

物語生成システムにおける文生成機構と文字表記機構

鎌田 まみ (Mami Kamada)
岩手県立大学

小方 孝 (Takashi Ogata)
岩手県立大学

1. はじめに

筆者らが現在開発を進めている統合物語生成システム(小方, 金井 2010; Akimoto & Ogata 2012)は, 概念構造と表層表現それぞれの生成機構に大分類されるが, 後者は表現メディアとして音楽, 映像, 言語を利用する。そのうち言語による文生成機構は, 概念構造の生成部分において生成されたその基本構成要素としての個々の事象(もしくは事象概念)を基本文に変換し, さらにそれを文や文章のための各種技法を通じて変形するという枠組みとなっている。熊谷 他 (2012)は, 幾つかの機能が複合された文生成機構の試作を報告した。またこれを応用システム KOSERUBE(今淵 他 2012)において使用した。本稿では, 物語生成システムのための文生成機構の現状を拡張した言語表記辞書の記述を加えて改めて整理し, 上記応用システムにおける利用とそこから得られた課題をまとめる。さらに, 新たに付加することを計画している文字表記機構へのアプローチを述べる。文字種類の多さとその表記の多様性は日本語の大きな特徴である。文字表記機構では, 文字表記の違いによる効果や特定の作家の文字表記の分析を行い, 受け手に特定の効果(読みやすい, 美しい等)を与える表記や, 特定の作家風の表記を行う方法を考察する。

2. 文生成機構の概要

文生成機構全体の処理手順を図1に示す(熊谷 他 2012)。物語の概念構造を入力として, 単文処理, 複文処理を経て文表現(文字列)を出力する。実装は Common Lisp で行っている。

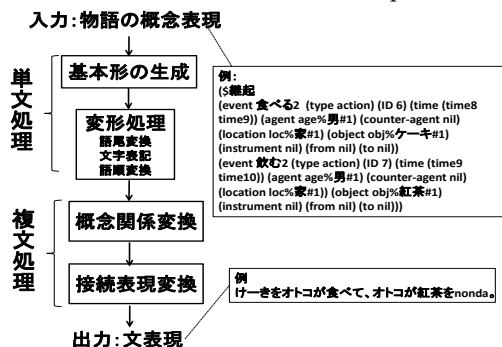


図1 文生成機構全体の流れ

2.1 入力となる物語の概念構造

概念構造は, 事象概念を終端節点として, それらを関係によって結合した一種の談話的構造を表す階層的な木構造により表現する。個々の事象概念は, 動詞概念と深層格からなる格フレーム形式で記述する。深層格は主に time, agent, counter-agent, object, instrument, location, from, to の8種類を用いる。これを構成する具体的な動詞の概念及び名詞の概念は, それぞれの概念辞書(Oishi et al. 2012)により提供される。なお, 深層格の値には, 名詞概念をインスタンス化した具体的な人物や物, 場所が格納される。データ上では, インスタンスの ID(固有記号)が深層格の値として記述される。図1の入力の概念構造においては, 「(loc%家#1で) age%男#1が obj%ケーキ#1を食べる」及び「(loc%家#1で) age%男#1が obj%紅茶#1を飲む」という意味を表すふたつの事象が, 「継起」という関係で結ばれている。

2.2 文生成処理の概要

文生成は, 大きく分けて単文処理と複文処理からなり, それぞれがさらに幾つかのモジュールに分かれる。まず, 単文処

理は概念構造に含まれる個々の事象概念を単文(文字列)に変換する。処理は大きく以下の3つに分かれる。

(1) **基本形の生成**: 事象概念から, 動詞概念辞書に登録された文型パターンを用いて基本的な単文構造を生成する。文型パターンとは, その動詞を述語とする文の基本的な文型を表したものであり, これだけで当該事象の最も原始的な意味は表現される。単文構造は, 文節ごとに区切られたリスト形式で表現される。

(2) **文の変形処理**: 単文構造に対して以下の3つの処理を順に行う。

① **語尾変換**: 単文構造中の動詞部分を抽出し, 特定の語尾変換規則を適用して変換した結果に置き換える。

② **文字表記変換**: 単文構造中の各概念名および助詞を, 後述の言語表記辞書に登録されている文字表記の中のひとつに置き換えた文を生成する。詳細は3節で述べる。

