

オノマトペと韻を踏んだ動詞を用いた文章の自動生成

古賀 祐太† 横井 健†

東京都立産業技術高等専門学校†

1. はじめに

オノマトペとは、擬音語・擬声語・擬態語の総称で、物事を感覚的に伝えることが可能な言葉である。物事を表現する際にオノマトペを使用することで、そのニュアンスを相手により分かりやすく伝えることが出来る。またそれに加えて、動詞とその動作のニュアンスを表現するオノマトペが韻を踏んでいる場合、文章全体の語呂が良くなり、相手が伝えたい物事のニュアンスをより理解しやすくなると考える。

そこで本研究では、動詞とオノマトペの音の近さと、意味の近さに着目して、それら両方が近いオノマトペと動詞を用いた文章を自動生成することを目的とする。

2. 提案手法

2.1 概要

提案する自動生成システムの概略図を図 1 に示す。韻を用いた文章を生成するシステムを作るにあたり、ポイントとなる処理が、①形態素解析、②意味比較・編集距離算出、③文章生成の 3 つで構成される。図 1 での形態素解析が①、②が意味比較、③が文章生成に対応している。各処理部の内容については 2.3 節で解説する。

処理の概要を簡単に述べると、入力文を形態素解析して各形態素の品詞情報の取得を行い、その中から動詞を抽出し日本語ワードネット[1]で動詞に含まれる意味を検索し、検索した動詞に意味属性を付与する。次に、オノマトペの一覧から接続し得る動詞を抽出し、被検索動詞と意味が一致・もしくは類似する物を探す。選択された属性に含まれるオノマトペと被検索動詞とで編集距離を算出し、距離が一番近いオノマトペを選択して、入力文に付加する。



図 1 自動生成システム概略図

2.2 対象とするオノマトペについて

オノマトペの特性について調査を行ったところ、「すかつ」や「びっしり」といった単語に

は、その語尾に「と」を付加すると、副詞になって動詞を修飾することが出来るという特性と、「がちがち」や「ぎーぎー」といった形態素などの単位を反復して使われる畳語の後に動詞が来て文章が成り立っているという特性が見受けられた。前者の副詞に変形できるという特性を用いて文章を作成すれば、動詞に装飾されてより直感的になると考える。本研究では副詞変形が可能なオノマトペに着目し、動詞との意味の近さを考慮して文章生成を行う。

オノマトペを用いて文章生成を行うにあたり、オノマトペを辞典[2]の意味分類別さくいんから収集し、そのオノマトペがどんな意味合いを持つか、動詞・形容詞を付加して作成した。オノマトペの一覧の一部を表 1 に示す。この一覧では、「ぎざぎざ」や「ぎーっ」等といったオノマトペにどのような動詞が付けられるかを予め定義しておく表でもあり、副詞に出来るオノマトペや畳語オノマトペもまとめられている。

表 1 オノマトペ一覧の一部

オノマトペ一覧	
属性	オノマトペ
切る	きざきざ
	ざくざく
	さくっ
	さくり
降る	ごろごろ
	こんこん
	ぎーぎー
	ぎーっ
	ざざっ

2.3 各処理の内容

① 形態素解析

形態素解析に用いる解析器は Igo-Ruby[3]を用いた。解析に用いる解析辞書は IPA 辞書を用いて解析を行った。ここでは、入力文を形態素に区切り、そこで出力された結果の内、品詞が動詞である形態素の抽出を行う。

② 意味比較・編集距離算出

意味比較を行うために、抽出された動詞とオノマトペ表にある動詞の両方を日本語ワードネット[1]に問い合わせ、その動詞の概念を調べる。ここで得られる概念情報は、主に同義語・関連 synet とのリンク・定義文・例文がある。今回の手法では、意味の近さを考慮するため、

An Automatic Text Generation Using Onomatopoeia and its Rhymed Verb

† Yuta Koga

† Takeru Yokoi

† Tokyo Metropolitan College of Industrial Technology

関連 synet のリンク情報を用いる。これにより、動詞が持つ概念と何らかの関係が存在する動詞を検索することが出来る。リンク情報には様々な種類があるが、その種類の中から上位語・下位語・引き起こし・関連の 4 つのリンク情報を用いることにより、動詞に意味が近い別の動詞を取得させることが可能になると考える。

