

文語体テキスト読書支援ツールの提案 Proposal for Reading Support Tool for Classical Japanese Text

山本 理紗子[†] 竹本朱里[†] 来住 伸子[†]
Risako Yamamoto Akari Takemoto Nobuko Kishi

1. はじめに

文語体で書かれた作品を現代の文体に変換し、音声読み上げ機能をつけることで日本語文語体の可読性の向上を目的としたツール、「文語体テキスト読書支援ツール」を提案し、試作している。このツールは、大正期の旧字・旧仮名・文語体の文章を対象とし、新字・新仮名への変換、文体変換、音声読み上げを行うものである。ツールで使用した文章変換と、音声読み上げ機能について、アンケートによるユーザー評価を実施し、可読性の向上に何が役立つかについて考察した。

2. 用語の説明と関連研究

本報告にて使用される用語や定義について説明する。文語体とは近代から明治・大正時代に使われた書き言葉とする。昭和 24 年の内閣告示で公布された当用漢字字体表により定められた漢字を新字、それ以前に用いられてきた漢字を旧字とする[2][3]。昭和 21 年の内閣訓令で交付されたかなづかいを新仮名、それ以前に用いられてきたかなづかいを旧仮名とする[2][4]。

小木曾ほかにより作成された歴史的日本語資料を対象とした形態素解析のための辞書[7]のうち、本研究では近代文語 UniDic[6]を使用した。形態素解析には、形態素解析システム MeCabを使用した[5]。



図 1 文語体テキスト読書支援ツールトップページ

3. ツールの設計

3.1. ツールの主な機能

文語体テキスト読書支援ツールは、旧字旧仮名で書かれた文章を新字新仮名の文章に変換する機能、文体を変換する機能、文章を読み上げる音声読み上げ機能の三つから構成されている。

3.2 ツール試作に使用したソフトウェア

ツールは html、JavaScript、Python、WebSpeechAPI[11]を使って開発した。テキスト変換のプログラム作成には Python を利用した。テキスト変換後のテキストは、JavaScript と WebSpeechAPI[11]を用いて音声読み上げができるようにした。利用できる声の種類は利用するブラウザによって異なる。

4. テキスト変換ツール

4.1 利用したデータ

4.1.1 変換対象としたテキスト

青空文庫[1]より、「京につける夕」[8]、「しるこ」[9]、「秋雨の追憶」[10]の 3 作品を選んだ。文語体の文章が青空文庫に存在し、本研究の対象とする年代である明治・大正時代に発表されていることからこれらの作品を選んだ。

4.1.2 変換に用いたデータ

1 種類のサイト、6 種類の辞書、1 種類の文献を参考にした。詳細は省略する。

4.2 変換ツール

変換ツールは、以下の 3 種類の手順でテキスト変換を行う。

- 手順 1: 記号を取り除く
- 手順 2: 旧字を新字に、旧仮名を新仮名に置き換える
- 手順 3: 品詞情報に基づく文体変換を行う

4.2.1 (手順 1) 記号を取り除く

青空文庫からダウンロードしたデータに含まれる文中の記号についての説明、ふりがな等のデータを、正規表現を使用して消去する Python のプログラムを作成した。

4.2.2 (手順 2) 旧字を新字に、旧仮名を新仮名に置き換える

文中に含まれる旧字・旧仮名をそれぞれ新字・新仮名に置き換えるプログラムを Python で作成した。文中に該当する文字列があれば対応する新字・新仮名を返す Json ファイルの形式の辞書を作成した。

4.2.3 (手順 3) 品詞情報に基づく文体変換を行う

近代文語の解析用辞書である近代文語 UniDic[6]を辞書として用いることにした。MeCab[5]と近代文語 UniDic[6]を用いて形態素解析し、形態素それぞれの品詞ごとに文語体から現代の言葉に置き換えた。文法に則り対応する現代語訳が異なるものについて場合分けを行なった。

[†] 津田塾大学 Tsuda University

5. ユーザー評価

5.1 評価の目的

文語体の文章、返還後の文章、変換後の文章に音声をつけたものを読んだ時の印象の違いや可読性の差を調べることを目的に、アンケートによるユーザー評価を行った。

5.2 評価方法

青空文庫に掲載されている「しるこ」[9]という作品から以下の 3 つの Web ページを用意した。①旧字旧仮名の文章のみのページ、②旧字旧仮名を新字新仮名に置き換え、文体変換をした文章のページ、③②の文章と②の文章を音声読み上げした音声ファイルを添付したページである。この 3 ページを図 2、図 3 のようによく似たデザインの Web ページで提供し、デザインの差によって印象の違いが生じないようにした。

