

# 物語文章における文字表記の分析と模倣

## Analysis and Imitation of the Japanese Letter Notation in Narrative Texts

鎌田 まみ<sup>\*1</sup>  
Mami Kamada

小方 孝<sup>\*2</sup>  
Takashi Ogata

<sup>\*1</sup> 株式会社電子工学センター  
Electronic Engineering Center

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

One of the characteristics of Japanese language is to have several types of character and techniques of their use can show each author's different thought and make various effects in readers. Focusing on these points, we are trying to prepare a letter notation management mechanism in our integrated narrative generation system. In this paper, we present the results of an investigation of the effects to readers by different letter notations and an analysis of letter notations in actual novels & their imitations by a sentence generation mechanism to abstract experimentally important elements relating to this issue.

### 1. まえがき

日本語の表記の大きな特徴は、漢字、ひらがな、カタカナ、ローマ字という異なる文字種類が使用されていることである。ある書かれたテキストを見た時、我々はその表記自体から、ある印象を、内容とは別の独立のものとして、受ける。あるいはまた、文字表記がその意味的側面にも影響を及ぼしていることを感じる。

日本語の文字表記を自然言語生成において扱おうとする場合、まず現象的な分析に基づくアプローチが可能であろう。ある特定の対象テキストを決め、その中で、漢字、ひらがな、カタカナ、ローマ字が出現する割合を分析し、それを模倣すれば、対象テキストと類似した印象をもたらす見聞を実現することが出来ると思われる。さらに精密にするためには、対象テキストにおける品詞ごとの文字表記の割合を分析し、それを模倣すれば良い。特定の作家の特定の作品乃至その部分と類似した見聞を持つ、別のテキストを作り出すことも可能だろう。(これに対して、非現象的な方法、すなわち、文字表記のあり方を現実のテキストにおける実態に求めるのではなく、論理的に探求された可能性に求めるという方法もある。)

しかしながら、どの品詞のどの語を、どの文字種類で表記するか、ということの中には、重要な問題が潜んでいる。例えば小説の中で、通常は漢字で表記されているべき文字がカタカナで表記されていたりすると、そのカタカナの言葉は、何らかの意味で、小説の中で重要な意味を担っていると思わせる。

一方、[村越 1998]は、日本語において、第一次的な文字表記と第二次的な文字表記とを区別する。前者は、筆者らの言葉では制度的な文字表記であり、人が普通それに従属する規範的な文字表記のあり方である。それに対して後者は、制度を意識化し、その規範から逸脱した文字表記のあり方であり、文字表記における狭い意味での修辞の領域に重なる。上述のカタカナや、三島由紀夫や大江健三郎に見られるような一種の「過剰」は、文字表記における修辞性の強い発動を意味すると考えられる。

上の現象的分析の中に、修辞的な分析を付加することの可能性を探ることは、本研究における一つの発展した主題になり得る。それは直接的には、あるテキストの文字表記の中に、「規範」からの「逸脱」の部分を探ることである。さらに、受容と使用の二つの立場から、その印象と効果を考察し、文生成システムの中に一種の戦略として組み込むことである。

この規範と逸脱の問題系は、単に文字表記のみに留まらず、筆者らの物語生成システム全体における、最も重要なメタレベ

連絡先: 鎌田まみ, 株式会社電子工学センター東京技術部, 東京都千代田区岩元町 2-1-3 和光ビル 9階, g031gmiri@yahoo.co.jp

ルの戦略指針に相当する。

以上に、文字表記を巡る筆者らの問題意識を述べた。上述のように、この主題は、「現象的方法と非現象的方法」や「規範と逸脱」のような、文字表記を超えた物語生成システム全般に敷衍すべきコンセプト乃至戦略と関連する。

本稿では以上を引き継ぐ考察は割愛し、物語生成システムにおける文字表記の調節を巡る現在の試みの一つを紹介する。特に、作家の文字表記の模倣による機構に向けた最初の試行を示す。

### 2. 文生成機構の試作

統合物語生成システム[小方 2010][Akimoto 2012]は、概念構造と表層表現それぞれの生成機構を含み、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。文生成機構全体の処理手順を図 1 に示す。Common Lisp で実装されている。

