NAIST-IS-DD0161020

博士論文

医療文書の自動点字翻訳における精度向上法

菅野　亜紀

2010 年 9 月 24 日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学) 授与の要件として提出した学位論文である.

菅野　亜紀

審査委員：

松本　裕治教授　　(主指導教員)
松本　健一教授　　(副指導教員)
新保　仁准教授　　(副指導教員)

医療文書の自動点字翻訳における精度向上法*

菅野亜紀

内容梗概

点字は視覚障害者のための触読表音文字であり，主要な情報獲得手段の一つである．医療現場では，近年の個別化医療の進展に伴い患者個人に対する医療情報の提供が必要になりつつあるが，点字での情報提供への取り組みは進んでいない．従来開発されてきた自動点字翻訳プログラムは，点訳ボランティアの支援という立場で開発されており，しかも医療現場での使用は考慮されていない．そこで本研究では，点字翻訳の専門家と同等の高い点訳精度を実現する方法の解明を目的とする．その際，特に医療機関での使用を可能にすることを目指した．

最初に，自動点字翻訳プログラムの評価用コーパスとして，新聞記事を基にした通常文書評価用コーパスと特定機能病院の患者向けの文書を含む医療文書を基にした医療文書評価用コーパスを作成した．同時に評価指標として，分かち書きの精度（F_1），漢字の読みや記号の変換の正解率に加えて，これらの積を点訳精度と定義した．次に，日本点字表記法2001年版に記載された点字表記規則を分析した．加えて，我々の自動点字翻訳プログラムeBrailleの点訳精度と他の自動点字翻訳プログラムのそれとを比較した．更に，プログラムの辞書へ医療用語や東洋医学用語を追加した場合の点訳精度，Support Vector Machineに基づく統計的学習モデルを導入した分かち書きについて評価実験を行い，点訳精度の向上に有効な手法の解明に取り組んだ．

日本点字表記法の分析の結果，点字表記規則は，文字種，品詞，活用形，出現形，読み，音韻変化，モーラ（拍）の数を指標として作成する必要があることが明らかとなった．これらを指標とした点字表記規則を作成しeBrailleの点訳エンジンに実装した結果，eBrailleの点訳精度は他の自動点字翻訳プログラムよりも有意に高かった．また，点訳精度の向上には点訳対象の文書に適した辞書の語彙構成にすることが有効であることが示された．更に，形態素解析結果とモーラの数を学習素性とした統計的学習モデルと

点訳エンジンの分かち書き出力を学習素性に追加した統計的学習モデルにルールベースの分類器を組み合わせた手法が, 医療文書の分かち書きへの適用に有用であることが示された.

キーワード

自動点字翻訳，辞書，医療用語，統計的学習モデル

*奈良先端科学技術大学院大学情報科学研究科情報処理学専攻学位論文, NAIST-IS-DD0161020, 2010 年 9月24日

Accuracy Improvement of Automatic Japanese-to-braille Translation for Medical Information*

Aki Sugano

Abstract

Braille is phonograms of six tactile raised dots which is used for a reading and writing system for the blind and the partially sighted. It is recognized as one of the important ways for them to get information in the world. With recent advancement of personalized medicine, to provide medical information to individual patient is becoming more necessary in medical institutions, however such efforts are not made very much. Previous braille translation programs were developed to help braille translation volunteers. Furthermore, they were not considered for use in medical offices. This study aims to elucidate effective methods for higher accuracy for a braille translation program, which especially aims at the use in medical institutions, so that even people with no knowledge of the braille writing system can easily create braille documents.

We first made ordinary text corpus from newspaper articles and medical text corpus to evaluate braille translation programs. We then defined accuracy for word segmentation (F_1), *Kana* and symbols translation and braille translation accuracy. To elucidate the factors for braille transcription rules, we analyzed the latest Japanese braille transcription rules 2001. Next, we analyzed the braille translation accuracy of our program, followed by the comparison with other braille translation programs. We then evaluated our program when the dictionary is expanded with medical words or words of oriental traditional medicine and performed experiments of statistical learning model for braille word segmentation to clarify effective approaches for higher accuracy.

The result of our analysis showed that Japanese braille transcription rules are based on the types of character systems, part of speech, conjugation forms, surface forms, *Kana* translation of *Kanji*, phonological changes and number of moras for every word. Using these elements, we implemented braille transcription rules on our braille translation engine. As a result, our braille translation program achieved the higher translation accuracy than those of other braille translation programs. In addition, our experimental results suggested that, to increase the translation accuracy, we should make the composition of the dictionary suitable for the texts to translate and an effective approach for word segmentation of medical texts was to combine two statistical learning models one of which included outputs of our braille translation engine as an additional feature by using a rule-based classifier.

Keywords:

Automatic braille translation, Dictionary, Medical words, Statistical learning model

*Doctoral Dissertation, Computational Linguistics, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0161020, September 24, 2010.

目 次

図目次

表目次

第1章

序　論

1.1　研究の背景

点字は，1825年に盲目のフランス人のLouis Brailleが，視覚障害者用の記述文字として創出した6個の点から成る触読文字である．1878年にパリで開催された国際会議（Universal Congress for the Amelioration of the Blind and Deaf-mutes）において，世界的に受け入れられた [1]．日本では，東京盲唖学校（当時）の教員であった石川倉次が，Louis Brailleの点字を日本語に対応させた日本語用の点字を考案し，1890年に東京盲唖学校が採用して現在の日本語の点字の基礎となった [2]．その後，1901年の官報に「日本訓盲点字」として公表されており，視覚障害者の文書による情報獲得の主要な文字として使われている．

現在の日本では，点字は6点1マスの大きさは，縦約6mm×横約3.5mmである．他に，縦約7mm×横約4mmのものもあり，これは中途視覚障害者の触読性向上を目的としている．点字文書には専用の点字用紙を用いる．この紙は通常の紙より粘着力があるため，凸点を作っても穴があきにくく，点字文書作成に適している．この紙に，点筆と専用の板（点字器）を用いて凸点を打つ．又は，点字プリンタで点字文書を作成することも行われる．さらに点字ディスプレイという視覚障害者用の触読ディスプレイがあり，点字の6点の凹凸を，ピンを機械的に上下させることで表示する．この点字ディスプレイを用いることで，点字を連続的に可変させた表示が可能である．点字ディスプレイはパソコンへの接続により電子データの表示も可能になり，視覚障害者用の装置やプログラム開発が進んでいる．前者の例には，点字ディスプレイとカラオケを連動させる「点字カラオケ[1]」という装置が，後者の例には点字自己学習用の触読点字e-learning [3]がある．

諸外国の点字への社会的な取り組みについて以下に記述する．まずEUに於いては，医薬品や医療材料全般への点字による情報提供が義務づけられている[4, 5]．さらにEUは，

[1] http://www.telesoft.co.jp/b/karaoke.htm

eTENというヨーロッパの公共の利益のための電子サービスを助成する目的で，RoboBrailleというプロジェクトへの助成も行っている [6, 7]．RoboBrailleとは，非営利目的の電子メールサービスである[2]．利用者が電子テキストデータを送信すれば，点字に翻訳したデータ又は音声を合成したデータが返信される．米国では，リハビリテーション法（Rehabilitation Act, 1973）により，政府刊行物の点字による提供が義務化されている．さらにリハビリテーション法第508条の中では，連邦政府の機関に対する障害者の電子情報技術へのアクセシビリティ向上のための実施項目が定義されており，これも義務とされている．具体例には，ウェブ上の情報やアプリケーションは点字ディスプレイによる表示を可能にするという義務である．

次に，我が国の点字への取り組みの現状を記述する．現在，一部の郵便局とレストラン，官公庁や自治体のウェブページで点字による情報提供への取り組みが実施されているが，義務化されている訳ではなく対応出来ているとはいえない．数少ない普及の例としては，一部の民間企業が業界内でアルコール飲料に「おさけ」や「びーる」と任意で点字表示しているものがある．点字表示と同様の試みで，シャンプーとリンスを区別する印も企業が考案し，これらは実際に我々が日頃から目にすることが可能なものである．

ここで病院での現状に話を移す．医療現場と視覚障害者向けのアクセシビリティ向上に向けて，「視覚障害者等に対する服薬指導について」（厚生省 1998 年 8 月 19 日 政医第 289 号，厚生省保健医療局）が出されている．その中では，個々の視覚障害者の状況に応じた適切な配慮，例えば，薬袋（薬を入れる袋）への点字による記載事項の表示，などの努力目標が通達として出された．これは医療分野以外の社会的状況と同様に，義務化されている訳ではない．ところで近年，患者との合意に基づく医療の進展に伴い，治療や検査等に関する説明を口頭と文書で行うようになり，多くの患者向けの文書が手渡されるようになった．例えば，患者へ渡す文書のうち患者ごとに個々で内容が異なるもの，インフォームドコンセント，クリニカルパスなどの疾患治療の説明書，入退院に関わる計画や指導書，検査や病状の説明書，副作用に関する説明書，各種同意書などの医療情報が該当する．さらに，今後普及していくと予想される高度先端医療では，個別化医療（テーラーメイド医療）が前提となるため，患者個人に対する医療情報の提供が重要となる．このような患者に対する疾患などの説明は，基本的に患者本人に対して行う．実際，病院では晴眼者（健常者）にその対応を行っている．しかし患者が視覚障害者の

[2] http://www.robobraille.org/

場合には，付き添い者と患者が一緒に診察室に入り同席する場合が多い．この場合の付き添いは，患者の家族であることが多い．しかし，2005 年に我が国で個人情報保護法[8]が全面的に施行された結果，法的には患者の家族でも第三者として扱うことになった．すなわち，いかなる場合も患者が同席を望んだ場合を除き，患者本人にのみ説明することが原則となった．その結果，説明の際に手渡される文書は点字である必要が生じた．その理由は，視覚障害の患者が後で読むことが可能な文書は，点字による印刷物だからである．これらの例は，個人情報保護につながる視覚障害者対応としての課題として，視覚障害を有する患者向けに点字文書を渡す必要性を示唆している．既に「平成 12 年度社会保険診療報酬改定等の概要」（2000 年 2 月 29 日 厚生省保険局医療課）において，点字等を用いた薬剤情報提供が，診療報酬として 10 点（100 円）又は 15 点（150 円）が加算可能となっている．しかし，診療報酬は 100 円又は 150 円にすぎないので，この費用で病院が点字文書を提供するのは困難である．仮にコストを度外視して病院外へ医療文書の点字翻訳を委託した場合には，患者の氏名などの個人情報に加えて，疾患，病状，薬などの情報を外部に出す必要が出てくる．これは，個人情報保護法に照らすと望ましくない状況である．更に，業務で多忙な病院職員が，点字を習得して対応するのは現実的には不可能である．しかも，我が国のみならず世界的に点訳者不足は深刻な問題である[9, 10]．以上の問題は，点字の知識がなくても，単に点字文書を作成できるフリーの自動点字翻訳プログラムがあれば解決する．実際，欧米では点字翻訳プログラムが開発，販売されている．勿論，我が国でも幾つかの点字翻訳プログラムが開発，販売されている．

ところで，1990 年代半ばに，パーソナルコンピュータの規格の一つで，50%以上のシェアを占めていた NEC による日本独自の PC98 規格のコンピュータの製造が中止される出来事があった．当時の視覚障害をはじめとする障害者向けのプログラムや周辺機器は全てこの PC98 規格であったため，多大な影響を受けるに至った．というのも，ユーザーとなる障害者の数は，汎用のプログラムが対象としている数に比較して少なく，速やかなプログラムの再開発は困難だからである．再度新たに開発するには，公的支援を得る必要があり，そのため時間もかかる．15 年以上を経て，現在では，全ての点字翻訳プログラムや周辺機器は Windows OS 上で動作可能になっている．

しかし，今の状況は Windows OS に動作環境が依存しているともいえる．そこで我々は，1990 年代半ばのような事態が生じても，殆ど影響を受けないようなプログラム形態が望ましいとの考えに至った．そして我々は，1997 年に開発された高岡らの自動点字翻

訳サーバ「eBraille0.81 [11]」を基に研究に取り組んだ．このプログラムは，形態素解析器の「ChaSen1.51」を利用したCGIアプリケーションであり，Webブラウザでの利用を前提としている．しかし，2001年に日本点字委員会が実施した日本点字表記法の改定[2]に未対応な部分を含め，点字翻訳精度（点訳精度）は高くない．とはいえ，プログラムのコンセプトに問題はなく，現行の日本点字表記法に対応させることで，点訳精度が向上すると予想された．加えて，更に点訳精度を高める方法を本研究で明らかにする．最後に，明らかにした知見をプログラムに実装することで，点字による医療文書を簡単に作成可能なプログラムを開発したいと考えて，この研究に取り組むことにした．

1.2 研究の目的

本研究の目的は点字翻訳の専門家と同等の高い点訳精度を実現する方法を明らかにすることである．具体的には，本研究により，

(1) 日本点字表記法に則った点字翻訳を実現可能な規則の作成とその実装方法
(2) 形態素解析の精度向上の実現方法
(3) 日本点字表記法のうち，複合語の「意味の境界」を用いた分かち書きの実現方法

の3点を明らかにする．
これらの解析結果を用いて，医療文書の高い点訳精度を実現したプログラムを開発する．

1.3 論文の構成

本論文の構成は，次の通りである．

この後，第2章では日本点字表記法の点字表記規則の概要を説明し，コンピュータプログラムで点字翻訳の実現が可能な方法について明らかにする．

第3章では，関連研究と他の自動点字翻訳プログラムについて説明する．

第4章では，自動点字翻訳プログラムの評価に必要なコーパスの作成と評価指標について記述する．

第5章では，我々の点字翻訳プログラムの設計について記述する．ここでは特に，日本語の点字表記規則である日本点字表記表記法2001年版の分析結果に基づく，点訳エンジンへの表記規則の実装の詳細を説明する．

続く6，7章では，点訳精度の向上に効果的な手法について解析した結果を示す．まず第6章では，自動点字翻訳プログラムが使用する辞書に焦点を置いて説明する．作成した3種類の拡張辞書と通常の辞書との分かち書き精度，読みの精度，点訳精度の比較解析結果から，辞書拡張による精度向上の効果を通常文書評価用コーパスと医療文書評価用コーパスの2種類のコーパスで解析し，明らかにする．

次に第7章では，統計的手法による分かち書きの精度向上を提案する．我々は各種の統計的学習モデルを作成し，6章までに作成した自動点字翻訳プログラムの分かち書きの精度と比較・精査して，統計的手法の応用による点訳精度向上の可能性を明らかにする．最後に第8章で結論を記述する．

第2章

日本語の点字

2.1　日本語の点字表記規則

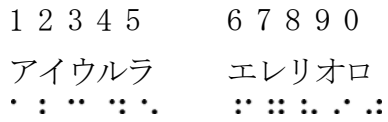
日本語の点字は，英語の点字と異なる．日本語の点字には，仮名文字（表音文字）に対応する点字や，仮名文字と区別するためにアルファベットの前に配置する符号，日本語で用いられる各種の記号を表す点字などがある．通常の日本語の文章は漢字仮名混じり文（表意文字と表音文字より成る）で記述するが，日本語の点字では音を表す表音文字を主として，前置点や分かち書きなどを用いて記述する[2]．前置点とは，数字，アルファベットなど別の文字体系や，日本語の言語音の濁音，半濁音，拗音，特殊音を表現するために，文字の前に配置する点字である．分かち書きとは，英語やヨーロッパ諸国の言語で単語と単語の間にスペースを入れるのと同様の記述で，点字文の理解を助けるために語句と語句の境界を空白で区切るもので，通常の日本語の文章にはない記述方法である [2, 12, 13]．このような日本語の点字表記規則は日本点字委員会が「日本点字表記法」（2001年）[2]として定めている．

2.1.1 仮名，数字，アルファベットの表記

次に，仮名や数字，アルファベットの点字表記について詳しく記述する．点字は，縦3点，横2列の6つの凸点の組み合わせで構成されており，日本ではこの単位に「マス」という名称が付けられている [2]．日本語の点字表記では，1マス6点のドットパターン63通りを3種類の文字体系と各種記号に適用する．そのため，50音の仮名は点字と1対1の対

応となっているが，数字やアルファベットに対応する点字は，仮名に対応する点字の一部と重複する．例えば，数字は，ア行とラ行の仮名の点字と同じである．

１２３４５　　６７８９０
アイウルラ　　エレリオロ



「ア[1]」に相当する点字に数符（）を前置すると「1」になる．すなわち「ア」は

と表記し，「1」は，数符の後に「ア」に相当する点字を続けて，

と表記する．
一方，アルファベットの点字は，下記の仮名の点字に対応する．

ａｂｃｄｅｆｇｈｉｊｋｌｍｎｏｐｑｒｓｔｕｖｗｘｙｚ
アイウルラエレリオロナニヌツタネテチノトハヒソフムマ

例えば「a」は，外字符（：）を前置して以下のように表記する．

以上の通り，点字の重複があるため，数字の後にア行又はラ行の仮名が続いて１つの単語を形成する場合は，誤読を防ぐためにつなぎ符という点字を配置する．例えば，「1

[1] 点字は平仮名と片仮名の区別がないため，本論文では，仮名を片仮名表記に統一する．

エン（1円）」の場合，「エ」は数字の「4」と同じ点字に対応するため，「1」と「エ」の間につなぎ符（ ⠤ ）を配置する．

1 エ ン (1円)

(3マス目の点字がつなぎ符を表す)

アルファベットに仮名が続いて単語を形成する場合は，全ての場合つなぎ符を挿入する．

X セ ン (X線)

(1マス目の点字が外字符を，2マス目は大文字を表す大文字符，
4マス目がつなぎ符を表す)

なお，仮名の次に数字やアルファベットが続いて単語を形成する場合は，数符や外字符があるため，つなぎ符は不要となる．また，数字とアルファベットで1単語を形成する場合も同様につなぎ符は不要である．

次に，仮名の点字表記について記述する．仮名の場合，50音の1文字に対応する点字は1マスで表し，濁音，半濁音や拗音，拗濁音，拗半濁音，特殊音は前置点を含めて2マスの点字が対応している．

濁音　　テ ン ジ（点字）　(3マス目の点字が濁点を表す)

半濁音　　サ ン ポ（散歩）　(3マス目の点字が半濁点を表す)

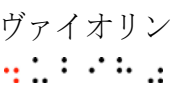
拗音　　キョリ（距離）　(1マス目の点字が拗音点を表す)

拗濁音　　ジュツ（術）　　(1マス目の点字が半濁音を表す)

半拗濁音　　デンピョー（伝票）（4マス目の点字が半拗濁音を表す）

特殊音（合拗音系の例）[2]

ヴァイオリン（1マス目の点字が合拗音系の前置点を表す）



加えて，一部の仮名は，発音する通りの文字に対応する点字で表記する．例えば，上記の「デンピョウ（伝票）」や「クウキ（空気）」のようなウ列の長音，「オウサマ（王様）」などのオ列の長音には，長音符を用いる．但し，オ列の長音であっても「オオカミ（狼）」など「オ」を用いる場合は，長音符を用いず「オ」を使用する．また，係助詞の「ハ」や格助詞の「ヘ」は，発音する通りに「ワ」や「エ」と書き表す．

2.1.2 点字の分かち書き

日本語の点字の分かち書きは, 原則として文節の単位で区切る. 日本点字表記法では, 文節の定義を「自立語または自立語に助詞や助動詞を添えたもの」と規定している [2]. 自立語は，内容語を中心とした語彙として意味をもつ単語であり, 名詞, 動詞, 形容詞, 形容動詞，副詞，連体詞，接続詞，感動詞を指す．但し，この規定には例外がある．例えば，動詞や形容詞の前の単語が，動詞の連用形又は形容詞の語幹である場合は続けて書く（以後，■は分かち書きの区切りとして空白を表す）．

・動詞の連用形 + 動詞「歩き続ける」

○ アルキツヅケル

× アルキ■ツヅケル

---

[2] 合拗音系は外来語に用いられる音を指す．他に，「イェ」，「キェ」等のイ列の仮名に小文字の「エ」を添えた開拗音系，「スィ」，「ティ」等サ行とタ行（「その他」と分類される）の特殊音がある．

・形容詞の語幹 + 形容詞「面白おかしい」

○ オモシロオカシイ

× オモシロ■オカシイ

また, 形式名詞や補助用言は自立語に該当するが, 音韻変化によって続ける場合がある.

・形式名詞「こと」

カク■コトヲ（書くことを）

・形式名詞の音韻変化

イワンコッチャ■ナイ（言わんこっちゃない）

・補助用言「いる」「いただく」

カイテ■イル（書いている）

カイテ■イタダク（書いていただく）

・補助用言の音韻変化

カイテル（書いてる）

さらには, 自立語の中でも長い文字列の複合語や固有名詞にも分かち書きが適用される. これらの自立語の分かち書きの手順は，次の通りである．はじめに，複合語や固有名詞の内部を構成要素に分割する．次に構成要素を，自立可能な意味の成分と副次的な意味の成分に分類する．そして，自立可能な意味の成分の前は区切り，副次的な意味の成分は前に続けて分かち書きする.

