

ブロック玩具に学ぶ文章生成ツール HaoriBricks

HaoriBricks: A Construction Set for Japanese Text



名古屋大学大学院工学研究科教授

佐藤 理史

京都大学大学院工学研究科電気工学第二専攻博士課程研究指導認定退学。博士（工学）。北陸先端科学技術大学院大学、京都大学を経て、2005年より現職。現在、言語処理学会副会長。

1 はじめに

形態素解析や構文解析といった日本語の解析系ツールはかなり整ってきたが、その一方で、日本語の生成系ツールは存在しないに等しい。我々は、短編小説の自動生成の研究において、ストーリー文法に従って文章を組み立てる GhostWriter と、その配下で文を組み立てる文生成器 Haori を作成した [1]。そして、その経験を踏まえ、現在、それぞれの改良版である GhostWriter2 と HaoriBricks を開発している。このうち、本稿では HaoriBricks を紹介する。

2 HaoriBricks の基本設計

我々が対象としている短編小説は、いわゆるショートショートと呼ばれるもので、その長さはおおよそ原稿用紙 20 枚（8000 字）以内が目安である。「日経星新一賞」[2] の制限字数は 1 万字で、これよりは少し長い。2016 年 3 月に開催した星新一賞への応募報告会で、SF・ファンタジー作家の長谷敏司氏より、「制限字数に近い分量の作品を応募しないかぎり入賞は望めない」とのコメントをいただいた。そのため、我々は、1 万字程度の小説を作ることを次の目標の 1 つに設定している。

この長さの文章を生成するプログラムを作るためには、それ相応の準備が必要である。その 1 つが、文章を生成するプログラムをできるだけ簡便に（少ない記述量で）書けるようにすることである。なぜなら、自動生成するテキストとして、1 万字という長さはとてつもなく

長いからである。

その実現のために、新たに設計したのが HaoriBricks であり、そこで採用したのが、ブロック玩具を模したブロックモデルである [3]。ブロック玩具では、基本となる部品が提供され、それを組み合わせることで、色々なモノを作ることができる。たとえば、ブロック玩具の代表格であるレゴでは、建物や乗り物、動物や植物、架空の宇宙船など、ほとんどありとあらゆるもののミニチュアモデルを作ることができる。HaoriBricks はこれに倣い、文章を構成する基本的な部品を提供し、それらを組み合わせることによって、任意の文章の合成を可能にすることを目標としている。その模式図を図 1 に示す。

さらに、いくつかの部品を組み合わせで作った大きな部品（マクロブロック）を、基本部品とまったく同じように使えるようにする。これはちょうど、プログラミングにおいてサブルーチンを定義することに相当する。この機能により、典型的な日本語表現のライブラリを実装し、記述量の削減を目論む。

3 具体例

図 2 に、「鈴木さんは車の調子が悪いことに気づいた」という文を生成するための HaoriBricks のコードを示す。それぞれのブロックは関数（正確にはプログラミング言語 Ruby のメソッド）として実装されている。このコードは、次の 14 個のブロックから構成されている。（スラッシュの後の数字は引数の数を表す。）

