

令和 2 年 4 月 15 日現在

機関番号：33306

研究種目：基盤研究(C) (一般)

研究期間：2017～2019

課題番号：17K01097

研究課題名(和文)全盲学生のボランティアを支援する授業資料等の点字への翻訳システムの提案

研究課題名(英文) Proposal of a translation system for Braille such as class materials to support volunteers of blind students

研究代表者

下村 有子 (Shimomura, Yuko)

金城大学・社会福祉学部・教授

研究者番号：70171006

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：全盲学生に必要な書籍の点字本をすみやかに提供することが目的である。そのために点字文法への変換はPython言語を用いて行った。また日本語から点字への変換は形態素解析ツールのMeCabを用いた。しかし、MeCabの形態素分割の精度が悪く、構文ツールそのものを見直しを行った。その結果、AIのDeep Learningを用いて翻訳プログラムを作成することとした。今までの構文解析に用いたデータや解析結果はAIの教師データとして用いることができる。AIのシステムはまだ構築が一部であるが、翻訳結果は99.86%の精度を出すことができた。今後引き続き研究を行い、完成させていく。

研究成果の学術的意義や社会的意義

覚障害学生にとって点字教科書や参考書はどうしても必要なものであるが、ソフト開発業者は収益が上がらないソフトウェアは開発しない。間違わずに点訳できるボランティア熟練者は少ない。点字文法は複雑で難しいうえに、時々変更される。熟練者は高齢であり、どんどん減っていく。コンピュータで100%翻訳させることが理想であるが、まだ行われていない。そのため今回は日本語形態素解析ツールやAIのDeep Learningを用いた。最新ITを用いることで、今までより、より100%に近づけることができた。

研究成果の概要(英文)：Our aim is to provide Braille students with the necessary Braille books promptly for blind students. For that, we used the Python language to convert it to Braille grammar. The conversion from Japanese to Braille was done using a parsing tool MeCab. However, the morpheme segmentation of MeCab was inaccurate, and we examined other methods that are not syntactic tools. As a result, we decided to create a translation program using AI Deep Learning. The data and analysis results used so far can be used as AI training data. Although the AI system we created is still part, the translation result was 99.86% correct. We will continue our research and aim to complete the system.

研究分野：福祉情報工学

キーワード：視覚障害学生支援 墨字翻訳

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

1. 研究開始当初の背景

全盲学生の支援の1つに教科書などを点字にすることがある。点字編集ソフトは「カナ文字」で入力すると、点字コードに変換してくれる単純ソフトウェアである。そのためボランティアは印刷された文字を「カナ文字」で入力し、「分かち書き」を行い、「点字文法に沿った書き方・文字」で入力する。点字編集ソフトは入力された文字を点字コードで保存し、点字を出力する。このような過程を経て点字本は全盲学生に渡される。図書館での点字図書作成時は2回の校正を行って出版されるが、学生の支援としては時間がかかりすぎる。全盲学生は大学入学が決まると教科書が決まる。授業開始までの1-2か月の間にボランティアは点字の本を作成する。大量のデータである。もちろん授業の講義(白板の文字や配布物など)なども同様で、ボランティアは翌週の授業までに、点字などの理解ができるデータにして全盲学生に渡し、全盲学生は内容理解や宿題をこなしていく。

教科書や参考書などの点字本は点訳ボランティアなどによって作成されるが、著作権法第37条(1)により著作物の全ては、点字による複製が認められている。そのため、どのような書籍でも点字図書にすることは可能になっている。テキストデータで渡される書籍も点字化は可能である。しかし、出版社はテキストデータを渡してくれない。そのため、印刷物からデータを作成している現状がある。

日本語点字は規則もコード配置も日本語独特のものである。そのため外国との比較は行えない。国内では点字入力・編集ソフトはいくつかが販売されているが、入力文字コードを点字文字コードに変換することが主目的であり、一部の簡単な点字規則に沿った変換を行うにとどまっている。また点字の分かち書きなどの規則も数年に1回検討され、変更も行われているために、ソフト開発者は踏み込んだ点字編集は行っていない。

私は1980年に点字と墨字(活字)の変換システムを研究して以来、36年間視覚障害者支援システムを研究している。以前の点字は点字用紙に直接手打ちであったが、1990年頃より点字コードが定められ、PCに入力されることとなってきた。PCへのデータ入力は人間の手で行われるために、入力ミスは必ず発生する。聞き違い、見違いによる人為的ミスが原因の中心である。単純ミスより思い込みミスが多い。

