ語+と』となっている文節は擬音語なので削除する。

例：ムシャムシャと・ぐんぐん

- (6) 接続詞は削除する。また、文頭にあり、接続詞的な役割を担う複合語に関しては、そのような複合語を集めデータベースを作成し、削除対象とした。

例：そんなわけで、それからというもの

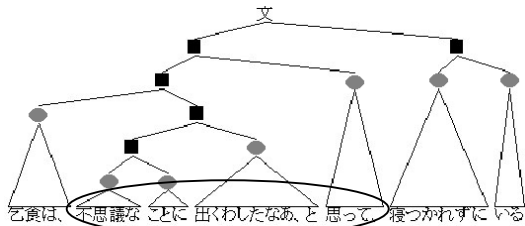


図5. 文の木構造

4. 実験と評価

これまで述べた手法により、実際に昔話を場面分けし、タイトルをつける。そしてシナリオ化する。その例として「かちかち山」を取り上げ、結果は図6のようになる。

かちかち山は9つの場面に分けることができ、以下に挙げるようなタイトルをつけることができる。

場面2：おじいさんが山へ豆を植えに行く

場面6：兎が茅山に行く

場面9：兎と狸が海へ出る

場面2はおじいさんが山へ行き、狸と出会う場面、場面6は兎が茅山で狸を懲らしめる場面、場面9は兎と狸がそれぞれ木の舟と泥の舟で海へ出る場面であり、それぞれの場面に適したタイトルであるといえる。

場面とそのタイトルを学生8名に評価してもらったところ、場面は90.1%、タイトルは61.8%のものが物語の展開と適合しているという結果を得た。場面は、物語の重要な要因である場所、登場人物、時間により決定され

たので、評価が高いと考えられる。一方、タイトルは登場人物の動作にのみ着目して決定したので、動作以外にタイトルにとって必要な要素があるために評価が低かった。しかし、シナリオ化にはほとんど影響がない。

○山の畑
爺さん「一粒まいたら、千粒になれ。」
二粒まいたら、二千粒になあれ」
爺さんが豆をまいている
狸が出て来て、切り株にちまこんと坐る
狸「一粒まいたら、腐ってしまえ。」
二粒まいたら、もろとももろと腐れ」

図6. かちかち山のシナリオ

シナリオは、作成者により様々なものが作成される。柱とセリフに関しては作成者が違っても変化は少ないが、ト書きは作成者により内容が様々になる。よって、シナリオの良し悪しを判定するのは難しい。

そこで、本システムでシナリオ化したものを、柱の立てた位置の適切さ、セリフの話者の特定という2点から著者が精度を調べた。柱のうち89.4%は適切な位置で立っており、セリフのうち89.8%は話者を特定できた。ト書きの評価は難しく、著者が自ら作成したシナリオと比較してみると、省略できる部分が若干あると思われる。

本システムは、場面決定ができ、シナリオ化は高い精度で行えるので、使用に十分耐えうるものである。さらに、シナリオに曖昧な箇所がないように、シナリオ化の精度を上げる必要もあると考えられる。

5. おわりに

本稿では物語の一つである昔話を用いて、場面分け、場面ごとのタイトル決定、そしてシナリオ化のシステムを提案した。

柱、セリフの適合率を上げ、昔話以外の様々な種類の物語にも対応できるようなシステムが構築されることが望まれる。

参考文献

- [1] ドコモ電子図書館
<http://www.digitallib.nttdocomo.co.jp/nihonb-unka/minwa/index.html>
- [2] NTTコミュニケーション科学研究所、「日本語語彙体系」、岩波書店
- [3] 日本語読解学習支援システム「あすなろ」
<http://hinoki.ryu.titech.ac.jp/index-j.php>