

模倣と置換に基づく超短編小説の自動生成

Automatic Generation of Very Short Stories by Imitation and Substitution

緒方 健人*¹ 佐藤 理史*¹ 駒谷 和範*²
 Kento Ogata Satochi Sato Kazunori Komatani

*¹名古屋大学大学院 工学研究科 電子情報システム専攻
 Graduate School of Engineering, Nagoya University

*²大阪大学 産業科学研究所
 The Institute of Scientific and Industrial Research, Osaka University

Machine may create works of art that human has never imagined. This paper proposes a method for automatic generation of very short stories within 140 characters. The method is based on imitation and substitution, where a text extracted from a story is modified by substitution of subordinate clauses. Because a subordinate clause is a syntactic unit with a predicate, the substitution produces a syntactically correct sentence with a different meaning.

1. はじめに

人工知能による自動創作の研究は、探求する価値がある。なぜならば、人間とは異なる創作法を組み込むことにより、コンピュータは、予期せぬ新たな芸術的価値を生み出す可能性があるからである。さらに、機械的な創作技術は、人間の創作の幅を広げることに役立つ可能性もある。

実際に、機械的な方法から、人間が発想を得る方法は存在する。たとえば、星新一の弟子である江坂遊は、新しい着想を得るため、雑誌や短篇集の作品名を品詞に解体し、機械的に組み合わせる方法を紹介している [江坂 1992]。この組み合わせの中から、面白い素材を見つけようというのである。これは、機械的な方法により、人間ならば思いつきにくい組み合わせを生成し、そこに新しい価値を見い出そうという考え方である。

2012年にスタートした「きまぐれ人工知能プロジェクト『作家ですよ』」[松原 2013]は、コンピュータに星新一のような小説を創作させることを目指している。星新一は、質の高いショートショートを数多く残したことで著名な作家である。ショートショートとは、およそ400字から8000字の小説を指し、多くは印象的なオチを持つ。このプロジェクトでは、物語の構造が明確であることや、創作に関する記述が存在することなどを理由に、参考とすべき作家に星新一を選んでいる。

本研究では、小説の自動生成の第1ステップとして、140字程度の超短編小説の生成を目指す。一般に、小説作品においては、文字数が長くなるほど物語の統一性を保つのは難しいため、当初から400字以上のショートショートを生成するのはハードルが高い。そこで、Twitter小説大賞と呼ばれるTwitter上の文学大賞 (<http://twonovel.net/>) の規定を参考とし、まずはTwitter小説 [内藤 2009] のような140字程度の超短編小説の生成を試みる。

本研究で採用する方法は、既存の作品から文章を抜き出し、その一部を別の作品のある部分で置換する方法である。この置換の単位として、「述語を中心としたまとまり」を表す節という言語単位を採用する。さらに、置換対象を文の骨格を除いた

若く美しく、つんとしていても楽しく、人間と同じに働く日。趣味での読書のとき、見た絶望という文字。いま、べたべたおせじを言うし、酒など飲んでいっぱいの日々が、ますます人気ができるはずよ。だが、ぼろを出し、お客はにか笑いし、簡単なうけ答えができたロボットと気がつく時がくる。今晚で終わりとと思うのねとつぶやく。(151字)

図1: 生成例

若く美しく、なにを見ても楽しく、甘くかくわしい年代のある日。赤い椅子に腰かけての読書のとき、ふと目にはいった絶望という文字。いま、こんなに楽しいんだし、これからだって幸福でいっぱいの日々が、永久につづくはずよ。だが、歳月がたち、さまざまながあり、信じていたものに裏切られる時がくる。これが絶望なのねとつぶやく。(157字)

図2: 模倣元の文章(『歳月』より)

従属節に限定することにより、置換後の文章が大きく破綻することを抑制する。先に述べた江坂遊の方法は、単語を単位として組み合わせ、その組み合わせより発想を得た人間が、創作を行なう方法であるが、本手法は、いわば、節を組み合わせ、それをそのまま文章として採用しようという方法とみなすこともできる。

以下、まず2節で、提案する超短編小説作成法の概略について述べる。3節では、本作成法の中核をなす、従属節の置換方法について述べる。4節では、実際に機械的に生成した作品の例を示し、現状と問題点について議論する。最後に5節で、今後の課題について検討する。