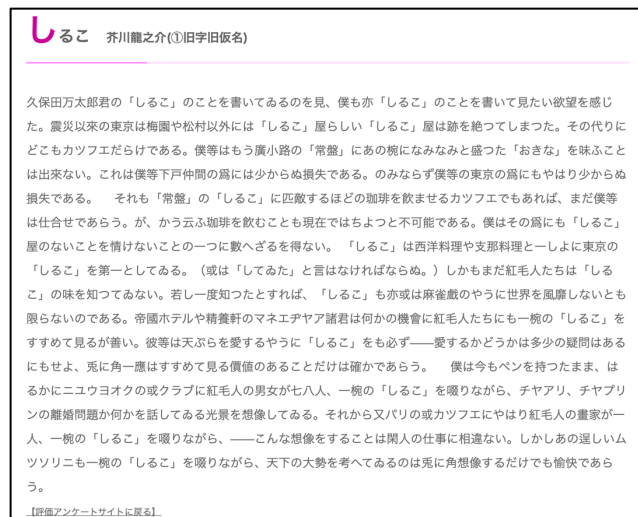


図 2 ①旧字旧仮名で書かれた文章

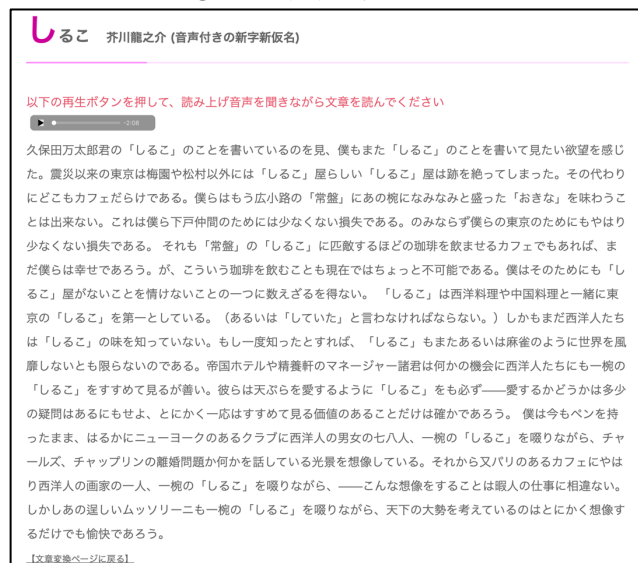


図 3 ③②の文章と音声読み上げ

5.3 アンケート結果

3 種類のデザインを利用した 16 名のユーザーに対し、アンケートを実施した。①の文章より②または③の文章の方

が読みやすい、と 16 名が回答したので、旧字旧仮名から新字新仮名への置き換えと文体変換は読みやすさの向上に有効であるとわかった。②と③のどちらが読みやすいかという問いでは②を選んだ人が 7 名、③を選んだ人が 9 名という結果になった。③を選んだ人の意見には、音声で聞くことが集中力の維持や漢字の読解の助けになったという意見があった。②を選んだ人の意見には、音声読み上げが読書の支援にならなかった、音声読み上げの機械的な声が読書の妨げになってしまったという意見があがった。

5.4 考察

新字新仮名への置き換えと文体変換は読みやすさの向上に有効であるとわかったが、音声読み上げで読みやすさが向上するか否かについては意見が別れた。音声読み上げがよりわかりやすい音声になるよう改良した後で、再評価が必要である。今回のユーザー評価で使用した文章は新字新仮名への変換と文体変換の両方を行ったものである。新字新仮名への置き換えのみを行った場合やこれに音声をつけたものなどより多くの組み合わせによる比較評価も今後必要である。

6. まとめ

「文語体テキスト読書支援ツール」を試作し、ユーザー評価を行った。ユーザー評価により、元の文語体の文章に比べ、本ツールで変換した文章の方が可読性は向上した。音声の読み上げ機能による可読性の向上はユーザーによって個人差があった。音声読み上げについては、聞き取りやすい抑揚についてさらに調査し、工夫を重ねたい。文章変換については、新字・新仮名への置き換えや文体変換に関し、より多くの組み合わせを用意しユーザーの好みに対応できるようにしたい。そのためには、辞書や文法処理の強化が必要である。また、任意のアップロードされたテキストファイルに対応できるように改良したい。

参考文献

- [1] 「青空文庫」 <https://www.aozora.gr.jp/index.html>
- [2] 「スーパー大辞林」三省堂、Apple ソフトウェア。
- [3] 文化庁「3. 当用漢字字体表 (昭和 24 年 4 月) 【国語審議会】」『国語施策沿革資料 1 2 (平成 9 年 1 月 17 日) 漢字字体資料集 (諸案集成 2・研究資料)』 https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/sisaku/enkaku/enkaku12.html
- [4] 文化庁「8. 昭和 21 年 11 月 内閣訓令第 8 号、内閣告示 33 号 現代かなづかい」『国語施策沿革資料 1 (昭和 55 年 3 月 31 日) 仮名遣い資料集 (諸案集成)』 https://www.bunka.go.jp/kokugo_nihongo/sisaku/joho/joho/sisaku/enkaku/enkaku1.html
- [5] 工藤拓『実践・自然言語処理シリーズ 第 2 巻 形態素解析の理論と実装』近代化学社、2018 年。
- [6] 国立国語研究所 コーパス開発センター「近代文語 UniDic」『UniDic』 https://unidic.ninjal.ac.jp/download_all#unidic_kindai
- [7] 小木曾智信、小町守、松本裕治「歴史的日本語資料を対象とした形態素解析」『自然言語処理』 20 巻 5 号、2013 年、p.727-748、 https://www.jstage.jst.go.jp/article/jnlp/20/5/20_727/article/char/ja
- [8] 「京に着ける夕」『青空文庫図書カード No.777』 <https://www.aozora.gr.jp/cards/000148/card777.html>
- [9] 「しるこ」『青空文庫 図書カード No.24452』 <https://www.aozora.gr.jp/cards/000879/card24452.html>
- [10] 「秋雨の追憶」『青空文庫図書カード No.437』 <https://www.aozora.gr.jp/cards/000076/card437.html>
- [11] 「MDN Web Docks Web Speech API」 https://developer.mozilla.org/ja/docs/Web/API/Web_Speech_API