・複合名詞

テンジ■トショカン（点字図書館）

（「図書」は自立可能な意味の成分，「館」は副次的な意味の成分）

日本点字表記法では，自立可能な意味の成分の区切りの指標に，モーラ（拍）[14]という言語音の単位を採用している [2]．モーラは基本的に仮名一文字に相当し，撥音「ッ」や促音「ン」も１つと数える．日本点字表記法は，自立可能な意味の成分は3拍以上，副次的な意味の成分は2拍以下が多いという知識に基づいて分かち書きの区切りを規定している [2]．2拍以下で自立可能な意味の成分の場合は，構成要素の自立性の強弱と意味の理解が容易になるか否かを指標にして，分かち書きの区切りを判断するよう規定している [2]．但し，これらの指標は明確に定義されていない．下記に2拍以下の自立可能な意味の成分を含む複合語の分かち書きの例を示す．

・続けて表記する場合

ナツヤスミ（夏休み）　　サイボーマク（細胞膜）

・区切って表記する場合

ボシ■ネンキン（母子年金）　　コーツー■ジコ（交通事故）

このように，分かち書きは規則で記述できる部分が大部分であるが，明確な規則として定義されていない部分がある．

2.2　コンピュータプログラムによる自動点字翻訳

この節では，コンピュータプログラムで日本語から点字への翻訳が可能な方法を明らかにする．

漢字仮名混じり文の日本語と点字は，文法は共通であるが表記体系が異なる．そこで，日本点字表記法に基づく仮名の変換や分かち書きの規則を作成し，これを利用することが必要である．その際，点字の分かち書きは，原則として文法の単位に基づいて規定されているため（2.1節），文法的な情報を得るために形態素解析が必要となる．形態素解

析は漢字を仮名に変換するためにも必要である．分かち書きについての曖昧な表記規則に対しては，ルールベースでの実現が不可能であるため，統計ベースや事例ベースで日本語と点字表記の語句の対応を学習させることが可能である．以上のことから，日本語の点字翻訳に適した基本的な手法は，形態素解析を用いて漢字仮名混じり文を仮名へ変換すること，そして品詞の情報に基づいた点字表記規則を作成し分かち書きを出力すること，と結論づけられる．分かち書き規則のうち，表記規則の記述が不可能な部分については，統計的な手法で正しい分かち書きを学習する必要がある．

最後に，点字の出力については，現在2種類の点字の出力方法が可能である．一つは計算機が可読な点字データのファイル形式を使用することである．点字データの汎用的なファイル形式にBASE形式がある．このBASE形式は，各種の点字プリンタ，点字ディスプレイや点字エディタでの使用が可能であり，官公庁のウェブページや視覚障害者へ点字や音声のデータを提供するネットワーク「サピエ」[15]で採用されている．また，BASE形式は北米点字コード（North America Braille Computer Code, NABCC）を採用しており，点字1マス6点のドットパターン64種類[3]にASCII文字が対応している．もう一つの点字の出力方法は，点字を画像として出力することである．これも点字のドットパターンに対応する画像を出力することで対応可能である．

[3] 空白を含めると 64 種類になる．

第3章

関連研究

3.1 先行研究

自動点字翻訳に関する先行研究は，点訳者を支援するための技術開発に重点を置いている．これらの研究のうち，プログラムを公開しているものを表 3.1に示す．ibukiTenC [16][1]とIBUKI-TEN [17][2]は点字エディタが付属しており，ユーザが点訳結果を修正可能である[3]．Onoら[18]は，ウェブベースの点字翻訳プログラムを開発した．これは，実装した点字表記規則とユーザの修正履歴を集積した事例ベースを利用して点字翻訳する．すなわち，ユーザに点字翻訳の知識があることが前提となっている．

鈴木ら [19]の点字翻訳プログラムは，自動処理部と対話処理部に分かれている．自動処理部では，漢字を仮名に変換するためのテーブル，ユーザ辞書，知識ベースを用いて点字を翻訳する．知識ベースは，分かち書きの区切りの決定，分かち書きの修正の提示の 2 種類がある．次に，後者の修正用の知識ベースを用いて，対話処理部で分かち書きの修正箇所をユーザに提示する．高木ら [20]は，漢字仮名混じり文の表層解析と if-then 形式のルールで分かち書きを行い，例外の分かち書きや分かち書きの曖昧な規定に対して事例ベースで補完するという手法を提案した．この事例ベースはユーザの修正履歴を基に作成される．以上のように，事例ベースや知識ベースを用いた手法は，ユーザの点字の知識を利用して点字翻訳を実現している．

---

1 http://www.ikd.info.gifu-u.ac.jp/ibukiTenC/

2 http://www.ikd.info.gifu-u.ac.jp/IBUKI-TEN/

3 IBUKI-TEN は 2007 年，ibukiTenC は 2009 年 3 月 12 日に開発が中止された．

Web 上で電子楽譜（MusicXML）を点字楽譜に翻訳するプログラムとしてBrailleMUSE (Braille MUsic Support Environment)がある[21]（表 3.1）．これは，五線譜上で表現される音符や和音等の楽譜固有の情報を点字（BASE 形式）に変換するプログラムであるため，漢字仮名混じりの文を点字翻訳する場合と異なる表記規則（「点字楽譜の手引き」文部省編，日本ライトハウス，大阪，1984）を適用している．

3.2 他の自動点字翻訳プログラム

日本語の自動点字翻訳プログラムには，無償のものと有償のものがある (表 3.2)．前者には，「お点ちゃん」 [22]，「点字自動翻訳システム」（Système de transcription automatique en braille）[23]，「CGI自動点訳」がある．後者は，国内ではEXTRA for Windows（有限会社エクストラ，静岡市）とブレイルブリッジ for Windows （ニュー・ブレイル・システム株式会社，東京都）が，国外ではDuxbery Braille Translatorが販売されている．以上のプログラムは，点字自動翻訳システム [23]とCGI自動点訳を除いて，全てプロプライエタリなプログラムで，Windows OSでのみ動作する．点字自動翻訳システム [22]は，インターネット上で点字翻訳するプログラムで，我々のeBraille 0.81を参考にして開発されたものである．点字の出力形式に関しては，ウェブベースのプログラムを除いて全てBASE形式ファイルに対応しているため，点字プリンタでの出力が可能である．点字自動翻訳システム [23]は，我々同様に点字画像をGIFイメージで出力している．CGI自動点訳もウェブベースのプログラムであるが，漢字を点訳できない．

日本語の点字は，主に品詞に対して表記規則を適用している（第2章）ことから，我々の自動点字翻訳プログラムの方針は，ルールベースを基本とする．これまで報告された点字翻訳プログラムの先行研究で用いられてきた事例ベースや知識ベースはユーザの点字修正の履歴を利用しているため，ユーザの点訳の知識に依存しているといえる．これでは事例ベースや知識ベースにユーザの点訳誤りが含まれ，プログラムの点訳精度を低下

させる可能性がある．また，我々のプログラムはユーザの点字知識が不要であることを前提としているため，これらの手法は適さない．

プログラムの点訳精度の比較対象としては，誰でも自由に使用可能なプログラムを用いる．具体的には，無料のプログラムibukiTenC [16]とその前身であるIBUKI-TEN [17]，お点ちゃん [22]，点字自動翻訳プログラム [23]を対象とする．

表 3.1. 先行研究で開発された自動点字翻訳プログラム

Program		IBUKI-TEN	ibukiTenC	Braille translation	BrailleMUSE
Developer		Dept. of Information Science Faculty of Engineering, Gifu Univ., Ikeda Lab.	Dept. of Information Science Faculty of Engineering, Gifu Univ., Ikeda Lab.	Satoshi Ono, Dept. of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima Univ.	Graduate School of Environment and Information Sciences, Yokohama National University, Gotoh Lab.
Language		Japanese	Japanese	Japanese	Music score
Operating System		Windows	Windows	Web-based	Web-based
Data input		Plain text	Plain text	Plain text	MusicXML
Japanese translation	*Kanji* to *Kana*	○	○	○	×
	Word segmentation (*Wakachigaki*)	○	○	○	×
English translation (Available grade)		×	×	×	undisclosed
Output format for print	*Kana* text (file)	○	○	○	×
	BASE/NABCC	○	○	×	○
	Images	×	×	○	×
	Mirror images	×	×	○	×
Music score translation		×	×	×	○
Remarks		ibukiTenC's predecessor. The development was stopped in 2007.	The development was stopped in March, 2009.	http://mediaeng.ibe.kagoshima-u.ac.jp/tenji/	The braille translation program for music score

これらのプログラムは2010 年以前に開発または公開された．

表 3.2.　個人又は企業が開発した自動点字翻訳プログラム

Program		お点ちゃん	点字自動翻訳システム	CGI自動点訳	EXTRA	ブレイルブリッジ	Duxbury Braille Translator
Developer		勝沼貞幸	厨子直人	中村正明	有限会社エクストラ	ニュー・ブレイル・システム株式会社	Duxbury Systems, Inc., USA
Language		Japanese	Japanese	Japanese	Japanese, English	Japanese	English and other 45 languages
Operating System		Windows	Web-based	Web-based	Windows	Windows	Windows
Data input		Plain text	Plain text	Plain text	Plain text, MSWord, 一太郎 HTML, PDF	Plain text	Plain text, MSWord, HTML, etc.
Japanese translation	*Kanji* to *Kana*	○	○	×	○	○	×
	Word segmentation (*Wakachigaki*)	○	○	×	○	○	×
English translation (Available grade)		×	×	×	1, 2	1, 2	1, 2
Output format for print	*Kana* text (file)	○	○	×	○	○	×
	BASE/NABCC	○	×	×	○	○	○
	Images	×	○	○	×	×	○
	Mirror images	×	○	○	×	×	undisclosed
Music score translation		×	×	×	×	×	undisclosed
Remarks			The program follows eBraille 0.81	http://hp.vector.co.jp/authors/VA014370/cgi/braille/index-j.htm			http://www.duxburysystems.com/

これらのプログラムは2010 年以前に開発または公開された.

第4章

研究材料

4.1 評価用コーパスの作成

4.1.1 通常文書評価用コーパス

自動点字翻訳プログラムの実現に向けた研究では，点訳精度と点訳上の問題点を精査する必要がある．そこで，自動点字翻訳プログラムのベンチマークコーパスが必要となるが，これまでにそのようなコーパスについての報告はない．先行研究では，プログラムの評価に情報処理のテキスト [20], コンピュータの書籍 [19], 毎日新聞社の点字毎日[16]を用いている．これらのうち，情報処理のテキストとコンピュータの書籍の電子データについての情報は報告されておらず入手が不可能であるが，点字毎日は，漢字仮名混じり文（墨字文）と点字文のデータが対で入手可能な新聞記事の生コーパスである[1].

以上のことから，我々は，本研究用に点字毎日の電子ファイルを入手し，新聞記事474文（2002 年 6 月の 12 記事，毎日新聞社）から成るコーパスを作成することにした．コーパス作成に際して, 墨字文と点字文の両ファイルに修正が必要であった. その理由は, 点字毎日の生コーパスが墨字文と点字文の文章が一致していない箇所が存在するためである[2]．また，点字表記が日本点字表記法 2001 年版 [2]に未対応な部分も存在したため,

---

1 点字文データは毎日新聞社から，墨字文データはニフティ株式会社の「フォーラム@nifty」から入手可能だったが，「フォーラム@nifty」のサービスは2007 年 3 月に終了した.

2 毎日新聞社の担当者に問い合わせたところ，墨字文と点字文は別に作成しているため，文章は完全に一致するわけではないとのことだった.

それらを日本点字表記法2001年版に対応させた．加えて簡単に確認可能にすべく，点字毎日の新聞記事の点字文ファイルのBASE形式から点字エディタ（点字編集システム4, テクノツール株式会社，川崎市）により，仮名表記に変換し，数符や外字符は「数」や「外」の文字で表現した．これらの方法で墨字文と点字文の完全なデータ対を作成し，合計5,191文（点字毎日，2002年4月〜7月の233記事）の233データ対からなる「通常文書評価用コーパス」を完成させた（図 4.1）．

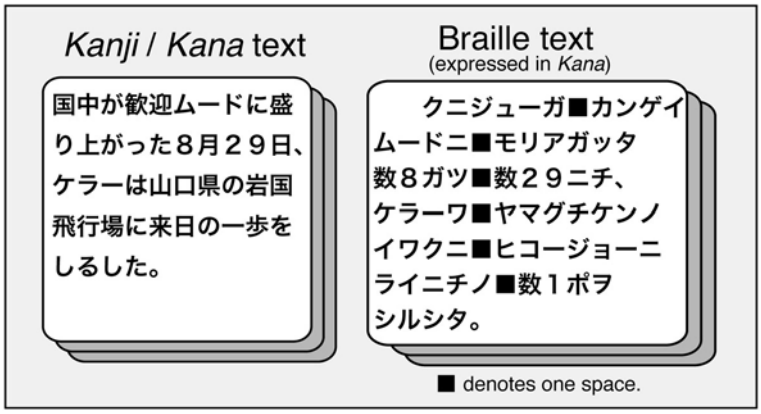


図 4.1. 通常文書評価用コーパスの例

4.1.2 医療文書評価用コーパス

医療文書の自動点字翻訳プログラムの実現のために，その点訳精度と点訳上の問題点を明らかにする必要がある. 医療文書は, 新聞記事には出現しないような疾患名や薬剤名, 医療機器の名前などの専門的な用語を含むことが予想される．そのため，医療文書のベンチマークコーパスが必要である．医療文書の墨字文と点字文が対になったコーパスは存在せず，作成する必要がある．

そこで，神戸大学医学部附属病院の全診療科の看護記録2007年4月分（112,917文）から無作為に選択した359文，同病院で使用中の患者向けの文書69文書2,536文，を電

子化した[3]．ここでいう患者向け文書とは，各種同意書，治療や病状の説明，検査票，問診表，治療計画書，入院計画書，退院計画書である．さらに，ICD10 対応電子カルテ用標準病名マスター V2.62 [24]に含まれる疾患名（約 2 万語）から無作為に選択した疾患名 130 件と，厚生労働省の難治性疾患克服研究事業の臨床調査研究分野対象となっている疾患名[25] 155 件[4]を含む文例 155 文を作成しコーパスに加えた．点字文の作成は，墨字文を複数の自動点字翻訳プログラムで点訳させ，数符や外字符などの前置点を付記した後，我々が点訳の誤りを修正して完成させた．以上の方法で，最終的に計 77 データ，3,180 文の「医療文書評価用コーパス」を作成した.

4.2　点訳精度の定義とその算出方法

従来，自動点字翻訳の研究においては分かち書きに焦点があてられており，プログラムの評価では分かち書きの区切りの精度を算出していた．しかし，漢字仮名混じり文の文章を正しく点字翻訳するには，漢字の読み，記号に相当する点字，数符や外字符，つなぎ符の配置が正しいことも重要である．そこで，我々は自動点字翻訳プログラムの評価のために，分かち書きの精度に加えて漢字や記号の読みの精度を算出し，更に，総合的な評価値として，点訳精度 Braille Translation Accuracy (BTA)を，分かち書きの精度と読みの精度の積と定義した.

$$\mathrm{BTA} = F_1 \times T \tag{4.1}$$

ここで，F_1は，分かち書きの精度を表し，Tは漢字，数字，記号の読みや長音，前置点，つなぎ符への変換の正解率[5]を表す．F_1は*F-measure*（適合率（P）と再現率（R）の調和

[3] 看護記録や患者向け文書に個人情報が含まれている場合，全て削除した上で提供を受けた.

[4] 難治性疾患克服研究事業の対象は 130 疾患であるが，疾患名の別称を 1 件と数えているため，155 件となっている.

[5] これを我々は「正音率」と命名した.

平均）である．P はプログラムが出力した分かち書きの数における正解の分かち書きの数の割合であり，R はプログラムの出力した正解の分かち書きの数の，コーパスの分かち書きの数における割合である．F_1 は下記の van Rijsbergen [26]の公式から派生した式を用いて計算する：

$$F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

$$\beta = 1 \tag{4.2}$$

T は，下記のように算出する．

$$T = \frac{k}{K} \qquad (k \le K) \quad (0 \le T \le 1) \tag{4.3}$$

$$T = \frac{K - (k - K)}{K} \qquad (k > K) \quad (0 \le T \le 1) \tag{4.4}$$

この時，k はプログラムが正しい読みを付与した分かち書きの数で，K はコーパスの分かち書きの数の総合数である．k に含める分かち書きは，その分かち書きの中の漢字や記号の読みが全て正しくなければならない．すなわち，一文字でも読み誤りがあると k に含めることはできない．この場合，プログラムが出力する分かち書きが小さいと読みの精度が高くなる傾向があるため，K をコーパスの分かち書きの総合数にしている．そのため，コーパスの分かち書きの数よりもプログラムの出力した読みの正解の数が多い場合は，その差をコーパスの分かち書きの数から差引いて計算した（4.4 式）．

点訳精度（BTA）は，2 通りの方法で計算した：評価に用いたコーパスの全データを一塊として計算した値（Total BTA）と各データファイルの点訳精度の平均±標準偏差 (mean BTA ± SD)である．点訳精度の具体的な計算手順は，以下の通りである（図 4.2）．

1. 自動点字翻訳プログラムにコーパスの墨字文ファイルを点訳させ，点訳結果を仮名（表音文字）で表記した点字文ファイルとして出力させる[6]．点字を仮名で出力する理由は，プログラムの点訳誤りを解析する際に，人の目で確認が可能なためである．

2. 自動点字翻訳プログラムの点字文ファイルとコーパスの点字文ファイル（仮名で表記された正解点訳の点字文）とを分かち書き単位で比較し，差分を出力する．

3. 出力された差分には，読み誤りと分かち書きの誤りの両方が含まれているため，人手で読みと分かち書きの誤りの数を数え，分かち書きの正解の数を確定する．

4. 分かち書きの正解の数，読みの正解の数を基に，前述の(4.1)の式で点訳精度を計算する．

以上の評価に際しては，我々が作成した評価補助プログラムを用いて，半自動で各種精度を算出した．

[6] eBraille の場合は，点訳の途中で仮名表記した点字文を生成する．他の自動点字翻訳プログラムで点訳結果を点字エディタの形式（BASE 形式）で出力する場合は，BASE ファイルを仮名表記に変換し，数符や外字符に相当する ASCII 文字を「数」や「外」に変換して仮名表記の点訳文ファイルを作成する．

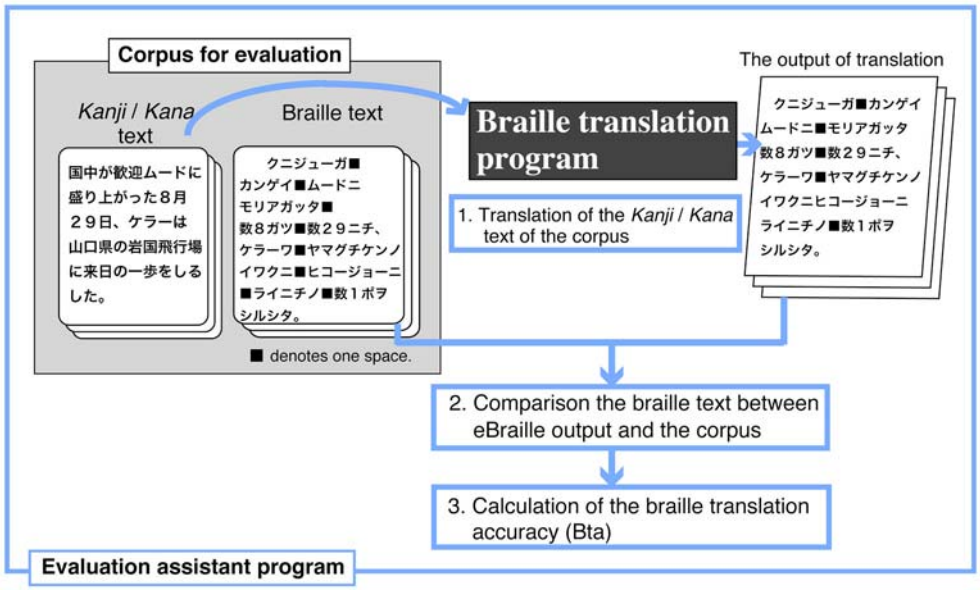


図 4.2.　自動点字翻訳プログラムの評価手順

第5章

自動点字翻訳プログラムeBraille

5.1 はじめに

この章では，高い点訳精度の自動点字翻訳プログラムの実現方法についての研究と，その実装について記す．我々は，1997年に開発された自動点字翻訳プログラムeBraille 0.81 [11]を基に点字表記規則の実装方法を明らかにし，実装後のeBrailleの評価実験を行う．はじめに， eBrailleの実装方法とユーザインタフェースを説明する（5.1.1節）．続いて，eBraille 0.81 [11]の点訳精度とその点訳誤りを解析し，問題点を明らかにする (5.1.2節). 次に，日本点字表記法に則した点字表記規則の実装方法を記述する（5.2節）．更に，点字表記規則を実装したeBrailleの評価実験の方法を記す(5.3節)．そして，形態素解析器ChaSenの更新（5.4.1節），eBrailleの点訳エンジンへの点字表記規則の実装（5.4.2節）と実験結果を記載する (5.4.3-5.4.5節)．そして，最後に結果をまとめる（5.5節）．

5.1.1 eBraille 0.81の実装とインタフェース

eBraille 0.81[11]は，コンピュータの OS に依存しないよう，Web ブラウザで利用する CGI プログラムとして構築された（図 5.1）. eBraille サーバには, HTTP デーモンとして Apache (version 2.2.3)を，形態素解析プログラムとして 1997 年公開の形態素解析器 ChaSen (version1.51) が組み込まれている.