近年、ひらがな変換や分かち書きを行ってくれる形態素解析ツールのChaSen(5)やMeCab(6)などが出現している。また、OCRソフトはデータ変換精度が上がってきており、ミスが減っている。このように点字を翻訳化するためのソフトがそろってきた。我々はこの手法を用いて、コンピュータが漢字かな混合文をカナ文に変換し、分かち書きを行い、点字文法ミスを修正していく。このシステムでは人手を介しない分、ミスの発生も少なく抑えることができる。

視覚障害者に対する修学支援の分野では、全盲の児童・生徒・学生の多くが良質の点字での学習資料の保障を強く望んでいると言われている。それに応えるには、正確な点字文書を迅速に製作するシステムの構築が不可欠である。しかしながら一方で、学習資料の点訳のほとんどは長年の経験を持つ中高年のボランティアが担っていることから、全く新規のシステムの導入には無理があり、従来の方式の継承も考慮する必要がある。こうしたことから、煩雑な処理をコンピュータに行わせる簡単な操作の点字訳システムの開発を計画した。

2. 研究の目的

本研究の目的は全盲学生に<早く><正しい>情報を渡すことである。全盲学生は大学に入学すると、教科書、参考書など多くの点字の本が必要になる。大学が決定し、授業が始まるまでにすべての必要な書籍の点字本が手に入らないと授業が受けられない。ボランティアはなるべく早く全盲学生に点字本を渡したい。しかし点字には点字文法があり、入力・変換ミスが発生する。我々はミス発生を最小限にするために、コンピュータが「テキストコード」化し、点字文法の「カナ文字に変換」し、「分かち書き」にし、「変換ミスや文法ミスを修正」し、「点字に翻訳」するシステムを提供する。全盲学生に少しでも早く正しい情報を渡し、健常学生と同等の修学が行えるようにするシステムである。

3. 研究の方法

このシステムの研究計画・方法は以下のとおりである。

- (1)形態素解析ツールの検討と、システム構築を行う。
- (2)印刷物をスキャナーで読み込む。形態素解析を用い日本語コードの間違いを訂正する。
- (3)形態素解析ツールを用い、カナに変換し、分かち書きを行う。
- (4)点字規則を用いて、点字文章に変換する。
- (5)点字コードに変換し、点字で提示する。

(6)全体システムを構築する。

4. 研究成果

(1)平成 29 年度の研究成果

形態素解析ツールの検討

形態素ツールは「MeCab」「Chasen」「JUMAN」「KyTea」などを比較検討した。「JUMAN」「KyTea」は処理が遅く、「実際にインストールして検討したものは「MeCab」「Chasen」である。最終的には「MeCab」を採用したが、理由は以下のとおりである。1 に、発音を出力してくれること、2 に出力結果の後の処理言語に Python を用いることである。特に、点字は音声言語であるために、発音の出力は 1 つ処理を省くことができた。実際は読みを発音に変換するプログラムの開発を予定していたためである。「MeCab」で開発を行った。

システム構築

システムの全体像を構築し、Python 屋 C の開発言語、処理内容等を検討し、徐々に開発を行った。

印刷物をスキャナーで読み込む。

スキャナーの検討を行った。検討するために購入したスキャナーは EPSON「GT-S560」と Fijitsu「Fi-7160b」である。「Fi-7160b」の添付の OCR スキャナー「ABBYY FineReader」は認識率もよいためこのスキャナーと OCR ソフトを用いることとした。読み取ったデータはいろいろなパターンを用意し、検討した。

(2)平成 30 年度の研究成果

形態素解析ツールを用い、カナに変換し、分かち書きを行う。

形態素解析ツールは MeCab に決め、MeCab で形態素に分割し、カナ（発音カナ）に変換した。また発音カナになった文章は、MeCab と連携ができる言語 Python で開発を進めた。Python で行ったシステム開発は、(1)MeCab の形態素分割と点字の分かち書きは違うため、点字の分かち書きにするためのシステム構築、(2)多くの点字規則があるために、それぞれの点字規則に従った文章にするためのシステム開発、(3)点字規則に沿った文章を点字コードに変換するシステム、である。

点字規則を用いて、点字文章に変換する。

いろいろな規則に沿って、Python 言語で変換を行った。

点字コードに変換し、点字プリンタで提示する。

点字コードを点字プリンタで印刷し、検証する計画であったが、点字にすると点字を読める人に検証を行ってもらう必要があるために、今回は点字コードを文字化する逆変換プログラムを作成し、印刷された発音カナ文を読むことで検証を行った。そのために開発サイドで検証ができ、謝金の発生を抑えることができた。