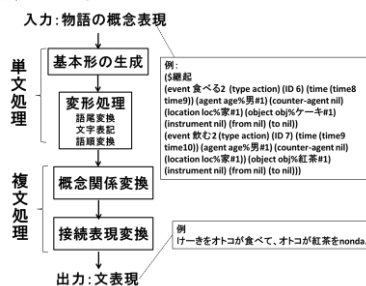


図 1 文生成機構の処理の流れ

物語の概念構造は、事象概念を終端節点とし、それらに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される。個々の事象概念は、動詞概念と深層格からなる格構造形式で記述される。深層格には主に time, agent, counter-agent, object, instrument, location, from, to の 8 種類がある。動詞概念、名詞概念共にそれぞれの概念辞書[Oishi 2012]により提供され、各深層格の値としては、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 の入力概念構造の例では、「(loc%家#1 で) age%少年#1 が obj%ケーキ#1 を 食べる」及び「(loc%家#1 で) age%少年#1 が obj%紅茶#1 を 飲む」という二つの事象が、「継起」の関係で結合されている。

文生成は、単文処理と複文処理からなり、前者は、以下の処理により、概念構造に含まれる個々の事象概念を単文(文字列)に変換する一まず、一事象概念を、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に変換する。これは文節ごとに区切られたリスト形式で表現される。次に、この単文構造に対して、語尾変換、文字表記変換、語順変換を施す。こうして出来た単文構造を文字列に変換し、それを概念構造中のもとの事象概念と置換する。これらの処理を入力構造におけるすべての事象概念に対して行う。

この結果を入力として複文処理を続ける。入力構造における中間節点(概念的関係)に応じて適当な接続規則を選択・適用し、その出力が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。具体的には、入力に含まれる各概念的関係を言語的關係(「順接」「逆接」等)に変換し、言語的關係で結ばれた二つの文に対して、その関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、木構造の最下層から最上層まで行い、最終的に一つの文字列として結合された物語文を出力する。

### 3. 文字表記決定機構の試み

文字表記決定機構は、構築した文字表記辞書を利用して上記単文構造における概念名や助詞の表記を決定する。言語表記辞書は、概念辞書に登録されたすべての動詞概念と名詞概念、一部の形容詞及び形容動詞概念、40種類の助詞に対応する漢字、ひらがな、カタカナ、ローマ字(ヘボン式)の4種類の表記を登録している。ひらがなとカタカナの読みは新仮名使いのみとする。詳細は[熊谷 2012]に示される。なお「ローマ字」という概念は日本語をラテン文字で表記するための特殊な方式であり、文字の類別としてはラテン文字とした方が良いと考えられるが、以下の試みはローマ字という言葉を使って行っており、ここではローマ字を踏襲する。

「まえがき」で、非現象的方法という言葉を使った。現象的方法が、現実の世界に存在する(この場合)文字表記の様式を、主に統計的に分析して利用する方法だとすれば、ここでの非現象的方法とは、文字表記の様式を現実世界における存在とは関わらずに措定し、利用する方法であり、学問的方法であると言うよりは、多分に芸術的・文学的方法であると言える。

このような意味での文字表記のための非現象的方法に分類される試行として、漢字・ひらがな・カタカナ・ローマ字の四種類の文字表記と品詞との間の可能な組み合わせを予め用意し、受け手にもたらされる効果の仮定によりその使用を決定する方式を試みた。受け手における効果の仮定作成は実際の人間による結果に基づいているので、純粋な意味での非現象的方法とは言えないが、この分類が現実に出現する様式に基づいたものではないという意味で、そのような傾向が強い。

現状の文生成機構の単文基本形生成により生成される文は、名詞、動詞、助詞の三種類の品詞のみからなる。そこで、これらの品詞毎に四種類の文字表記(漢字、ひらがな、カタカナ、ローマ字。但し助詞は漢字を除く三種類)の何れかを指定すると、37通りの組み合わせ(文字表記パターンと呼ぶ)が得られる。例えば、名詞・動詞を漢字、助詞をひらがなとすれば、「河童が冒険へ出立する」のようになり、名詞を漢字、動詞をカタカナ、助詞をひらがなとすれば、「河童が冒険へシュツツスル」のようになる。