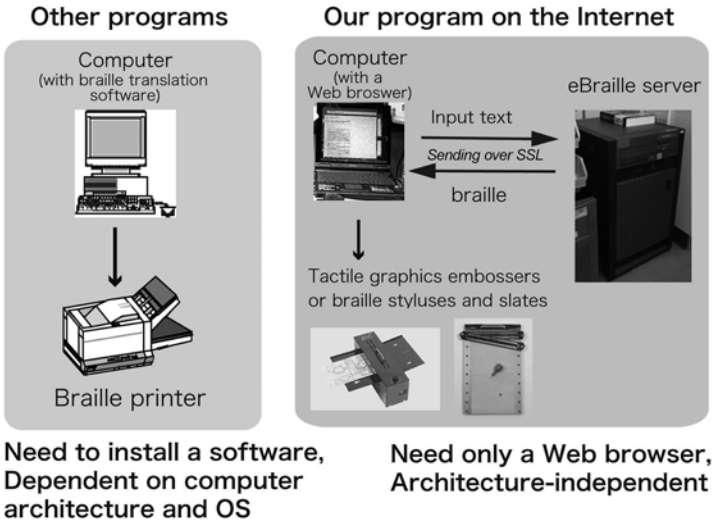


図 5.1. eBraille の特徴

eBraille の点訳処理は，主に ChaSen と eBraille に実装された読みと分かち書きの表記規則が担っている．具体的には，入力文（日本語の漢字仮名混じり文）の形態素解析，分かち書き，助詞「は」と「へ」の表音文字（「ワ」，「エ」）への変換，数符と外字符の挿入，点字への変換，というプロセスを経る．点字への変換は 2 段階を経る．はじめに，仮名や符号 1 字に対して 6 桁の数字から成る点字 1 字へ変換する．6 桁の数字は，点字 6 点の点の有無を 0 又は 1 で表したドットパターンである．次に，この 6 桁の数字を基にして，pbm 形式の画像に変換する．最後に pbm から GIF イメージへ変換する．この点字の GIF イメージは，立体コピー作成機や点字器を用いた点字文書の作成に用いることができる．

eBrailleのユーザインタフェースについては，ユーザがウェブページのテキストフォームに漢字仮名混じり文を入力して，eBrailleサーバへ転送した後に下記の2段階を経るものである（図5.2）：

(1) eBrailleサーバから，入力文を仮名に変換した文（表音文字）が提示される．ここでユーザは，漢字の読みと分かち書きを修正可能である．

(2) ユーザが仮名文をeBrailleサーバへ転送する．eBrailleサーバが，仮名文字とそれらに対応する点字のGIFイメージを提示する．

なお（2）の段階では，プルダウンメニューによる「鏡像にしない」「鏡像にする」を選択可能となっている．「鏡像にしない」は触読の際の通常の点字を表示する．そして，「鏡像にする」では触読する面の裏を表示しており，「鏡像にしない」点字の左右を反転させている（図5.2）．これは，人が点字器を用いて点字文書を作成する際に裏面を上にして凸点を作成するため，鏡像の点字イメージの通りに点を打つことを考慮したものである．

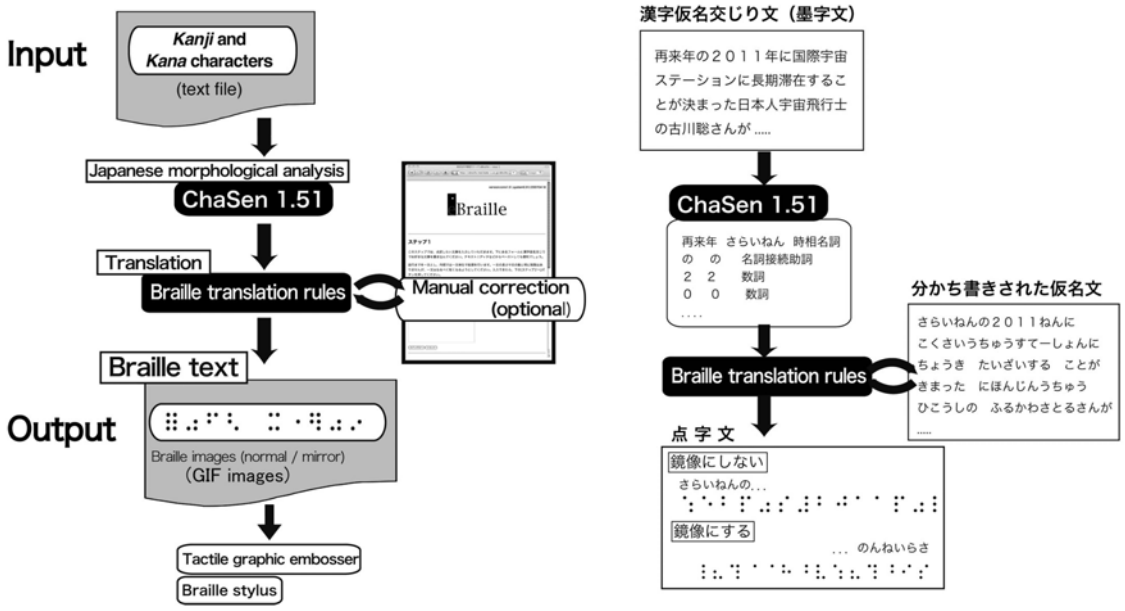


図 5.2. eBraille 0.81 の概要

5.1.2　eBralle 0.81の点訳精度

この節では，eBraille 0.81の点訳精度と点訳誤りを解析し，その問題点を明らかにする．

はじめに，通常文書評価用コーパスのうち474文（4.1.1節）を用いて，4.2節の手順でeBraille 0.81の点訳精度を求めた．その結果，eBraille 0.81のTotal BTAは37.77 BTAで，mean BTA ± SDは，37.83 ± 12.00だった（表 5.1）．次に，eBraille 0.81の点訳の誤りを精査し，点訳誤りの原因をChaSenの解析に起因するもの，ChaSenの解析とeBrailleの点字表記規則の両方に起因するもの，eBrailleの点字表記規則に起因するもの，に分類した（表 5.2）．ChaSenの解析に起因する点訳誤りは，ChaSenが付与する読みに関連するものとChaSenの品詞体系に関連するものに細分類が可能であった．ここでChaSenの品詞体系に関連する点訳誤りとは，ChaSenの解析結果の品詞に関連する点訳誤りを指す．ChaSenの品詞体系には点字表記規則の適用が困難な品詞が含まれている．例えば，ChaSenが普通名詞と判定した形態素（表 5.2のNo.9，10，11，12）の分かち書きは，前を続ける場合と区切る場合があるため，普通名詞にのみ適用する分かち書きの規則では対応できない．これらの普通名詞に適用可能な日本点字表記法[2]の表記規則を下記に示す．はじめに，「10対0（正解：「10■対■0」）」の「対」（表 5.2のNo.9）に対しては，

自立語は前を区切って書き表す.

が適用される．この場合，普通名詞「対」が接続詞的な役割を持つため，「対」の後も区切るという解釈が前提となる．「要素以外（正解：「要素■以外」）」の「以外」（表 5.2のNo.11）には，

形式名詞は，自立語であるから前を区切って書き表す

があり，「以外」が形式名詞に該当することが前提となっている．「豊中■市の（正解：「豊中市の」）」「市」（表 5.2のNo.12）に対しては，

接頭語や接尾語で副次的な意味の成分は，自立可能な意味の成分に続けて書き表す.

が適用可能である．これは，「市」が接尾語に該当するとの解釈が前提となる．最後に「旧イスラエル（正解：「旧■イスラエル」）」の「旧」（表 5.2のNo.10）には，

接頭語や造語要素であっても，後ろの成分に対して連体詞的な関係を持ち，意味の理解を助ける場合には…（中略）…区切って書き表す．

が適用可能である．この場合，接頭語「旧」は「イスラエル」と連体詞的な関係を持つことが前提となっている．以上のことは，点字表記規則を普通名詞に適用するには，普通名詞になんらかの区別が必要であることを示唆している．他の例では，「コンディショニングあるいは■リコンディショニング（正解：「コンディショニング■あるいは■リコンディショニング」）」の「あるいは」の分かち書きの誤りがある（表 5.2のNo.14）．これは「あるいは」が「助詞-名詞接続助詞」と解析されたことに起因する誤りである．しかし，日本点字表記法では，助詞は前を続けて表記するように規定している[2]ため，品詞と表記規則の対応には問題ない．実際，「助詞-名詞接続助詞」に分類されている「や」や「やら」この規則の適用で正しい分かち書きを出力可能である．このように，ChaSen1.51の品詞体系には点字表記規則の適用が困難な品詞が存在する．

表 5.1. eBraille 0.81 の点訳精度

	Word segmentation (F-measure)	Correct translation of *Kanji*, etc.(%)	Braille translation accuracy (BTA)
Total score	61.82	61.10	37.77
mean ± SD	61.06 ± 9.60	60.42 ±11.68	37.83 ±12.00

表5.2　eBraille .081の点訳誤りパターンの分類

No.	誤りの原因			誤りの種類	点訳誤りパターン	点訳誤りの例		
						墨字	eBrailleの出力	正解の出力
1	ChaSen			読み	ChaSenの読み誤り	時には	ジニワ	トキニワ
2	ChaSen			読み	連濁への未対応	筋力不足	キンリョクフソク	キンリョクブソク
3	ChaSen			読み	「2人」を「2ニン」と読む	2人	数2ニン	フタリ
4	ChaSen			読み	ChaSenの形態素解析の誤り	この前方	コノマエガタ	コノ■ゼンポー
5	ChaSen		点字表記規則	読み 分かち書き	ChaSenの形態素解析の誤り	子供の数1817万人	コドモノスウ1817マンニン	コドモノ■カズ■1817マンニン
6	ChaSen		点字表記規則	分かち書き	判定詞、形容詞の一部「ある」の前を続ける	必要である	ヒツヨウデアル	ヒツヨウデ■アル
7	ChaSen		点字表記規則	分かち書き	方言	思いまんねん	オモイ■マン■ネン	オモイマンネン
8	ChaSen		点字表記規則	分かち書き	感動詞の音韻変化	うわあ	ウワ■ア	ウワア
9	ChaSen	ChaSen (POS)	点字表記規則	読み 分かち書き	ChaSenの読み誤り 普通名詞「対」の前後を続ける	10対0	数10ツイ数0	数10■タイ■数0
10		ChaSen (POS)	点字表記規則	分かち書き	普通名詞「旧」 の後を続ける	旧イスラエル	キュウイスラエル	キュー■イスラエル
11		ChaSen (POS)	点字表記規則	分かち書き	普通名詞「以外」の前を続ける	要素以外で	ヨウソイガイデ	ヨーソ■イガイデ
12		ChaSen (POS)	点字表記規則	分かち書き	普通名詞「市」の前を区切る	豊中市の	トヨナカ■シノ	トヨナカシノ
13		ChaSen (POS)	点字表記規則	分かち書き	動詞「なさい」 の前を区切る	しなさい	シ■ナサイ	シナサイ
14		ChaSen (POS)	点字表記規則	分かち書き	助詞-名詞接続助詞「あるいは」の前を続ける	コンディショニングあるいはリコンディショニング	コンディショニングアルイハ■リコンディショニング	コンディショニング■アルイワ■リコンディショニング
15			点字表記規則	読み	長音変換を行わない	事情	ジジョウ	ジジョー
16			点字表記規則	読み	数に7行、ラ行の単語が続く場合に、つなぎ符を挿入しない	10リラ	数10リラ	数10__リラ
17			点字表記規則	読み	「あるいは」の「は」の読み誤り	あるいは	アルイハ	アルイワ
18			点字表記規則	読み	「%」を読み仮名に変換する 小数点の後に数符を挿入する	21．4%	数21．数4パーセント	数21．4外p [*1]
19			点字表記規則	読み	一度を1ドと読まない	もう一度	モー■イチド	モー■1ド
20			点字表記規則	読み	漢数字を数字に変換しない	六月	ロクガツ	数6ガツ
21			点字表記規則	読み	ピリオドが欠落する	1．25キロの	数125キロノ	数1．25キロノ
22			点字表記規則	分かち書き	名詞-サ変名詞「する」を続ける	通用した	ツウヨウシタ	ツーヨー■シタ
23			点字表記規則	分かち書き	数の前を続ける	通用した10リラ	ツーヨーシタ数10リラ	ツーヨー■シタ■数10__リラ
24			点字表記規則	分かち書き	指示詞「その」の後の名詞を続ける	その年には	ソノトシニワ、	ソノ■トシニワ、
25			点字表記規則	分かち書き	アルファベットと助詞の間を続ける	BSと	BSト	外大大BS■ト
26			点字表記規則	分かち書き	カギ括弧開の前を続ける	…作曲「グレの歌	…サッキョク「グレノ■ウタ	…サッキョク■「グレノ■ウタ
27			点字表記規則	分かち書き	丸括弧開の前を区切る	1200円（税別）	数1200エン■｜ゼイベツ｜	数1200__エン｜ゼイベツ｜
28			点字表記規則	分かち書き	複合動詞（連続した動詞）を区切る	聴き取れ	キキ■トレ	キキトレ
29			点字表記規則	分かち書き	複合名詞（連続した名詞）を続ける	インフレ危機	インフレキキ	インフレ■キキ
30			点字表記規則	分かち書き	形式名詞「の」の前を区切る	取り組んだのが	トリクンダ■ノガ	トリクンダノガ
31			点字表記規則	分かち書き	副詞的名詞「よう」を区切る	参加できるよう	サンカデキル■ヨウ	サンカ■デキルヨー
32			点字表記規則	分かち書き	接尾辞「ない」を前の語と続ける	珍しくない	メズラシクナイ	メズラシク■ナイ
33			点字表記規則	分かち書き	接尾辞-動詞性接尾辞「いる」の前を続ける	しているのも	シテイル■ノモ	シテ■イルノモ
34			点字表記規則	分かち書き	「…せずして」を「せず■して」と区切る	見ずして	ミズ■シテ	ミズシテ
35			点字表記規則	分かち書き	「…するなかれ」を「する■なかれ」と区切る	見るなかれ	ミル■ナカレ	ミルナカレ

品詞の名称はChaSen1.51に準拠している．

ChaSen(POS), 品詞体系に起因する誤り; 数, 数符; 外, 外字符;　大, 大文字符;

*1 「外p」, 「%」に相当する点字.

次に，eBraille0.81の点字表記規則に起因する点訳の誤りを記述する．読みでは，長音符への変換，つなぎ符の挿入，助詞-副助詞「アルイハ」から「アルイワ」への変換，記号「%」に相当する点字への変換, 数字を含む単語の表記に失敗していた（表5.2のNo. 15-19, 20）．これらは全て日本点字表記法で規定されている表記規則である[2]が，eBraille0.81に実装されていなかった．また，分かち書きの誤りでは，動詞「する」の前を区切る，という2001年の日本点字表記法で改定された規則に未対応であった（表 5.2のNo. 22）．加えて,「自立語の前を区切る」という基本的な分かち書きの規則（表 5.2のNo.23, 24),

「アルファベットと助詞の間は区切る」という規則，複合動詞や複合名詞の分かち書きも実装されていなかった(表 5.2のNo.25，28，29)．eBraille 0.81に実装されている点字表記規則は，読みに関しては3個，分かち書きに関するものは12個のみである(表 5.3)．これらのことから，日本点字表記2001年版の点字表記規則への対応と表記規則の追加が必須であることが明らかとなった.

表 5.3. eBraille .081に実装されている点字表記規則

No	表記の種類	規則適用の対象(品詞)	点字表記規則
1	読み	助詞	「ハ」を「ワ」に変換
2	読み	助詞	「へ」を「エ」に変換
3	読み	数字，アルファベット	数符、外字符の挿入
4	分かち書き	全ての品詞	原則として前後を区切る(デフォルト)
5	分かち書き	記号	前と後を続ける(デフォルト)
6	分かち書き	記号−括弧開	後を続ける
7	分かち書き	記号−括弧閉	前を続ける
8	分かち書き	判定詞(「です」)	前を続ける
9	分かち書き	助動詞(「ので」)	前を続ける
10	分かち書き	名詞−接尾	前と後を続ける
11	分かち書き	名詞−サ変接続	後ろを続ける
12	分かち書き	名詞−数	前と後を続ける
13	分かち書き	助詞	前と後を続ける
14	分かち書き	接頭辞	後ろを続ける(デフォルト)
15	分かち書き	接尾辞	前を続ける(デフォルト)

5.2 自動点字翻訳の精度向上方法

前節では，点訳精度の向上には，ChaSenの品詞分類を点字表記規則に適用可能にすることと，日本点字表記法 2001年版 [2] に則した点字表記規則の追加実装の必要性が明らかになった．これらを実現するための方法を記述する.

はじめに，2007年当時のChaSenの最新版，version 2.3.3 [27]の品詞体系をChaSen1.51と比較し，ChaSenの更新で品詞分類の問題が解決可能かを調査した．その結果を基にしてChaSenの更新の可能性とeBrailleへの組み込みの方法を明らかにした．

次に，eBraille0.81の点訳誤りを基に実装すべき点字表記規則を解明した．加えて，日本点字表記法の表記規則について分析し，実装する点字表記規則を決定した．点字表記規則を決定する具体的な手順は下記の通りである．

1. eBrailleの点訳誤りを修正可能な点字表記規則を日本点字表記法から抽出する．
2. 日本点字表記法で，点字表記規則の適用の際に用いる基準（指標）を調査する．
3. 日本点字表記法における点字表記規則の優先順位を解明する．
4. 作成した点字表記規則に矛盾がないかを調査し，規則の修正又は追加を行う．

なお，日本点字表記法 [2] は2編で構成されており，その内容は以下の通りである．

第1編　点字の表記

1. 点字の記号
2. 語の書き表し方
3. 語の区切り目の分かち書きと自立語や固有名詞内部の切れ続き
4. 文の構成と表記符号の用法
5. 書き方の形式と点字化のための配慮
6. 古文の書き表し方
7. 漢文の書き表し方

第2編　参考資料

1. 点字の表記に関するキーワードの解説
2. 点字の意義と歴史
3. 点字記号一覧
4. 情報処理用点字表記の解説

この中で，現代の仮名遣いで記述された文章の点字翻訳に関する表記規則は，第1編の2と3および4の一部に記述されている．我々が点訳対象とする文書は，通常文書や医療文

書であるため，これらの基本的な規則を分析の対象とした．そして，分析結果からeBrailleに実装する点字表記規則を明らかにした．次に，eBrailleに点字表記規則を実装した．

5.3 実験方法

バージョンを更新した ChaSen と前節の分析結果に基づいて作成した点字表記規則の有効性を，eBraille の点訳精度の解析から明らかにした．点訳精度の計算は，4.2 節と同様に行った．はじめに，通常文書評価用コーパス 1,983 文を用いて eBraille0.81 と新しい eBraille の点訳精度を比較した．次に，通常文書評価用コーパス 5,191 文を用いて，新しい eBraille と他の自動点字翻訳プログラム，点字自動翻訳システム [23]，お点ちゃん（version 4.1） [22]，ibukiTenC (version 0.65) [16]，IBUKI-TEN（version 0.56）[17]の点訳精度を比較解析して評価した．

5.4 結果

5.4.1 ChaSenの更新

ChaSen2.3.3はIPA品詞体系に基づいており，ChaSen1.51との品詞体系とは大きく異なる（表 5.4）．例えば，eBraille0.81の点訳誤りで ChaSen 1.51が「普通名詞」と判定した単語は，ChaSen 2.3.3では「名詞-接続詞的」，「接頭詞-名詞接続的」，「名詞-非自立-副詞可能」,「名詞-接尾-地域」のいずれかに分類される（表 5.5）. このことから, ChaSen2.3.3の品詞体系を利用して，細分化された品詞に点字表記規則を適用することにより，正しく点訳できる可能性がある．そこで，eBrailleで用いる形態素解析器をChaSenを2.3.3へ更新した．この更新に伴い，eBraille0.81に実装されていた点字表記規則をversion 2.3.3の品詞体系に対応させた．

表 5.4. ChaSen1.51 と 2.3.3の品詞体系の比較

ChaSen1.51		ChaSen2.3.3			
名詞	普通名詞	名詞	一般		
名詞	サ変名詞	名詞	固有名詞		
名詞	固有名詞	名詞	固有名詞	一般	
名詞	地名	名詞	固有名詞	人名	
名詞	人名	名詞	固有名詞	人名	一般
名詞	数詞	名詞	固有名詞	人名	姓
名詞	形式名詞	名詞	固有名詞	人名	名
名詞	副詞的名詞	名詞	固有名詞	組織	
名詞	時相名詞	名詞	固有名詞	地域	
動詞		名詞	固有名詞	地域	一般
形容詞		名詞	固有名詞	地域	国
副詞	様態副詞	名詞	代名詞		
副詞	程度副詞	名詞	代名詞	一般	
副詞	量副詞	名詞	代名詞	縮約	
副詞	頻度副詞	名詞	副詞可能		
副詞	時制相副詞	名詞	サ変接続		
副詞	陳述副詞	名詞	形容動詞語幹		
副詞	評価副詞	名詞	数		
副詞	発言副詞	名詞	非自立		
接頭辞	名詞接頭辞	名詞	非自立	一般	
接頭辞	動詞接頭辞	名詞	非自立	副詞可能	
接頭辞	イ形容詞接頭辞	名詞	非自立	助動詞語幹	
接頭辞	ナ形容詞接頭辞	名詞	非自立	形容動詞語幹	
接続詞		名詞	特殊		
連体詞		名詞	特殊	助動詞語幹	
助詞	格助詞	名詞	接尾		
助詞	副助詞	名詞	接尾	一般	
助詞	引用助詞	名詞	接尾	人名	
助詞	名詞接続助詞	名詞	接尾	地域	
助詞	述語接続助詞	名詞	接尾	サ変接続	
助詞	終助詞	名詞	接尾	助動詞語幹	
助動詞		名詞	接尾	形容動詞語幹	
感動詞		名詞	接尾	副詞可能	
特殊	句点	名詞	接尾	助数詞	
特殊	読点	名詞	接尾	特殊	
特殊	空白	名詞	接続詞的		
特殊	括弧開	名詞	動詞非自立的		
特殊	括弧閉	名詞	引用文字列		
特殊	記号	名詞	ナイ形容詞語幹		
接尾辞	名詞性述語接尾辞	動詞	自立		
接尾辞	名詞性名詞接尾辞	動詞	非自立		
接尾辞	名詞性名詞助数辞	動詞	接尾		
接尾辞	形容詞性述語接尾辞	形容詞	自立		
接尾辞	形容詞性名詞接尾辞	形容詞	非自立		
接尾辞	動詞性接尾辞	形容詞	接尾		
判定詞		副詞	一般		
指示詞	名詞形態指示詞	副詞	助詞類接続		
指示詞	連体詞形態指示詞	接頭詞	名詞接続		
指示詞	副詞形態指示詞	接頭詞	動詞接続		
		接頭詞	形容詞接続		
		接頭詞	数接続		
		接続詞			
		連体詞			
		助詞	格助詞		
		助詞	格助詞	一般	
		助詞	格助詞	引用	
		助詞	格助詞	連語	
		助詞	接続助詞		
		助詞	係助詞		
		助詞	副助詞		
		助詞	間投助詞		
		助詞	並立助詞		
		助詞	終助詞		
		助詞	副助詞／並立助詞／終助詞		
		助詞	連体化		
		助詞	副詞化		
		助詞	特殊		
		助動詞			
		感動詞			
		記号	一般		
		記号	句点		
		記号	読点		
		記号	空白		
		記号	アルファベット		
		記号	括弧開		
		記号	括弧閉		
		その他			
		その他	間投		
		フィラー			
		非言語音			
		語断片			

表 5.5. ChaSen 1.51と2.3.3の品詞タグの違いの例

解析対象	ChaSen1.51	ChaSen2.3.3
10**対**0	名詞-普通名詞	名詞-接続詞的
旧イスラエル		接頭詞-名詞接続
要素**以外**で		名詞-非自立-副詞可能
豊中**市**の		名詞-接尾-地域
コンディショニング**あるいは**リコンディショニング	助詞-名詞接続助詞	接続詞
日程**や**時間が		助詞-並立助詞
雨**やら**雪やら		助詞-並立助詞

5.4.2 点字表記規則の実装

日本点字表記法を分析した結果，点字表記規則を適用する際の指標は下記の通りであった.