オノマトペの持つ意味と入力文の動詞の意味が調査できると、関連する意味合いの比較を行う必要がある。意味比較には、入力文に含まれている動詞の意味と、オノマトペの一覧で属性として登録してある動詞とで一致する場合、その動詞属性に含まれている複数のオノマトペと、入力文の動詞とで編集距離を算出し、値が一番小さいオノマトペを選択する。編集距離を求める時に動詞が漢字を含んでいる場合に、平仮名に変換してから距離を求める。ここで用いている編集距離とは、2 つの文字列がどの程度異なっているかを示す数値で、一方の文字列を挿入・削除・置換を行いもう一方の文字列に変形するのに必要な最小回数である。編集距離を求める際にオノマトペの末尾に「と」を予め連結して、副詞に変形させてから距離を求める。選択されたオノマトペは次の文章生成部に用いられる。

③ 文章生成部

数多くの日本語文章を調査した結果、副詞が動詞を前方から修飾する事が分かった。ここでは、編集距離が一番短かったオノマトペを入力文の動詞の前方に配置する。このオノマトペは副詞に変形させているので、文章の動詞の前に配置することにより音が近いオノマトペが動詞を修飾し、文章全体の語呂が良くなると考える。

3. 実験

3.1 実験方法

文章を自動生成するにあたり、オノマトペ辞典からオノマトペを 1 つの動詞・形容詞に対して約 20 個程度収集し、それぞれのオノマトペに動詞・形容詞の属性を付与した。今回提案した自動生成システムに入力文を 3 つ入れて出力した。入力文は、十数文字で構成されているシンプルな文章を手動で作成したものとした。

3.2 結果と考察

入力文と提案システムから出力された文章、入力文に含まれている動詞とオノマトペの編集距離を求めたものを表 2 に示す。入力された文で動詞になる部分「カット」の前に「すかっと」が挿入されていて、すっきりとしたイメージが、「どかっと」が「破壊」の前に挿入されていて物を崩すように破壊しているのが読み取れる。

実験でそれぞれ得られた出力文が相手に伝わったときにどんな状況が想像出来るかという観点で考察を行った。1 つ目の結果では、入力文よりも出力文のほうが、伸びてしまった髪の毛を床屋で切ってもらった際にどんな風に切ってもらったのかを他人に分かりやすく簡潔に伝えることができていますと推測される。また 2 つ目の結果では、木の柱を破壊するという事を単に相手に伝えるだけでなく、どのように折ったのかを説明できているので、その状況を簡単に把握することができると考えられる。

しかし、3 つ目の結果のように入力文から抽出された動詞と動詞別オノマトペ表とで編集距離を求めた際に、距離が 3 で動詞の原型を留めないオノマトペが選択されてしまい、文章の意図が掴みにくくなる可能性がある。これは、各オノマトペに動詞が付与されているが、動詞に属する各オノマトペが似たような音を持つという特性があると考えられる。動詞を意味解析した際に単体の動詞のみを考慮するのではなく動詞を複数用いて意味解析部での編集距離を求めてオノマトペを選択する必要があると考える。

表 2 入出力結果

入力文	出力文	距離
髪の毛をカットした	髪の毛をすかっとカットした	1
木の柱を破壊する	木の柱をどかっと破壊する	3
彼が朝食を食べている	彼が朝食をべろっと食べている	3

4. まとめ

本研究では、動詞とオノマトペの音の近さと、意味の近さに着目して、それら両方が近いオノマトペと動詞を用いた文章の自動生成を行った。その結果、韻を踏んだオノマトペが動詞を修飾することに成功し、よりニュアンスが伝わりやすくなったことを確認することが出来た。

今後は、文章に追加されたオノマトペで韻を踏まなくなる可能性があるものについて、複数の動詞を用いて検討を行う。

参考文献

- [1] Francis Bond, 日本語 Wordnet, 2012-09-28 <http://nlpwww.nict.go.jp/wn-ja/>, (参照 2014-01-10).
- [2] 小野正弘, “擬音語・擬態語 4500 日本語オノマトペ辞典”, 小学館, 2011, 701p.
- [3] Kyowichi Nishi, kyow/igo-ruby, 2010, <https://github.com/kyow/igo-ruby>, (参照 2014-01-10).