③ **語順変換**: 単文構造中の文節を並び替える。但し, 動詞は必ず文末に置く。

(3) **文変換**: 単文構造を文字列に変換する。この結果を概念構造中の元の事象概念と置換する。

以上の一連の流れの実行例を図2に示す。これらの処理をすべての事象概念に対して行い, 最終的に入力された概念構造のすべての事象概念が単文に置換されたものを出力とする。

```
(event 食べる2 (agent age%男#1) (object obj%ケーキ#1) (location loc%家#1))
↓ [基本形の生成]
((agent age%男#1) (p が)) ((object obj%ケーキ#1) (p を)) ((v 食べる))
↓ [変形処理]
((object けーき) (p を)) ((agent オトコ) (p が)) ((v 食べた))
↓ [文変換]
けーきをオトコが食べた。
```

図2 単文処理の過程

以上の結果を入力として, 複文処理を行う。その中間節点(概念的関係)に応じて適当な接続規則を選択・適用する。出力は物語全体の文表現となる。概念的関係とは, 因果関係や継起関係のような事象接続の意味的關係である。処理は以下の2つに分かれる。

(4) **概念構造から言語関係への変換**: 入力に含まれる各概念的関係を, 言語的關係(「順接」「逆接」等)に変換する。

(5) **接続詞・接続助詞の変換**: 言語的關係で結ばれた2つの文に対して, その言語關係に対応する接続規則を適用して, 接続表現(接続助詞あるいは接続詞)を付加する。これを, 木構造の下層から最上位階層まで行い, 最終的にひとつの文字列として結合された物語文を出力する。

概念的関係から言語的關係への変換例と複文処理の出力例を図3に示す。

```
($並列 "オトコがけーきを食べた。" "オトコが紅茶を nonda。")
↓ [概念的関係から言語関係への変換]
($順接 1A "オトコがけーきを食べた。" "オトコが紅茶を nonda。")
↓ [接続詞・接続助詞の変換]
オトコがけーきを食べて、オトコが紅茶を nonda。
```

図3 複文処理の過程

3. 言語表記辞書と文字表記決定機構

日本語の文字種は多様で, 同じ意味の言葉に対して様々な文字表記が可能である(例: 「男」→「男」「おとこ」「オトコ」「otoko」他)。文字表記決定機構は, 概念辞書における概念名や文型パターンにおける助詞の表記を決定することで, その表現的多様性を実現する。各概念および助詞に対応する文字表記の可能な種類を記述した言語辞書(言語表記辞書)を構築しており, 文字表記決定機構はそれを利用する。

3.1 言語表記辞書

言語表記辞書は, 概念辞書に登録された動詞概念と名詞概念の名前(動詞概念 4920件, 名詞概念 80181件)及び40種類の助詞に対する漢字, ひらがな, カタカナ, ローマ字(へ

ボン式)の4種類の表記を登録する。ひらがなとカタカナの読みは新仮名使いとする。図4に示すようなリスト形式で記述され、先頭に概念名を記述し、その後に漢字、ひらがな、カタカナ、ローマ字の順で各々の表記を列挙する。

(紅茶 (紅茶) (こうちゃ) (コウチャ) (koucha))
(二世 (二世) (にせい) (ニセイ ニセ) (nisei nise))

図4 言語表記辞書の記述例

3.2 処理手順

言語表記辞書を用いた表記決定機構の処理手順を述べる。単文構造を入力とし、そこに含まれる各要素(インスタンス名・助詞・動詞概念名)に順番に次の流れで処理を施す—①対象の要素がインスタンス名の場合、対応する(instance-of 関係にある)名詞概念名を取得し、それを検索キーに言語表記辞書を参照して表記データを獲得する。助詞及び動詞概念名の場合は、それをそのまま検索キーとして同様に言語表記辞書から表記データを獲得する。②その中からひとつの表記を後述する方法により選択する。③選択した表記を単文構造中のもとの要素と置換する。処理②において、複数ある表記の中からひとつを選択する方法であるが、現状の機構では、ランダムで決定する方法と品詞毎に特定の表記を予め指定しておく方法の2種類を用意する。

4. KOSERUBE における文生成機構の利用

KOSERUBE と呼んでいるシステムは、岩手県に因んだ登場人物や場所や物品が登場する民話風物語を生成・上演する、統合物語生成システムの一応用システムである(今淵 他 2012)。これは統合物語生成システム全体を含むが、民話風の比較的分かりやすい子供向け物語を生成する娯楽的機構を意図しており、使用する機能の種類や順序の限定・単純化、物語を面白くするための各種演出や工夫を含んでいる。ストーリー生成機構としては主にプロップに基づく機構(Imabuchi & Ogata 2012)を利用して民話的なストーリーが容易に展開されるようにしている。生成された物語の表現は簡易アニメーションを含む画像自動編集、音楽自動生成に加えて文生成によって行っている。単純な文の生成でもそれなりの効果はあると考え、現状での文生成機構のひとつの実験の場としても位置付けている。