上の各文字表記パターンが受け手に与える効果を調べるために、次のような実験を行った(詳細は[鎌田 2013])。まず、実験に用いる例文として、10種類の事象概念の各々から単文基本形を生成し、それぞれ上述の37種類の文字表記パターンに

よる変換を行った。すなわち、ひとつの事象概念から37種類の例文が生成され、合計370の例文が得られる。そして、各例文に対して被験者が、「読みやすさ」「文の意味の理解度」「字面の美しさ」「字面の面白さ」という4種類の項目を採点する。「読みやすさ」は2段階、それ以外は3段階の選択式である。実験は、例文を横書きで表示した場合と縦書きで表示した場合に分けて行い、それぞれ10名の被験者(重複なし)が行った。

評価項目(効果)毎に、各文字表記パターンにおける得点を順位付けし、これを利用して、ユーザが指定した要求に応じて文字表記パターンを選択するプログラムを作成した。図2に示す物語を入力とした生成例を幾つか示す。なお、4節で示す各生成例でもこの物語を入力として使用する。

親が病で死ぬ。カッパが悲しむ。踊子が沢蟹を捜す。沢蟹が噂を漏らす。玉子がカッパに欠如する。カッパが玉子で悩む。カッパの頭を抱える。詩人がカッパを旅へ連れて行く。カッパが冒険に備える。カッパが冒険へ出立する。カッパが御所野縄文公園に向かう。カッパが猿に気付く。カッパが猿を助ける。カッパが猿の希望に応じる。カッパが炭を猿から盗む。カッパが地下道へ降りる。カッパが敵地に向かう。踊子が真実を語る。狐師の嘘偽りが露見する。カッパが宮殿を御所野縄文公園に建てる。カッパが宮殿に住む。詩人が沢蟹を容赦する。詩人が狐師を容赦する。カッパが踊子と結婚する。

図2 入力した物語(概念構造を基本文に変換したもの)

図3は、「字面の美しさ」を高めるように指定した生成例である。「名詞:漢字、動詞:漢字、助詞:ひらがな」という文字表記パターンが選択された。かなり規範的な表記になっていると思われる。

親が病で死ぬ。河童が悲しむ。踊子が沢蟹を捜す。沢蟹が噂を漏らす。玉子が河童に欠如する。河童が玉子で悩む。河童の頭を抱える。詩人が河童を旅へ連れて行く。河童が冒険に備える。河童が冒険へ出立する。河童が御所野縄文公園に向かう。河童が猿に気付く。河童が猿を助ける。河童が猿の希望に応じる。河童が炭を猿から盗む。河童が地下道へ降りる。河童が敵地に向かう。踊子が真実を語る。狐師の嘘偽りが露見する。河童が宮殿を御所野縄文公園に建てる。河童が宮殿に住む。詩人が沢蟹を容赦する。詩人が狐師を容赦する。河童が踊子と結婚する。

図3 「字面の美しさ」を高く指定した生成例

一方、図4は「字面の面白さ」を高めるように指定した生成例である。「名詞:漢字、動詞:ローマ字、助詞:ひらがな」という文字表記パターンが選択され、かなり逸脱的な表現が得られた。

親が病で sinu。カッパが kanasimu。踊子が沢蟹を sagasu。沢蟹が噂を morasu。玉子がカッパに ketuzyosuru。カッパが玉子に nayamu。カッパの頭を kakaeru。詩人がカッパを旅へ tureteiku。カッパが冒険に sonaeru。カッパが冒険へ syuttatusuru。カッパが御所野縄文公園に mukau。カッパが猿に kiduku。カッパが猿を tasukeru。カッパが猿の希望に ouziru。カッパが炭を猿から nusumu。カッパが地下道へ oriru。カッパが敵地に mukau。踊子が真実を kataru。狐師の嘘偽りが rokensuru。カッパが宮殿を御所野縄文公園に tateru。カッパが宮殿に住む。詩人が沢蟹を yousyasuru。詩人が狐師を yousyasuru。カッパが踊子と kekkonsuru。