(1) 読みの表記規則の適用に用いる指標
品詞，文字種，文字（出現形）

(2) 分かち書きの表記規則の適用に用いる指標
品詞，活用形，単語の出現形，文字種，音韻変化，モーラ（拍），
単語の自立性の強弱，発音や意味の境界

更に，分かち書きの表記規則には，(2)の指標を，連続する２つの形態素に適用する場合が存在した．我々は，上記の指標とその適用範囲に従って点字表記規則を作成した（表5.6）．但し，（2）の指標のうち，単語の自立性の強弱と発音や意味の境界については，

日本点字表記法[2]で明確に定義されていないため，指標の対象外にした．以下に作成した点字表記規則の例を示す．

・ 読みの表記規則の例

(1) 品詞，文字を指標とした規則
係助詞「ハ」を「ワ」に変換する　（表5.6のNo.2）

(2) 文字種を指標とした規則
数字やアルファベットの前に数符，外字符を前置する（表5.6のNo.115）

・ 分かち書きの規則の例

(1) 品詞，文字種を指標とした規則
「記号-アルファベット」に「助詞」，「助動詞」，「名詞-一般」，「名詞-固有名詞-地域-一般」，「名詞-サ変接続」のいずれかが続く場合は，その間を区切る（付録，表A.3のNo. 110）

(2) 活用形を指標とした規則
「動詞-自立」又は「動詞-非自立」の連用形に「動詞-自立」が続く場合は，間を続けて表記する（付録，表A.2のNo.52）

(3) 単語の出現形を指標とした規則
出現形が「%」で読みが「パーセント」の形態素は，前を続け，後ろを区切る（付録，表A.2のNo.52）

(4) 音韻変化を指標とした規則

「名詞-一般」に「名詞-サ変接続」が続き，「名詞-サ変接続」の形態素の読みが「カワリ」や「スキ」の場合，連濁の読み「ガワリ」，「ズキ」に変換し，2形態素の間を続けて表記する（付録，表A.1のNo. 6，No.20）

(5) モーラ（拍）を指標とした規則

2モーラ以下の「名詞-一般」又は「名詞-サ変接続」に「名詞-一般」が続く場合，その間を続けて表記する（付録，表A.1のNo.8）

日本点字表記法[2]では「形式名詞」や「補助動詞」「補助形容詞」といった，ChaSen 2.3.3の品詞体系にない品詞を用いていた．この場合，「補助動詞」や「補助形容詞」の具体例とそれらの表記例を参照し，ChaSenの品詞で対応可能なものを用いて点字表記規則を作成した．具体的には，「形式名詞」には「名詞-非自立-副詞可能」や「名詞-非自立-形容動詞語幹」，「名詞-接尾-一般」，「名詞-接尾-副詞可能」を対応させて表記規則を作成した（付録，表A.1のNo. 29，31，34，40）．「補助動詞」には「助動詞」を，「補助形容詞」には「名詞-ナイ形容詞語幹」を対応させて表記規則を作成した（付録，表A.3のNo 90，表A.1のNo. 45）．

次に，我々が作成した点字表記規則の矛盾とその対応の例を記載する．長音符への変換規則は，隣接する2文字の条件を「ウ段」又は「オ段」と「ウ」の組み合わせとしていた．しかし，この条件では「スウェーデン」の「ウ」も長音符へ変換されるため，変換の対象外となる条件を追加した．具体的には，(1)「ウ段」又は「オ段」と「ウ」の後にア行の小文字が続く場合と，(2) 動詞の語尾の「ウ」（例：「イウ（言う）」，「オモウ（思う）」，(3) 特定の単語（例：「抑鬱（ヨクウツ）」「憂鬱（ユウウツ）」「ソウル」）を長音変換の対象外にした．なお，ChaSenをversion2.3.3に更新したため，その辞書IPADICの「発音」が使用可能になったことから，「発音」を点字表記規則に組み込むことを検討した．しかし，点字で「オカアサン（お母さん）」，「オニイサン（お兄さん）」，「オネエサン（お姉さん）」，「オオキイ（大きい）」と表記すべきア列，イ列，エ列，オ列の長音が，IPADICの「発音」では「オカーサン」，「オニーサン」，

「オネーサン」「オーキイ」と表記されていた. そのため点字表記規則にはIPADICの「発音」を使用しなかった.

更に，点字文書の汎用的な書式である，１行に点字を32文字以内に出力する規則を追加した（表 5.6のNo.114）.

最後に，ChaSen が1形態素と解析する単語，例えば，「について」「として」という連語を，日本点字表記法で規定している分かち書きの単位に従い「に■ついて」や「と■して」に分割する規則を追加した（付録，表A.1のNo.76）.

以上のように各種の指標を用いて，細分化された品詞のうち57種類に対して日本点字表記法に則した点字表記規則を作成した（付録A）.

表 5.6. 作成した点字表記規則の概要

No	規則の種類	規則適用の基準	点字表記規則
1	読み	文字	ウ段,オ段の音に続くウを長音符に変換する
2	読み	品詞，文字	読み仮名の変換（係助詞「ハ」,「日本」,「一度」）
3	読み	文字	辞書の読みを修正
4～111	読み，分かち書き	品詞，活用形，単語の出現形，文字種，音韻変化，モーラ（言語音の単位）	品詞毎に付与した表記規則に従って，分かち書きを決定する，又は形態素の読みを変換する(*)
112	分かち書き	文字	分かち書き後のスペースの数の調整
113	読み	文字種	数字を含む単語の読みの変換（読み誤り，音韻変化への対応）
114	書式	文字	1行32文字以内に調整
115	読み	文字種	数符，外字符，大文字符の挿入
116	読み	文字種，文字	数字とア行，ラ行の仮名の間につなぎ符を挿入
117	読み	文字種，文字	アルファベットの文字体系として扱う記号 ×÷＝＜＞＋－’％＃＆＊＼

(*)，品詞毎の表記規則は108個作成した（詳細は付録Aを参照）.

日本点字表記法の分析結果を基にして作成した点字表記規則を実装し，点訳エンジン Kobe University Intelligent eBraille Engine for ChaSen (KUIC) と命名した．KUICはCGIプログラム群であり，ChaSen 2.3.3[27] の形態素解析結果を入力として，読み仮名の修正，分かち書き，点訳処理の補正を行い，分かち書きされた仮名文を出力する．最後に，分かち書きされた仮名文を点字へ変換する．なお，形態素解析結果は点字表記規則で用いる情報である，形態素の出現形，読み，品詞コード，活用形コードを出力するよう設定した．

次に，KUICの分かち書きの処理について記載する．品詞毎に設定した規則（付録A）を適用し，形態素の前後にスペースを付加するか否かを決定する．基本的なアルゴリズムはeBraille 0.81を基にしている．下記にアルゴリズムを示す：

1. n番目の形態素の品詞コードを参照する．（nの初期値は1）
2. 品詞コードに設定された分かち書きの規則を参照する．
   規則が設定されている場合，規則に従って形態素の前後（あるいは，前か後ろのいずれか）に半角スペース2個又は4個を付加する／付加しない．規則が設定されていない場合は，形態素の前後（あるいは，前か後ろのいずれか）に半角スペース1個を付加する．
3. 分かち書き規則を適用した形態素を出力文字列のバッファに追加する．
4. n←n+1
5. 停止条件「文末である」を満たすまで，1から4を繰り返す．

この後，形態素間の半角スペースを全角スペースに変換して分かち書きを確定する．スペースの変換規則は，(1) 半角スペース1個は消去，(2) 半角スペース2個以上は，2個につき全角スペース1個に変換，(3) 句点（「．」「。」），「!」，「？」の後ろは強制的に全角スペース2個に変換，の3種類である．

分かち書きの規則を適用した後は，点訳処理の補正を行う．具体的には，数量や順序を表す単語の読み仮名や全角スペースの挿入ミスの修正を行う．数量などの単語の読みでは，変換対象「1ニン（1人）」や「2ニチ（2日）」とそれらの読み仮名「ヒトリ」や「フツカ」の対応表を作成して，これを参照することで読みを変換する．全角スペースは，括弧閉の前や読点の後という条件に従い，プログラムがその数を調節する．最後に，前置点やつなぎ符を挿入して，点字翻訳された仮名文が完成する（表5.6のNo. 112-117）．

点字翻訳文の仮名文字や各種の符号を点字へ変換する際は，eBraille0.81と同様の手法で，数字6桁から成る点字6点のドットパターンを用いた（5.1.1節）．このドットパターンは北米点字コードに従い，BASE形式のASCII文字へ変換可能である．そこで，点訳結果をBASE形式ファイルで出力する機能を追加した．その結果，点字プリンタでの印刷が可能となった.

5.4.3 点字表記規則を実装したeBrailleとeBraille 0.81の比較

この節ではeBralle0.81と点訳エンジンKUICを組み込んだ新しいeBraille（version 1.50）の点訳精度を比較して，実装した点訳表記規則の有効性を評価する.

eBraille 0.81では，形態素解析結果から品詞の情報のみを利用して，1形態素の前後の分かち書きを決定していた. 一方, KUICでは, 分かち書きの規則に２つの形態素の品詞, 活用形，出現形，文字種，音韻変化，又はモーラ（拍）[14]の数，を指標とし，それらの組み合わせを条件として分かち書きを決定する.

通常文書評価用コーパス1,983文（点字毎日，85記事）を用いて，新しいeBrailleとversion 0.81の点訳精度を解析した．点字表記規則をKUICへ実装した結果，eBraille 1.50のTotal BTAは，version 0.81の約3倍の点訳精度に達した（表 5.7）．また，両者の点訳精度の分布は, 著しく異なっていた（表 5.7）. eBraille 1.50の分かち書き精度のTotal F_1はversion 0.81の約1.6倍に，読みの精度のTotal scoreは約1.8倍に達した（表5.8，表5.9）.

表5.7. eBraille1.50 と0.81の点訳精度の比較

BTA score	Number of articles with the BTA scores	
	eBraille 0.81	New eBraille
95-100	0	25
90-95	0	42
85-90	0	12
80-85	0	6
75-80	0	0
70-75	0	0
65-70	0	0
60-65	0	0
55-60	1	0
50-55	3	0
<50	81	0
Mean BTA ± SD for each data	30.40 ± 10.92	92.60 ± 4.13
Total BTA (one file of 85 data)	30.58	92.45

表 5.8. eBraille1.50 と 0.81 の分かち書き精度の比較

score	eBraille 0.81	New eBraille
Mean F_l ± SD for each data	59.97 ± 9.44	94.97 ± 3.00
Total F_l	57.86	94.88

表 5.9. eBraille1.50 と 0.81 の読みの精度の比較

score (%)	eBraille 0.81	New eBraille
Mean ± SD for each data	51.71 ± 11.38	97.47 ± 2.00
Total score	52.85	97.44

5.4.4　eBrailleと他の自動点字翻訳プログラムとの比較

この節ではeBraille1.50と他の自動点字翻訳プログラムの点訳精度を比較解析し，実装した点字表記規則の有効性を解明する．

表 5.10にeBrailleの点訳精度と他の自動点字翻訳プログラムとの比較を示す．点訳精度は，新聞記事全233データを一塊として解析した全計算値（Total BTA），各記事の点訳精度の平均と標準偏差(SD)，の2通りで示した．評価には，通常文書評価用コーパス5,191文を用いた．評価の結果，eBrailleのTotal BTA は91.76 BTAだった（表 5.10）．233データの点訳精度の分布では，85 BTA以上が85%以上であり，85 BTA未満の点訳精度の記事は医療に関する記事とインタビュー等の記事であった．これらの記事には，専門用語や方言，口語が他の記事よりも多く含まれているのが特徴であり，精査した結果，これらが点訳精度に影響していた．

次に，eBrailleと他の点字翻訳プログラムの点訳精度を比較した．点訳精度の統計解析では repeated measures ANOVAを行い，ポストテストにTukey法を用いた．その結果，eBrailleの点訳精度が他の全ての自動点字翻訳プログラムよりも有意に高かった(図 5.3)．特に，我々のプログラムを模倣して作られた点字自動翻訳システム [23] の点訳精度は，他の全ての自動点字翻訳プログラムよりも有意に低い点訳精度であった（図 5.3）．以上の結果から，eBrailleは実用上必要にして十分な点訳精度を有していることが示された．

表 5.10. 通常文書評価用コーパス 233 データ(5,191 文)の点訳精度の比較

BTA score	Number of articles for each program with the BTA scores				
	eBraille	点字自動翻訳システム	お点ちゃん	ibukiTenC	IBUKI-TEN
95-100	59	0	25	2	18
90-95	114	0	121	25	99
85-90	31	0	55	77	78
80-85	23	1	20	70	29
75-80	5	4	8	40	6
<75	1	228	4	19	3
Mean BTA ± SD for each data	91.78 ± 4.96	56.39 ± 10.97	90.26 ± 4.90	83.49 ± 6.31	89.22 ± 4.72
Total BTA (one file of 233 data)	91.76	56.20	90.48	83.93	89.05

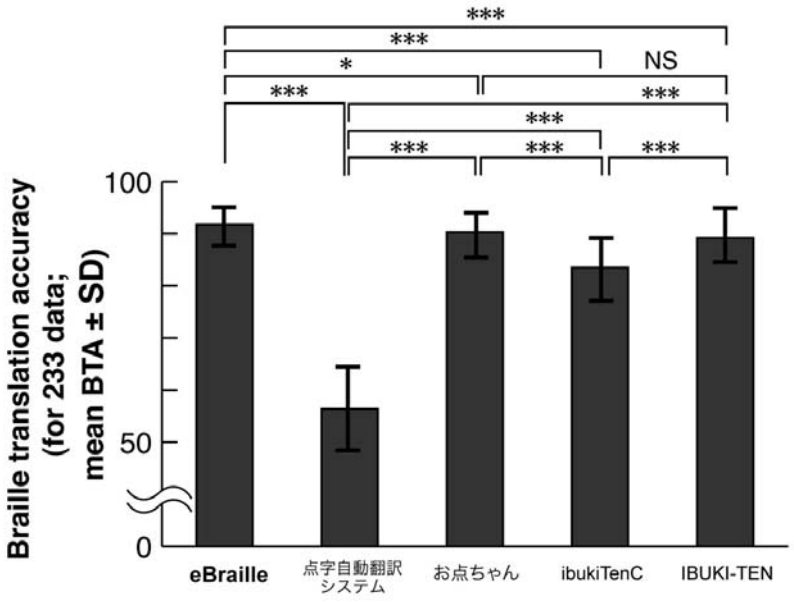


図 5.3. 通常文書評価用コーパスの点訳精度
*, $p < 0.05$; ***, $p < 0.001$; NS, Not significant.

プログラム毎の点訳精度の違いの原因を明らかにするため，点訳精度と自動点字翻訳プログラムが用いる辞書の語彙数の関係を調べたところ, 両者の間に相関はなかった (図5.3，図5.4)．そこで，点訳精度の違いは各プログラムの点訳処理に起因する可能性を考え，それぞれの辞書に登録された語句を解析した.

まず，お点ちゃんでは，人名（姓名)，地名をはじめとする固有名詞や複合名詞に対して，単語の読みをあらかじめ分かち書きして辞書に収録していた. 加えて，各単語に「レベル値」という数値を付与しており，これを利用して漢字の読みや，分かち書きの際の単語間の結合の強弱に関する優先順位を決定していた [22]．次に，ibukiTenC や IBUKI-TEN の辞書には，言語学の分野の品詞体系に則った単語に加えて，点訳に有利になるように形態素を複数まとめた単語や文節も収録している．これらの収録語には，分かち書きの区切り，読みに関する規則等が付与されており，点訳に利用されている [16, 17].

このように eBraille と他の点訳プログラムが点訳時に利用している辞書の構成は大きく異なり，このことは点訳アルゴリズムが異なることを示唆している．実際，ibukiTenC [16]，IBUKI-TEN [17]と eBraille とでは点訳処理法が異なる．ibukiTenC と IBUKI-TEN では，プログラムに組み込まれている文節解析システム ibuki により漢字仮名混じり文を文節単位に分割し，辞書又はプログラムに含まれる点訳規則で文節を更に分割した後に点訳する．それに対して eBraille では，ChaSen を用いて文章を文節よりも小さい単位である形態素に分割し，点訳エンジン KUIC で形態素を単独で，又は複数まとめあげて分かち書きの単位にして点訳する.

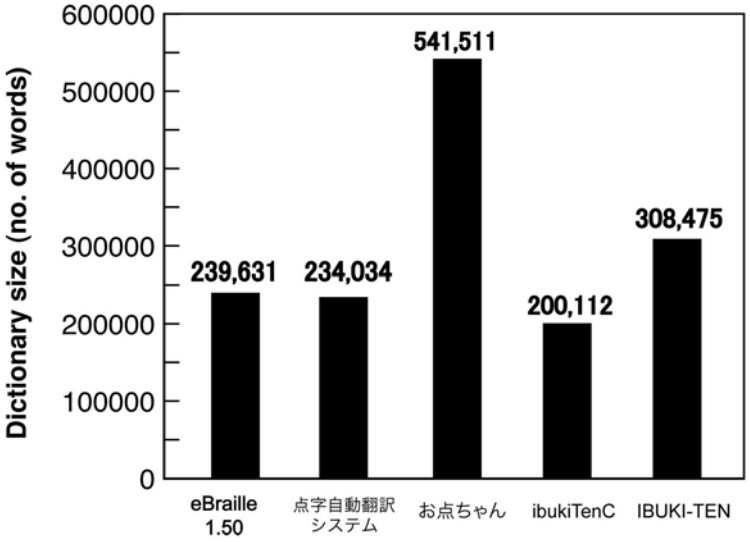


図 5.4. 自動点字翻訳プログラムが用いる辞書の規模

5.4.5 eBrailleの実用性

次に，eBrailleによる自動点字翻訳後の各ブラウザでのCGIプログラムの動作と点字イメージの表示，および点訳後に生成される点字データの標準フォーマットであるBASE形式のファイルの可用性，について検討した．その結果，CGIプログラムの動作と点字翻訳後の点字イメージの表示は，全てのWebブラウザで同様に動作と表示が可能であった．また，プログラムが生成したBASE形式ファイルは，日本で用いられているほぼ全ての点字エディタ，"T・エディタ[1]"，"うぃんびー[2]"，"IBUKI-TEN [17]"，"点字編集システム4"，"EXTRA (有限会社エクストラ，静岡市)"，"ブレイルスター (ニュー・ブレイル・システム株式会社，東京)"（全てWindows上にて動作）で利用可能なことを確認した．

点字文書の印刷では，点字の GIF イメージは，立体コピー作成機（PIAF, Quantum Technonology Pty Ltd., シドニー，オーストラリア）で点字化が可能なことを確認した．

---

1 http://www6.ocn.ne.jp/~t-editor/

2 http://homepage2.nifty.com/winb/index.html

この方法で作成した点字文書は，実際に視覚障害者が支障なく触読可能なことを確認済みである．また，BASE 形式のファイルは，点字プリンタ（DOG-Multi，日本テレソフト株式会社，東京）を使って印刷が可能なことを確認した．

次に，eBrailleの点字翻訳の速度を点訳ボランティアが翻訳した場合と比較した．点訳ボランティアの点字翻訳の速度は個人差があるが，最も速い人で書籍1冊を3日間で翻訳すると言われる[3]．eBrailleの場合，夏目漱石の「坊ちゃん」（88,272字）を点字翻訳させたところ，大学のネットワーク環境を介した場合で33分43秒であった（MacBook Air, Intel Core2Duo 2.13GHz, 2GB RAM, 128GB HDD）．また，eBrailleをスタンドアロンで使用して点字翻訳させた場合は，5分49秒であった．

eBrailleをインターネットに公開したところ，ページビューの数は増加傾向にある（図5.5）．特に，新聞，テレビやラジオの各種メディアで紹介された2008年10月，2009年2月，3月，2010年2月，3月は急増している（表5.11）．また，eBrailleの利用実績としては，視覚障害者向け触地図作成システムで，出発地と目的地の自動点字翻訳に使用されていること[28]に加えて，2010年度から兵庫県立点字図書館でeBrailleを利用した点字印刷のサービスが提供されている[4]．

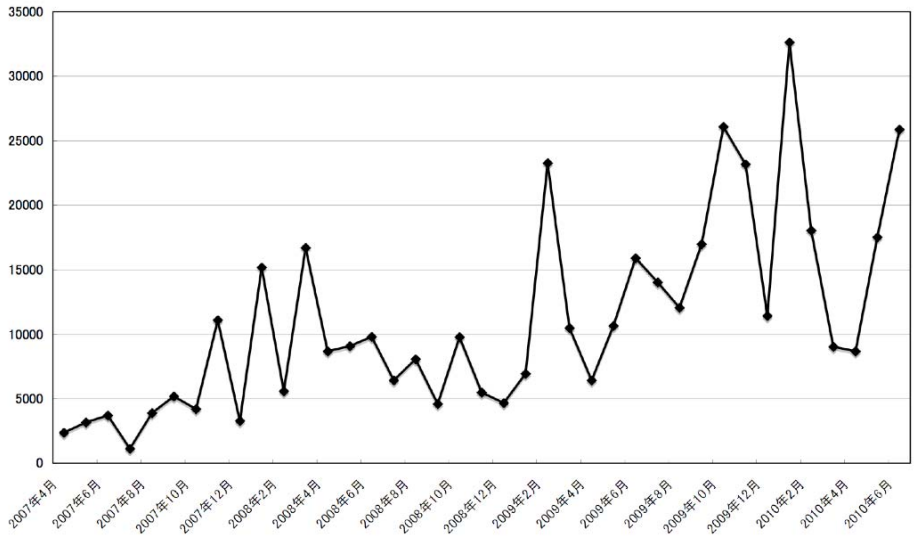


図 5.5. eBraille のページビューの数

[3] 日本点字図書館（東京都）の場合，最速で 10 冊を 1 ヶ月で点字翻訳する．

表 5.11. eBrailleが紹介されたメディアの一覧

年月日	種類	内容
2008年10月22日	神戸新聞	“医療用語の自動点字翻訳システム開発 神戸大”
2009年2月3日	NHKテレビ	「ニュースKOBE発」
2009年2月3日	NHKテレビ	「兵庫ニュース 845」
2009年2月7日	NHKテレビ	「おはよう日本」
2009年2月16日	NHKラジオ第一	「関西ラジオワイド・関西きょうのニュース」
2009年3月22日	点字毎日(点字版) 毎日新聞	“神戸大が点訳 ソフトを公開”
2009年3月26日	点字毎日(活字版) 毎日新聞	“医療情報点訳ソフト開発”
2010年2月25日	神戸新聞	“グッドデザインひょうご” ユニバーサルデザイン部門賞
2010年3月5日	産経新聞	“神大のプログラム受賞”

5.4.6 医療現場での eBraille の使用実績と評価

eBraille を使って神戸大学医学部附属病院の外来案内および入院案内を点訳し，点字プリンタで印刷して点字版の案内パンフレットを作成した．作成した点字版の案内パンフレットは，病院の総合案内，中央受付，外来窓口，入院窓口，眼科外来受付，眼科病棟，患者情報センターに配置し，誰でも閲覧可能にした．次に，点字文書を患者に提供可能になったことを病院内の診療科長等会議で説明すると同時に看護師長会議でも周知し，病院内での認知を高めるようにした．点字による情報提供の開始と運用は，看護部，医事課，我々の部門が共同で実施しており，看護部と我々の部門が窓口となって病院内からの点字文書作成の依頼に対応するための体制を整えている．

次に，点字版の外来案内と入院案内を用いて，eBrailleとその他の自動点字翻訳プログラムの点訳精度を評価した．その結果，他の自動点字翻訳プログラムよりもeBrailleの点訳精度（Total BTA）が高かった（表 5.12）．しかし，新聞記事を基に作成した通常文書評価用コーパスの点訳精度91.76BTA（表5.10のTotal BTA）よりも低かった．これは，

4 http://www.universal-hyogo.jp/cgi-bin/whatsnew/whatsnew.cgi?mode=preview&select=100331094926&file=whatsnew

診療科名や「月」，「火」，という曜日の点訳に失敗したことが原因であった．この結果から，医療文書の点訳には医療用語の辞書への追加が有効であることが示唆された.

表 5.12.　点字版の外来案内・入院案内の点訳精度の比較

	eBraille	点字自動翻訳システム	お点ちゃん	ibukiTenC	IBUKI-TEN
Total BTA for hospital brochures	89.01	61.31	88.58	79.83	83.98

我々の自動点字翻訳プログラムと点字版外来案内と入院案内の有用性と意義について，実物の点字版外来案内と入院案内を用いて神戸視力障害センターの教官と学生，合計30名によるアンケート評価を実施した．回答者の点字学習歴（使用歴）と回答者の点字習熟度は以下の通りだった.

