

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 15 日現在

機関番号：14501

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24590609

研究課題名(和文)医療文書の自動点字翻訳を実現する辞書と固有表現点訳システムの研究開発

研究課題名(英文) Study for the accuracy improvement and a supporting program for Japanese braille translation of medical documents

研究代表者

菅野 亜紀 (SUGANO, AKI)

神戸大学・医学部附属病院・特命助教

研究者番号：20457039

交付決定額(研究期間全体)：(直接経費) 4,200,000円

研究成果の概要(和文)：医療用語を多数含む医療文書の自動点字翻訳時の点訳精度向上法の解明を目指して、我々が研究開発した自動点字翻訳プログラムが使用している辞書や形態素解析ソフトを更新し点訳エンジンを改良した。また、固有表現の点訳修正支援法の確立を目標として医療文書を解析し、病院職員を対象とした点訳修正支援インターフェースを作成した。

研究成果の概要(英文)：To improve the accuracy of our automatic Japanese braille translation program for medical documents which are provided to patients in hospital, we added more medical terms to the dictionary, replaced the morphological analyzer in our braille translation program and revised our braille translation engine. In addition, we made an interface for our braille translation program to choose the correct braille translation for named entity, that is proper nouns such as personal names or organization names in medical documents.

研究分野：医療情報学

キーワード：自動点字翻訳 医療文書 辞書

1. 研究開始当初の背景

我々は、患者に手渡す医療文書の自動点訳を目的として、自動点字翻訳プログラム eBraille を研究開発し、インターネット上で公開している (<http://ebraille.med.kobe-u.ac.jp/>)。これまでに電子カルテから医療用語を抽出して辞書に追加し、医療文書の点訳精度を向上させてきた。加えて eBraille の点訳エンジンの改良や新プログラムの研究などの基礎的な技術研究に取り組んできた。これらの研究の結果、eBraille の医療文書の点訳精度は他の点字翻訳プログラムと比較して高くなった。今後、更に点訳精度を向上させるには、点訳処理の前段階に行っている形態素解析の際に医療文書用に最適化すること、辞書へ医療用語を更に追加すること、が考えられる。前者については、eBraille が使用している辞書の収録語の形態素生起コストの最適化である。形態素生起コストは、文書内の単語の生起頻度を基に形態素解析済みのコーパス(言語資料)を用いて統計的に算出した値をいう。現在は、新聞記事を基にしたコーパスで形態素生起コストの値を算出しているが、医療文書の正確な点訳には、形態素生起コスト算出に用いる医療文書を基にしたコーパスが必要であると考えられる。他に、従来とは別の形態素解析ソフトの使用を検討し、医療文書の点訳精度を解析する方法もある。後者については、辞書に未収録の医療用語を追加することで点訳精度の向上が可能である。

また、自動点字翻訳などの言語処理においては固有表現(人名や組織名などの固有名詞)が問題となる場合が多い。固有表現は、漢字の読み誤りの主要因となっており、辞書への単語の追加では本質的な解決が不可能である。プログラムでの対応も困難であり、eBraille の自動点字翻訳の過程で人間が漢字の読みを確認・修正することが実用的といえる。実際に本学の医学部附属病院では、eBraille で点訳した点字文書を患者へ渡す前に申請者が精査して固有表現等の点訳誤りを修正している。

そこで本研究では、形態素解析の辞書収録語の最適な形態素生起コストの算出、形態素解析ソフトの更新を検討し医療用語の点訳精度向上を目指すと共に、人名、組織名、薬剤・薬品名、疾患名などの固有表現の出現頻度等を解析し、読みの修正を促す支援インタフェースを研究開発する。以上の研究成果を eBraille プログラムに追加することで、医療文書の点訳精度の更なる向上が期待されると考えた。

2. 研究の目的

医療現場での自動点字翻訳プログラムの実用化と点字文書作成の効率化が点字文書作成の普及に向けた課題であると考え、本研究では、医療用語を多数含む医療文書の自動点字翻訳時の点訳精度向上法の解明および

固有表現(人名や組織名などの固有名詞)の点訳修正支援法の確立、を目指す。

3. 研究の方法

(1) はじめに、我々が作成した自動点字翻訳プログラム eBraille を形態素解析ソフト MeCab(<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>)に対応させたバージョンを作成した。その理由は、従来 eBraille が使用していた形態素解析ソフト ChaSen(<http://chasen.naist.jp/hiki/ChaSen/>)を用いており、よりも解析速度が速い等の利点があるためである。次に、MeCab を用いた eBraille と ChaSen を用いた eBraille で医療文書の点訳結果を比較する実験を行った。なお、点訳精度の算出は既報の手法(Sugano et al. *Int. Res.*, 20(5), 2010)を用いた。