図4 「字面の面白さ」を高く指定した生成例

図4のような文字表記が実際に人間が書いた物語や文章の中に現れることは非常に稀であり、すなわち現象的な分析から、このような結果が得られることは殆どないと考えられる。あるとすれば、詩のような通常の言語使用とは異なる原理による性格が強いテキストの場合であろう。このような非現象的方法の追求は、文字表記のいわば表現主義的な方向の追及として、意味があると思われる。

### 4. 文学作品の文字表記の模倣の考察

一方、現象的方法についての考察も行っている。ここに示すのは、文学作品における文字表記の傾向を模倣する方法へのアプローチである。特定の文学作品(小説)における文字表記の特徴を、物語生成システムの出力結果で模倣する方法について模索する。ここで示すのは、作品中に現れる単語の文字表記の割合を品詞別に算出し、その割合を模倣して文字表記選択を行う方法である。

#### 4.1 分析方法及び実験用文字表記変換プログラム

作品の分析方法を述べる。まず、テキスト分析ソフト KH Coder (<http://khc.sourceforge.net/>)を用いて特定の小説の一部のテキストデータから品詞別に単語を抽出する(品詞毎に単語

が分類される。また各単語の出現回数も得られる。現状の文生成機構が扱う品詞は、名詞、動詞、形容詞、形容動詞、助詞の五種類であるため、それ以外の品詞は「その他」に一括する。各品詞に含まれる単語数を数え(同一単語が作品中に複数回現れる場合も1と数える)、それをもとに品詞毎に文字表記の割合を算出する。

次に、この割合を利用して文字表記変換を行う実験用の文生成プログラムについて説明する。物語の概念構造及び各品詞の文字表記割合(各品詞につき、漢字、ひらがな、カタカナ、ローマ字の各文字表記の割合を百分率(%)で指定)を入力し、次の方法で文字表記変換及び文生成を行う。まず、入力された概念構造に含まれる各事象概念を単文基本形に変換し、その結果に含まれる全単語を品詞毎に分けてリストアップする。そして、入力の文字表記割合に基づき、次の方法により各単語に種類の文字表記を割り当てる。まず、文字表記割合に従い、各品詞(に含まれる単語)における各文字表記の使用数を、「品詞に含まれる単語総数 × 各文字表記割合」(小数点以下四捨五入)という計算式により決定する。例えば、物語中に10種類の名詞が含まれ、名詞の文字表記の割合が「漢字:50%、ひらがな:50%、他 0%」と指定された場合、5つの単語を漢字、残る5つをひらがな、とする。この結果(文字表記毎の使用数)に従って、各単語にひとつの文字表記を割り当てる(どの単語にどの文字表記を割り当てるかはランダムで決定)。この結果に従い文字表記変換処理を行い、文を出力する。

以上の分析方法及びプログラムを用いて、4つの作品を対象に分析・実験を行った。対象作品は、「蜘蛛の糸」(芥川龍之介、1918年、青空文庫)、「暁の寺」(三島由紀夫、1970年、新潮社)、「abさんご」(黒田夏子、2013年、『文藝春秋一月号』)、「みずから我が涙をぬぐいたまう日」(大江健三郎、1972年、講談社)である。そのうち「蜘蛛の糸」と「みずから我が涙をぬぐいたまう日」による結果を以下に示す。

#### 4.2 「蜘蛛の糸」の模倣

図5に、「蜘蛛の糸」の分析対象とした部分を示す。表1は、その分析結果である(各品詞について、文字表記毎の単語数及び割合を表す)。なお、表1中の角括弧で示した値は、「その他」の品詞に分類される単語の文字表記を反映させるための工夫として、それらをシステムが扱う五種類の品詞に割り振って算出した値である。各文字表記について、品詞毎の単語数の割合を計算し、その割合に応じてその他に含まれる単語を配分した。次に述べる生成例は、この補正後の値を用いたものである。