2. 模倣と置換に基づく小説生成

2.1 模倣による生成

図1に、本論文で提案する方法で作成可能な作品の一例を示す。この作品はぎこちないところもあるが、「人間のように働くロボットが、いつかぼろを出し、お客に正体がバレてしまう姿を想像する」というストーリーとなっている。

連絡先: 緒方健人, 名古屋大学大学院 工学研究科 電子情報システム専攻, 〒4648603 愛知県名古屋市千種区不老町 C3-1(631) IB 電子情報南棟 159, 052-789-4435, k.ogata@nuee.nagoya-u.ac.jp

若く美しく、【なにを見ても】楽しく、【甘くかくわしい年代のある】日。【赤い椅子に腰かけての】読書のとき、【ふと目にはいった】絶望という文字。いま、【こんなに楽しいんだし】、【これからだって幸福で】いっぱいの日々が、【永久につづく】はずよ。だが、【歳月がたち】、【さまざまなことがあり】、【信じていた】【ものに裏切られる】時がくる。【これが絶望なのねと】つぶやく。

図 3: 模倣元における従属節

実は、この作品は、星新一の作品『歳月』(星新一 ショートショート 1001[星 1998] 第 2 巻 pp563) から抜き出した文章の一部を、別な表現に置換することによって作成した。図 2 に元の文章を示す。元の文章は、「若いころに想像した絶望を、いつか実際に体験する時がくる」というストーリーである。

この例に示したように、我々は、ある作品の一部を抜き出し、さらにその一部を改変することにより、新たな作品をつくり出す方法を採用する。もちろん、これは剽窃以外のなものでもないが、機械による小説の創作は、まずは、このレベルの習作からスタートする必要があると考える。

2.2 節の置換

模倣元(図 2)から新しい作品(図 1)をつくり出す過程では、模倣元の一部の改変した。ここで変更した部分は、**従属節**である。従属節とは、**節**と呼ばれる言語単位の下位分類である。

節とは、簡単に言えば「述語を中心としたまとまり」[益岡 1992] のことであり、統語的にも意味的にも、ある程度のまとまりを備えた言語単位である。節は、文の骨格をなす**主節**と、主節に対して従属的な関係で結び付く**従属節**に大きく分けられる。たとえば、下記の例では、「高津さんは苦手だ」の部分の主節、「朝早く起きるのが」の部分に従属節である。

高津さんは【朝早く起きるのが】苦手だ。

図 3 に、模倣元に含まれる従属節を、上記の例文と同様の形式で示す*1。この図と図 1 を見比べれば、【と】で囲われた従属節の部分のみが変更されていることが確認できる。

従属節の置換先には、別の星新一の作品『ポッコちゃん』([星 1998] 第 1 巻 pp18-20) から抽出した従属節を用いている。図 4 に、『ポッコちゃん』から抽出した従属節のうち、置換に用いたものを示す。たとえば、図 3 の従属節「なにを見ても」は、図 4 の従属節「つんとしていることは」を変形した、「つんとしているても」で置換されている。なお、模倣元の従属節を『ポッコちゃん』のどの従属節で置き換えるかの決定は、自動的には行わず、筆頭著者が行なった。

2.3 生成手順

本研究で採用する生成手順をまとめると、次のようになる。

1. 模倣元の作品 O を定める。
2. 作品 O から、その一部を抜き出す。
3. 抜き出した文章の従属節を認定する。認定された従属節を、 $\{X_1, X_2, \dots\}$ とする。
4. 従属節の置換に利用する作品 R を定め、そこに含まれる従属節を抜き出す。抜き出した従属節を、 $\{Y_1, Y_2, \dots\}$ とする。

*1 ここでは、従属節を、節境界検出プログラム CBAP[丸山 2004]を用いて認定しているため、認定誤りが含まれている。

つんとしていることは
人間と同じに働く
酒など飲む
見たところでは
簡単なうけ答えができるだけ
趣味だ
ぼろを出しては
ロボットと気がつく
お客はにが笑いて
べたべたおせじを言わないし
ますます人気が出て
今晚で終りと思って

図 4: 置換に利用した従属節 (『ポッコちゃん』から抽出)