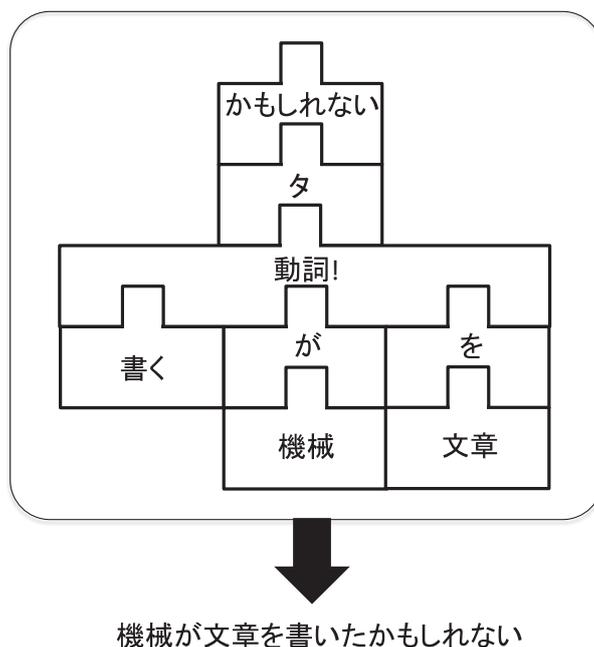


図1 HaoriBricksの模式図

ブロックを組み合わせて文を合成する。HaoriBricksは、そのための基本ブロックを提供する。

1. **主題 /2** 提題構造を作る。
2. **は /1** 副助詞「は」を付与する。
3. **:鈴木さん** 名詞「鈴木さん」を作る。
4. **タ /1** 時制を過去(有標)に設定する。
5. **動詞 /n** 動詞を述語とする述語構造を作る。
6. **:気づく** 語「気づく」を作る。その品詞は5.により動詞と設定され、活用型は子音動詞型と自動推定される。
7. **に /1** 格助詞「に」を付与する。
8. **補足節こと /1** 形式名詞「こと」を伴う補足節を作る。
9. **形容詞 /n** 形容詞を述語とする述語構造を作る。
10. **:悪い** 語「悪い」を作る。その品詞は9.により形容詞と設定され、活用型はイ形容詞型と自動推定される。
11. **が /1** 格助詞「が」を付与する。
12. **の /2** 「AのB」の構造を作る。
13. **:車** 名詞「車」を作る。

14. **:調子** 名詞「調子」を作る。

これらのブロックのうち、コロンで始まるブロックが内容語に対応し、それ以外のブロックは、機能語に対応するブロックか、あるいは、文の構造を規定するブロックである。これらのブロックは、いずれも同等の働きを持つブロックと置き換え可能である。

ある特定の文を作って、それを出力するだけであれば、このように文をブロックで組み立てる必要はない。ブロックで組み立てる利点は、プログラムで条件によって異なる文を生成したり、組み立てたブロックを再利用したりすることができる点にある。たとえば、先のコードに、ブロック「ない」を追加すれば、「鈴木さんは車の調子が悪いことに気づかなかった」を生成することができる。さらに、これにブロック「丁寧」追加して、「鈴木さんは車の調子が悪いことに気づきませんでした」を生成することもできる(図1)。そして、こうして組み立てた任意のコードを、新たなマクロブロックとして定義することが可能である。

```
x0 = haori{ 主題(は(:鈴木さん),
           タ(動詞(:気づく,
                 に(補足節こと(形容詞(:悪い,
                               が(の(:車, :調子))))))))) }

print x0.surface_string, "\n"
# => 鈴木さんは車の調子が悪いことに気づいた

x1 = haori{ ない(x0) }
print x1.surface_string, "\n"
# => 鈴木さんは車の調子が悪いことに気づかなかった

x1 = haori{ 丁寧(x1) }
print x1.surface_string, "\n"
# => 鈴木さんは車の調子が悪いことに気づきませんでした
```

図2 HaoriBricks コードの例

ブロックの組み合わせは、関数合成と同じ要領。コードは `haori{ }` で囲う。
表層文字列は、`surface_string` メソッドで生成される。

4 Basic パッケージ

現時点において、HaoriBricks の原子ブロック（あらかじめ定義されている組み込みブロック）は、実は 22 種類しかない。この原子ブロックと Ruby の機能を使って、日本語の主要な機能語と構造に対応するブロックを定義している。この方式は、プログラミング言語 Lisp に倣っている。なお、機能語の辞書は内部に組み込まれており、文法情報はこの辞書から供給される。内容語の辞書は外付けで用意することもできるが、一部の例外的な語を除けば、事実上不要である。

現在、編纂を進めているのが Basic パッケージと名付けたブロック集合で、約 1,200 種類のブロックが定義されている。このパッケージは、通常の日本語の文章に現れる機能語と構造をカバーすることを目指している。より具体的には、益岡・田窪の『基礎日本語文法－改訂版－』[4] の例文（約 1,100 件）を違和感なく記述できることを目標としている。

小説には複数人の会話も含まれる。以前のシステムではこの扱いに苦労したため、HaoriBricks では、文章

の書き手や発話の話し手を陽に制御する機構（これも、ブロックとして実装される）をあらかじめ組み込んだ。現時点では、話し手によって言い方を変えるために専用のプログラムを書かなければならないが、この点に関しても自動化の研究を進めている [5]。

参考文献

- [1] 佐藤理史．『コンピュータが小説を書く日－AI 作家に「賞」は取れるか』．日本経済新聞出版社，2016.
- [2] 日経星新一賞．<http://hoshiaward.nikkei.co.jp>
- [3] 佐藤理史．HaoriBricks: ブロック玩具に学ぶ日本語文章生成ライブラリ．言語処理学会第 23 回年次大会発表論文集，pp.20-23, 2017. <https://youtu.be/hRtlRB2wa9E>
- [4] 益岡隆志，田窪行則．『基礎日本語文法－改訂版－』．くろしお出版，1992.
- [5] 夏目和子，刀山将大，佐藤理史．発話文自動生成のための日本語表現文型辞書の作成．言語資源活用ワークショップ 2016, pp.126-136, 2017.