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っしゃいました。池の中に映っている蓮の花は、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。するとその地獄の底に、健陀多と云う男が一人、ほかの罪人と一しょに蠢いている姿が、御眼に止まりました。この健陀多と云う男は、人を殺したり家に火をつけたり、いろいろ悪事を働いた大泥坊でございますが、それでもたった一つ、善い事を致した覚えがございます。と申しますのは、ある時この男が深い林の中を通りまると、小さな蜘蛛が一匹、路ばたを這って行くのが見えました。そこで健陀多は早速足を挙げて、踏み殺そうと致しましたが、「いや、いや、これも小さいながら、命のあるものに違いない。その命を無暗にとると云う事は、いくら何でも可哀そうだ。」と、こう急に思い返して、とうとうその蜘蛛を殺さず助けてやったからでございます。御釈迦様は地獄の容子を御覧になりながら、この陀多には蜘蛛を助けた事があるのを御思い出しになりました。そうしてそれだけの善い事をした報には、出来るなら、この男を地獄から救い出してやろうと御考えになりました。幸い、側を見ますと、翡翠のような色をした蓮の葉の上に、極楽の蜘蛛が一匹、美しい銀色の糸をかけて居ります。御釈迦様はその蜘蛛の糸をそっと御手に御取りになって、玉のような白蓮の間から、遙か下にある地獄の底へ、まっすぐそれを御下しなさいました。

図5 「蜘蛛の糸」の一部

表1 「蜘蛛の糸」の分析結果(品詞別の文字表記割合)

	名詞	動詞	形容詞	形容動詞	助詞	その他
漢字	128(94.1) [162(90.0)]	71(67.6) [90(53.9)]	18(81.8) [23(71.9)]	10(90.9) [13(86.7)]	0	60(39.2)
ひらがな	8(5.9) [18(10.0)]	34(32.4) [77(46.1)]	4(18.2) [9(28.1)]	1(9.1) [2(13.3)]	26(100.0) [59(100.0)]	93(60.8)
カタカナ	0	0	0	0	0	0
ローマ字	0	0	0	0	0	0
総単語数	136[180]	105[167]	22[32]	11[15]	26[59]	153

角括弧「[]」:補正後の値/丸括弧「()」:割合(%)

図6は、表1に示した文字表記の割合を用いて表現した生成文である。元のテキストと生成結果それぞれの文字表記の割合を文字単位で算出すると表2のようになる。なお、表中の「その他」は句読点や括弧等の記号である。ひらがなで約5%の差が出た以外は、おおむね近い割合となった。

親が病で死ぬ。かっぱがかなしむ。かっぱは怖い。踊子が沢蟹を獲す。踊子は活発だ。沢蟹が噂を漏らす。沢蟹が結婚をよげんする。玉子がかっぱに欠如する。かっぱが玉子に悩む。かっぱが頭を抱える。詩人がかっぱの旅へ連れて行く。詩人は心が優しい。かっぱが冒険にそなえる。かっぱが冒険へ出立する。かっぱが御所野縄文公園にむかう。かっぱが猿に気付く。かっぱが猿を助ける。かっぱが猿の希望に公応する。かっぱが炭を猿よけぬすむ。かっぱが地下道へ降りる。かっぱが敵地にむかう。踊子が真実をかたる。狐師のうそいつわりがるけんする。狐師はたくましい。かっぱが宮殿を御所野縄文公園に建てる。かっぱが宮殿にすむ。詩人が沢蟹をよろしやる。詩人が狐師をよろしやる。かっぱが踊子とけっこんする。

図6 表1の文字表記割合に基づく生成例

表2 文字単位での各文字表記の個数及び割合

	漢字	ひらがな	カタカナ	ローマ字	その他	総数
元テキスト (図5)	242 (30.1%)	503 (66.5%)	0 (0%)	0 (0%)	58 (7.2%)	803
生成結果 (図6)	98 (29.8%)	202 (61.4%)	0 (0%)	0 (0%)	29 (8.8%)	329

