

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

菅野亜紀¹, 大田美香¹, 村井勇介¹, 三浦研爾¹, 相良かおる², 松浦正子³,
池上峰子³, 前田英一⁴, 松本裕治⁵, 大島敏子³, 高岡 裕¹

1 神戸大学大学院医学研究科内科系講座小児科学分野ゲノム医療実践学部門,

2 西南女学院大学保健福祉学部栄養学科, 3 神戸大学医学部附属病院看護部,

4 神戸大学医学部附属病院医療情報部,

5 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座

E-mail: sugano@med.kobe-u.ac.jp

要旨 我々は、誰でも簡便に点字文書の作成を可能にする自動点訳プログラム「eBraille」を開発しているが、病院で視覚障害者へ渡す文書の点訳には更なる改良が必要だった。その理由は、自動点訳プログラムが使用する辞書中の検査名、疾患名、薬剤名などの医療用語が不十分なためである。加えて、(1)「骨(ほね)」を「骨(こつ)」と読むような例、(2)数多くの複合名詞の存在、が原因で汎用の辞書では正しい点訳が困難だった。そこで今回、医療用自動点訳プログラム「医 Braille」を医療用語約 8,000 語の辞書への追加により開発した。本発表では、解析した点訳精度と今後の課題について報告する。

キーワード 点訳, 医療用語, 辞書

1. 背景

先端医療は個別化医療に向かって進展しており、全ての患者に対して必要な情報提供をしなければならない。我々は、視覚障害者に対して点字による医療情報の提供を実現すべく、自動点字翻訳プログラム” eBraille”を開発した。その結果、一般的な文書の自動点字翻訳では実用的なレベルを達成したが、医療文書の点訳精度はその約 8 割であった。その原因は、辞書に未登録の語（なかでも医療関連の用語）に起因する漢字等の読みの誤りに加えて、複合名詞等の分かれ書きの失敗だった。これらのことから、システムの辞書への医療用語の追加の必要性が示唆された[1]。

2. 医療用語の選定と辞書への追加

“eBraille”で使用している辞書（IPADIC）の拡張には、医療用語 30,146 語を収録している ComeJisyo[2]から選択した。ComeJisyo に収録された用語は、殆どが名詞（病名、薬品名、機器の名前、化学物質等）であり、読み仮名が 7 文字以上の長い複合名詞は全体の約 6 割を占めていた。日本点字表記法[3]によると、7 拍以上の長い複合名詞は触読の際の記憶の単位として適さず、複合語全体の意味の理解を考慮して分かち書きの有無を判断するように定めている[3]。そのため、複合名詞を構成要素に分割し、IPADIC の収録語と重複しない語だけを抽出した。次に、抽出した医療用語に浅原らの手法[4]で形態素生起コストを付与し拡張辞書を作成した。そして“医 Braille”と命名した。

追加した医療用語のうち、漢字は「類骨（ルイコツ）」「偽牲（ギセイ）」「咬筋（コウキン）」等、漢字 2～3 文字の語が大多数であった。読み仮名が 7 拍以上の語に「腕橈骨筋（ワントウコツキン）」等では更に小さな単位に分割不可能だったため、そのまま辞書に収録した。カタカナの医療用語は、人名が疾患名となったものもあり、姓と名に分けて収録した。

3. 医療文書の評価実験

IPADIC に医療用語 8,170 語を辞書へ追加し、“医 Braille”で医療文書の点訳精度を解析した。評価用コーパスは、本学附属病院の看護記録や患者向け文書、電子カルテ用標準病名マスター[5]、厚生労働省指定の特定疾患名を基に、墨字文（漢字仮名混じり文）と点字文（仮名表記）の対から成る医療文書コーパス 2,065 文を作成し評価に用いた。なお、点訳精度は分かち書き（単語や語句の境界を空白で区切る表記）の精度である F-measure と漢字等の読みの正解率の積、と定義し、