・点字学習歴（使用歴）

1 年未満	16.7%
5 年以上 10 年未満	50.0%
10 年以上	33.3%

・点字習熟度

初級（数字のみ読むことができる）	30.0%
中級（数字と仮名 50 音が読める）	46.7%
上級（数字・全ての音や特殊記号が読める）	23.3%

5 段階評価で満点 5 点の評価の結果，Web ベースの点訳プログラムの必要性については平均 4.3 点，点字知識のない人でも簡単に点訳できるプログラムの有用性は平均 4.4 点，点字版外来案内の有用性は平均 4.8 点，点字版入院案内の有用性は平均 4.7 点であった．これらの結果は我々の自動点字翻訳プログラムが有用であり，点字版外来案内と入院案

内が視覚障害者に役立つものであることを示している.

5.5 まとめ

この章では，eBraille0.81 の点訳誤りと日本点字表記法の点字表記規則を分析し，点訳精度を向上させる方法を明らかにした．分析の結果，より細分化された品詞体系を採用している ChaSen 2.3.3 への更新と，日本点字表記法に則した点字表記規則の追加が必要と判明した．更に，点字表記規則は文字種，品詞，活用形，出現形，読み，音韻変化，モーラ（拍）の数を指標としていること，又，これらの指標を２つの形態素に適用する場合があることを明らかにした.

次に，我々は ChaSen を 2.3.3 へ更新し，点字表記規則を点訳エンジン KUIC へ実装して eBraille の点訳精度を通常文書評価用コーパスで解析した. 解析の結果, 新しい eBraille（version1.50）の点訳精度（Total BTA）は version 0.81 の約 3 倍向上した．また，eBraille の点訳精度（mean BTA ± SD）は，比較に用いた全ての点字翻訳プログラムよりも有意に高かった．他の自動点字翻訳プログラムでは，点字翻訳に特化させた品詞体系や辞書の構成で点訳精度を向上させていた．対して我々のプログラムは，点訳エンジンである KUIC に実装したルールベースでの翻訳により，高い点訳精度を実現していることが明らかになった.

今回，点訳エンジン KUIC により，プログラムの冗長性として旧来の点字エディタ上での利用も可能なパスを作り，点字プリンタでの印刷を可能にした．そして，大学病院で実際に eBraille を利用して点字版の外来案内と入院案内を作成し, 病院内に設置した. 視覚障害者へのアンケート結果から, eBraille の開発は実際に視覚障害者のアクセシビリティの向上に役立つことが示唆された．更に，点字版の外来案内と入院案内を用いて eBraille の点訳精度を計算したところ, 通常文書評価用コーパスを用いての点訳精度よりも低く，医療用語（診療科名）の点字翻訳に失敗していた．このことから，eBraille では辞書に語彙を追加することで，点訳精度の向上が可能なことが示唆された.

第6章

辞書の拡張と点訳精度

6.1　はじめに

我々は，日本点字表記法（2001年版）[2]に対応した点字表記規則を点訳エンジンKUICに実装した．その結果，新聞記事の点訳では高い点訳精度を達成した（第5章）．しかし，医療用語を含む文書の点訳精度は，新聞記事のそれと比べて不十分であった．その原因は医療用語の点訳の誤りである．我々は，医療現場で使用可能な自動点字翻訳プログラムを目指しているため，プログラムの辞書へ医療用語を追加することで，医療文書の点訳に対応することとした．更に，鍼灸やマッサージを含む東洋医学の専門的な文献の点訳のニーズがあるため，東洋医学用語を追加した辞書を作成した．高精度の機械翻訳システムには大規模な対訳辞書が必要であることから[29]，医療用語辞書と東洋医学用語辞書で点訳精度の向上が予想される．そこで，この章では点訳対象の文書２種類を用いて，辞書の規模と点訳精度の関連を明らかにする．具体的には，通常文書評価用コーパスと医療文書評価用コーパスを用いて，辞書の規模が異なる4種類のプログラムの点訳精度を解析し，点訳精度の向上に有効な辞書を解明する．

6.2　IPADIC

eBrailleはChaSenの辞書IPADIC [30]を使用している. IPADICはIPA品詞体系に基づく日本語辞書で, ChaSenの開発者らが管理してきた. 我々は, version 2.7.0 [30]を使用している. 図 6.1にIPADICの辞書定義ファイルの内容の例を示す. IPADICには, 各単語の見出し語に加えて, 品詞, 読み, 発音, 形態素生起コストが付与されている. 形態素生起コストは, 形態素解析済みのデータであるRWCPコーパス [31]から学習した単語の出現確率を基に計算した数値である. この数値が小さいほど, 形態素が出現しやすいことを示している [30].

図 6.1. IPADIC の辞書定義ファイルの内容

6.3 医療用語辞書

医療用語辞書は, ComeJisyo V1 [32]を基にして作成した. この医療用語辞書は, 看護領域を中心とした用語 30,146 語を収録しており, 疾患や怪我の名称, 検査名, 症状, 解剖学的な体の部位の名称, 薬品・薬剤名, 機器・道具の名称, 化学物質の名称を含んでいる. そして, これらの名詞には, 出現形 (見出し語), 品詞, 品詞細分類, 原形, 読み, 発音に加えて, 文脈に関するタグと辞書独自のタグが付与されている. この辞書は, 医療用語の形態素解析で, 複合語をそのまま出力させることを目的として作成された. そのため, 文字列の長い複合語が収録されている [32]. 例えば, 読み仮名が 7 文字以上の複合名詞は辞書全体の約 6 割を占める. 日本点字表記法では, 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, 文全体の意味を速く理解することが困難となる

ため，これを自立可能な構成要素で区切るように定めている [2]．そこで我々は，文字列の長さに関わらず複合名詞を構成要素に分割した．そして，IPADIC の収録語と重複しない単語を人手で抽出し，医療用語辞書に用いた．表 6.1 に ComeJisyo に収録された長い文字列の医療用語とその構成要素への分割の例を示す．読み仮名が 7 文字以上の単語であっても「腕橈骨筋（ワントウコツキン）」や「没食子酸（ボッショクシサン）」のように，これ以上小さい単位に分割することが不可能な名詞は，そのまま辞書に収録した．また，医療従事者の間でのみ使用されるような略語は追加の対象外とした．ComeJisyo のタグは，出現形と原形，発音と読みは全く同じであったため，我々の医療用語辞書には，出現形，品詞，品詞細分類，読みを用いた．以上の結果，8,170 語を抽出した．最後に，抽出した医療用語を IPADIC に追加し，RWCP コーパスを用いて浅原ら [33]の手法で形態素生起コストを算出した．そしてこの辞書を用いた eBraille を「eBraille-M」と命名した．

表 6.1. 医療用語の抽出例

Words	*Kana* translation
亜急性 | 感染性 | **心内膜炎**	アキュウセイ | カンセンセイ | **シンナイマクエン**
下腹 | **神経叢** | 損傷	カフク | **シンケイソウ** | ソンショウ
周期性 | **好中球** | 減少症	シュウキセイ | **コウチュウキュウ** | ゲンショウショウ
上部 | 心臓型 | 総 | 肺静脈 | 還流 | 異常症	ジョウブ | シンゾウガタ | ソウ | ハイジョウミャク | カンリュウ | イジョウショウ
膿瘍性 | 横行 | 結腸 | **憩室**	ノウヨウセイ | オウコウ | ケッチョウ | **ケイシツ**

各単語は構成要素に分割した．太字の構成要素は IPADIC にない単語であるため，辞書に追加した．

6.4 東洋医学用語辞書

東洋医学用語辞書の作成には，古医書や東洋医学文献の専門出版社から辞書を提供してもらい，これを使用した．この辞書は，編集者が漢方や鍼灸の専門書，古典，学術書等の編集の過程で，ワードプロセッサ用に単語を登録して作成した辞書と，東洋医学の研

究者が個人的に作成した辞書を含んでおり，合計 25,756 語を収録している．辞書のタグは，見出し語，読み，品詞であり，収録語の漢方薬の一部には調剤に使用する生薬の名前が付されていた．この辞書も長い文字列の名詞を含んでいるため，構成要素へ分割した（表 6.2）．構成要素には，辞書内の短い文字列の名詞と調剤のタグに書かれた生薬の名前を用いた．専門性が高く，構成要素の特定が困難な用語については，東洋医学の文献研究の専門家に判定を依頼して，複合名詞を分割した．なお，この辞書には「ICU」や「RNA」などの西洋医学の用語も含まれていたが，東洋医学の分野で使用される可能性が低いため，これらの用語を除外した．その結果，東洋医学の理論や思想，漢方薬や人名を含む 17,290 語を抽出した．これらの東洋医学用語を IPADIC に追加して，医療用語辞書と同様の方法で形態素生起コストを算出した．この辞書を用いた自動点字翻訳プログラムを「eBraille-TM」と命名した．

なお，作成した医療用語辞書と東洋医学用語辞書の収録語を比較したところ，両者で重複する単語が301個あった．これらの語は，漢方薬名，疾患名，体の部位，解剖学用語，体調や症状を表す名詞だった．また「ぶどう膜炎」と「くる病」を除いて全て漢字のみの名詞であり，約65%が漢字2文字，約32%が漢字3文字，約3%が漢字1文字だった．

表 6.2.　東洋医学用語の抽出例

Words	*Kana* Translation
芍薬 | **地黄** | 湯	シャクヤク | **ジオウ** | トウ
絡傷 | 出血 | 証	**ラクショウ** | シュッケツ | ショウ
陽虚 | 発熱	**ヨウキョ** | ハツネツ
婦人 | **良方** | 大全	フジン | **リョウホウ** | タイゼン
胞瞼 | **風熱** | **外襲** | 証	**ホウケン** | **フウネツ** | **ガイキョウ** | ショウ

各単語は構成要素に分割した．太字の構成要素は IPADIC にない単語であるため，辞書に追加した．

6.5　allBrailleの辞書

医療用語 8,170 語と東洋医学用語 17,290 語と IPADIC を統合し，形態素生起コストを算

出して，辞書を作成した．この辞書を使用する自動点字翻訳プログラムを「allBraille」と命名し，eBraille, eBraille-M, eBraille-TM と共に点訳精度の比較解析に供した（表 6.3）．

表 6.3. 評価に用いたプログラムの辞書の構成

Dictionary	**eBraille**	**allBraille**	**eBraille-M**	**eBraille-TM**
ipadic 2.7.0（239.631 words）	●	●	●	●
Medical words dictionary（8,170 words）	–	●	●	–
Oriental traditional medicine dictionary（17,290 words）	–	●	–	●

6.6 実験方法

通常文書評価用コーパス5,191文（233データファイル）と医療文書評価用コーパス3,180文（77データファイル）の2種類を用いて，辞書の規模が異なる4種類のプログラム，eBraille，allBraille，eBraille-M，eBraille-TM，の分かち書き精度（F_1）と漢字や記号の読みの正解率（正音率），点訳精度をデータファイル毎に算出し，mean ± SDで比較解析した（図 6.2）．また，医療文書の点訳精度の向上に有効な辞書を解明するため，各プログラムの点訳の誤りを比較した．

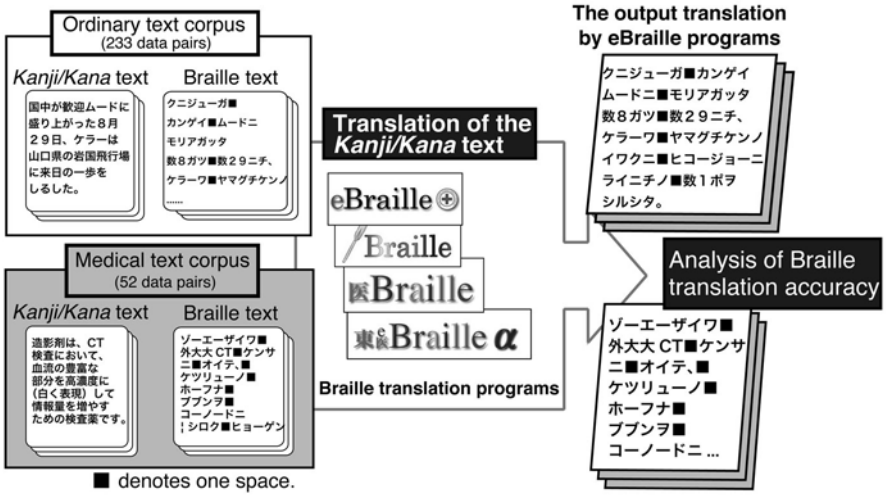


図 6.2. eBraille，allBraille，eBraille-M，eBraille-TMの評価手順

6.7 結果

はじめに，通常文書評価用コーパスでの結果を記述する．4種類の自動点字翻訳プログラムの分かち書き精度，読みや記号の変換の精度（正音率），点訳精度の結果を表6.4に示す．これら3種類の精度の平均値を比較したところ，語彙数の最も多い辞書を使用するallBrailleは，eBraille-Mよりも低かった（表6.4）．更に，allBrailleとeBraille-Mの分かち書き精度，読みや記号の変換の精度，点訳精度のpaired *t*-test を行った結果，両者の正音率には有意差がなかったが，分かち書き精度と点訳精度では，eBraille-Mが有意に高かった（図6.3a，b，c）．eBraille-Mの分かち書き精度と点訳精度が高い理由は，通常文書評価用コーパス中に医療や医学に関する記事があり，これらの精度がeBraille-Mの方が高かったためだった.

表 6.4.　通常文書評価用コーパスの各種精度

Score	eBraille	allBraille	eBraille-M	eBraille-TM
分かち書き精度（F_1）	94.78 ± 3.44	94.81 ± 3.41	**94.93** ± 3.37	94.72 ± 3.38
読みの精度（%）	96.79 ± 2.73	96.97 ± 2.54	**96.99** ± 2.42	96.93 ± 2.67
点訳精度（BTA）	91.78 ± 4.96	91.98 ± 4.75	**92.11** ± 4.63	91.85 ± 4.85

各精度の数値は，233 データファイルの mean ± SD を示す.

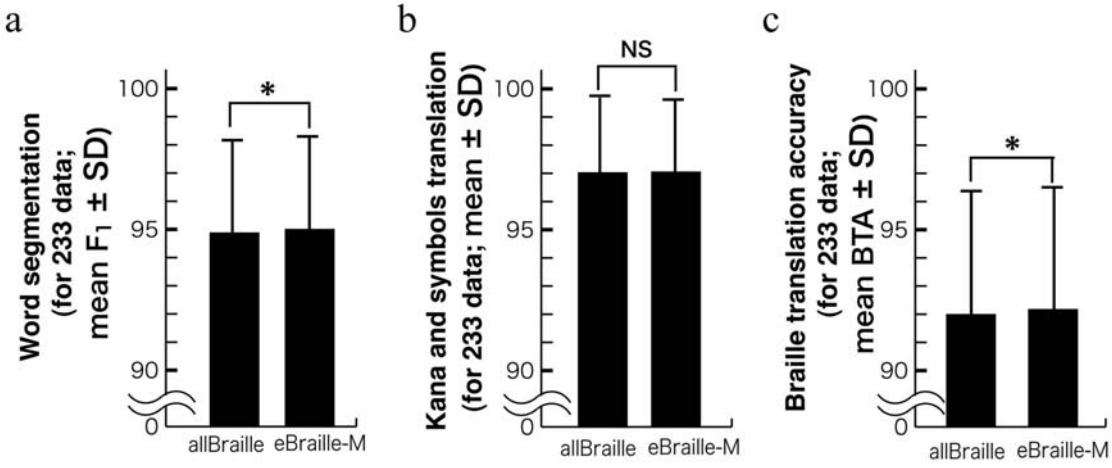


図 6.3.　allBraille と eBraille-M の通常文書評価用コーパスの各種精度
a，分かち書き精度（F_1）；b，読みの精度；c，点訳精度.
*, $p < 0.05$; NS, Not significant.

医療文書評価用コーパスでの分かち書き精度，読みの精度，点訳精度を表 6.5 に示す．4 種類の自動点字翻訳プログラムを各種精度の平均値毎に比較したところ，いずれも eBraille-M の精度が他のどのプログラムよりも高く，次に allBraille の精度が高かった．allBraille と eBraille-M の分かち書き精度, 読みや記号の変換の精度, 点訳精度の paired *t*-test を行った結果，通常文書の評価結果と同様に，両者の正音率には有意差がなかったが，分かち書き精度と点訳精度では eBraille-M が有意に高かった（図 6.4a，b，c）．

表 6.5.　医療文書評価用コーパスの各種精度

Score	eBraille	allBraille	eBraille-M	eBraille-TM
分かち書き精度 (F_1)	92.73 ± 5.64	95.77 ± 2.93	**96.06** ± 2.74	93.73 ± 4.64
読みの精度 (%)	94.77 ± 4.07	96.20 ± 2.99	**96.32** ± 2.62	96.05 ± 3.94
点訳精度 (BTA)	88.06 ± 8.29	92.17 ± 4.89	**92.57** ± 4.60	89.21 ±7.17

各精度の数値は，77 データファイルの mean ± SD を示す.

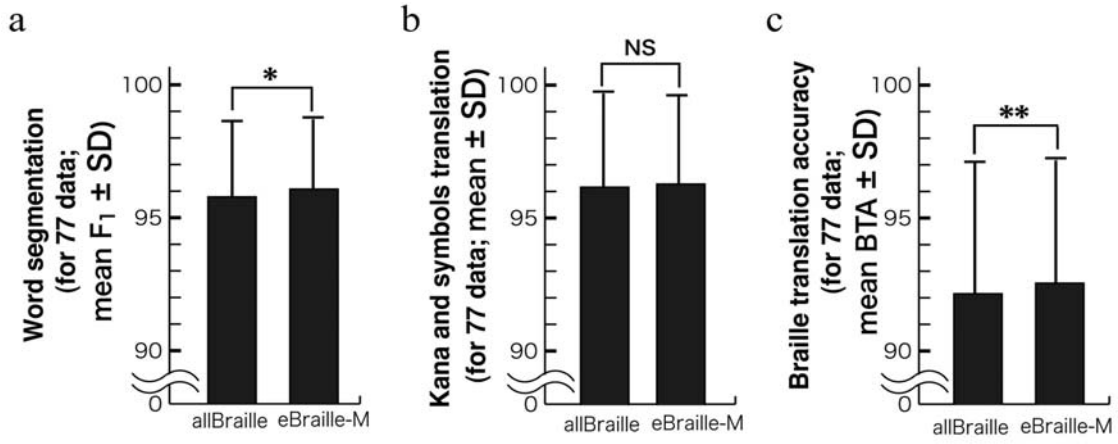


図 6.4.　allBraille と eBraille-M の医療文書評価用コーパスの点訳精度
a，分かち書き精度（F_1）; b，読みの精度 ; c，点訳精度.
*, $p < 0.05$; **, $p < 0.01$; NS, Not significant.

次に，各自動点字翻訳プログラムによる医療文書評価用コーパスの点訳の誤りを精査した．分かち書きでは，eBraille と eBraille-TM が正解の分かち書きよりも細かく区切っていた．その例として，「脳動静脈瘻（ノー■ドージョーミャクロー)」，「下唇赤唇部癌（カシン■セキシンブガン)」という漢字で構成される疾患名の複合名詞や「キシロカイン（キシロカイン)」，「ミエロパチー（ミエロパチー)」，「はきけ（ハキケ)」を代表とする仮名の名詞があった．

・脳動静脈瘻（正解：ノー■ドージョーミャクロー）

eBraille	ノー■ドージョーミャク■　（「瘻」に相当する出力無し）
eBraille-TM	ノー■ドー■ジョーミャク■ロー

・下唇赤唇部癌（正解：カシン■セキシンブガン）

eBraille	シタクチビル■アカ■クチビルブ■ガン
eBraille-TM	シタクチビル■アカ■クチビルブ■ガン

・キシロカイン（正解：キシロカイン）

eBraille	キシ■ロ■カイン
eBraille-TM	キシ■ロ■カイン

・ミエロパチー（正解：ミエロパチー）

eBraille	ミエ■ロ■パ■チー
eBraille-TM	ミエ■ロ■パ■チー

・はきけ（正解：ハキケ）

eBraille	ワ■キケ
eBraille-TM	ワ■キケ

上記の仮名の名詞は，eBraille と eBraille-TM の辞書に収録されていないために分かち書きに失敗したと考えられた．一方 allBraille はこれらの複合語や名詞を正しく分かち書きしたが，「貯痰音（チョタンオン)」や「十二指腸潰瘍（ジューニシチョー■カイヨー)」，「針刺入部（ハリシニューブ)」では，「チョタン■オト」，「ジュー[数符] 2 シチョーカ

イヨー」，「シンシ■ニューブ」と分かち書きを失敗していた．医療用語を追加した eBraille-M は，「貯痰音」と「十二指腸潰瘍」，「針刺入部」のいずれも正しく分かち書きしていた（下記の赤で記載された「■」や「｜」と赤字は，誤りの箇所を表す）．

・ allBraille の分かち書き誤りを eBraille-M が正しく出力した例

墨字	allBraille	eBraille-M（正解）
貯痰音は	チョタン■オトワ	チョタンオンワ
心臓は拍動	シンゾー■ワハクドー	シンゾーワ■ハクドー
尿混濁	ニョー｜コンダク	ニョー■コンダク

最後に，漢字や記号の読みについて精査した．まず eBraille の点訳では，「脳動静脈瘻」の「瘻」，「嘔気」の「嘔」や「心嚢前縦隔」の「隔」の漢字の読みが出力されなかった．eBraille-M と allBraille では，「両下肢紅斑（リョーカシ■コーハン）」，「間質性（カンシツセイ）」の読みを正しく出力したが，辞書中に医療用語を追加していない eBraille-TM では，「リョーカシベニ■ブチ」，「マシツセイ」と，誤った読みを出力していた．allBraille は，辞書中に含まれる単語であっても「心窩部（シンカブ）」，「歯肉炎（シニクエン）」や「貯痰音（チョタンオン）」の読みを，「ココロ■ブ」，「ハニクエン」，「チョタン■オト」と誤っていたが，eBraille-M では正しく点訳していた．

・ allBraille の読み誤りを eBraille-M が正しく出力した例

墨字	allBraille	eBraille-M（正解）
口蓋帆	コーガイホ	コーガイハン
嘔気	オーケ	オーキ
易感染症	イカンセンショー	エキカンセンショー
心嚢	ココロ■	シンノー

（allBraille は「心嚢」の「嚢」を出力できていない）

allBraille と eBraille-M の各々の点訳誤りの数のうち，医療用語の誤りの数が占める割合は，allBraille では約 42%，eBraille-M では約 14%だった．このことから，allBraille が医療用語の点訳に失敗していることが，eBraille-M よりも各種の精度が低い原因であることが示唆された．また，4 種類の eBraille はいずれも同じ点字表記規則を実装していることから，プログラム間の点訳の違いは，辞書の語彙構成，形態素生起コストや形態素解析に起因すると考えられた.

