

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年 5月28日現在

機関番号：14501

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21790497

研究課題名（和文） 病院向け自動点訳プログラムの開発

研究課題名（英文） Automatic Braille Translation Program for Hospital Use

研究代表者

菅野 亜紀（SUGANO AKI）

神戸大学・医学部附属病院・学術推進研究員（特務助教）

研究者番号：20457039

研究成果の概要（和文）：患者に手渡す医療文書の自動点字翻訳の実現を最終的な目標として、我々が開発した自動点字翻訳プログラムの医療用語点訳精度向上の研究を実施した。具体的には、医療文書コーパス（電子的な言語資料）を作成し、辞書の収録語と自動点字翻訳プログラムの点訳精度の関係の解析を行って点訳精度向上の手法を明らかにした。

研究成果の概要（英文）：We studied accuracy improvement of our automatic Japanese braille translation program for medical documents which are provided to patients in hospital. We made medical text corpus to analyze braille translation accuracy and analyzed the relationship between the dictionary and the accuracy of our program to clarify the method for accuracy improvement.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1800,000	540,000	2,340,000
2010年度	900,000	270,000	1,170,000
2011年度	600,000	180,000	780,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：医歯薬学

科研費の分科・細目：境界医学・医療社会学

キーワード：点字、自動点訳、医療文書、病院、辞書、自然言語

## 1. 研究開始当初の背景

アクセシビリティ向上という社会的要請にとめない、医療現場に対し「視覚障害者等に対する服薬指導について」（平成10年8月19日 政医第289号、厚生省保健医療局）の通達が出ており、「平成12年度社会保険診療報酬改定等の概要」では「視覚障害者に対する点字等を用いた薬剤情報提供が薬剤情報提供料として加算可能」と明記された。そこで我々は病院での点字文書提供システ

ム構築の研究（萌芽研究 平成19年度～平成20年度、研究代表者：大島敏子、課題番号：19659563）で自動点字翻訳プログラム「eBraille」を開発し、新聞記事の自動点訳の評価では実用レベルを達成した。しかし、病名や薬剤名などを含む文書の点訳精度は不十分なことが判明していた。その主な原因は、医療用語の漢字の読み誤りと長い文字列の名詞の分かち書き（点字表記規則に基づく単語の区切れ）の誤りである。前者には、eBraille が使用する辞書へ医療用語を追加

することが効果的であることが判明していた。後者に対しては、自然言語処理技術を利用したプログラムの開発で分かち書きの精度向上が期待できる。以上の2点を eBraille に組み込むことで医療文書の自動点訳に対応可能となると、点字知識のない病院職員が簡便に点字文書を作成し、視覚障害を有する患者に提供可能となる。その結果、情報格差の是正と医療安全の一端を担えろと考え、本研究提案を着想するに至った。

## 2. 研究の目的

本研究は、我々が開発した自動点字翻訳プログラム eBraille を基に、患者に手渡す医療文書の自動点字翻訳を実現するプログラムを開発し、病院での実用化（病院内での自動点訳と点字文書提供の実現）を最終目標とする。研究期間内に、(1)医療文書を基にした、自動点字翻訳プログラムの点訳精度評価用コーパス（言語資料）を作成し、(2)医療用語の自動点字翻訳を実現する追加プログラムやモデルを開発した後に、点訳精度の解析と病院内での点字文書運用の試行を実施する。

## 3. 研究の方法

(1) 自然言語処理プログラムの研究において、その精度を評価するためにはコーパスが必要である。しかし、自動点字翻訳プログラムの点訳精度を評価するための汎用的なコーパスについては報告されていない。更に、本研究では医療用語が含まれるコーパスが必要である。そこで、本学医学部附属病院で配布している患者向けの医療文書を基に、点訳精度評価用コーパス（医療文書評価用コーパス）を作成した。なお、このコーパスは漢字仮名混じり文（墨字文）と点字文の対で構成されており、これらを用いて点訳精度を算出した。

(2) 医療用語を追加した辞書と点訳精度との関係を解析した。辞書に追加された医療用語は、看護領域を中心とした汎用の辞書（ComeJisyo）や患者向け文書等から抽出した用語で辞書に未登録のものである。これらの用語から成る西洋医学の医療用語辞書に加えて、東洋医学の研究者が作成した辞書（使用許可を取得済み）を基に東洋医学用語辞書を作成した。これら2種類の辞書を個別に、又は、統合して点訳精度や自動点字翻訳プログラムの点訳誤りを解析した。

(3) 疾患名や薬剤名等の医療用語は、長い文字列の複合名詞があるため、分かち書きの区切り誤りが原因で eBraille の点訳精度が低下する場合があった。そこで、分かち書き精

度向上の手法を解明することを目的として、汎用のチャンカープログラム（YamCha）を用いて、正しい分かち書きを統計的に学習するモデルを試作した。次に、この学習モデルの分かち書きの評価や解析に関する予備実験を行い、モデルの有効性や問題点を検討した。