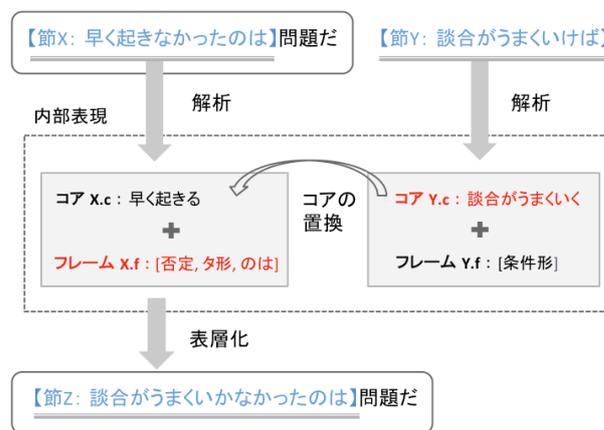


図 5: 従属節の置換の全体像

5. 従属節 X_i を Y_j で置換する。このとき、必要に応じて Y_j を適切な形に変形する。

この生成手順のステップ 5 を自動化するためには、以下の 2 つを実現する必要がある。

- a. 従属節 X_i に対して、置換候補となりうる節 Y_i を選択する
 - b. 従属節 X_i を Y_j で置換する際に、適切に Y_i を変形する
- このうち、本論文では、後者の実現法を次節で述べる。

3. 従属節の置換

3.1 全体像

従属節の置換の全体像を、図 5 に示す。この図では、文「早く起きなかったのは問題だ」に含まれる従属節 X 「早く起きなかったのは」を、従属節 Y 「談合がうまくいけば」で置換する場合を示している。

ここで、従属節 X のあった場所に従属節 Y を代入し、主節「問題だ」とつなげるために、従属節 X の情報を利用する。この例では、従属節 X は否定表現を含み、述語の活用形は「タ形」であり、末尾に助詞「の」「は」が存在する。これらの情報を、従属節 Y 「談合がうまくいけば」に適用することで、従属節 Z 「談合がうまくいかなかったのは」を生成する。

このように従属節の置換では、まず従属節を解析して内部表現へと変換し、内部表現の状態を組み換えを行い、あらためて表層化する手順を踏む。

3.2 従属節の解析と内部表現

本研究では、従属節を、**コア**と**フレーム**という2つの要素に分解する。コアとは、おおよそ従属節の内容部であり、フレームとは、従属節のタイプを表す情報である。より具体的には、コアは、節の先頭から述語部分までの形態素列で、フレームは、以下の3つの情報の組として表現する。

- 否定表現の有無
- 述語の活用形
- 付属語

たとえば、従属節 X 【早く起きなかったのは】は、次のように分解する。

コア $X.c$ [早く起きる]
フレーム $X.f$ [否定, タ形, のは]

なお、コアの末尾の述語は、基本形に変換することとする。

従属節の表層形(形態素列)を、上記のような内部表現に変換することは、比較的たやすい。すなわち、まず従属節の末尾の付属語、否定表現を認定した後、これらを切り離し、次に残った形態素列の末尾の述語の活用形を認定し、これを基本形に変換すればよい。付属語の認定は、mecab、CBAP、BCCWJを参考に、助詞、非自立名詞、連語などを認定する。述語の活用形や否定表現は、益岡・田窪文法 [益岡 1992] に準拠する。ただし、判定詞・ナ形容詞の連体形は、動詞・イ形容詞との整合性を考慮して、基本形とみなす*2。

3.3 コアの置換

コアの置換では、従属節 X のコア $X.c$ を、別の従属節 Y のコア $Y.c$ で置き換える。コアの置換は、一般に、かなり自由に行なえるが、従属節 X のフレーム $X.f$ の種類によっては、置換先のコア $Y.c$ が制限される場合がある。

形容詞や判定詞はそれ自身で状態を表すため、「連用形+ながら」、「連用形+つつ」、「連用形+につれて」などの形式をとることができない。このため、フレームが「ながら」、「つつ」、「につれて」を持つ場合、そのコアを、述語が形容詞や判定詞であるコアでは置換できないという制限を設ける。このような制限により、たとえば、「本当の話にながら」、「勤勉につつ」、「苺が赤くにつれて」のような文法的に許容されない表現の生成を抑制する。