システムを起動すると、タイトル画面が表示され、ユーザが①主人公と敵役、②物語の長さまたは型、③語り手を選択する。その後、システムが生成した物語が、紙芝居風に上演される(図5)。基本的に、事象がひとつずつ順番に文・画像・音楽により表現される。文表現は、画面下の字幕として表示されると同時に、音声読み上げソフトによって朗読される。実際に使用している文生成機構は、2節で述べた機構に幾つかの拡張を加えたものである。



図5 物語上演画面

文表現の自然さや面白さのために行った工夫を図6にまとめる。方言や代名詞、動詞言い換え機能については、KOSERUBE 独自の辞書を作成し、それを用いて変換を行う。また、語り手のキャラクターを選択することで文表現が変化する工夫も行った(表1)。

KOSERUBE の実行例の文表現における文法的誤りや不自然な点、分かり難い点を洗い出した結果を表2にまとめる。

一方、アンケート調査とミュージアムでの展示を行った(詳細は今淵 他 (2012)を参照)が、文表現に関しては、特に語り手の設定による表現の違いの改善が話題になった。上記表2の

方言:動詞・名詞・形容詞・助詞を岩手の方言に変換(「死ぬ」→「めおどす」)。

動詞言い換え:「帰還する」→「帰る」のように簡単な表現に変換。

助詞言い換え:「へが」と「へは」の変換。

人物の呼び方:「概念名」、「名称」、「概念名と名前(例:カッパの太郎)」の3種類を用意。例えば、初登場の人物は概念名と名前前で呼ぶ。

語尾変換:通常は丁寧態(～ました。)、予言により語られる事象は未来形(「～でしょう」を付与)、等。

複文化:スクリプトにより詳細化された事象列の接続表現変換・複文化(順接「～して、」)。

場所記述追加:場所(location)変化時のみ、文に場所(～)を追加(それ以外は場所を含まない)。

描写・説明文挿入:初めて登場した人物・物・場所についてそれらの説明や描写文を挿入する。

代名詞化:ひとつ前の事象に出現した登場人物(agent)や物(object)が連続で現れた場合、それを代名詞に変換する。

省略:ひとつ前の事象に出現した登場人物(agent)や場所が連続で現れた場合、それを省略する。

入れ子文:入れ子文の繋ぎに文を挿入(例:青年は「～すること」を子供に命令した)。

予言・回想:予言により語られる事象文の語尾に「～でしょう」を付加。また、回想や予言による変換時の導入文テンプレートを挿入(例:これから～による回想が始まります)。

図6 文表現の自然さや面白さのための工夫

表1 語り手による文字表記の違い

語り手	文生成
座敷わらし	表記をすべてひらがな
おばあさん	漢字のある名詞・動詞は漢字とし他はひらがな
宮沢賢治	漢字がある名詞・動詞の表記は漢字とし他はカタカナ
猿	一文の表記をすべてカタカナかすべてローマ字の何れか(一文毎に何れかをランダムで選択)
南部鉄器	一文毎に名詞、助詞、動詞の各表記を漢字、ひらがな、カタカナ、ローマ字の中からランダムで選択

表2 文表現に関連する問題点

省略機能	<ul style="list-style-type: none"> ・物品の省略が行われていない。 ・人に対する省略の不備。 ・前の文で出現していない人物が出現した際、場所が省略される場合があり、違和感が生じる。
代名詞機能	<ul style="list-style-type: none"> ・場所の省略が行われていない。 ・変換した語が元々何を指していたのか分かりにくい部分がある。例:「龍泉洞で彼が鮭を調べて、宮沢賢治がそれを捜しました。」 ・代名詞変換後の文の後の文に代名詞が適用される前の人物名が再度出て来る。例:「宮沢賢治が鬼の館からイギリス海岸まで来ました。彼が被害を知りました。宮沢賢治が冒険に備えました。」(しかしこの方が分かり易い場合もある。) ・描写文における代名詞変換の不備。 ・場所の単語が「それ」に変換される場合がある。
その他	<ul style="list-style-type: none"> ・単語に一般的でない表記の文字が使われる場合がある。 ・語順変換を行った際、意味の通らない文になる場合がある。例:「イギリス海岸で嘘偽りがアデルイのばれました。」 ・描写や説明の文が不足していて、意味が分かりにくい場合がある。(登場人物の行動の理由等。) ・場所が文の最初に現れることが多い。 ・後の文で出現しない人物が一回だけ出現する場合がある(文の問題ではない)。 ・単文の場合、読点がない。