6.8 まとめ

医療文書の点訳における特徴的な点訳の誤りは, 医療用語の名詞や複合名詞に見られた. また，点訳誤りの原因は，辞書の収録語に依存する場合に加えて形態素解析に起因する場合があると考えられた.

実験に用いた通常文書評価用コーパスは，医療や医学に関する記事を含んでおり，医療文書評価用コーパスは，西洋医学を中心とした，病院で使用する医療用語を多く含む文書を基に作成された．eBraille-TMとallBraille では東洋医学用語を多く含む辞書を使用しており，allBrailleはそれに加えてeBraille-Mの辞書と同じ医療用語を追加した辞書を使用している．にもかかわらず，eBraille-TMとallBrailleの両方で点訳精度がeBraille-Mよりも有意に低い結果だった. eBraille-Mの方が他のeBrailleプログラムよりも有意に点訳精度が高かったことから，対象の文書に適した語彙構成の辞書が自動点字翻訳プログラムの点訳精度を向上させることが示された.

第7章
点字の分かち書きへの統計的手法の利用

7.1　はじめに

我々は，日本点字表記法（2001 年版）[2]に準拠した点訳エンジン KUIC により，eBraille の点訳精度を向上させた (第 5 章)．次に，我々は点訳対象に適合した辞書の語彙構造が点訳精度の向上に有効であることを示した．そして，我々の医療用語辞書が医療文書の点訳精度を向上させることを明らかにした（第 6 章）．しかし，日本点字表記法には，これまでの方法で対応が困難な点字表記規則がある．具体的には，(1）単語の自立性の強弱，(2）意味や発音の境界を指標とする分かち書き規則である（第 5 章）．しかし，これら 2 種類の指標は曖昧な定義になっている．よって，点訳エンジン KUIC の表記規則としての実装は不可能である．そこで，この章では新たに別の方法で分かち書きへの対応を試みる．

曖昧な 2 種類の指標は，主に長い文字列の複合語や固有名詞を対象としており，その分かち書きの原則は，単語を構成要素に分割し「自立可能な意味の成分の前は区切り，副次的な意味の成分は続ける」として，意味の境界に焦点を当てている．自然言語処理で意味を取り扱う場合，基本的に意味タグ（語義）が付与された辞書やコーパスを用意する方法を用いる [34]．しかし，辞書やコーパスの作成コストは非常に大きい [35]ことに加えて，タスクに適したアノテーションの付与は困難である．別の方法として統計的手法による意味の解析の適用が考えられる [35]．統計的学習モデルの適用が可能な理由

は，日本語の分かち書きは，chunk という任意の文字列をその機能ごとに分類する一連の手続とみなす事ができるためである[36, 37]．自然言語処理に適用されてきた統計的手法のアルゴリズムには，隠れマルコフモデル (HMM) [38-42]や Maximum Entropy (ME) [43, 44]，Support Vector Machine（SVM）[45]がある．HMM と ME は多数のタスクに用いられてきた．その例を挙げると，形態素解析，POS タグ付け，構文解析，固有表現抽出，音声認識，機械翻訳，情報検索がある [46-51]．これらのアルゴリズムと比較して，SVM は多くの素性を用いた高次元データでも汎化能力が高く，文書分類や日本語の係り受け解析等に非常に有用であることが報告されている[37]．そこで，SVM に基づく統計的学習モデルを作成し，分かち書きの精度におよぼす影響を明らかにする．

7.2 統計的学習モデル

SVM を利用した統計的学習モデルには，汎用的なチャンカーである YamCha [52]を用いた．この YamCha は，人手による規則を用いた検出器よりも高い性能を発揮することが報告されている[53].

7.3 実験方法

この実験では，文章を形態素（token）に分割した後，形態素を分かち書き（chunk）にまとめあげるチャンキングをタスクとし，統計的学習モデルに正しい分かち書きを学習させる．

統計的学習モデルの学習コーパスには，我々の作成した医療文書評価用コーパス (4.1.2 節) から，大学病院の看護記録と患者向けの文書（治療説明書や検査票等），電子カルテ用標準病名マスター[24]の疾患名を含む文，難治性疾患克服研究事業の対象疾患名[25]を含む文，の合計 648 文（4,233 chunks）を選択して用いた．これらのデータは，

eBraille-M での分かち書き精度が他のデータよりも低いために選択した.

次に，選択した文章を eBraille-M の辞書（6.3 節）と ChaSen [27]を用いて形態素解析した．そして，解析結果の形態素毎に正解の分かち書きのラベルを付与した．分かち書きのラベルは，以下の 2 種類を用いた.

B　分かち書き（chunk）の先頭であることを示す
I　分かち書き（chunk）の一部であることを示す

統計的学習モデルは，学習素性を基にこれらの分かち書きのラベルを学習する.
学習の素性には, ChaSen の形態素解析結果のうち, 見出し語の出現形, 読み仮名, 発音, 見出し語の基本形，品詞の全階層，活用形を用いた．加えて，モーラ（拍）の数として各形態素の文字数を素性に用いた．モーラは言語音の単位の一つ[14]で，日本点字表記法ではこれを分かち書きの区切れの一基準と定めている（2.1.2 節）．以下に，実験で用いた学習素性とその例の一覧を示す：

素性	例
見出し語	山田
読み	ヤマダ
発音	ヤマダ
基本形	山田
品詞	名詞
品詞細分類 1	固有名詞
品詞細分類 2	人名
品詞細分類 3	姓
活用形	* （入力値なし）
モーラ（拍）の数	3

なお，「〜によって」，「〜として」など，ChaSen が 1 形態素として解析する「助詞-格助詞-連語」は，点訳エンジン KUIC へ実装した表記規則では「〜に■よって」，「〜と■して」と 1 文字目の助詞の後を区切って 2 tokens に分割している（5.4.2 節）ことから，学習コーパスでも同様に，連語を 2 tokens に分割し，各 token の品詞と品詞細分類はいずれ

も「助詞」-「格助詞」-「連語」を付与した．連語以外の形態素は，ChaSen の解析結果をそのまま用いた．そのため，ChaSen が出力した形態素が分かち書きの単位よりも大きい場合，例えば，複合名詞を構成要素に分割せずに出力した場合や，形態素の境界と分かち書きの境界が異なる場合でも ChaSen の出力をそのまま token として使用した．

分かち書きに有効な学習素性を解明するため，合計 10 種類の素性のうち，「発音」，「モーラ（拍）の数」を加減した組み合せで 4 種類の学習コーパスを用意し，統計的学習モデルを作成した．「発音」は長音の表記が ChaSen の解析結果と日本点字表記法[2]とで異なる部分があるため（5.4.2 節），「モーラ（拍）の数」は複合名詞の分かち書きの基準の一つであるため（7.1 節），加減する学習素性の対象とした．

・統計的学習モデル A（Statistical Learning Model A, SLM-A）
10 種類の素性から「発音」と「モーラ（拍）の数」を削除した学習コーパス

・統計的学習モデル B（SLM-B）
10 種類の素性から「モーラ（拍）の数」のみを削除した学習コーパス

・統計的学習モデル C（SLM-C）
10 種類の素性から「発音」のみを削除した学習コーパス

・統計的学習モデル D（SLM-D）
10 種類の全ての素性を使用した学習コーパス

統計的学習モデルの評価では，学習コーパスを医療文書評価用コーパスの内容別に 11 ファイルに分け，10 ファイルを訓練セット，残り 1 ファイルを評価セットとして 11 組の学習セットで交差検定を行い，分かち書きの精度（F_l）を算出した．eBraille-M の評価は，11 ファイルの分かち書きの精度の平均を算出し，4 種類の統計的学習モデルの分かち書き精度と比較した（表 7.1）．更に，統計的学習モデルと eBraille-M の分かち書きの誤りを比較して解析した．

表 7.1.　評価したプログラムのシステム構成

プログラム	システム構成	使用しない素性
eBraille-M	KUIC, ChaSen, IPADIC + Med dic.	ー
SLM-A	ChaSen, IPADIC + Med dic.	発音，モーラの数
SLM-B	ChaSen, IPADIC + Med dic.	モーラの数
SLM-C	ChaSen, IPADIC + Med dic.	発音
SLM-D	ChaSen, IPADIC + Med dic.	なし

KUIC, 点訳エンジン; ChaSen, ChaSen 2.3.3; IPADIC, ipadic 2.7.0; Med dic., 医療用語辞書; SLM, 統計的学習モデル

7.4　結果

4 種類の統計的学習モデルと eBraille-M の分かち書き精度を比較した結果，全ての統計的学習モデルが eBraille-M よりも高い分かち書き精度を示した（表 7.2）．このことから，統計的学習モデルが分かち書きに有用であることが示された．SLM-A から SLM-D の 4 種類の統計的学習モデルの比較では，10 種類の学習素性から「発音」のみを除いた SLM-C が最も高い精度を示した（表 7.2）．

表 7.2.　eBraille-M と 4 種類の統計的学習モデルの分かち書き精度

	Precision (%)	Recall (%)	F-measure (F_1)
eBraille-M	93.43	92.84	93.13
SLM-A	93.24	93.23	93.21
SLM-B	93.25	93.27	93.23
SLM-C	**93.49**	**93.23**	**93.34**
SLM-D	93.43	93.25	93.31

SLM, 統計的学習モデル

次に，eBraille-M と SLM-C の分かち書きの誤りを比較した．SLM-C は，「視野■欠損」や「歯科■口腔」といった複合名詞の分かち書きを正しく出力していた．しかし，点訳エンジン KUIC に実装されている表記規則によって eBraille-M が正しく出力している分かち書きの出力に失敗していた（以後，赤で記載された「■」や「｜」は誤りの箇所を表す）．

・eBraille-M の誤りを SLM-C が正しく分かち書きした例

SLM-C（正解）	eBraille-M
視野■欠損	視野｜欠損
歯科■口腔	歯科｜口腔
示指■ガングリオン	示指｜ガングリオン
酸■ホスファターゼ	酸｜ホスファターゼ
肺■アスペルギルス症	肺｜アスペルギルス症

・SLM-C の誤りを eBraille-M が正しく分かち書きした例

SLM-C	eBraille-M（正解）
ビタミン｜D	ビタミン■D
CT ｜にて	CT■にて
吸引■にて	吸引にて
夕食｜以後	夕食■以後
角膜■混濁｜等	角膜■混濁■等

SLM-C が複合名詞の分かち書きを正しく出力したことから，複合名詞 750 個（2,238 chunks）のみの分かち書き精度（F_1）を eBraille-M と SLM-C で比較した．その結果 eBraille-M の分かち書き精度は 88.83 で，SLM-C は 93.08 であった．以上の結果から，SLM-C は，複合名詞の分かち書きには有効であるが，点訳エンジン KUIC の表記規則で正しく出力する分かち書きを誤ることが示された．

7.5 ルールベースと統計的学習の融合

7.5.1 eBraille-M の出力を利用した 2 種類の統計的学習モデル

前節の結果から, ルールベースの eBraille-M と統計的学習モデル SLM-C の両者の正しい分かち書きを組み合わせると高い分かち書き精度を実現する可能性が示唆された. eBraille-M と SLM-C の出力の組み合わせには, 次の 2 種類の方法が考えられる. 一つは, eBraille-M の出力と SLM-C の出力から正しい分かち書き出力を選択する一種のスタッキング[54]の手法である. もう一つは, eBraille-M の分かち書き出力を学習素性として新たに統計的学習モデルを作成する方法である. 我々は, これらの 2 種類の方法での分かち書き精度を算出し, SLM-C を単独で用いた場合の分かち書き精度と比較した.

まず, Semi-stacking モデルとして, eBraille-M の出力と SLM-C の出力を選択する統計的学習モデルを作成した. このモデルの学習素性には, SLM-C と同様に, ChaSen の形態素解析結果である, 見出し語, 読み, 基本形, 品詞と品詞細分類を含む全階層, 活用形とモーラの数を用いた. 加えて, 形態素毎に eBraille-M と SLM-C の分かち書き出力を付与した. 具体的には, eBraille-M の分かち書き出力の開始を eBM-B, 分かち書き出力の一部を eBM-I とし, SLM-C の出力では, 分かち書きの開始を SLMC-B, 分かち書きの一部を SLMC-I とした. Semi-stacking モデルが学習する正解のラベルは eBraille-M 又は SLM-C とした.

次に, SLM-C の学習素性に eBraille-M の分かち書き出力（B 又は I）を追加し, 統計的学習モデル（SLM+eBM）を作成した. SLM+eBM が学習する正解の分かち書きのラベルは, SLM-C と同様に B 又は I とした.

以上の 2 種類の統計的学習モデルの分かち書き精度を, 7.3 節と同様に交差検定で算出し, SLM-C の精度との比較に用いた.

7.5.2 分かち書き精度と誤りの比較

ChaSen の形態素解析結果とモーラの数を学習素性とした SLM-C，eBraille-M と SLM-C の出力から正しい出力を学習する Semi-stacking モデル，eBraille-M の分かち書き出力を学習素性に組み込んだ SLM+eBM の 3 種類のモデルの分かち書き精度を比較した．その結果，医療文書 648 文の分かち書き精度は，SLM+eBM が最も高かった（表 7.3）．しかし，複合名詞 750 個のみの分かち書き精度は SLM-C が最も高かった（表 7.4）．このことから，SLM+eBM と SLM-C の分かち書き誤りのパターンが異なることが示唆された．

表 7.3. SLM-C と 2 種類の統計的学習モデルの医療文書の分かち書き精度

	Precision (%)	Recall (%)	F-measure (F_1)
SLM-C	93.49	93.23	93.34
Semi-stacking M	94.19	93.96	94.07
SLM+eBM	**94.57**	**94.49**	**94.53**

表 7.4. SLM-C と 2 種類の統計的学習モデルの複合名詞の分かち書き精度

	Precision (%)	Recall (%)	F-measure (F_1)
SLM-C	**92.51**	**93.67**	**93.08**
Semi-stacking M	90.23	90.89	90.54
SLM+eBM	91.34	92.54	91.92

そこで，SLM+eBM と SLM-C の分かち書きの誤りを精査し，各モデルに共通する誤りの数，あるいは eBraille-M と共通する誤りの数を分かち書きの単位で比較した．SLM+eBM の誤りの数は 136 個（136 chunks），SLM-C の誤りの数は 155 個で，両者に共通の誤りの数は 76 個だった（表 7.5）．SLM+eBM と SLM-C に共通する誤りの数が各モデルの誤りの数に占める割合は，いずれも 50%前後であった．

表 7.5.　SLM+eBM, SLM-C, eBraille-M の共通の誤りの数

eBraille-M（179）	SLM+eBM（136）	SLM-C（155）	共通の誤り
○	○	○	67
／	○	○	76
○	○	／	126
○	／	○	70

（　）内の数値は各モデルあるいはプログラムの誤りの数を表す.

次に，SLM+eBM と SLM-C の誤りのパターンを精査したところ，SLM+eBM は一続きに表記する名詞，助詞やアルファベットに関わる分かち書きを正しく出力したが，SLM-C は失敗していた．一方，SLM-C は，複合名詞の内側を 2 モーラで区切る分かち書きを正しく出力していたが，SLM+eBM は誤って出力していた．下記に SLM+eBM と SLM-C の分かち書きの例を示す.

・SLM-C の誤りを SLM+eBM が正しく分かち書きした例

SLM+eBM（正解）	SLM-C
ヘアピン	ヘア■ピン
直腸瘻	直腸■瘻
ビタミン■D	ビタミン|D
CT■にて	CT |にて
吸引にて	吸引■にて
夕食■以後	夕食|以後
角膜■混濁■等	角膜■混濁|等

・SLM+eBM の誤りを SLM-C が正しく分かち書きした例

SLM+eBM	SLM-C（正解）
視野|狭窄	視野■狭窄
歯科|口腔	歯科■口腔
皮膚|剥離	皮膚■剥離
皮下|組織	皮下■組織
眼|オンコセルカ症	眼■オンコセルカ症
腱|移行術	腱■移行術
酸|ホスファターゼ■欠損症	酸■ホスファターゼ■欠損症

さらに，SLM+eBM が出力した正しい分かち書きが，eBraille-M に実装した表記規則と合致していることから，SLM+eBM は eBraille-M の表記規則を学習していることが明らかになった.

一方， SLM+eBM と eBraille-M に共通する誤りの数は 126 個で，SLM+eBM の誤りの数の約 90%を占めていた．このことから，SLM+eBM が eBraille-M の分かち書きの誤りをも学習している可能性が示された. 但し, SLM+eBM は eBraille-M の誤りの数の約 30%を正しく分かち書きしていた. 具体的には, SLM+eBM は記号の後を区切る分かち書き，名詞や複合名詞の分かち書き，名詞に動詞「ある」，又は，形容詞「ない」が続く分かち書きを正しく出力した.

・eBraille-M の誤りを SLM+eBM が正しく分かち書きした例

SLM+eBM（正解）	eBraille-M
1.■手術	1. |手術
FDG-PET	FDG-■PET
貯痰音	貯痰■音
視野■欠損	視野|欠損
ただれ■あり	ただれ|あり
訴え■なし	訴え|なし

SLM+eBM は点訳エンジン KUIC に実装した表記規則と同様の分かち書きを出力することから，KUIC の表記規則を学習していることが明らかになった．また，eBraille-M の分かち書き誤りの約 30%を修正することが示された．同時に，SLM+eBM が eBraille-M の誤りを学習する場合があり，2 モーラで区切る複合名詞の分かち書きについては，eBraille-M の出力を学習素性としない SLM-C の方が精度が高いことが示された．これらの結果から，点訳エンジン KUIC の表記規則を学習した SLM+eBM と，複合名詞の分かち書き精度の高い SLM-C の双方の正しい分かち書き出力を組み合わせると精度が更に向上することが予想された.

7.5.3　分類器の導入

分類器の作成

医療文書で最も分かち書き精度が高かった SLM+eBM と複合名詞の分かち書きで最も成績が良かった SLM-C の分かち書き出力を組み合わせて正しい分かち書きを実現する目的で, 2 種類の分類器を作成した. まず, スタッキングの手法 [54] を応用して, SLM+eBM と SLM-C の分かち書き出力のうち，正しい分かち書き出力を学習する分類器 A を作成した．分類器 A の学習コーパスの素性には，ChaSen の形態素解析結果である見出し語の出現形，読み仮名，見出し語の基本形，品詞の全階層，活用形に加えて，各形態素のモーラの数，eBraille-M の分かち書き出力，SLM+eBM の分かち書き出力，SLM-C の分かち書き出力を用いた．正解のラベルは SLM+eBM 又は SLM-C とし，両者の出力がいずれも正しい形態素に対しては，SLM+eBM を正解とした.

次に，SLM+eBM と SLM-C の分かち書き出力をルールで選択する分類器 B を作成した．分類器に実装するルールは，前節の結果を基に作成した．具体的には，名詞や「〜以後」,「〜等」を含む名詞の連続に対しては SLM+eBM の出力を，複合名詞は SLM-C の出力を選択するように条件を設定した．更に，他に設定すべき条件を明らかにするため，これらの 2 種類の SLM が,「名詞」に対して出力した BI ラベルの誤りの数を精査した．その結果,「名詞-数」の誤りの数は，SLM+eBM の方が少なかった（表 7.6）．そ

こで,「名詞」の連続に「名詞-数」が含まれる場合に SLM+eBM の出力を選択する条件を追加した.

表 7.6. 「名詞」に対する SLM+eBM と SLM-C の出力誤りの数

品詞	SLM+eBM	SLM-C
名詞-一般	71	59
名詞-サ変接続	10	9
名詞-数	**4**	**14**
名詞-固有名詞-一般	1	0
名詞-固有名詞-人名-一般	0	0
名詞-固有名詞-人名-姓	0	0
名詞-固有名詞-人名-名	0	0
名詞-固有名詞-組織	0	0
名詞-固有名詞-地域-一般	3	2
名詞-固有名詞-地域-国	0	1
名詞-非自立-一般	1	1
名詞-非自立-形容動詞語幹	0	0
名詞-非自立-助動詞語幹	0	2
名詞-非自立-副詞可能	5	1
名詞-接続詞的	0	0

品詞	SLM+eBM	SLM-C
名詞-接尾-一般	4	8
名詞-接尾-形容動詞語幹	0	0
名詞-接尾-サ変接続	1	1
名詞-接尾-助数詞	1	1
名詞-接尾-助動詞語幹	1	1
名詞-接尾-人名	2	1
名詞-接尾-地域	0	0
名詞-接尾-特殊	0	0
名詞-接尾-副詞可能	0	2
名詞-代名詞-一般	1	1
名詞-特殊-助動詞語幹	0	0
名詞-ナイ形容詞語幹	0	0
名詞-副詞可能	1	1
名詞-形容動詞語幹	0	0

以下に分類器 B に実装したルールを示す.

・名詞又は複合名詞（「名詞」又は「接頭詞」のみを含む形態素の連続）の分かち書きは, 下記の条件に従う.

1. 名詞（連続する 2 形態素が「接頭詞」と「名詞」, 又は「名詞」と「名詞-接尾」の場合）は, SLM+eBM の出力を選択する.
2. 「名詞-数」を含む場合, 又は, 出現形に「以後」や「等」を含む場合は, SLM+eBM の出力を選択する.
3. 1 と 2 以外は, SLM-C の出力を選択する.

・名詞と複合名詞以外は, SLM+eBM の出力を選択する

以上の 2 種類の分類器（図 7.1）を用いた場合の分かち書き精度を算出し, SLM+eBM, SLM-C の分かち書き精度と比較した.