#### 4.3 「みずから我が涙をぬぐいたまう日」の模倣

同様に、「みずから我が涙をぬぐいたまう日」を題材とした結果を示す。図7は分析対象とした部分である。後半からカタカナ交じりの文になっているという特徴がある。表3にこの分析結果を示す。

はじめ海をわたって行った機会にすぐさまあの人が見え、青年村長時代からの妻を離別すると、正式に結婚して谷間につれ戻った女だということだ。あの人には母親を終生森の奥に縛りつける算段をすませると、自分はまたあらためて中国、満州国へと出かけてはそこに永く滞在して、なにごとか活動したんだ。そこで当然に母親の電報は養父の側の手づるにむけられたのだったろう。そしてついに倉のなかで、あの人と母親とが高い声をあげて口論するということが、はじめておこった。それも、おれがその後の谷間の噂のたすけをかりてやっと思ひあいをくつきりさせながら、あらためて記憶しなおしたものだ、母親は、早く向う側ヘタリツカナクレバ殺サレ、と、泣いて泣くのだった。あの方は激怒して、ナニヲイカ、国体ヲ国体トモ思ワス者ノ血ヲヒタイ、貴様ゴトキニ育テサセタメニコウナツタ、イマハ早く射殺サレテ、ナントカ遺骨ダケデモ戻ルヨウ、戦死アツカインシテモラオウト苦慮シテルノダゾ、と怒鳴りかえす。アナタハ、アノ子ガ向ウヘタリツクマニ殺サセヨウトシテルノデスカ！アナタハ自分ノ子供ガ、\* \* \* \* \*ノユニ下賤ナ権力家ノ命令ガ、ウシロカラ撃チ殺サレルヨウ、頼ムタメニ電報ヲウツテルノデスカ、アノ子ガ向ウヘタリツクヨウトシテヒトリー所懸命、走ッテイル、ソノウシロカラ撃タセルノ！と母親がなお涙まじりに叫びたてている、という具合なのだ。

図7 「みずから我が涙をぬぐいたまう日」の一部

表3 「みずから我が涙をぬぐいたまう日」の分析結果(品詞別の文字表記割合)

	名詞	動詞	形容詞	形容動詞	助詞	その他
漢字	47(92.2) [52(89.7)]	15(45.5) [17(34.7)]	3(100.0) [3(100.0)]	2(100.0) [2(100.0)]	0	7(15.6)
ひらがな	1(1.9) [2(3.4)]	13(39.4) [25(51.0)]	0	0	17(56.7) [32(62.7)]	28(62.2)
カタカナ	3(5.9) [4(6.9)]	5(15.1) [7(14.3)]	0	0	13(43.3) [19(37.3)]	10(22.2)
ローマ字	0	0	0	0	0	0
総単語数	51[58]	33[49]	3[3]	2[2]	30[51]	45

角括弧「[]」:補正後の値/丸括弧「()」:割合(%)

表3に示した文字表記の割合を用いて表現した生成文を図8に示す。また、表2と同じように文字単位で文字表記の割合を比較した結果を表4に示す。漢字やカタカナで6%前後の差が出た。一方、見た目の印象では、生成文の方がカタカナが少ない印象を受ける。元のテキストは後半にカタカナが多用され集中する特徴的な文字表記となっており、その部分の印象が強い

ためであると考えられる。また、「向コウ」や「頼ム」のように、漢字とカタカナを組み合わせた表記が現在の文字表記辞書に登録されていないことも関連すると考えられる(文字表記辞書の拡張という課題が生じる)。

親が病で死ぬ。カッパがカナシム。カッパハ怖い。踊子が沢蟹を捜す。踊子ハ活発だ。沢蟹が噂を漏らす。沢蟹が結婚をよげんする。玉子がカッパにけつじよする。カッパが玉子になやむ。カッパが頭をかかえる。詩人がカッパをたびへ連れて行く。詩人ハ心が優しい。カッパが冒険にそなえる。カッパが冒険へしゅつたつする。カッパが御所野縄文公園にむかう。カッパがサルに気付く。カッパがサルをたすける。カッパがサルの希望に応じる。カッパが炭をサルヨリぬすむ。カッパが地下道へオトリル。カッパが敵地にむかう。踊子が真実をかたる。狐師の嘘偽りがロケンスル。狐師ハ遅しい。カッパが宮殿を御所野縄文公園に建てる。カッパが宮殿に住む。詩人が沢蟹をようしやる。詩人が狐師をようしやる。カッパが踊子と結婚する。