3.4 従属節の表層化

従属節の表層化では、コア $Y.c$ とフレーム $X.f$ から、それらを合成した従属節の表層形を作成する。図5に示した例では、コア $Y.c$ [談合がうまくいく] とフレーム $X.f$ [否定, タ形, のは] のペアを表層化することで、従属節 Z 【談合がうまくいかなかったのは】を生成している。

内部表現の表層化は、次の2つのステップで行う。

1. コアの述語を、フレームの否定表現と活用形に基づき、適切な表層形に変換する。
2. 得られた表層形の末尾に、フレームの付属語を追加する。

このうち、最初のステップは、述語の種類によって特殊な処理が必要な場合が存在する。以下では、動詞、判定詞・ナ形容詞の場合について、活用形の変換方法を述べる。

*2 動詞・イ形容詞では、体言に接続する活用形は基本形である。

動詞の場合

動詞は、ほとんど全ての活用形へと変換可能であるが、否定表現を含む場合には注意が必要である。

例えば、動詞「起きる」の連用形を否定を付加する場合、「起きなく」と「起きず」の2種類の選択肢が存在する。本研究では、動詞の否定に2種類の選択肢がある場合は、全て「ず」系へ変換する。具体的には、動詞「起きる」に否定を付加する場合は、以下のような表層形を生成する。

基本形	起きない
条件形	起きなければ
連用形	起きず
タ形	起きなかった
タ系条件形	起きなかったら
テ形	起きずに
タリ形	起きなかったり

ただし、動詞「ある」は例外とする。動詞「ある」に否定を付与する場合、基本形は「ない」、条件形は「なければ」、連用形は「なく」、タ形は「なかった」、タ系条件形は「なかったら」、テ形は「なくて」、タリ形は「なかったり」とする。

判定詞・ナ形容詞の場合

本研究では、活用形の変換において、判定詞とナ形容詞は同様に扱う。先に述べたように、動詞・イ形容詞との整合性を優先したため、フレームに記述する活用形として、連体形を設定していない。そのため、コアの述語が判定詞・ナ形容詞の場合は、フレームの付属語に応じて、適切な表層形に変換する。具体的には、付属語が「ことが」のような形式名詞を含む場合、連体形(ナ形)へ変換する。付属語が「けれど」「が」「と」のような接続表現に対しては、基本形(ダ形)へ変換する。付属語が「なら」に対しては、語幹に変換する。以下に例を示す。

[勤勉だ] + [基本形, ことが] ⇒ 【勤勉なことが】
[勤勉だ] + [基本形, けれど] ⇒ 【勤勉だけれど】
[勤勉だ] + [基本形, なら] ⇒ 【勤勉なら】

4. 生成例

3節で提案した手法を用いて、従属節の置換を行った例を表1に示す。この表では、図3に含まれる従属節を、星新一の作品『永遠の青春』([星 1998] 第3巻 pp1304)の従属節の中から無作為に選んだ従属節で置換した例を、4例、示した。

表1の例では、ほとんどの従属節が文法的に正しい形で生成されている。従属節を一度内部表現へ変換し、その後、表層化するという2段階を踏むことで、置換後も文法的に違和感のない文を生成することに成功している。

一方で、例2と例3の【手にもってい】や、例4の【青年が部屋のなかでぼんやりとしてい】など、不自然な箇所も存在する。「手に持ってい」は、「手に持っていて」とするのが自然であり、「青年が部屋のなかでぼんやりとしてい」は「青年が部屋のなかでぼんやりとしていて」とするのが自然である。このように、いくつか特定の動詞に関しては、さらに、特別な規則を設定し、より自然な従属節を生成する必要がある。

