

文章のリズムを考慮した小説執筆支援システムの作成

齊藤 雄大[†] 長谷川 大 佐久田 博司

青山学院大学 理工学部 情報テクノロジー学科[‡]

1 はじめに

1.1 背景・目的

自分の意図するところを短時間に効果的な文章として作成することは、文章を書き慣れた人間でも難しい作業である。特に小説は読み手の興味を惹きつけ続けることが必要で、さらに難易度は高いと言える。

小説を書く上で問題となる要因はいくつかある。誤字脱字や表記揺れといった表層的なものから、文章のリズム、すなわちいかに読者の「文章を読む」という作業を妨げないか、といった深層的なものまで様々である。誤字脱字や表記揺れといった問題は視覚的に見つけることができ、主観の入り込む余地もなく、また支援するソフトも多い。しかし文章のリズムは視覚的に確認し得るものではなく、個人個人の主観が入りやすいため客観的な判断が難しい。そのため小説を趣味で書いている人達、中でもより良い文章を書きたい人達からは、文章のリズムを客観的に判定できるシステムの作成が望まれている。

そこで本研究は、執筆者が意図する文章のリズムに合う様な修正案を提示する小説執筆支援システムを作成することを目的とする。

1.2 関連システム・研究

小説執筆を支援するものとしては佐久間・小方らの研究 [1] が挙げられる。この研究ではプロットを生成することで物語のストーリー作成を支援しており、文章修正を支援しようとしている本研究の目的とは異なる。

また、本研究のように文章をチェックするシステムとしては文体診断ロゴーン¹が挙げられる。このシステムは入力された文章を形態素解析し、文体の特徴を診断してくれるシステムである。しかし文章のリズム感には触れられておらず、また出力される結果も評価及び詳細までで改稿へのアドバイスは生成されない。また、小説用とは明記されていない。

2 予備調査

システムを作成するにあたって、小説のリズムに影響する要因を調べる為の予備調査を行った。実験内容は、10代~50代の男女12名に、小説から抜粋した4行程度の文章を10個読んで貰い、それぞれについて文章のリズムについての質問に回答して貰うというものである。質問は7段階リッカートスケールによって作成した。

この調査で得られた回答を元に、文章の文字数を始め、リズムの速さに影響を与えそうな要素とリズムの速さの被験者間平均とで相関を調査した。その結果、1文あたりの平均(以下平均)文字数、平均単語数、平均文節数、平均助詞数、平均修飾語数、平均漢字数、平均ひらがな数の6項目において有意な相関が出た。

文章のリズムを判定するためにはその文章のリズムがどれほどの速さであるかを数値化する必要がある。そこで、予備調査でリズムの速さに影響があると判明した要因にそれぞれの相関係数を掛け、その和を取った。また、今回のシステムでは形態素解析だけで行える平均文字数(Ac)、平均単語数(Aw)、平均文節数(As)、平均助詞数(Ap)の4要素についてのみ扱うこととした。

このようにして定められたリズムの速さLを求める式は次の通りである。

$$\begin{aligned} \text{リズムの速さ } L &= \text{平均文字数 } (Ac) \times Rc \\ &+ \text{平均単語数 } (Aw) \times Rw \\ &+ \text{平均文節数 } (As) \times Rs \\ &+ \text{平均助詞数 } (Ap) \times Rp \end{aligned}$$

この計算式で用いられているRc, Rw, Rs, Rpとは、それぞれAc, Aw, As, Apと被験者評価との相関係数である。

3 システム概要

システムの処理の流れは以下の通りである。

1. テキストファイルを読み、MeCab²によって形態素解析を行う

² <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

[†] Yudai Saito (15809037ys@gmail.com)

[‡] Department of Integrated Information and Technology, College of Science and Engineering, Aoyama Gakuin University

¹ <http://logoon.org/>

2. 解析結果を元に文章全体のリズムの速さを算出する
3. 算出されたリズムの速さを元に、文章全体のリズムレベルを判定し、出力する
4. リズムの速さの算出に用いた文字数等の要素に関してもリズムレベルを判定する
5. ユーザ希望のリズムレベルと比較し、必要であれば各要素についてアドバイスを生成する