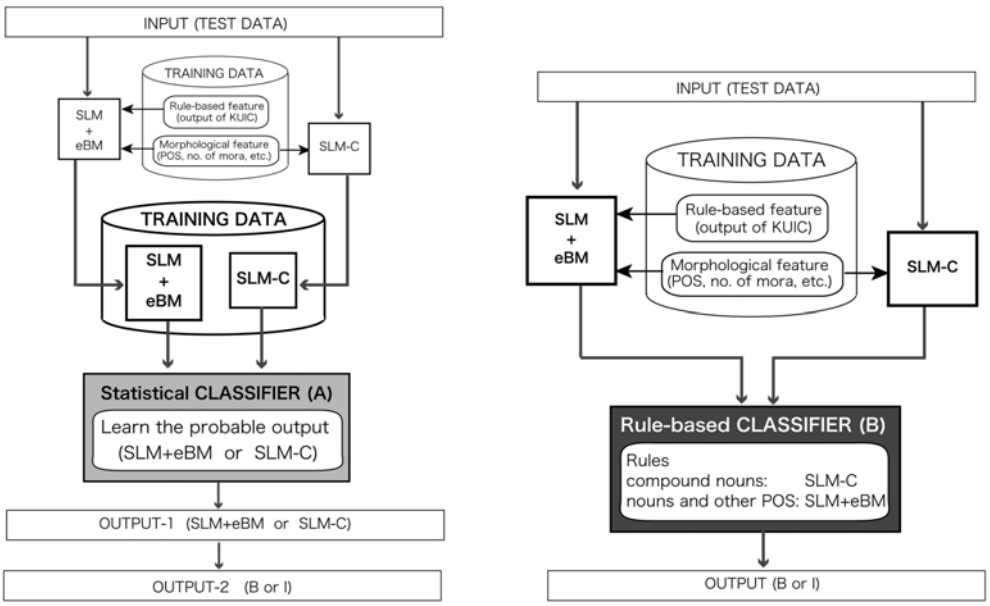


図 7.1. スタッキングを応用した分類器 A とルールベースの分類器 B

統計的学習モデルと分類器の分かち書き精度の比較

統計的学習を用いた分類器 A とルールベースの分類器 B の分かち書き精度を SLM+eBM, SLM-C の精度と比較した．その結果，医療文書 648 文と複合名詞のいずれの場合も，分類器 B が最も高い精度を示した（表 7.7，図 7.2，表 7.8，図 7.3）．

表 7.7.　医療文書 648 文の分かち書き精度

	Precision (%)	Recall (%)	F-measure (F_1)
SLM+eBM	94.57	94.49	94.53
SLM-C	93.49	93.23	93.34
Classifier A	94.54	94.49	94.51
Classifier B	**94.90**	**95.22**	**95.05**

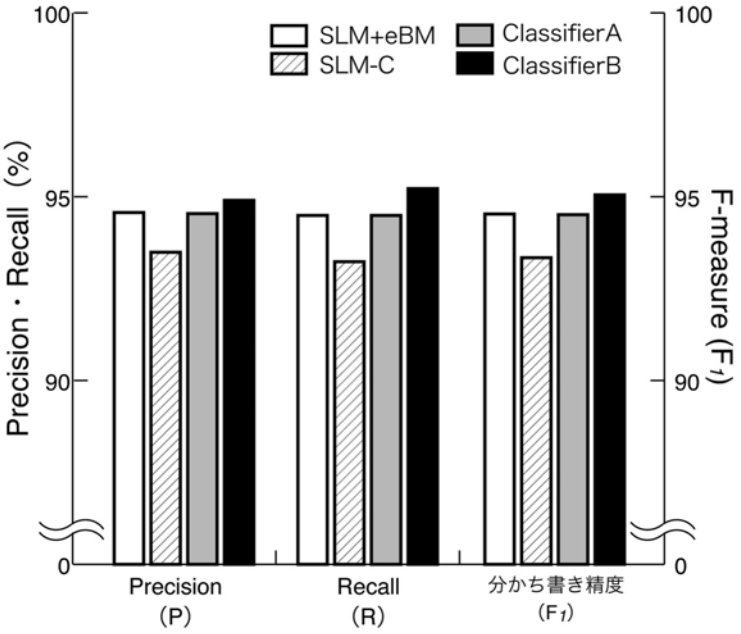


図 7.2.　医療文書 648 文の分かち書き精度

表 7.8.　複合名詞 750 個の分かち書き精度

	Precision (%)	Recall (%)	F-measure (F_1)
SLM+eBM	91.34	92.54	91.92
SLM-C	**92.51**	93.67	93.08
Classifier A	91.24	92.53	91.87
Classifier B	92.41	**94.84**	**93.58**

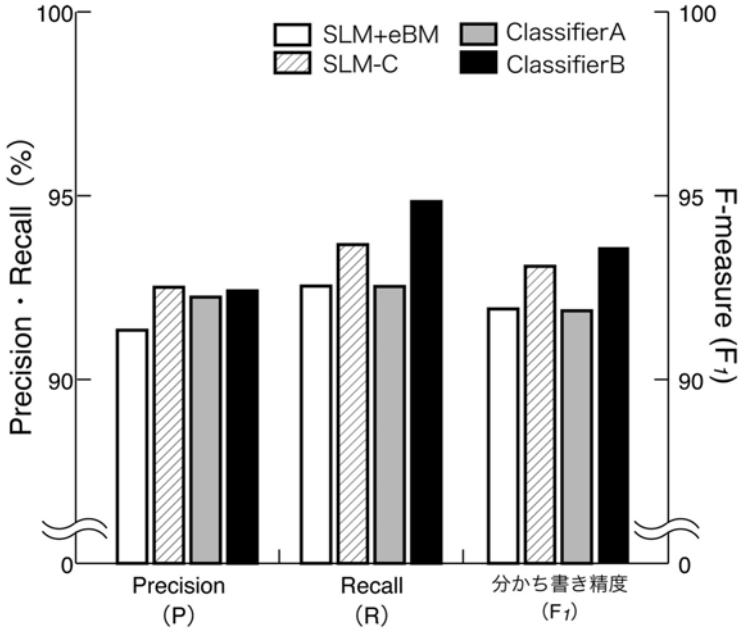


図 7.3.　複合名詞 750 個の分かち書き精度

分類器 A の適合率，再現率，分かち書き精度（F_1）は，SLM+eBM の各種精度とほぼ同じであった（表 7.7，表 7.8）ことから，分類器 A は SLM-C の正解のラベルを学習できていないことが示唆された．また，分類器 A の学習コーパスの全形態素数 8,336 tokens のうち，SLM-C を正解とする形態素数の割合が約 0.6%だったことから，学習コーパスの全形態素数における SLM+eBM と SLM-C の各々の正解ラベル数の割合の差が大きいことが，学習に影響を及ぼした可能性が示唆された．一方，分類器 B は，品詞を基準としたルールに従い，SLM+eBM 又は SLM-C の分かち書き出力から正しい出力を選択した．

SLM+eBM	SLM-C	分類器 B（正解）
皮膚|剥離	皮膚■剥離	皮膚■剥離
腱|移行術	腱■移行術	腱■移行術
TSH|受容体■異常症	TSH■受容体■異常症	TSH■受容体■異常症
チン■氏帯	チン氏帯	チン氏帯
小腸癆	小■腸癆	小腸癆
6 月■6 日	6 月|6 日	6 月■6 日
夕食■以後	夕食|以後	夕食■以後
角膜■混濁■等	角膜■混濁|等	角膜■混濁■等
各■項目	各|項目	各■項目

以上の結果から，点訳エンジン KUIC の表記規則で対応が不可能な分かち書きには，KUIC の表記規則を学習素性に組み込んだ統計的学習モデル SLM+eBM と，形態素の情報のみを学習素性とした SLM-C の 2 種類のモデルの出力をルールベースの分類器 B で選択する手法が有効であることが明らかになった．

7.5.4　分かち書き誤りの重篤度の比較

分かち書き誤りの重篤度については，先行研究や文献に体系的な分類は見当たらない．そこで，eBraille-M および SLM+eBM，SLM-C，分類器 B の 3 つを組み合わせた場合での分かち書き誤りを精査し，語句の読みや意味の理解の誤りの原因となる分かち書き誤りの種類を特定した．重篤度の高い誤りは，(1)文字体系の境界に関する誤りと(2)語の区切れに関する誤りであった．

(1) 文字体系の境界に関する誤り

例：「CT|にて」　（正解：「CT■にて」）

「に」は「l (L)」，「て」は「q (Q)」と同じ点字を使用するため，区切れがないと「CTLQ」と読む．

(2) 語の区切れに関する誤り

（品詞の判定誤りが原因で，語句の境界を誤った場合）

例：「PML で■す.」（正解：「PML■です. 」）

「す」を動詞「する」の「文語基本形」と誤って判定した．

次に，これらの誤りの数が eBraille-M 又は分類器を用いた場合の分かち書き誤りの数に占める割合を比較した．その結果，(1)文字体系の境界に関する誤りの数は，eBraille-M の分かち書き誤りの総合数の約 20%，分類器 B の場合はその誤りの総合数の約 15%だった．また，(2)語の区切れに関する誤りの数は，eBraille-M の分かち書き誤りの総合数の約 20%，分類器 B の場合はその誤りの総合数の約 13%だった．以上の結果から，分類器 B を用いると，重篤度の高い分かち書き誤りの割合が減少することが示された．

なお，複合名詞の分かち書き誤りの重篤度は，触読の速度や習熟度によって個人差があると予想される．例えば，「デンキノコギリ（電気鋸）」を触読する場合，点字の触読の速度が遅い人は「デンキノ」を「電気の」と読み誤る可能性があるため，「デンキ■ノコギリ」と区切った方が，早く意味を理解することが可能であると言われる．このことから，複合名詞の分かち書き誤りは比較の対象外とした．

7.6 まとめ

この章で我々は，機械学習の SVM のアルゴリズムを利用した統計的学習モデルを作成し，分かち書きの精度と分かち書きの誤りを解析した．統計的学習モデル（SLM-C）は，ルールベースの eBraille-M よりも高い分かち書き精度を示す一方で，eBraille-M の点訳エンジン KUIC の表記規則が正しく出力する分かち書きを誤って出力していた．このことから，eBraill-M と SLM-C の正しい出力を組み合わせる手法が分かち書き精度の向上に有効であることが示唆された．eBraille-M と SLM-C の組み合わせには，eBraille-M の分かち書き出力を SLM-C の学習素性に追加した SLM+eBM が有用であることが示された．そして，SLM+eBM は，KUIC の表記規則を学習して eBraille-M と同様の分かち書きを出力した．この結果は，ルールベース（KUIC の出力素性）と統計的学習の融合（SLM+eBM）が分かち書きに有効であることを示している．

一方，複合名詞のみの分かち書き精度の比較では，eBraille-M の出力を素性に用いた SLM+eBM よりも形態素解析結果とモーラの数のみを学習素性とした SLM-C の方が精度が高かった．そこで，2 種類の統計的学習モデル（SLM+eBM，SLM-C）に分類器を組み合わせる手法を導入した．統計的学習による分類器 A とルールベースの分類器 B を用いた手法を比較した結果，ルールベースの分類器 B が，医療文書 648 文と複合名詞のいずれの場合も最も高い分かち書き精度を示した．さらに，分類器 B は，SLM+eBM と SLM-C を単独で用いた場合よりも分かち書き精度が高かった．以上の結果から，２種類の統計的学習モデル（SLM+eBM，SLM-C）とルールベースの分類器（B）を組み合わせることが分かち書きに有効であることが明らかになった．そして，これらの統計的学習モデルと分類器を eBraille に組み込み分かち書きに適用することで，分かち書き精度と点訳精度の向上が実現可能であることが示された．

第8章

結　論

我々は，医療文書の点訳において点字翻訳の専門家と同等の高い点訳精度を実現する方法の解明を目的として，点訳精度を向上させる手法を示し，その有効性を解析した．そして，以下の３点を明らかにした．

まず第5章では，従来のeBraille0.81[11]の点訳誤りと日本点字表記法[2]を分析した．その結果，より細分化された品詞体系を採用しているChaSen 2.3.3[27]へ更新する必要があること，プログラムに実装する点字表記規則は，文字種，品詞，活用形，出現形，読み，音韻変化，モーラ（拍）の数を指標として作成し，これらの指標を２つの形態素に適用する必要があることを明らかにした．ChaSenを更新し，点字表記規則を点訳エンジンKUICに実装した結果，通常文書評価用コーパスでのeBrailleの点訳精度はversion 0.81の約3倍に達した．また，eBrailleと他の自動点字翻訳プログラムの点訳精度を比較したところ，eBrailleの点訳精度が有意に高かった．

次に，点訳精度の向上には，点訳対象の文書に適した辞書の語彙構成にすることが有効であることが明らかになった．第6章で辞書の語彙構成を変えた4種類の自動点字翻訳プログラム，eBraille，allBraille，eBraille-M，eBraille-TM，の点訳精度を通常文書評価用コーパスと医療文書評価用コーパスで評価した結果を比較したところ，いずれのコーパスの場合も，点訳精度の平均値はeBraille-Mが他の全てのプログラムよりも高かった．

allBrailleでは，語彙数が最も多い辞書を使用している．その辞書は，eBraille-Mと同様の医療用語を含んでいるが，allBrailleの分かち書き精度と点訳精度はeBraille-Mよりも有意に低く，医療用語の点訳に失敗していた．以上の結果から，辞書の語彙数と点訳精度の高さは比例しないこと，医療文書の点訳にはeBraille-Mの辞書の語彙構成が適していることが示唆された．

最後に，SVMを利用した統計的学習モデルが分かち書きの精度に及ぼす影響を解析した．複合名詞の分かち書きは，単語の意味の境界を指標として表記するよう規定されている [2]が，意味の境界の定義は記されていない．このような不明確な分かち書きの規則を，eBrailleの点訳エンジンKUICの表記規則としてプログラムに実装することは不可能である．そこで，第7章では，統計的手法の中でも汎化能力の高いSVMを利用したチャンカーYamCha [52]を用いて統計的学習モデルの評価実験を行った．実験の結果，複合名詞の分かち書きには，形態素解析結果とモーラの数を学習素性とした統計的学習モデル（SLM-C）が，複合名詞以外の分かち書きには，KUICの表記規則を学習素性に追加した統計的学習モデル（SLM+eBM）が有効であることが示された．また，これらの2種類の統計的学習モデルの出力をルールベースの分類器で選択する手法が分かち書きに有効であることを明らかにした．

今後の課題として，更なる精度の向上が挙げられる．先行研究には，点字翻訳で実用上必要とされる点訳精度についての議論は見当たらない．しかし，一つの指標として，毎日新聞社の点字毎日の場合[1]，校正後の各種精度，すなわち，分かち書き精度（F_J）のmean ± SD は 98.87 ± 0.68，読みの精度（%）は 99.13 ± 0.68，点訳精度（BTA）は 98.02 ± 1.12[2]である．今後，辞書に用語を追加し，大規模な学習コーパスで統計的学習モデルを作成して分類器に組み込むことで，各種の精度向上に取り組みたい．

---

1 点字毎日の記事 474 文（2002 年 4 月）．毎日新聞社によると，記事の点字文は，墨字文を自動点字翻訳プログラムで点訳した後，人手で誤りを訂正している．

2 点字毎日の記事の点字文は，2001 年版の日本点字表記法で追加された表記規則，動詞「する」の前を区切る分かち書きに未対応であったため，この規則への未対応の部分を誤りに含めずに精度を算出した．

謝辞

本研究の遂行と本論文の作成にあたり，懇切なるご指導と御校閲を賜りました松本裕治教授と神戸大学大学院医学研究科の高岡 裕特命准教授に謹んで御礼申し上げます. 特に，高岡博士からは，本研究遂行の機会と多くの有益な示唆やアイデアを頂きました．心より厚く御礼申し上げます.

審査委員として本論文に対して有益なコメントを下さいました，松本健一教授，新保仁准教授に厚く御礼申し上げます．特に新保准教授には，第7章の実験設定で有益な示唆を頂いたことに深謝いたします.

また，本研究遂行にあたりご助言を賜りました，神戸大学大学院医学研究科の大田美香博士に深く感謝いたします.

さらに，西南女学院大学の相良かおる博士からはComeJisyoを，六然社の寄金丈嗣氏からは東洋医学用語辞書を，奈良先端科学技術大学院大学の浅原正幸博士からは拡張辞書の作成に必要なツールを提供頂きました．御礼を申し上げます．また，神戸大学医学部附属病院の大島敏子看護部長兼副病院長，松浦正子副看護部長，花岡澄代副看護部長，高橋京子副看護部長，池上峰子看護師長，医療情報部長の前田英一特命教授から，多くの協力を頂いたことに深く感謝いたします.

加えて，国立神戸視力障害センターの厚生労働教官小田 剛氏，神戸大学大学院医学研究科の村井勇介氏，三浦研爾氏にも協力頂きました．これらの諸氏の協力にも感謝します.

参考文献

[1] Javier Jiménez, Jesús Olea, Jesús Torres, Inmaculata Alonso, Dirk Harder and Konstanze Fischer. Biography of Louis Braille and invention of the braille alphabet. *Survey of Ophthalmology*, 54(1): 142-149, 2009.

[2] 日本点字委員会. 日本点字表記法 2001 年版. 株式会社大活字, 東京, 2001.

[3] 大田美香, 小田 剛, 三浦研爾, 菅野亜紀, 高岡 裕. 点字自己学習用の『触読点字 e-learning』の開発. 電子情報通信学会信学技報 109(467), pages 1-4, 2010.

[4] Official Journal of the European Union. Directive 2004/27/EC of the european parliament and of the council of 31 March 2004 amending Directive 2001/83/EC on the Community code relating to medicinal products for human use. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:136:0034:0057:EN:PDF.

[5] European Commission. Guideline on the readability of the label and package leaflet of medicinal products for human use (Update according to Directive 2001/83/EC as amended by Directive 2004/27/EC), Revision 1, 12 January 2009. http://ec.europa.eu/enterprise/sectors/pharmaceuticals/files/eudralex/vol-2/c/2009_01_12_readability_guideline_final_en.pdf.

[6] European Commission. eTEN programme. http://ec.europa.eu/information_society/activities/eten/index_en.htm.

[7] News - Free braille translation service and software. *Journal of Visual Impairment and Blindness*, 10(8): 504-506, 2007.

[8] 個人情報の保護に関する法律 （平成十五年五月三十日法律第五十七号). http://law.e-gov.go.jp/htmldata/H15/H15HO057.html.

[9] Anne L. Corn and Robert S. Wall. Training and availability of braille transcribers in the United States. *Journal of Visual Impairment and Blindness*, 96(4): 223-232, 2002.

[10] Robert Wall Emerson, Anne L. Corn and Marry Ann Siller. Trends in Braille and large-print production in the United States: 2000-2004. *Journal of Visual Impairment and Blindness*, 100(3): 137-51, 2006.

[11] 五十嵐大和, 高岡裕. ChaSen を利用したインターネット点字翻訳サーバ -日本語点字翻訳サーバ eBraille の開発-, 第 18 回医療情報学連合大会論文集, pages 814-815, 1998.

[12] James Marshall Unger. Japanese braille. *Visible Language*, 18(3): 254-266, 1984.

[13] Mitsuji Kadota. Japanese braille tutorial. Oct, 1997. http://www.hi.sfc.keio.ac.jp/access/arc/NetBraille/etc/brttrl.html.

[14] Nikolai Sergeyevich Trubetzkoy. *Grundzüge der Phonologie (Principles of Phonology)*. University California Press, Berkeley, CA. 1958/69.

[15] 特定非営利活動法人 全国視覚障害者情報提供施設協会. サピエ（視覚障害者情報提供ネットワーク）. https://www.sapie.or.jp/.

[16] 高田和典, 江本倫基 , 脇田貴之, 山田佳裕, 池田尚志. 自動点字翻訳編集システム ibukiTenC －点字毎日との比較による精度評価－. 言語処理学会第 13 回年次大会論文集, pages 218-221, 2007.

[17] 兵藤安昭, 横平貫志, 早川哲史, 村上裕, 池田尚志. 誤り箇所指摘機能をもたせた点字翻訳編集システム IBUK-TEN. 電子情報通信学会論文誌, J84-D-1(7):1102-11, 2001.

[18] Satoshi Ono, Yoshinobu Hamada, Yoshitsugu Takagi, Seiichi Nishihara and Kazunori Mizuno. Interactive Japanese-to-Braille translation using case-based knowledge on the web. in Riichiro Mizoguchi and John Slaney ed., *PRICAI 2000 Topics in Artificial Intelligence*, pages 634-646, Springer-Verlag, Berlin, 2000.

[19] 鈴木恵美子, 小野智司, 平岡大樹, 狩野均, 西原清一. 知識ベースに基づく点字翻訳のための日本語文節区切り手法, 情報処理学会研究報告, 97(69):141-146, 1997.

[20] 高木喜次, 小野智司, 宮下和雄, 西原清一, “表装解析に基づく点字用日本語分かち書きへの事例ベースの適用,” 情報処理学会研究報告(自然言語処理研究会), 99-NL-129, 1999.

[21] Toshiyuki Gotoh, Reiko Minamikawa-Tachino and Naoyoshi Tamura. A web-based braille translation for digital music scores. In *Proceedings of the 10th international ACM SIGACCESS conference on computers and accessibility*, pages: 259-260, 2008.

[22] 勝沼貞幸. お点ちゃん. 2003. http://www17.plala.or.jp/otenchan/.

[23] 厨子直人. 点字自動翻訳システム（Système de transcription automatique en braille）. 2002. http://www.muzik.gr.jp/tenji/default.asp.

[24] 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

[25] 難病情報センター. http://www.nanbyou.or.jp/index.html.

[26] Cornelis Joost van Rijsbergan. Information Retrieval. Butterworths. London, 1979.

[27] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka and Masayuki Asahara. Morphological Analysis System ChaSen version 2.3.3 Manual. Technical report, Nara Institute of Science and Technology, 2003.

[28] 渡辺哲也, 山口俊光, 南谷和範, 大内進, 宮城愛美, 岩下恭士. 視覚障害者が触知可能な触地図作成システムの開発. 電子情報通信学会信学技報 108(488), pages 13-18, 2009.

[29] 北村美穂子, 松本裕治. 言語資源を活用した実用的な対訳表現抽出. 自然言語処理 13(1), pages 3-25, 2006.

[30] 浅原正幸, 松本裕治. ipadic version 2.7.0 ユーザーズマニュアル. 奈良先端科学技術大学院大学情報科学研究科松本研究室, 2003

[31] Real World Computing Partnership. RWC Text Database, 1995.

http://www.rwcp.or.jp/wswg/rwcdb/text/.

[32] 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, pages 938-939, 2008.

[33] 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, 43(3), pages 685-695, 2002.

[34] 松本裕治. 自然言語処理における意味研究. 第 24 回人工知能学会全国大会論文集, 2G1-OS3-3, 2010.

[35] 北研二, 中村哲, 永田昌明. コーパスに基づくアプローチ. 森北出版株式会社, 東京, 1996.

[36] Taku Kudo and Yuji Matsumoto, Chunking with Support Vector Machines. In *Proceeding of 2nd Meeting of the North American Chapter of the Association for Computational Linguistics*, pages192-199, 2001.

[37] 工藤拓, 松本裕治, "Support Vector Machine を用いた Chunk 同定," 自然言語処理, 9(5), pages 3-21, 2002.

[38] Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6): 1554-1563, 1966.

[39] Leonard E. Baum and John A. Eagon. An inequality with applications to statistical estimation for probabilistic function of a Markov process and to a model for ecology. *Bulletin of the American Mathematical Society*, 73: 360-363, 1967

[40] Leonard E. Baum and George R. Sell. Growth functions for transformations on manifolds. *Pacific Journal of Mathematics*, 27(2): 211-227, 1968.

[41] Leonard E. Baum, Ted Petrie, George Soules and Norman Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *The Annals of Mathematical Statistics*, 41(1): 164-171, 1970.

[42] Leonard E. Baum, An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes, *Inequalities*, 3:1-8, 1972.