5. 検討

本研究において、2.3節で示した生成手順のステップ5のうち、従属節の置換の自動化は、ある程度実現できたと考える。

表 1: 置換例

例 1
若く美しく、【マコちゃんと会っても】楽しく、【気の進む】日。【お飲みになったの】読書のとき、【その前に飲んでおいた】絶望という文字。いま、【人口問題に関連するし】、【小さなピンを受け取って】いっぱいの日々が、【手に持っている】はずよ。だが、【男の帰り】、【一気に飲みほし】、【ききめがなかった】【いやに簡単に言う】時がくる。【本当に不老長寿のからだになれるのね】とつぶやく。
例 2
若く美しく、【いやに簡単に言っても】楽しく、【小さなピンを受け取る】日。【その前に飲んでおいたの】読書のとき、【気の進んだ】絶望という文字。いま、【青年が部屋のなかでぼんやりとしているし】、【男の帰って】いっぱいの日々が、【一気に飲みほす】はずよ。だが、【お飲みになり】、【手に持ってい】、【本当に不老長寿のからだになれる】【マコちゃんと会う】時がくる。【人口問題に関連するのねと】つぶやく。
例 3
若く美しく、【人口問題に関連しても】楽しく、【ききめがない】日。【小さなピンを受け取っての】読書のとき、【マコちゃんと会った】絶望という文字。いま、【お飲みになるし】、【青年が部屋のなかでぼんやりとしていて】いっぱいの日々が、【一気に飲みほす】はずよ。だが、【手に持ってい】、【男の帰り】、【その前に飲んでおいた】【気の進む】時がくる。【いやに簡単に言うのねと】つぶやく。
例 4
若く美しく、【その前に飲んでおいても】楽しく、【手に持っている】日。【本当に不老長寿のからだになれるの】読書のとき、【人口問題に関連した】絶望という文字。いま、【一気に飲みほすし】、【男の帰って】いっぱいの日々が、【いやに簡単に言う】はずよ。だが、【小さなピンを受け取り】、【青年が部屋のなかでぼんやりとしていて】、【マコちゃんと会った】【お飲みになる】時がくる。【気の進むのねと】つぶやく。

現在は、従属節のコアのみを置換しているが、フレームも置換できる場合も存在する。本研究で設定したフレームには 52 種類の ID をいずれかを付与しており、同一 ID のフレームは、ほぼ同じ種類のフレーム（たとえば、時間節）を意味する。このようなフレーム間の置換を、今後実装する予定である。

一方、置換候補として適切な従属節を選択することは、現時点では全く実現できていない。このため、先に示した実行例では、無作為に置換先の従属節を選択したが、この方法では、従属節の置換により、意味的に成立しない文を生成してしまう可能性が高い。

置換候補となりうる従属節を適切に選択する一つの方法として、置換元の従属節と類似した従属節を選ぶという方法が考えられる。この方法は、意味的に破綻しない文を生成できる可能性は高いが、元の文の意味を大きく変えることは困難である。すなわち、模倣元と良く似た作品しか生成できないと考えられる。

模倣元に含まれる従属節をすべて置換する必要はない。適切な置換候補が存在するもののみ置換することにしてもよい。さらに、必ずしも置換候補を一つに限定せず、いくつかの候補を

人間に提示し、その中から人間が最も適切だと思われるものを選択することにしてもよい。置換候補の選択には、多くのオプションがあり、これらについて今後検討していく予定である。

2.3 節の生成手順ステップ 2 で、模倣元の文章を抜き出す部分は、現在、

- 作品 O の先頭の 5 文を選ぶ
- 作品 O の末尾の 5 文を選ぶ
- 作品 O から、無作為に 5 文を選び、出現順に並べる

の 3 つの方法を採用し、その中でよさそうなものを人間が選ぶことで実現している。これは、暫定的な方法であり、ストーリーとして完結する 5 文を抜き出すよい方法を考える必要がある。

超短編小説といえども、全自動創作への道は険しい。

謝辞

本研究では、星新一ショートショート 1001（新潮社）のテキストデータを使用した。データを提供していただいた、星ライブラリ、および新潮社に感謝します。

参考文献

- [江坂 1992] 江坂 遊: 小さな物語の作り方 (2011), 樹立社。
- [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, 2013 年度人工知能学会全国大会 論文集。
- [内藤 2009] 内藤 みか, 安達 瑤 b, 新城 カズマ, 小林 正親, 渡辺 やよい, 吉井 春樹, 泉 忠司, 黒崎 薫, 柁野 浩一, 円城塔, Twitter 小説 140 字の物語, デイスクヴァー・トゥエンティワン, 2009。
- [星 1998] 星 新一, 星新一ショートショート 1001, 新潮社, 1998。
- [益岡 1992] 益岡 隆志, 田窪 行則: 基礎日本語文法・改訂版 (1992), くろしお出版。
- [丸山 2004] 丸山 岳彦, 柏岡 秀紀, 熊野 正, 田中 英輝, 日本語節境界検出プログラム CBAP の開発と評価, 自然言語処理 vol. 11, No. 3, 39-68, 2004。