問題は文表現機構一般の問題であり今後の発展のための課題群を形成する。一方後者はそれ自体としては KOSERUBE 特有の演出の問題であるが、問題を一般化すれば、語り手の違いや物語に対する位置付けに応じた生成方略の制御の問題につながる。これは、本研究が特に物語言語機構において参照するジュネットの物語言語論(Genette 1972)における「態」の範疇と関連する問題である(上田, 小方 2004)。統合物語生成システムの枠組みの中では、物語言語機構において上記態に相当する物語における語り手及び聴き手の位置付け(例えば語り手は物語内容の事象群を事後的に語るのか、予言的に語るのか等)や属性を付与し、これが物語内容の物語言語化の具体的なあり方を規定するというモデルとなっているが、この種の情報は物語表現機構にも影響を与える。本来「物語言語」という概念は本研究で言う「物語表現」も包含した概念である。KOSERUBE の第一版では予め語り手を数種設定してその中に語り手の文表現の方略を予め組み込んだが、上述のような諸情報に基づいてそれをもっと柔軟に編成して行けるような制御

方法の組み込みが今後の文表現機構(及び音楽や映像の表現機構)における重要な課題である。

5. 効果に基づく表記選択

以上に述べた文表現機構のうち、文字表記変換のための一手法として、受け手に与えたい効果に基づき表記選択を行う方法の構想を述べる。これは KOSERUBE において語り手による表記の違いを強調する方法としても利用する。文字表記の違いが受け手に与える効果を調べるための実験の方法と結果を説明し、それに基づく方法案を述べる。

5.1 実験の方法と結果

(1) 実験方法

ひとつの事象概念を意味する単文の基本形に対して、動詞・名詞・助詞それぞれに、4 種類の文字表記の何れかを指定した 37 パターン(表 3 にその一部を示す)の文字表記変換を行った文を用意する。10 種類の事象概念から、合計 370 文を用意した。何れの文も、名詞がふたつ存在し、先頭が主語(「人」となる。

表 3 文字表記の組み合わせ(一部)

パターン	組合せ	番号
全てひらがな	全てひらがな【おとこがつよきをみせる】	1
全てカタカナ	全てカタカナ【オトコガツヨキミセル】	2
全てローマ字	全てローマ字【otoko ga tuyoki wo miseru】	3
漢字+ひらがな	名詞・動詞を漢字【男が強気を見せる】	4
	名詞を漢字【男が強気みせる】	5
	動詞を漢字【おとこがつよきを見せる】	6
漢字+カタカナ	名詞・動詞を漢字【男が強気ヲ見せる】	7
	名詞を漢字【男が強気ヲミセル】	8
	動詞を漢字【オトコガツヨキヲ見せる】	9
漢字+ローマ字	名詞・動詞を漢字【男ga強気wo見せる】	10
	名詞を漢字【男ga強気wo miseru】	11
	動詞を漢字【otoko ga tuyoki wo 見せる】	12

被験者に対して、例文を 1 文ずつスクリーンに 3~5 秒提示し、その後 4~5 秒の間に被験者が以下の 4 つの評価項目に回答する。回答方法は、①は 2 段階、それ以外は 3 段階の選択式である。

- ① 読みやすさ{1. 読みづらい, 2. 読みやすい}
- ② 文の意味の理解度{1. 文の意味が理解できない, 2. 文の意味が少し理解できた, 3. 文の意味がすぐ理解できた}
- ③ 字面の面白さ{1. 字面がつまらない, 2. 平凡に感じる, 3. 字面が面白い}
- ④ 字面の美しさ{1. 字面が美しくない, 2. 平凡に感じる, 3. 字面が美しい}

これを 1 文から 370 文まで全て繰り返す。なお、時間内に回答出来なかったところは空欄にするように指示した。全文の回答後、被験者が③と④に関してどのような基準で評価したかを記述する。なお、実験は同一の 370 文を横書きで表示した場合と縦書きで表示した場合の二通り行った。横書きの文は、10~20 代の男性 7 人、女性 3 人、縦書きの文は、同じく男性 8 人、女性 2 人がそれぞれ行った(重複なし)。