(2) eBraille で使用する汎用辞書 IPADIC を再構築し、歯科領域の専門用語を追加した。具体的には、辞書内の複合名詞で、元々一語として登録されている単語を分割して再登録した。また、医療情報システム開発センターが提供している歯科病名マスター、歯科手術・処置マスター等に含まれる用語から、MeCab を用いて辞書に未収録の用語を決定した。加えて、MeCab が単語の区切りに失敗する用語も抽出し、これらの語を辞書に追加した。そして、その辞書を eBraille が利用可能なように実装した。

(3) 医療文書を収集し、文書に含まれる単語の中で組織名など、病院職員が容易に読みや区切りを修正可能と考えられる固有表現を対象に形態素解析し、固有表現の修正支援のインタフェースを試作した。

4. 研究成果

(1) MeCab 版 eBraille の研究開発では、eBraille の点訳エンジンである KUIC のプログラム内の品詞 ID を MeCab の出力に合わせる修正が必要だった。ChaSen 版 eBraille と MeCab 版 eBraille の点訳について、点字毎日の記事 474 文で比較解析したところ、従来の ChaSen 版 eBraille では誤訳となっていた一部の単語の品詞や読みが MeCab 版で正しく解析されていた。一方で、MeCab が半角の英数字や記号を誤って解析していたため、これらについては eBraille へのプログラム追加で対応した。また、MeCab 版 eBraille の方が点訳精度が高かった。

(2) eBraille の辞書の再構成により、複合名詞(例:「神戸大学」)がその構成要素である名詞(例:「神戸」および「大学」)に分割されたため、辞書の仕様に起因する誤訳を防ぎ、正確な自動点字翻訳を可能にした。歯科領域の用語は 1,638 語追加し eBraille が利用可能なように実装した。なお、この

eBraille をスタンドアロン型で組み込んだ、視覚障害の患者に歯科治療に関する点字文書や口腔内の状態を触図で提供可能なシステム Dentact を、大阪大学歯学部森崎市治郎教授の研究チームと共同で研究開発した。現在、このシステムは大阪大学歯学部附属病院で実運用されており、患者への情報提供に使用されている。

(3) 収集した医療文書に含まれる単語の中で組織名など病院職員が容易に読みや区切りを修正可能と考えられる固有名詞を対象に解析した結果、出現頻度が低い、又は読みが困難な人名については、病院情報システムの電子カルテの患者情報との連携が有効であることが予想された。また、固有表現の点訳修正支援プログラムとして、プルダウンで修正候補を表示させるインタフェースを試作した。このプログラムの使いやすさについて病院職員に意見を求めたところ、修正候補の数や修正完了までの画面の数の多さについて指摘を受けた。

本研究では、MeCab 版 eBraille を開発し、従来の eBraille よりも点訳精度が高くなった。MeCab には単語生起コスト推定モジュールが付属している。今回、MeCab 版の方が点訳精度が高かったが、より規模の大きいコーパスを用いて、医療文書で形態素生起コストを最適化した辞書を用いた ChaSen 版との比較が課題となった。

また、固有表現の点訳修正支援の操作性に関しては、ユーザによって意見が異なる場合があるため、より多くの意見を収集する必要がある等が今後の検討課題となった。

なお、(2) で研究開発した歯科医療情報提供システム Dentact は、2013 年度のグッドデザイン賞（公益財団法人日本デザイン振興会）を受賞した。また、この研究成果を第 15 回日本クリニカルパス学会学術集会で発表したところ、学術集会優秀賞と座長賞を受賞した。

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 7 件)

- ① 財間 達也、村上 旬平、三浦 研爾、萱野 亜紀、高岡 裕、森崎 市治郎、歯科領域における、視覚障害のある人への合理的配慮について、信学技報(電子情報通信学会技術研究報告)、査読無、114 巻、217 号、2014、45-49
- ② 梅田 由紀恵、萱野 亜紀、池上 峰子、関口 紗代、大田 美香、松浦 正子、熊岡 穰、前田 英一、高岡 裕、点字文章表現に適した構文構造の解析、査読無、114 巻、217 号、2014、51-53