ここでは3,4,5の項目について説明する。

3.1 リズムレベルの算出

まず、リズム感Lを条件分岐によって最も適したリズムレベルに振り分ける。リズムレベルは1~7までの7段階であり、1が最も遅く7が最も速い。また、各レベルの値の範囲は予備調査のデータを元に設定した。

Ac, Aw, As, Apについても同じように設定した条件分岐式を用いて、リズムレベルを判定する。

3.2 アドバイスの生成

次に、前項で算出したリズムレベルとユーザ希望のリズムレベルとを比較し、アドバイスを生成するかどうかを判定する。一致している場合はその旨を出力する。

アドバイスを生成する場合は、Ac, Aw, As, Apの4要素について、それぞれ希望のリズムレベルに近づけるようなアドバイスを6段階の中から選択し、出力する(図1)。なお、その要素がユーザ希望のリズムレベルと一致している場合はアドバイスを生成しない。

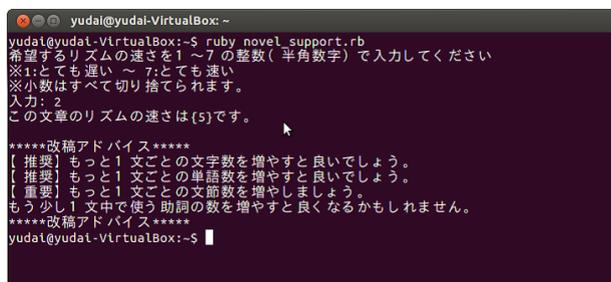


図1 アドバイスの出力画面

4 評価実験

2章のLを求める計算式が有用であるかを確認するために、次のような評価実験を行った。

4.1 実験概要

10代~20代の男女11名に5行程度の小説文章を10個読んで貰い、それぞれについてリズム感を7段階で判定して貰い、システムの評価と一致するかを確かめた。なお文章は全て、予備実験の際とは異なるものを使用した。この時被験者には、1つ文章を読む度に対応する設問へと回答するよう指示をした。また、各文章について絶対評価で回答するよう注意を促した。

4.2 結果と考察

システム出力のリズムレベル及び被験者が回答したリズムレベルの平均は図2のようになった。

被験者平均のリズムレベルをそれぞれ四捨五入したところ、10個中4個の文章においてシステム出力の値と一致した。このことから、システム出力のリズムレベルとユーザ希望のリズムレベルとが一致する確率は40%であると分かった。

評価実験の結果を分析したところ、1文あたりの読点数はシステム出力のリズムレベルとは強い負の相関($R = -0.753$)を示しているものの、被験者評価とは非常に弱い負の相関($R = -0.333$)となっていた。読点が増えると必然的に文章が長くなり、それによってリズムレベルが下がるはずである。しかし被験者評価はそれに従っていない。このことから被験者は読点を無視している、もしくは何か別の用途、例えば一部の読点を句点の代わりなど、リズム感に影響するようなものとして扱っている可能性があると考えられる。

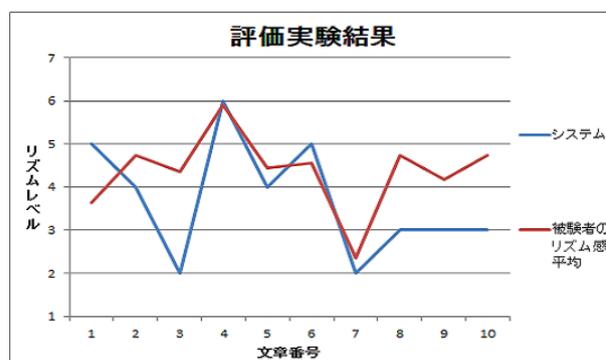


図2 評価実験結果

5 おわりに

本研究では文章のリズム感に関するアドバイスを生成する小説執筆支援システムの開発を行った。この結果、ユーザは文章のリズムという新たな観点からシステムの支援を受けることが可能となったが、希望するリズムレベルとシステム出力との一致度は40%と、高い確率とはいえない結果となった。今後の課題としては一致確率の上昇が挙げられる。その例としては、読点の用途解析や今回扱わなかった平均漢字数などの要素の導入を行うことが挙げられる。

参考文献

- [1] 佐久間友子, 小方孝: プロットの物語内容論を利用したストーリー生成支援システムとその考察, 人工知能学会 2005,2005