(2) 実験結果

横書きでは各項目に対して総計 3650~3689 件(空欄除く)、縦書きでは同じく総計 3465~3558 件の回答が得られた。このデータを、次の 3 種類の方法により統計的に分析した(何れも MS Excel を利用)。

- A) 文字表記の組み合わせ毎に各評価項目の各回答の割合を計算。
- B) 分析 A の結果のそれぞれを降順にソート(各評価項目の各回答についての順位付け)。
- C) 分析 A の結果をもとに、各評価項目について、他の評価項目それぞれとの関係を散布図上にプロットし、相関関係を分析。

A の分析結果においては、全般的に縦書きの文の方が横書きの文よりも評価が低くなる傾向が見られたが、それ以外には特記するような違いは見受けられなかった。以下の考察は横書きの結果に基づき行う。表 4 に横書きにおける A の分析結果の一部を示す。評価項目の①②④に関しては、組み合わせ番号 4(名詞:漢字, 動詞:漢字, 助詞:ひらがな)の得点が最も高かったのに対して、③(面白さ)では、組み合わせ番号 7(名詞:漢字, 動詞:漢字, 助詞:ひらがな)の得点が最も高かった。分析

B では、各評価項目の上位に、漢字を多く含む組み合わせが多かった。また、分析 C からは、「読みやすさ」とその他の評価項目それぞれとの間に正の相関関係があることが見受けられた。以上の結果から、基本的には漢字を用いた一般的な表記が全般的に高い評価を得られるが、面白さに関してはそれから少し外れるような表記が効果的であると考えられる。

表 4 表記結果の一部(横書き)

組合せ番号	読みやすさ			文の意味の理解度			面白さ			美しさ		
	1	2	3	1	2	3	1	2	3	1	2	3
1	22.0%	78.0%	1.0%	15.0%	84.0%	39.0%	48.0%	13.0%	28.3%	40.4%	31.3%	31.3%
2	36.0%	64.0%	2.0%	31.0%	67.0%	41.0%	49.0%	10.0%	34.0%	46.4%	19.6%	19.6%
3	53.5%	46.5%	46.9%	23.5%	29.6%	47.2%	48.3%	4.5%	39.3%	33.7%	27.0%	27.0%
4	14.0%	87.0%	0.0%	1.0%	99.0%	33.0%	47.0%	20.0%	4.0%	19.0%	77.0%	77.0%
5	14.1%	85.9%	3.0%	5.1%	91.9%	31.3%	48.5%	20.2%	6.2%	38.1%	55.7%	55.7%
6	16.0%	84.0%	1.0%	21.0%	78.0%	29.0%	53.0%	18.0%	31.0%	44.0%	25.0%	25.0%
7	17.0%	83.0%	1.0%	11.1%	87.9%	22.4%	49.0%	28.6%	27.4%	46.3%	26.3%	26.3%
8	19.0%	81.0%	4.0%	22.0%	74.0%	24.5%	50.0%	25.5%	25.8%	50.5%	23.7%	23.7%
9	28.0%	72.0%	4.0%	35.0%	61.0%	39.4%	41.4%	19.2%	54.2%	33.3%	12.5%	12.5%
10	17.0%	83.0%	3.0%	20.0%	77.0%	31.0%	49.0%	20.0%	37.4%	41.4%	21.2%	21.2%

5.2 文生成機構における利用の構想

以上の結果を、文生成機構における文字表記選択に利用する方法の構想を述べる。分析 B の結果に基づき、各評価項目の最高点(項目①は 2, その他は 3)における割合の上位 10 種類の組み合わせを、図 7 に示すようなリスト形式で並べたデータを用意する(以降「表記効果 DB」と呼ぶ)。この各行は、各品詞に対する表記文字を指定する。

((noun kanji) (particle hiragana) (verb kanji))
 ((noun katakana) (particle hiragana) (verb kanji))
 ((noun kanji) (particle katakana) (verb hiragana))
 ((noun kanji) (particle hiragana) (verb katakana))
 ((noun kanji) (particle hiragana) (verb hiragana))
 ((noun hiragana) (particle hiragana) (verb kanji))
 ((noun kanji) (particle katakana) (verb kanji))
 ((noun kanji) (particle romaji) (verb kanji))
 ((noun kanji) (particle romaji) (verb hiragana))
 ((noun kanji) (particle katakana) (verb katakana))