## 4. 研究成果

(1) 本学医学部附属病院の看護部の協力で提供された、病状や手術に関する説明書、入院診療計画書、退院療養計画書等を用いて医療文書評価用コーパスを合計 5462 文作成した。その際、各種文書に含まれている個人情報情報を削除した。コーパスの点字文は、墨字文を eBraille で点訳し、人間が点訳の誤りを修正したものである。なお、この点字文は片仮名で表記され、かつ分かち書きされているため、この点字文と自動点字翻訳プログラムの点訳結果を比較して点訳精度を算出した。

(2) 辞書と点訳精度の関係を解析するため、次の4種類の辞書を準備した：eBraille に組み込んでいる形態素解析器「茶筌」と共に配布されている汎用の辞書、医療用語 8170 語のみを追加した辞書、東洋医学用語 17290 語のみを追加した辞書、全ての用語を追加した辞書である。そして、これらの辞書を使用する4種類の eBraille として、新聞記事などの一般的な文書点訳用の eBraille、医療文書点訳用の eBraille（「医 Braille」）、東洋医学文献等の点訳用の eBraille（「東医 Braille」）、全辞書を用いた eBraille、を作成した。次に、我々が作成した医療文書評価用コーパスを用いて、4種類の eBraille の点訳精度を解析した。その結果、医療文書の点訳では、辞書の規模（収録語数）に係らず医療用語のみを追加した辞書を用いた場合に点訳精度が最も高かった。このことから、点訳対象の文書に適した語彙構成の辞書が点訳精度を向上させることが示唆された。なお、医 Braille と東医 Braille は eBraille と共にウェブページに公開し、誰でも使用可能にしている。

医療用語の汎用辞書等から抽出された用語は 8170 語であった。加えて、専門的な医療用語を追加すべく本学医学部附属病院の美容外科分野の協力で、合計 1205 文の論文や総説を解析した。eBraille が使用する辞書の収録語と照合し未知語（辞書に未登録の語）を 53 語抽出し、医療用語辞書に追加した。

(3) 分かち書きの学習モデルの評価と精度解析の予備実験を行うため、648 文から成る学習用のコーパスを作成した。このコーパスは医療文書を形態素解析したもので、学習モ

デルは品詞や読みなどの情報を基に正しい分ち書きを学習する。この実験の結果、形態素の情報を使用した学習モデルは、複合名詞の中でも、仮名2文字目での区切りが正解となる分ち書きの精度が高いことが示唆された。また、大規模な学習用のコーパスを作成して、学習に用いる情報を検討する必要性も判明した。

医療文書評価用コーパスを用いて eBraille の点訳誤りを解析した結果、分ち書きの誤りの原因の一つに汎用辞書の収録方式の影響が考えられた。その収録方式とは、複合名詞が一語として登録されていることである。自動点字翻訳プログラムに正しい分ち書きを出力させるためには、これらの複合語を分割して辞書に再登録する必要がある。加えて汎用辞書の仕様に従い、複合語分割後に収録語の形態素生起コストを付与する必要があるが、そのためにはコスト付与のための形態素解析済みコーパスが必要である。また、今後も医療用語は増加する可能性が高く、辞書への用語追加など取り組みの継続が必須であり、今後の課題となった。

本研究で研究中の eBraille は、兵庫県より平成 21 年度の「グッドデザインひょうご」選定を受け、「ユニバーサルデザイン賞」を受賞した。加えて、財団法人日本産業デザイン振興会が主催しているグッドデザイン賞（平成 22 年度）を受賞した。

#### 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 8 件）

- ① 三浦研爾、菅野亜紀、長野 仁、山瀬健治、大田美香、小田 剛、後藤修司、西尾久英、松尾雅文、前田英一、西本 隆、高岡 裕、オントロジーによる漢方概念の研究、漢方と最新治療、査読有、20 巻、2 号、2011、161-167
- ② Aki Sugano, Mika Ohta, Tsuyoshi Oda, Kenji Miura, Shuji Goto, Masako Matsuura, Eiichi Maeda, Toshiko Ohshima, Yuji Matsumoto, Yutaka Takaoka, eBraille: a web-based translation program for Japanese text to braille, Internet Research, 査読有, Vol.20, No.5, 2010, 582-592
- ③ 菅野亜紀、大田美香、三浦研爾、松浦正子、松本裕治、大島敏子、高岡 裕、統計的学習モデルによる分ち書き解析器の自動点訳での有効性の解析、信学技報（電子情報通信学会技術研究報告）、査読無、109 巻、467 号、2010、5-8
- ④ 高岡 裕、菅野亜紀、大田美香、自然言語処理技術の医療現場への応用：視覚障害を有