[43] Edwin Thompson Jaynes. Information Theory and Statistical Mechanics. *Physical Review Series II*, 106 (4): 620–630, 1957.

[44] Edwin Thompson Jaynes. Information Theory and Statistical Mechanics II. *Physical Review Series II*, 108 (2): 171–190, 1957.

[45] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag Berlin, 1995.

[46] Julian Kupiec. Robust part-of-speech tagging using a hidden markov model. *Computer Speech & Language*, 6(3): 225-242, 1992.

[47] Lawrence Rabiner and Biing-Hwang Juang. Fundamentals of Speech Recognition. Prentice Hall Signal Processing Series. Prentice-Hall, Inc., Englewood Cliffs, NJ. 1993.

[48] Ronald Rosenfeld. Adaptive statistical language modeling: a maximum entropy approach. Ph.D Thesis, School of Computer Science, Carnegie Mellon University, CMU-CS-94-138. 1994.

[49] Adam L. Berger, Stephen A. Della Pietra, Vincent J. Della Pietra. A maximum entropy approach to natural language processing, *Computational Linguistics*, 22(1): 39-71, 1996

[50] Adwait Ratnaparkhi. A maximum entropy model for part-of-speech tagging. In *Proceeding of the Conference on Empirical Methods in Natural Language Processing 1996 (EMNLP 1996)*, pages 133–142, 1996.

[51] Eugene Charniak. A maximum-entropy-inspired parser. In *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference 2000 (NAACL 2000)*, pages 132–139, 2000.

[52] Taku Kudo. YamCha: Yet Another Multipurpose CHunk Annotator, http://chasen.org/~taku/software/Yam Cha/.

[53] 注連隆夫，内元清貴，土屋雅稔，高木俊宏，宇津呂武仁，佐藤理史，井佐原均．機械学習を用いた日本語複合辞のチャンキング，（社）情報処理学会研究報告，2005–NL-170, pages 135-142, 2005.

[54] David H. Wolpert, Stacked Generalization. *Neural Networks*, 5(2): 241-259, 1992.

研究業績

学術論文誌

1. 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*, 20(5), October 2010.

国際会議発表

1. Aki Sugano, Kenji Miura, Mika Ohta, Mineko Ikegami, Sumiyo Hanaoka, Eiichi Mæda, Masayuki Asahara, Yuji Matsumoto, Masako Matsuura, Masafumi Matsuo, Toshiko Ohshima, Yutaka Takaoka. Development of Japanese-into-Braille Translating Program for Medical Information "eBraille." *Asia Pacific Association for Medical Informatics 2009* (APAMI 2009). November, 2009 (Hiroshima, Japan)

著書

1. 菅野亜紀，テーラーメイド医療とバリアフリー．高岡 裕，久野慎一，大田美香，清野 進，高井義美（編），井村裕夫（監修），「実践ゲノムの最前線」第二部第六章, pages 289-290, 六然社, 2009.

技術報告論文

1. 菅野亜紀，寄金丈嗣，相良かおる，三浦研爾，大田美香，大島敏子，松本裕治，高岡 裕: 辞書の語意構成と点訳精度の解析．電子情報通信学会信学技報 110(53), 1-4, 2010
2. 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かち書き解析器の自動点訳での有効性の解析．電子情報通信学会信学技報 109(467), 5-8, 2010

3. 菅野亜紀，花岡澄代，相良かおる，浅原正幸，三浦研爾，大田美香，松本裕治，大島敏子，高岡 裕: 自動点訳サーバ eBraille を用いた病院内バリアフリー対応の試み. 電子情報通信学会信学技報 108(488), 19-24, 2009
4. 菅野亜紀，三浦研爾，浅原正幸，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 の医療文書点訳精度の向上に向けた IPADIC の最適化.情報処理学会研究報告 2008-NL-184, 55-60, 2008
5. 菅野亜紀，大田美香，三浦研爾，松浦正子，池上峰子，前田英一，大島敏子，松本裕治，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 9(5), 93-98, 2007
6. 菅野亜紀，高岡 裕，米田隆一，乾 健太郎，松本裕治 : 機械学習による点字表記の解析-点訳支援システム構築にむけて- 電子情報通信学会信学技報 102(254), 9-14, 2002

受賞

1. グッドデザインひょうご「ユニバーサルデザイン賞」（兵庫県），2010 年 3 月 2 日
   対象 : 自動点字翻訳プログラム「eBraille」

研究業績（本論文と直接関係しないもの）

学術論文誌

1. Yutaka Takaoka, Mika Ohta, Atsuko Takeuchi, Kenji Miura, Masafumi Matsuo, Toshiyuki Sakaeda, Aki Sugano, Hisahide Nishio: Ligand orientation governs conjugation capacity of UDP-glucuronosyltransferase 1A1. *J Biochemistry* doi:10.1093/jb/mvq048, 2010.
2. Mika Ohta, Aki Sugano, Shuji Goto, Surini Yusoff, Yushi Hirota, Kotaro Funakoshi, Kenji Miura, Eiichi Maeda, Nobuo Takaoka, Nobuko Sato, Hiroshi Ishizuka, Naoki Arizono, Hisahide Nishio, Yutaka Takaoka: Full-length sequence of mouse acupuncture-induced 1-L (*Aig1l*) gene including its transcriptional start site. *Evidence-based Complementary and Alternative Medicine* doi:10.1093/ecam/nep121, 2009.
3. Gunadi, Kenji Miura, Mika Ohta, Aki Sugano, Myeong Jin Lee, Yumi Sato, Akiko Matsunaga, Kazuhiro Hayashi, Tatsuya Horikawa, Kazunori Miki, Mari Wataya-Kaneda, Ichiro Katayama, Chikako Nishigori, Masafumi Matsuo, Yutaka Takaoka, Hisahide Nishio: Two novel mutations in the *ED1* gene in Japanese families with X-linked hypohidrotic ectodermal dysplasia. *Pediatric Research*, 65(4), 453-457, 2009.
4. Yutaka Takaoka, Mika Ohta, Akihiko Ito, Kunihiko Takamatsu, Aki Sugano, Kotaro Funakoshi, Nobuo Takaoka, Nobuko Sato, Hiroshi Yokozaki, Naoki Arizono, Shuji Goto, and Eiichi Maeda: Electroacupuncture Suppresses Myostatin Gene Expression: Cell Proliferative Reaction in Mouse Skeletal Muscle. *Physiol Genomics* 30(2), 102-110, 2007.

国際会議

1. Aki Sugano, Hitoshi Nagano, Kenji Miura, Mika Ohta, Yutaka Takaoka: Characteristics of logical structures in acupuncture and Kampo medicine (Japanese traditional medicine) analyzed by ontology construction. Society for Acupuncture Research 2010 International Conference. March 19-21, 2010 (North Carolina, USA)
2. Yutaka Takaoka, Mika Ohta, Sachiko Ikemune, Aki Sugano, Shuji Goto, Toshikazu Miyamoto, Masafumi Matsuo: Electroacupuncture suppress the Myostatin gene expression and prevent the disuse muscle atrophy: Study for the mice model of hindlimb suspension. Society for Acupuncture Research 2010 International Conference. March 19-21, 2010 (North

Carolina, USA)

3. Aki Sugano, Shuji Goto, Tsuyoshi Oda, Yusuke Murai, Mika Ohta, Masafumi Matsuo, Yutaka Takaoka: Development of Acupoints Database Search System on the Internet. Society for Acupuncture Research 2010 International Conference. March 19-21, 2010 (North Carolina, USA)
4. Mika Ohta, Aki Sugano, Shuji Goto, HisahideNishiio, Masafumi Matsuo, Yutaka Takaoka: Full-length sequence of mouse acupuncture-induced 1-L (Aig1l) gene including its transcriptional start site and its expression analysis. Society for Acupuncture Research 2010 International Conference. March 19-21, 2010 (North Carolina, USA)
5. Kenji Miura, Mika Ohta, Aki Sugano, Yutaka Takaoka: 3D-structure of acupuncture-induced 1-L (Aig1l) protein analyzed by homology modeling. Society for Acupuncture Research 2010 International Conference. March 19-21, 2010 (North Carolina, USA)
6. 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). November 22-24, 2009 (Hiroshima, Japan)
7. Yutaka Takaoka, Kenji Miura, Aki Sugano, Mika Ohta, Atsuko Takeuchi, Masafumi Matsuo, Hisahide Nishio: *in silico* estimation for the conjugation capacity of mutant UDP-glucuronosyltransferase 1A1. The 10th International Conference Bioinfo2009. November 4-6, 2009 (Busan, Korea)
8. Mika Ohta, Sachiko Ikemune, Aki Sugano, Kenji Miura, Shuji Goto, Hisahide Nishio, Toshikazu Miyamoto, Yutaka Takaoka: Prevention of muscle atrophy by electroacupuncture in a murine hindlimb suspension model. The 36th Congress of the International Union of Physiological Sciences (IUPS2009). July 27-August 1, 2009 (Kyoto, Japan)
9. Yutaka Takaoka, Kenji Miura, Hisahide Nishio, Atsuko Takeuchi, Aki Sugano, Mika Ohta: *in silico* estimation for the enzyme activity of mutant UDP-glucuronosyltransferase 1A1. CBI Annual Meeting 2008 International Symposium. Octber 22-24, 2008 (Tokyo, JAPAN)
10. Keiko Taki, Kenji Miura, Mika Ohta, Aki Sugano, Bing Wang, Tetsuo Nakajima, Tetsuya Ono, Yoshihiko Uehara, Yoichi Oghiso, Junji Magae, Mitsuru Nenoi, Yutaka Takaoka: New Bioinformatics program for the extraction of biological intelligence from transcriptome: in case of low-dose-rate radiation research. CBI Annual Meeting 2008 International

Symposium. Octber 22-24, 2008 (Tokyo, JAPAN)

11. Yutaka Takaoka, Mika Ohta, Kunihiko Takamatsu, Aki Sugano, Nobuo Takaoka: Identification and characterization of differentially expressed genes in electro-Acupuncture treated muscle. WFAS 2001 International Symposium on Acupuncture, December, 2001(Singapore)

技術報告論文

1. 大田美香, 小田 剛, 三浦研爾, 菅野亜紀, 高岡 裕: 点字自己学習用の『触読点字 e-learning』の開発. 電子情報通信学会信学技報 109(467), 1-4, 2010
2. 小田 剛, 菅野亜紀, 三浦研爾, 庄田浩基, 前田英一, 大田美香, 高岡 裕: 点字自己学習用 e-learning システムの開発. 電子情報通信学会信学技報 108(488), 25-29, 2009
3. 三浦研爾, 菅野亜紀, 庄田浩基, 小田 剛, 大島敏子, 大田美香, 高岡 裕: 点字自己学習用 e-learning の開発とその課題. 電子情報通信学会信学技報 108(470), 83-88, 2009
4. 高岡　裕, 大田美香, 菅野亜紀: クリニカルゲノムインフォマティクス : 新興分野人材養成に於ける教育法と今後の課題. 電子情報通信学会信学技報 108(354), 17-22, 2008

付録

A. eBraille の点訳エンジンに実装した，品詞に適用する点字表記規則

表 A.1. 名詞に適用する点字表記規則

No.	対象となる品詞 (m_i)	分かち書き規則		
		前の形態素 (m_{i-1})	他の条件	処理内容
4	全ての品詞	—	—	原則として前と後ろを区切る（デフォルト）
5	名詞−一般	名詞−一般	—	前を区切る
6	名詞−一般	名詞−一般	形態素m_iの読みが「カワリ」	読みを「ガワリ」へ変換する（連濁） 前を続ける
7	名詞−一般	名詞−非自立−助動詞語幹	—	前を区切る
8	名詞−一般	「名詞−一般」又は名詞−サ変接続」	形態素m_{i-1}が2文字以下	前を続ける
9	名詞−固有名詞−人名−一般	—	—	前を区切る（デフォルト）
10	名詞−固有名詞−人名−一般	名詞−固有名詞−一般	—	前を続ける
11	名詞−固有名詞−人名−姓	—	—	前を区切る（デフォルト）
12	名詞−固有名詞−組織	—	—	前を区切る（デフォルト）
13	名詞−固有名詞−地域−一般	名詞−一般	—	前を区切る
14	名詞−副詞可能	名詞−一般	—	前を区切る
15	名詞−サ変接続	名詞−一般	—	前を区切る
16	名詞−サ変接続	名詞−一般	形態素m_iの読みが「フソク」	読みを「ブソク」へ変換する（連濁） 前を続ける
17	名詞−サ変接続	名詞−非自立−助動詞語幹	—	前を区切る
18	名詞−サ変接続	名詞−一般」又は名詞−サ変接続」	前の形態素が2文字以下	前を続ける
19	名詞−接尾−形容動詞語幹	名詞−一般	—	前を続ける
20	名詞−接尾−形容動詞語幹	名詞−一般	形態素m_iの読みが「スキ」	読みを「ズキ」へ変換する（連濁） 前を続ける
21	名詞−数	—	—	前を区切る（デフォルト） 後ろを区切る（デフォルト）
22	名詞−数	名詞−数、接頭詞−数接続、記号−一般、記号−句点、記号−読点、記号−アルファベット、記号−括弧開のいずれか	—	前を続ける
23	名詞−数	名詞−数 or 接頭詞−数接続	—	原則として前を区切る
24	名詞−数	記号−一般, 記号−句点, 記号−読点, 記号−アルファベット or 記号−括弧開	—	前を続ける
25	名詞−数	—	形態素m_iの読みが「イチ」,「ニ」…「キュ」,「ゼロ」	読みを1, 2…9,0に変換する
26	名詞−数	—	形態素m_iの読みが「セン」,「マン」,「オク」	原則として後ろを区切る
27	名詞−数	「名詞−数」以外	形態素m_iの読みが「、」	前を続ける 後ろを区切る
28	名詞−非自立−一般	—	形態素m_iの読みが「ノ」又は「ン」	前を続ける
29	名詞−非自立−副詞可能	—	—	前を区切る（デフォルト）
30	名詞−非自立−助動詞語幹	—	—	前を続ける（デフォルト）
31	名詞−非自立−形容動詞語幹	—	—	前を続ける （デフォルト） 後ろを続ける（デフォルト）
32	名詞−特殊−助動詞語幹	—	—	前を続ける （デフォルト） 後ろを続ける（デフォルト）
33	名詞−接尾−一般	—	—	前を続ける（デフォルト）
34	名詞−接尾−一般	—	形態素m_iの読みが次のいずれか：「ゾー」「ゲン」「ボツ」「ジャク」「キョー」「ミマン」「ナシ」「トー」「テイド」「ユカリ」「クカン」「アツカイ」「ミスイ」「ツヅキ」「イエン」「サワギ」「イトー」「イセイ」「イホク」「ナカバ」「マガイ」「フキン」「マミレ」「アワセ」「フカ」「シリーズ」「イナイ」「ハツ」「ネガイ」「トドケ」	前を区切る
35	名詞−接尾−地域	—	—	前を続ける（デフォルト）
36	名詞−接尾−サ変接続	—	—	前を続ける（デフォルト）
37	名詞−接尾−助動詞語幹	—	—	前を続ける （デフォルト） 後ろを続ける（デフォルト）
38	名詞−接尾−形容動詞語幹	—	—	前を続ける（デフォルト）
39	名詞−接尾−副詞可能	—	—	前を続ける（デフォルト）
40	名詞−接尾−副詞可能	—	形態素m_iの読みが、「ゼンゴ」「イッパイ」「イゴ」「イライ」「アタリ」のいずれか	前を区切る
41	名詞−接尾−助数詞	—	—	前を続ける（デフォルト）
42	名詞−接尾−助数詞	—	形態素m_iの読みが「パーセント」	読みを「%」に変換する 前を区切る
43	名詞−接尾−特殊	—	—	前を続ける （デフォルト） 後ろを続ける（デフォルト）
44	名詞−接続詞的	—	—	前を区切る （デフォルト）
45	名詞−ナイ形容詞語幹	—	—	前を区切る（デフォルト）

表 A.2. 接頭詞，動詞，形容詞，形容動詞，接続詞に適用する点字表記規則

No.	対象となる品詞 (m_i)	分かち書き規則		
		前の形態素 (m_{i-1})	他の条件	処理内容
46	接頭詞-名詞接続	ー	ー	後ろを続ける（デフォルト）
47	接頭詞-名詞接続	名詞-一般	ー	前を区切る
48	接頭詞-名詞接続	ー	形態素m_iの読みが、「ゲン」「ゼン」「キュー」「シン」「カク」「モト」のいずれか	原則として後ろを区切る
49	接頭詞-動詞接続	ー	ー	後ろを続ける（デフォルト）
50	接頭詞-形容詞接続	ー	ー	後ろを続ける（デフォルト）
51	接頭詞-数接続	「記号-括弧開」以外の品詞	形態素m_iの読みが、「ダイ」	前を区切る 後ろを続ける
52	動詞-自立	「動詞-自立」又は「動詞-非自立」	形態素m_{i-1}の活用形が連用形	前を続ける
53	動詞-自立	名詞-一般	形態素m_iの読みが、「ナイ」又は「ナサイ」	前を続ける
54	動詞-自立	名詞-一般	形態素m_iの読みが「アリ」	前を区切る
55	動詞-自立	名詞-サ変接続	形態素m_iの読みが「ニ」	前を続ける
56	動詞-自立	名詞-サ変接続	形態素m_iの読みが「アリ」	前を区切る
57	動詞-自立	助詞-接続助詞	形態素m_iの読みが「マス」	前を続ける
58	動詞-自立	助動詞	形態素m_iの読みが「シ」	前を続ける
59	動詞-自立	記号-アルファベット	形態素m_iの読みが「ニ」	前を区切る
60	動詞-非自立	「動詞-自立」又は「動詞-非自立」	形態素m_{i-1}の活用形が連用形	前を続ける
61	動詞-非自立	形容詞-自立	ー	前を続ける
62	動詞-接尾	ー	ー	前を続ける
63	形容詞-自立	動詞-自立	形態素m_iの読みが「ナカレ」	前を続ける
64	形容詞-自立	名詞全て	形態素m_iの読みが「ナカッ」「ナキ」「ナク」「ナイ」「ナシ」「ナケレ」のいずれか	前を続ける
65	形容詞-非自立	ー	ー	前を続ける
66	形容詞-非自立	助詞-接続助詞	形態素m_iの読みが「イイ」「ヨイ」「ホシイ」のいずれか	前を区切る
67	形容詞-非自立	助動詞	形態素m_iの読みが「イイ」「ヨイ」のいずれか	前を区切る
68	形容詞-接尾	ー	ー	前を続ける
69	副詞-助詞接続可能	副詞-助詞接続可能	形態素m_iが2文字以下	前を続ける
70	接続詞	ー	ー	前を区切る（デフォルト）
71	接続詞	「記号-句点」、「記号-読点」、「記号-括弧開」のいずれか	ー	前を続ける

表 A.3. 助詞，助動詞，感動詞，記号に適用する点字表記規則

No.	対象となる品詞 (m_i)	分かち書き規則		
		前の形態素 (m_{i-1})	他の条件	処理内容
72	助詞-格助詞-一般	ー	ー	前を続ける（デフォルト）
73	助詞-格助詞-一般	ー	形態素m_iの読みが「ヘ」	読みを「エ」に変換する
74	助詞-格助詞-引用	ー	ー	前を続ける（デフォルト）
75	助詞-格助詞-連語	ー	ー	前を続ける（デフォルト）
76	助詞-格助詞-連語	ー	形態素m_iの読み1文字目が「ニ」「ト」「ヲ」のいずれか	「ニ」「ト」「ヲ」の後ろを区切る
77	助詞-接続助詞	ー	ー	前を続ける（デフォルト）
78	助詞-接続助詞	ー	形態素m_iの読みが「モノノ」又は「オヨビ」	前を区切る
79	助詞-係助詞	ー	ー	前を続ける（デフォルト）
80	助詞-係助詞	ー	形態素m_iの読みが「ハ」	読みを「ワ」に変換する
81	助詞-副助詞	ー	ー	前を続ける（デフォルト）
82	助詞-間投助詞	ー	ー	前を続ける（デフォルト）
83	助詞-並立助詞	ー	ー	前を続ける（デフォルト）
84	助詞-終助詞	ー	ー	前を続ける（デフォルト）
85	助詞-副助詞/並立助詞/終助詞	ー	ー	前を続ける（デフォルト）
86	助詞-連体化	ー	ー	前を続ける（デフォルト）
87	助詞-副詞化	ー	ー	前を続ける（デフォルト）
88	助詞-特殊	ー	ー	前を続ける（デフォルト） 後ろを続ける（デフォルト）
89	助動詞	ー	ー	前を続ける（デフォルト）
90	助動詞	ー	形態素m_iの読みが「アラ」「アリ」「アル」「アレ」「アッ」のいずれか	原則として前を区切る 原則として後ろを区切る
91	助動詞	助詞全体	形態素m_iの読み1文字目が「ナ」	前を区切る
92	助動詞	形容詞-自立	形態素m_iの読み1文字目が「ナ」	前を区切る
93	感動詞	助動詞	ー	前を続ける
94	記号-一般	ー	ー	前を続ける（デフォルト） 後ろを続ける（デフォルト）
95	記号-一般	ー	形態素m_iの出現形が「！」又は「？」	前を続ける 後ろに点字2マスの空白を挿入する
96	記号-一般	ー	形態素m_iの出現形が「・」	前を続ける 後ろを続ける
97	記号-一般	ー	形態素m_iの出現形が「%」形態素m_iの読みが「パーセント」	読みを「%」に変換する 前を続ける 後ろを区切る
98	記号-一般	「記号-アルファベット」以外	形態素m_iの出現形が「:」	後ろを区切る
99	記号-一般	記号-括弧開	形態素m_iの読みが「…」	前を続ける
100	記号-一般	ー	形態素m_iの読みが「…」	後ろを区切る
101	記号-句点	ー	ー	前を続ける（デフォルト） 後ろを区切り、2マス挿入（デフォルト）
102	記号-読点	ー	ー	前を続ける（デフォルト） 後ろを区切る（デフォルト）
103	記号-空白	ー	ー	前を続ける（デフォルト） 後ろを続ける（デフォルト）
104	記号-アルファベット	ー	ー	後ろを続ける（デフォルト）
105	記号-アルファベット	ー	形態素m_iの読みが出力されない	形態素m_iの出現形を読みとする
106	記号-アルファベット	名詞-一般	ー	前を区切る
107	記号-括弧開	ー	ー	前を区切る（デフォルト） 後ろを続ける（デフォルト）
108	記号-括弧開	ー	形態素m_iの読みが「(」	前を続ける
109	記号-括弧閉	ー	ー	前を続ける（デフォルト）
110	助詞、助動詞、名詞-一般、名詞-固有名詞-地域-一般、名詞-サ変接続	記号-アルファベット	ー	前を区切る
111	名詞-固有名詞-人名-姓、名詞-固有名詞-組織、名詞-非自立-副詞可能、名詞-ナイ形容詞語幹	記号-括弧開	ー	前を続ける