図 7 読みやすい表記組み合わせの上位順のリスト

このデータを利用して文字表記を自動選択する機構を、3 節で述べた文字表記決定機構の中に組み込む。まず、システム(統合物語生成システム全体あるいは文生成機構)実行時にユーザが、目標とする効果を、4 種類のパラメータ(5.1 節の評価項目に対応)の中からひとつを選択する(例:「字面の美しさ」)。その後、文字表記機構の最初の処理において、表記効果 DB から、選択されたパラメータに対応する項目の順位付けデータ(表記組み合わせ 10 パターン)を獲得し、その中のひとつを任意に選択する。ここで獲得されたデータによって、各品詞の表記文字が決まる。なお、この選択は物語に含まれる各事象に対して毎回行われる。その後、文字表記決定機構における文の各要素に対する文字表記選択処理(3.2 節の処理②)において、上記(2)の処理で選択した規則により特定の文字表記が選択される。

目標として「読みやすさ」を選択した場合のシステム動作をハンドシミュレーションした。「読みやすさ」を選択すると、図 7 に示した順位付けデータが得られる。この中から一文ごとに無作為に選択し、言語表記辞書に基づく表記変換を手作業で行った結果、図 8 に示す文が得られた。入力としたのは、4 節で紹介した「KOSERUBE」によるひとつの生成文である。

鬼の館デ親ガ病デしぬ。鬼の館デ河童ガガナシム。波板海岸デ赤鬼ガ山の美神を捜す。波板海岸デ山猫ガ山の美神を赤鬼にオシエル。江刺藤原の郷で赤鬼ガトプ。江刺藤原の郷 de 赤鬼 ga 河童 wo 追う。いしがみのおかでかっぱがきゅうでんを建てる。石神の丘で河童が宮殿に住む。イギリスカイガンでアカオニがヤマノカミに謝罪する。イギリス海岸 de 山の神 ga 赤鬼 wo 拘置所 ni とじこめる。石神の丘で河童が花嫁を山の神から授かる。

図 8 ハンドシミュレーションによる出力(1)(省略有)

全体としては、一文毎に「読みやすさ」が異なるパターンが混在しているため、煩雑で不統一な印象を与える。逆に、ひとつの物語の中で文字表記のパターンが変化する点やカタカナやローマ字が混ざる等の不自然な点には面白さを感じられる。今回の調査は一文毎に受け手の評価を得るものであったが、文のまとまり全体としての印象も重要であり、今後の検討課題となる。最も「読みやすさ」の評価が高いパターンだけを使用すると次のようになる(図 9)。

鬼の館で親が病で死ぬ。鬼の館で河童が悲しむ。波板海岸で赤鬼が山の美神を捜す。波板海岸で山猫が山の美神を赤鬼に教える。江刺藤原の郷で赤鬼が飛ぶ。江刺藤原の郷で赤鬼が河童を追う。石神の丘で河童が宮殿を建てる。石神の丘で河童が宮殿に住む。イギリス海岸で赤鬼が山の神に謝罪する。イギリス海岸で山の神が赤鬼を拘置所に閉じ込める。石神の丘で河童が花嫁を山の神から授かる。

図 9 ハンドシミュレーションによる出力②(省略有)

6. 作家の作品分析に基づく表記選択

もうひとつの利用法として、特定の作家の文字表記の模倣を考えている。手始めとしてここでは芥川龍之介のひとつの作品の文字表記を分析し、その結果から特定の作家の文字表記を模倣する方法を検討する。

6.1 分析の方法と結果

題材は短編小説「蜘蛛の糸」(芥川龍之介, 1918 年)とし、青空文庫からテキストデータを入力した。フリーソフトウェア「KH Coder」(<http://khc.sourceforge.net/>)を用い、テキストデータから品詞別に単語を抽出した。その後、名詞、動詞、形容詞、形容動詞、副詞、助詞それぞれについて、表記文字毎の個数を数えた。そして、各品詞に対する表記文字の種類毎の割合を計算した。品詞別の単語抽出を行った結果、名詞 136、動詞 105、形容詞 22、形容動詞 11、副詞 51、助詞が 26 個得られた。品詞毎の文字表記の割合は以下の通りである——名詞:漢字 94.1%, ひらがな 5.9%/動詞:漢字 67.6%, ひらがな, 32.4%/形容詞:漢字 81.8%, ひらがな, 18.2%/形容動詞:漢字 90.9%, ひらがな 9.1%/副詞:漢字 29.4%, ひらがな 70.6%/助詞:漢字 100.0%, ひらがな 0%。(カタカナとローマ字は使用されていない。)