する患者へのサービス向上と社会復帰支援、言語処理学会第 16 回年次大会発表論文集、査読無、2010、868-871

- ⑤ 菅野亜紀、大田美香、村井勇介、三浦研爾、相良かおる、松浦正子、池上峰子、前田英一、松本裕治、大島敏子、高岡 裕、病院での医療文書点訳に特化した自動点訳プログラム「医 Braille」の開発、言語処理学会第 16 回年次大会発表論文集、査読無、2010、615-618
- ⑥ Aki Sugano, Kenji Miura, Mika Ohta, Mineko Ikegami, Sumiyo Hanaoka, Eiichi Maeda, Masayuki Asahara, Yuji Matsumoto, Masako Matsuura, Masafumi Matsuo, Toshiko Ohshima, Yutaka Takaoka, Development of Japanese-into-Braille Translating Program for Medical Information “eBraille”, Proceedings of Asia Pacific Association for Medical Informatics 2009 (APAMI 2009), 査読有, 2009, 19
- ⑦ Kenji Miura, Aki Sugano, Hitoshi Nagano, Mika Ohta, Masafumi Matsuo, Yutaka Takaoka: A Trial of Ontology Construction for Acupuncture and Moxibustion. Asia Pacific Association for Medical Informatics 2009 (APAMI 2009), 査読有, 2009, 56

〔学会発表〕（計 14 件）

- ① 菅野亜紀、自動点字翻訳に用いる辞書の語彙構成と点訳精度の解析、第 31 回医療情報学連合大会（第 12 回日本医療情報学会学術大会）、2011 年 11 月 21 日、鹿児島市民文化ホール（鹿児島県）
- ② 菅野亜紀、自動点字翻訳プログラム～点字による医療情報提供を目指して～、日本音響学会東海支部・平成 23 年度技術講習会（招待講演）、2011 年 11 月 18 日、愛知工業大学 本山キャンパス（愛知県）
- ③ 三浦研爾、東医 Braille：漢方文献向けの自動点字翻訳プログラム、平成 23 年度日本東洋医学会関西支部例会、2011 年 10 月 30 日、大阪国際交流センター（大阪府）
- ④ Aki Sugano, Providing Medical Information in Braille: Research and Development of Automatic Braille translation program for Japanese “eBraille”, International Conference on Medical Informatics and Biomedical Engineering (ICMIBE 2011), 2011 年 10 月 27 日, Ramada Resort Benoa (Bali, Indonesia)
- ⑤ 菅野亜紀、点字による情報提供に向けた自動点字翻訳プログラム eBraille の研究開発、第 12 回日本医療情報学会看護学術大会、2011 年 7 月 17 日、神戸商工会議所会館（兵庫県）
- ⑥ 菅野亜紀、統計的学習モデルに分類器を組み合わせた自動点字翻訳プログラムの分ち

書き精度の解析、2010年度HCGシンポジウム、2010年12月17日、宮崎フェニックスシーガイアリゾート（宮崎県）

⑦ 高岡 裕、eBraille: 日本語自動点訳プログラムと英英点字翻訳プログラム、2010年度HCGシンポジウム、2010年12月17日、宮崎フェニックスシーガイアリゾート（宮崎県）

⑧ Aki Sugano, Characteristics of logical structures in acupuncture and Kampo medicine (Japanese traditional medicine) analyzed by ontology construction, Society for Acupuncture Research 2010 International Conference, 2010年3月19日, Sheraton Chapel Hill Hotel (USA)

⑨ 菅野亜紀、統計的学習モデルによる分かち書き解析器の自動点訳での有効性の解析、第52回福祉情報工学研究会、2010年3月12日、NTT武蔵野研究開発センター（東京都）

⑩ 高岡 裕、自然言語処理技術の医療現場への応用：視覚障害を有する患者へのサービス向上と社会復帰支援、言語処理学会第16回年次大会、2010年3月11日、東京大学（東京都）

⑪ 菅野亜紀、病院での医療文書点訳に特化した自動点訳プログラム「医 Braille」の開発、言語処理学会第16回年次大会、2010年3月10日、東京大学（東京都）

⑫ Aki Sugano, Development of Japanese-into-Braille Translating Program for Medical Information “eBraille”, Asia Pacific Association for Medical Informatics 2009 (APAMI 2009), 2009年11月22日, 23日, International Conference Center Hiroshima (Japan)

[その他]

自動点字翻訳プログラム eBraille の Web ページ

<http://ebraille.med.kobe-u.ac.jp/>

## 6. 研究組織

### (1) 研究代表者

菅野 亜紀 (SUGANO AKI)

神戸大学・医学部附属病院・学術推進研究員（特務助教）

研究者番号：20457039

### (2) 研究協力者

高岡 裕 (TAKAOKA YUTAKA)

神戸大学・医学部附属病院・准教授

研究者番号：20332281

大島 敏子 (OHSHIMA TOSHIKO)

神戸大学・大学院医学研究科・非常勤講師

研究者番号：80403244

松浦 正子 (MATSUURA MASAKO)

神戸大学・医学部附属病院・看護師

研究者番号：30379440

花岡 澄代 (HANAOKA SUMIYO)

神戸大学・医学部附属病院・看護師

研究者番号：10437486

池上 峰子 (IKEGAMI MINEKO)

神戸大学・医学部附属病院・看護師

研究者番号：80379453

一瀬 晃洋 (ICHINOSE AKIHIRO)

神戸大学・医学部附属病院・特命准教授

研究者番号：90362780

前田 英一 (MAEDA EIICHI)

神戸大学・医学部附属病院・特命教授

研究者番号：70322196