- ③ 高岡 裕、財間 達也、香川 泰俊、村上 旬平、三浦 研爾、大田 美香、前田 英一、萱野 亜紀、森崎 市治郎、点字と点図による歯科疾患管理文書提供システムの研究開発、医療情報学、33(Suppl.)、2013、710-712
- ④ 萱野 亜紀、梅田 由紀恵、三浦 研爾、大田 美香、塚本 紗代、池上 峰子、花岡 澄代、一瀬 晃洋、松浦 正子、前田 英一、高岡 裕、自動点字翻訳時の地名の読み精度向上のためのアルゴリズムの研究、査読有、医療情報学、33(Suppl.)、2013、1238-1239
- ⑤ 高岡 裕、萱野 亜紀、財間 達也、村上 旬平、森崎 市治郎、歯科領域に特化したスタンドアロン型の自動点字翻訳プログラムの研究開発、査読無、障害者歯科、34 巻 3 号、2013、334
- ⑥ 萱野 亜紀、三浦 研爾、大田 美香、喜多 伸一、山口 俊光、渡辺 哲也、前田 英一、高岡 裕、自動点字翻訳プログラムの触地図用途向け最適化、査読無、信学技報(電子情報通信学会技術研究報告)、112 巻、472 号、2013、285-288
- ⑦ 高岡 裕、村上 旬平、飯塚 潤一、喜多 伸一、渡辺 哲也、萱野 亜紀、塚本 紗代、城 茂治、森崎 市治郎、医療における視覚障害者への情報保障、査読有、医療情報学、32(Suppl.)、2012、260-261

[学会発表] (計 10 件)

- ① 萱野 亜紀、財間 達也、村上 旬平、三浦 研爾、高岡 裕、森崎 市治郎、視覚障害の患者用パス: 歯科診療情報提供システムの構築と運用、第 15 回日本クリニカルパス学会学術集会、2014. 11. 14、清風荘(福井県)
- ② 財間 達也、村上 旬平、三浦 研爾、萱野 亜紀、高岡 裕、森崎 市治郎、歯科領域における、視覚障害のある人への合理的配慮について、電子情報通信学会 第 74 回福祉情報工学研究会、2014. 9. 13、神戸大学楠地区キャンパス(兵庫県)
- ③ 梅田 由紀恵、萱野 亜紀、池上 峰子、関口 紗代、大田 美香、松浦 正子、熊岡 穰、前田 英一、高岡 裕、点字文章表現に適した構文構造の解析、電子情報通信学会 第 74 回福祉情報工学研究会、2014. 9. 13、神戸大学楠地区キャンパス(兵庫県)
- ④ 財間 達也、村上 旬平、三浦 研爾、萱野 亜紀、高岡 裕、森崎 市治郎、視覚障害のある人への「合理的配慮」実現に向けた歯科領域での取り組み、第 18 回日本医療情報学会春季学術大、2014. 6. 6、岡山コンベンションセンター(岡山県)
- ⑤ 高岡 裕、財間 達也、香川 泰俊、村上 旬平、三浦 研爾、大田 美香、前田 英一、萱野 亜紀、森崎 市治郎、点字と点図による歯科疾患管理文書提供システム

の研究開発、第 33 回医療情報学連合大会、
2013. 11. 23、神戸ファッションマート（兵庫
県）

- ⑥ 菅野 亜紀、梅田 由紀恵、三浦 研爾、
大田 美香、塚本 紗代、池上 峰子、花
岡 澄代、一瀬 晃洋、松浦 正子、前田
英一、高岡 裕、自動点字翻訳時の地名の
読み精度向上のためのアルゴリズムの研
究、第 33 回医療情報学連合大会、
2013. 11. 22、神戸ファッションマート（兵
庫県）
- ⑦ 高岡 裕、菅野 亜紀、財間 達也、村上
旬平、森崎 市治郎、歯科領域に特化した
スタンドアロン型の自動点字翻訳プログ
ラムの研究開発、第 30 回日本障害者歯科
学会総会および学術大会、2013. 10. 13、神
戸国際会議場（兵庫県）
- ⑧ 高岡 裕、点字と触図による視覚障害者支
援システムの研究開発、ICT イノベーション
フォーラム 2013、2013. 10. 1、幕張メッ
セ国際会議場（千葉県）
- ⑨ 菅野 亜紀、三浦 研爾、大田 美香、喜
多 伸一、山口 俊光、渡辺 哲也、前田
英一、高岡 裕、自動点字翻訳プログラ
ムの触地図用途向け最適化、電子情報通信学
会 第 67 回福祉情報工学研究会、
2013. 3. 12、福岡工業大学（福岡県）
- ⑩ 菅野 亜紀、神大病院の視覚障害者への情
報保障体制、第 32 回医療情報学連合大会
ワークショップ 2 障害者（主に視覚障害
者）に対する医療での情報保障（招待講演）、
2012. 11. 16、朱鷺メッセ新潟コンベンシ
ョンセンター（新潟県）

〔その他〕

自動点字翻訳プログラム eBraille
<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・特命助教
研究者番号：20457039

(2) 研究分担者

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

一瀬 晃洋 (ICHINOSE, Akihiro)
神戸大学・医学部附属病院・准教授
研究者番号：90362780

(4) 研究協力者

大田 美香 (OHTA, Mika)
神戸大学・医学部附属病院・学術研究員
研究者番号：20274706