6.2 文生成機構における利用の構想

単純に、文表現で使用する文字表記の割合を、上の分析結果から得られた割合と同程度にすることで、「蜘蛛の糸」風の文字表記が生成可能と考え、次の方法を考案した。

①文生成機構の入力となる物語の概念構造全体に含まれる、名詞概念及び動詞概念に対応する要素の総数を計算する。

②分析結果の割合に基づき、名詞と動詞の各要素に対する文字表記(漢字及びひらがな)の使用数を下の式により計算する(何れも少数第一位は四捨五入)。

- ◇ [名詞|動詞]の漢字使用数 =
[名詞|動詞]の漢字割合 × [名詞|動詞]の要素数
- ◇ [名詞|動詞]のひらがな使用数 =
[名詞|動詞]の要素数 - [名詞|動詞]の漢字使用数

③計算で示された要素の数だけ交互に、各言語表記辞書を参照し、漢字とひらがなの表記に変換する。この処理は事象概念の数だけ繰り返す。5.2 節と同じ物語を題材として、上記方法のハンドシミュレーションを行った(図 10)。

- ①物語に含まれる要素の数を計算 → 名詞が 53 個、動詞が 17 個
- ②各表記で表す要素の数を計算する(小数点第一位を四捨五入)

漢字で表す概念数

- 名詞: $94.1\%(\text{漢字の使用率}) \times 53(\text{名詞の数}) = 50$
- 動詞: $67.6\%(\text{漢字の使用率}) \times 17(\text{動詞の数}) = 11$

ひらがなで表す概念数

- 名詞: $53(\text{名詞の数}) - 50(\text{漢字表記になる要素数}) = 3$
- 動詞: $17(\text{動詞の数}) - 11(\text{漢字表記になる要素数}) = 6$

- ③言語辞書を参照し、計算結果の要素数だけ交互に表記を変換する。

((location 鬼の館)(p で))((agent おや)(p が))((instrument 病)(p で))((v 死ぬ))

図 10 ハンドシミュレーションの処理経過

ハンドシミュレーションで生成した文と「蜘蛛の糸」を比較する(図 11, 図 12)。目で見た印象では異なる印象を与える文となった。しかし、上記の計算は「蜘蛛の糸」全体を対象としたものであり、すべての部分でその結果になったとは考えにくい。そこで改めて、生成された文と同程度の名詞と動詞が記述されている箇所を「蜘蛛の糸」から選び、その部分についてだけ上記の方法で計算した結果に基づく、「蜘蛛の糸」と物語生成システムによる文字表現を、それぞれ図 13 と図 14 に示す。なお、「蜘蛛の糸」で対象とした部分には名詞は 32、動詞は 17 個記述されており、文字表記の割合は名詞:漢字 93.5%, ひらがな 6.5%, 動詞:漢字 76.9%, ひらがな 23.1%である。また、印象が多少異なる原因の一つは、「蜘蛛の糸」には形容詞や

形容動詞や副詞や敬語が多くの場合ひらがなで記されているが、生成された文にはそれらの品詞等がないためと思われる。

すると、一生懸命にのぼった甲斐があって、さきまで自分がいた血の池は、今ではもう暗の底にいつの間にかかくれて居ります。それからあのぼんやり光っている恐い針の山も、足の下になってしまいました。この分でのぼって行けば、地獄からぬけ出すのも、存外わけがないかも知れません。健陀多は両手を蜘蛛の糸にからみながら、こへ来てから何年にも出した事のない声で、「しめた。」と笑いました。

図 11 「蜘蛛の糸」の一部抜粋

鬼の館でおやが病で死ぬ。おにのやかたで河童がかなしむ。なみいたかいがんで赤鬼が山の美神を捜す。波板海岸で山猫が山の美神を赤鬼におしえる。江刺藤原の郷で赤鬼が飛ぶ。江刺藤原の郷で赤鬼が河童をおう。江刺藤原の郷で河童が鉄器を置く。江刺藤原の郷で河童が赤鬼からにげる。イギリス海岸で河童がイギリス海岸に到着する。イギリス海岸で勇者が報酬を山の神によきゆうする。