検証の結果、MeCab の形態素分割の精度が高くなく、プログラムで修正できるものは修正を行ったが、最終的に 1%のミスが発生が認められた。また、発音かな文字の変換ミスは 0.2%であり、大きなミスではなかった。

(3)平成 31 年度の研究成果

昨年度の検証結果は、MeCab の形態素分割の精度が良くないと認められた。形態素解析の Chasen もさらに悪い結果だったため、別の方法での検討を行った。その結果、AI の Deep Learning での翻訳を試みた。Deep Learning の得意な分野の 1 つに機械翻訳がある。このシステムが近年充実し、平成 31 年に Transformer という精度のよい翻訳システムが発表された。そのため今回開発中の MeCab 用のワークステーションのメモリを 128GB に増設し、GPU を搭載し開発を行った。ソフトウェアは以下のとおりである。OS:Win10、GPU 環境: cuda Ver.9.0、Tool kit : cudnn Ver.9.0 by NVIDIA、開発環境: Visual studio Community 2017、Python 3.6 by Python Software Foundation from Anaconda、Deep Learning 環境設定: Tensorflow ver.1.10、翻訳ライブラリ: Tensor2Tensor library ver. 1.9 by Google Including Transformer。

この AI システムでは、今までの形態素解析のデータを教師データとすることによって、今までの研究成果を生かすことができた。Deep Learning は全体システムの 3 か所で用いることとし、今回は点字データを墨字データに変換する箇所での実験を行った。その結果、99.86%の正解率であった。100%にするための改良とあと 2 つの変換については、今後引き続いて研究を行っていく。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 Yuko Shimomura, Hiroyuki Kawabe, Hidetaka Nambo and Shuichi Seto	4. 巻 1
2. 論文標題 The translation system from Japanese into Braille by using MeCab	5. 発行年 2018年
3. 雑誌名 Proceeding of the 12th International Conference on Management Science and Engineering Management	6. 最初と最後の頁 1125-1134
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yuko Shimomura, Hiroyuki Kawabe, Shuichi Seto, Hidetaka Nambo	4. 巻 1
2. 論文標題 Development of a system that beginner volunteers can translate Japanese into Braille	5. 発行年 2018年
3. 雑誌名 Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference	6. 最初と最後の頁 1-5
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yuko Shimomura, Hiroyuki Kawabe, Hidetaka Nambo, Shuichi Seto	4. 巻 11
2. 論文標題 Constraction of Restoration System for Old Books Written in Braille	5. 発行年 2017年
3. 雑誌名 Proceeding of the Eleventh International Conference on Management Science and Engineering Management	6. 最初と最後の頁 469, 477
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yuko Shimomura, Hiroyuki Kawabe, Shuichi Seto, Hidetaka Nambo	4. 巻 18
2. 論文標題 Braille translation system in Japan -Past and Now-	5. 発行年 2017年
3. 雑誌名 Proceeding of the 18th Asia Pacific Industrial Engineering and Management Systems	6. 最初と最後の頁 C6-27, C6-31
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Yuko Shimomura
2. 発表標題 The translation system from Japanese into Braille by using MeCab
3. 学会等名 The 12th International Conference on Management Science and Engineering Management (国際学会)
4. 発表年 2018年

1. 発表者名 Yuko Shimomura
2. 発表標題 Development of a system that beginner volunteers can translate Japanese into Braille
3. 学会等名 The Asia Pacific Industrial Engineering & Management Systems Conference (国際学会)
4. 発表年 2018年

1. 発表者名 Yuko Shimomura
2. 発表標題 Constraction of Restoration System for Old Books Written in Braille
3. 学会等名 The Eleventh International Conference on Management Science and Engineering Management
4. 発表年 2017年

1. 発表者名 Yuko Shimomura
2. 発表標題 Braille translation system in Japan -Past and Now-
3. 学会等名 The 18th Asia Pacific Industrial Engineering and Management Systems Conference
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
連携研究者	川邊 弘之 (kawabe hiroyuki) (60249167)	金城大学・社会福祉学部・教授 (33306)	
連携研究者	南保 英孝 (nanbo hidetaka) (30322118)	金沢大学・電子情報学系・准教授 (13301)	
連携研究者	瀬戸 就一 (seto shuichi) (90196973)	金城大学短期大学部・ビジネス実務学科・教授 (43304)	