図 8 表 3 の文字表記割合に基づく生成例

表 4 文字数単位での各文字表記の個数及び割合

	漢字	ひらがな	カタカナ	ローマ字	その他	総数
元テキスト (図 7)	135 (22.7%)	237 (39.8%)	184 (30.9%)	0 (0%)	39 (6.6%)	595
生成結果 (図 8)	91 (27.2%)	134 (40.0%)	81 (24.2%)	0 (0%)	29 (8.6%)	335

分析対象(図 7)は、前半部分と後半部分で文字表記の傾向が異なるため、図 7 の 9 行目、「～とって泣くのだった。」までを前半、それ以降(「あの方は激怒して、ナニヲイウカ、～)を後半として区切って同様の分析を行い、その結果を用いて生成を試みた。表 5 と表 6 にそれぞれ前半部分と後半部分の分析結果を示す。図 9 は表 5 の割合を用いた生成例であり、図 10 は表 6 の分析結果を用いた生成例である。

表 5 図 7 前半の品詞別文字表記割合

	名詞	動詞	形容詞	形容動詞	助詞	その他
漢字	30(100.0) [34(100.0)]	7(38.9) [8(34.0)]	3(100.0) [3(100.0)]	1(100.0) [1(100.0)]	0	5(38.5)
ひらがな	0	10(55.6) [13(56.5)]	0	0	15(93.8) [19(90.5)]	7(53.8)
カタカナ	0	1(5.5) [2(8.7)]	0	0	1(6.2) [2(9.5)]	1(7.7)
ローマ字	0	0	0	0	0	0
総単語数	30[34]	18[23]	3[3]	1[1]	16[21]	13

角括弧“[]”:補正後の値/丸括弧“( )”:割合(%)

表 6 図 7 後半の品詞別文字表記割合

	名詞	動詞	形容詞	形容動詞	助詞	その他
漢字	21(77.8) [23(69.7)]	9(47.4) [10(38.5)]	1(100.0) [1(100.0)]	1(100.0) [1(100.0)]	0	3(11.1)
ひらがな	0	4(21.1) [6(23.0)]	0	0	7(28.0) [11(27.5)]	6(22.2)
カタカナ	6(22.2) [10(30.3)]	6(31.5) [10(38.5)]	0	0	18(72.0) [29(72.5)]	18(66.7)
ローマ字	0	0	0	0	0	0
総単語数	27[33]	19[26]	1[1]	1[1]	25[40]	27

角括弧“[]”:補正後の値/丸括弧“( )”:割合(%)

親が病で死ぬ。河童が悲しむ。河童は怖い。踊子が沢蟹を捜す。踊子は活発だ。沢蟹が噂を漏らす。沢蟹が結婚をよげんする。玉子が河童にけつじよする。河童が玉子になやむ。河童が頭をかかえる。詩人が河童を旅へツレテイク。詩人は心が優しい。河童が冒険にそなえる。河童が冒険へシユツツスル。河童が御所野縄文公園にむかう。河童が猿に気付く。河童が猿を助ける。河童が猿の希望におうじる。河童が炭を猿からぬすむ。河童が地下道へおとり。河童が敵地にむかう。踊子が真実をかたる。狐師の嘘偽りが露見する。狐師は遅しい。河童が宮殿を御所野縄文公園に建てる。河童が宮殿に住む。詩人が沢蟹をようしやる。詩人が狐師をようしやる。河童が踊子と結婚する。