評価補助プログラム[6]で点訳結果を解析した。評価の結果，辞書拡張前よりも点訳精度が向上し（図 1），「屈筋断裂」，「亜急性裂肛」等のような多くの医療用語で読みと分ち書きの両方が改善した。

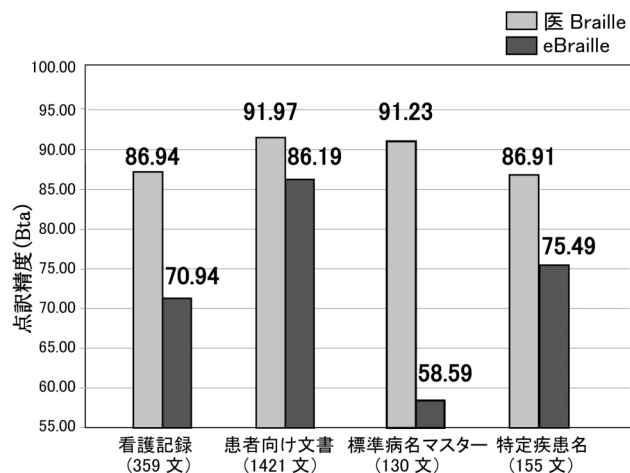


図 1. コーパス中の医療文書の種別毎の点訳精度

4. 考察

今回，医療分野で eBraille を活用するため，西洋医学の医療用語を中心に辞書を拡張した“医 Braille”を開発した。開発の結果点訳精度は飛躍的に向上したが，一部の複合名詞の医療用語では正しい分ち書きの出力に失敗した。このような失敗例の分ち書きの処理は、意味の理解や発音上の切れ目等を考慮する必要がある [3]，eBraille に実装した点字表記規則では対応不可能である。そこで，今後は統計的学習モデルに基づく分ち書き解析器を開発し，eBraille に組み込む予定である。なお，視覚障害者の中に多いあんま・指圧師，はり師・きゅう師の有資格者向けに，東洋医学文献の点訳に特化した“東医 Braille”も開発し公開している。以上の点訳プログラムは，我々の研究室のポータルサイト (<http://suzume.med.kobe-u.ac.jp/>) からアクセス可能である。

・ 謝辞

本研究は，科学研究費補助金若手研究(B)（課題番号 21790497），萌芽研究（課題番号 19659563），基盤研究(C)（課題番号 21592692），ひょうご科学技術協会奨励研究助成（21S092）による研究成果の一部である。

文 献

- [1] 菅野亜紀, 大田美香, 三浦研爾, 松浦正子, 高橋京子, 池上峰子, 前田英一, 松本裕治, 大島敏子, 高岡裕. “自動点訳サーバ eBraille の開発,” ヒューマンインタフェース学会研究報告集 vol.9, no.5, pp.93-98, 2007.
- [2] 相良かおる, 浅原正幸, 小野正子, 小作浩美. “形態素解析エンジン MeCab 用看護用語ユーザ辞書の作成と公開,” 医療情報学(Suppl.) 28, pp.938-939, 2008.
- [3] 日本点字表記法 2001 年版, 日本点字委員会 (編), 日本点字委員会, 2001.
- [4] 浅原正幸, 松本裕治. “形態素解析のための拡張統計モデル,” 情報処理学会論文誌, vol.43 no.03, pp.685-695, 2002.
- [5] MEDIS 標準マスター, 病名マスター(ICD10 対応電子カルテ用標準病名マスター), 財団法人医療情報システム開発センター, http://www.medis.or.jp/4_hyojyun/medis-master/index.html
- [6] 高岡裕, 菅野亜紀, 五十嵐大和, 三浦研爾, 大田美香, 小田剛, 松浦正子, 松本裕治, 大島敏子. “日本語自動点訳サーバ eBraille の改良と評価,” 医療情報学(Suppl.) 27, pp.1210-1213, 2007.