

201027028A

厚生労働科学研究費補助金
感覚器障害研究事業

多機能高精度自動点訳エンジンの開発

平成 22 年度 総括研究報告書

研究代表者 石川 准
平成 23 (2011) 年 5 月

201027028A

厚生労働科学研究費補助金

感覚器障害研究事業

多機能高精度自動点訳エンジンの開発

平成22年度 総括研究報告書

研究代表者 石川 准

平成23（2011）年 5月

目 次

I. 総括研究報告	
多機能高精度自動点訳エンジンの開発	----- 1
石川 准	

厚生労働科学研究費補助金（感覚器障害研究事業）

総括研究報告書

多機能高精度自動点訳エンジンの開発

研究代表者 石川 准 静岡県立大学 国際関係学部 教授

研究要旨

本研究は、近年の以下のような新しいニーズの高まりに対応した多機能高精度自動点訳エンジンの開発を目的とする。

- ・各種ドキュメントが有する構造情報、レイアウト情報、テキスト情報等を生かした高精度自動点訳の実現
- ・大量コーパスを用いる今日的形態素解析エンジンの自動点訳への応用による自動点訳の変換精度の向上
- ・点字コーパスによる点訳分かち書き誤りの発見と修正

研究分担者

宮本 修（筑波技術大学 障害者高等教育研究支援センター 特任助教）

A. 研究目的

高度化するニーズに答える多機能高精度自動点訳システムを開発する。

既存の自動点訳エンジンの形態素解析アルゴリズムの性能を解析する。

大量のコーパスデータを用い、自動学習機能を有する今日的な形態素解析手法を解析し、それを応用して、確率論的方法論に基づく新しい自動点訳エンジンを開発する。

B. 研究方法

22年度は以下の研究を行った。

1. 新しい形態素解析エンジンの利用可能性についての研究

読みの誤り、分かち書きの誤り等の点訳誤りを項目別に自動的に算出できるツールを開発し、市販・公開されている点訳ソフトウェアの性能を客観的に評価する。

研究代表者がこれまで開発してきた

2. 読みくだしの精度向上の研究

複数の読み候補の中からどの候補を選ぶべきかという問題は、前後の品詞や用語などの情報を元にはっきりと判断できる場合もある。しかし、長年の研究にも関わらずまだ十分に確立していない知識フレームワークなどの人口

知能の研究領域に及ぶ技術を必要とするものもある。

現在の自動点訳システムの限界を超えている複数候補の選択に関しては、複数候補の存在箇所を利用者に示し、人手で選択候補の中から適切な候補を選択し直すという、機械と人手による共同作業を効率よく作業できるシステムが現実的である。その際に点字コーパスの情報を利用することで、正しい読みの可能性の高いものから順に候補リストを提示する手法を開発する。

専門分野ごとにコーパスを整備し、ドキュメント種別判別の精度を高め、各専門分野特有の読み方に対応し点訳精度を向上させる。

3. 点訳分かち書き誤りの発見の研究

全国視覚障害者情報提供施設協会が運営する視覚障害者情報総合ネットワークサピエの点字図書データを用いて大量点字コーパスを作成する。

点字コーパスを用いた分かち書き誤り検出手法を開発する。

(倫理面への配慮)

研究の過程で知り得た個人情報の守秘義務を遵守した。

C. 研究結果

点字コーパスによる点訳分かち書き誤り検出技術を開発し、その性能を評価し良好な結果を得た。実験によれば、点訳エンジンの特性にもよるが、23%から46%の分かち書き誤りを修正することができた。これは本研究の最大の成果である。

一方、最新の形態素解析エンジンを

改良して試作した自動点訳エンジンは、現段階では既存の点訳エンジンの変換精度に達しなかった。高い変換精度を達成するには、形態素解析の性能に加えて、コーパスまたは辞書の性能、例外処理など、多くの労力を要する。20万語の医学等専門辞書を整備し、ドキュメントの内容ごとに専門辞書を切り替えて用いることで点訳精度が向上することを確認している。

D. 考察

今後点訳コーパスによる点字分かち書き修正機能を市販の自動点訳エンジン等を実装していく。また自動点訳を利用しない点訳者に対しても、分かち書き誤りチェッカーを開発して提供すれば、点訳校正を効果的に支援できる。

さらに点字コーパスは読み下しの誤りを減らすためにも利用できる。点字コーパスにまったく存在しない読みと多く存在する読みでは後者を選択する、あるいは複数候補の提示順を点字コーパスの情報をもとに決定するなどの応用が考えられる。複数候補提示インタフェースを既存の点訳システムに組み込む実装はすでに行っており、現在候補リストを点字コーパスに基づき提示する実装に着手している。

E. 結論

点訳コーパスによる分かち書き誤り判別技術には新規性があり、学術的価値があると考えられる。

一方読みの精度向上には、辞書の整備と形態素解析アルゴリズムの改良を地道に継続する必要がある、分かち書

き誤りの修正のような顕著な効果を短期間で実現する方法はいまのところ見当たらない。ただし、点字コーパスは読みの確からしさの評価にも応用できると期待できるので、今後既存点訳システムに実装してその有効性を検証する計画である。

F. 健康危険情報

該当しない。

G. 研究発表

1.論文発表

宮本 修，コーパスを使った点訳の分かち書き誤り修正システム，情報処理学会論文誌 投稿中

(口頭発表)

石川 准，宮本 修，多機能高精度自動点訳エンジンの開発，平成 22 年度感覚器障害研究・研究成果発表会，KKR ホテル東京

H. 知的財産権の出願・登録状況

なし