図 12 システムにより生成された文(省略有)

ある日の事でございます。御釈迦様は極楽の蓮池のふちを、独りでぶらぶら御歩きになっていらした。池の中に咲いている蓮の花は、みんな玉のようにまっ白で、そのまん中にある金色の蕊からは、何とも云えない好い匂が、絶間なくあたりへ溢れて居ります。極楽は丁度朝なのでございましょう。やがて御釈迦様はその池のふちに御佇みになって、水の面を蔽っている蓮の葉の間から、ふと下の容子を御覧になりました。この極楽の蓮池の下は、丁度地獄の底に当って居りますから、水晶のような水を透き徹して、三途の河や針の山の景色が、丁度覗き眼鏡を見るように、はっきりと見えるのでございます。

図 13 「蜘蛛の糸」より対象とする部分

鬼の館でおやが病で死ぬ。おにのやかたで河童がかなしむ。なみいたかいがんで赤鬼が山の美神を捜す。波板海岸で山猫が山の美神を赤鬼におしえる。江刺藤原の郷で赤鬼が飛ぶ。江刺藤原の郷で赤鬼が河童をおう。江刺藤原の郷で河童が鉄器を置く。江刺藤原の郷で河童が赤鬼からにげる。イギリス海岸で河童が英吉利海岸に到着する。英吉利海岸で勇者が報酬を山の神に要求する。英吉利海岸で山の美神が真実を語る。英吉利海岸で勇者の嘘偽りが露見する。石神の丘で河童が宮殿を建てる。石神の丘で河童が宮殿に住む。英吉利海岸で赤鬼が山の神に謝罪する。英吉利海岸で山の神が赤鬼を拘置所に閉じ込める。石神の丘で河童が花嫁を山の神から授かる。

図 14 システムにより生成された文

7. おわりに

統合物語生成システムのための文生成機構の現状を、拡張した言語表記辞書の利用を加えて整理し、応用システム KOSERUBE における利用も含め、得られた課題をまとめた。物語言説における「態」に相当する語り手と聴き手に基づく制御の必要性についても述べた。新たに付加することを計画している文字表記機構へのアプローチを述べた。すなわち、文字表記が受け手に与える効果に関する調査と特定の作家の小説における文字表記の分析を行い、受け手に特定の効果を与える文字表記を用いた文の生成方針及び特定の作家風の文字表記を模倣する方法について考察した。

参考文献

- Akimoto, T. and Ogata, T. (2012). "Macro Structure and Basic Methods in the Integrated Narrative Generation System by Introducing Narratological Knowledge." *Proc. of the 11th IEEE International Conference on Cognitive Informatics & Cognitive Computing*, pp.253-262.
- Genette, G. (1972). *Discours du récit, essai de méthode, Figures III*. Paris: Seuil. (花輪光, 和泉涼一 訳 (1985). 物語のディスクール. 水声社.)
- Imabuchi, S. and Ogata, T. (2012). "A Story Generation System based on Propp Theory: As a Mechanism in an Integrated Narrative Generation System." *Lecture Notes in Artificial Intelligence 7614* (Isahara, H. and Kanzaki, K. Eds.), pp.312-321. Springer-Verlag.
- 今淵祥平, 小野淳平, 遠藤順, 栗澤康成, 道又龍介, 鎌田まみ, 秋元泰介, 小方孝 (2012).『いわての民話 KOSERUBE』—プロップによるストーリー生成システムをベースに文・音楽・視覚表現の生成を統合したシステム—, 日本認知科学会第 29 回大会発表論文集, pp.733-742.
- 熊谷真哉, 船越宗, 秋元泰介, 小方孝 (2012). 言語辞書の構築と簡易物語文生成機構. 人工知能学会全国大会(第 26 回)論文集, 1J1-OS2-8.
- 小方孝, 金井明人 (2010). 物語論の情報学序説—物語生成の思想と技術を巡って—. 学文社.
- Oishi, K., Kurisawa, Y., Kamada, M., Fukuda, I., Akimoto, T. and Ogata, T. (2012). "Building Conceptual Dictionary for Providing Common Knowledge in the Integrated Narrative Generation System." *Proc. of the 34th Annual Conference of the Cognitive Science Society*, pp.2126-2131.
- 上田浩史, 小方孝 (2004). 視点と態による物語言説の多様性. 人工知能学会全国大会(第 18 回)論文集, 2D1-05.