図 9 表 5 の文字表記割合に基づく生成例

親ガ病デ死ぬ。カッパガかなしむ。カッパハ怖い。踊子ガサワガニヲサガス。踊子ハ活発だ。サワガニガウヲサヲ漏らす。サワガニガ結婚ヲよげんする。玉子ガカッパニケツジヨスル。カッパガ玉子ニヤム。カッパガ頭ヲカカエル。詩人ガカッパヲタビヘ連れて行く。詩人ハ心が優しい。カッパガボウケンニ備える。カッパガボウケンヘ出立する。カッパガゴシヨノジヨウモンコウエンニムカウ。カッパガ猿ニキツク。カッパガ猿ヲタスケル。カッパガ猿ノ希望ニ応じる。カッパガ炭ヲ猿ヨリ盗む。カッパガ地下道ヘおとり。カッパガ敵地ニムカウ。踊子ガ真実ヲ語る。狐師ノ嘘偽リガロケンスル。狐師ハ遅しい。カッパガ宮殿ヲゴシヨノジヨウモンコウエンヘタテル。カッパガ宮殿ニム。詩人ガサワガニヲ容赦する。詩人ガ狐師ヲ容赦する。カッパガ踊子ト結婚する。

図 10 表 6 の文字表記割合に基づく生成例

前半と後半を分けた分析結果を利用することによって、特に後半部分の結果(図 10)で元のテキストにかなり類似した印象を

与える特徴的な表現が得られた。但し、元の文章では比較的主として・集中して現れているひらがなが分散して現れることから異なる印象も生れている。さらに近付けるためには、このような文字表記毎の集中・分散の程度も考慮する必要がある。何れにせよ、模倣を目標とする場合は、同じ一つの作品と言っても、文字表記出現の特徴が異なる部分毎に分割した分析が必要となる場合がある。

## 5. あとがき

統合物語生成システムの中の文生成機構に新たに付加することを計画している文字表記調節機構のためのアプローチについて述べた。具体的には、幾つかの小説における文字表記の分析を行い、特定の作品風の文字表記を模倣する方法を試み、生成文を題材文と比較して評価・考察を行った。

さらに模倣の精度を高めるために考慮しなければならないことが幾つか明らかになった。分析及び生成の両方において文表現の要素をより精細に扱うことが必要であるが、その中には、上述した対象テキストの特徴に基づく分割、品詞毎の出現の集中や分散の程度その他、文自体の長さや句読点の数、含まれる品詞の種類等が含まれる。また、品詞等の文要素を区別せずに、単純に文字単位で各文字表記の量が一致するように選択を行うということも試みたが[鎌田 2013]、この場合も品詞毎の文字表記と組み合わせる必要が生じる。また、特定の単語自体のその作品中での意義や重要性がその文字表記と関連している場合もあり、そうすると単語という意味的側面の問題との関連も考慮する必要が生じて来る。規範と逸脱の観点からは、ある特定の単語の文字表記を例えば一度変化させることは、修辭的に重要な意味を持っている可能性があると考えられる。

ある作品の文字表記のあり方を模倣するという目的設定は、その窮極の達成目標や評価基準を定めにくい、以上のように、言語における意味的側面とは異なる側面、あるいは狭義の意味的・概念的領域の外側に(それと接する部分もあるとはいえ)広がる領域の存在を意識化させるという効果を持つ。また、文字表記のみならずその他の物語生成の側面に関しても、模倣を通じて行うという現象的アプローチは、物語生成に関連する様々な知見の発掘につながる可能性がある。ある作品の文字表記の様式を窮極的に真似しようとする、結局言語表現におけるその他の側面(文の長さ、句読点の使用、品詞の種類、単語の長さ、等々)も真似しなければならなくなる。

## 参考文献

- [Akimoto 2012] Akimoto, T. and Ogata, T.: 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, 253-262, 2012.
- [鎌田 2013] 鎌田まみ・小方孝:物語生成システムにおける文生成機構と文字表記機構, 言語処理学会第 19 回年次大会発表論文集, pp.698-701, 2013.
- [熊谷 2012] 熊谷真哉・船越宗・秋元泰介・小方孝: 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, IJ1-OS2-8, 2012.
- [村越 1998] 村越行雄: 複雑化された日本語の文字表記一文文化論一, 跡見英文学, Vol.12, 11- 35, 1998.
- [小方 2010] 小方孝・金井明人: 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, 2010.