

## 論文内容の要旨

博士論文題目 Automated Grammatical Error Correction Using Statistical Machine Translation Techniques with Revision Log of Language Learning SNS  
(語学学習 SNS の添削ログを用いた統計的機械翻訳技術による文法誤り訂正)

氏名 水本智也

### (論文内容の要旨)

文法誤り訂正や誤り検出といった自然言語処理による言語学習者支援の研究が盛んに行なわれるようになってきた。しかしながら、自然言語処理による学習者支援を行なう上で重要である大規模な学習者コーパスがないといった問題がある。本論文では、まず、語学学習 SNS の添削ログから大規模な第2言語学習者コーパスの作成を行なった。このコーパスは簡単に大規模に手に入れることができ、さまざまなトピックやスタイルをカバーしており、学習者と教師両方にとって重要な知識源となる。

また、これまでの第2言語学習者の作文誤り訂正では誤りを限定して行なうことが多かったが、実際の学習者の誤りには様々な誤りが含まれている。そこで、本論文では統計的機械翻訳の手法を誤り訂正に応用して、全ての誤りを対象に日本語/英語の文法誤り訂正を行なった。日本語学習者の文には誤りやひらがなが多く含まれているため、単語分割に失敗してしまうという問題もある。本論文では、単語分割の問題を解決するために文字単位を用いた手法を提案した。文字単位の手法を用いることで、誤り訂正の精度が向上することを確認した。

英語に対しても、日本語同様に統計的機械翻訳の手法を用いて文法誤り訂正を行ない、本手法がどの誤りタイプに有効であるかを調査した。また、複単語表現を考慮した文法誤り訂正、品詞や構文情報を考慮したりランキングが英語文法誤り訂正に有効であることを示した。

氏名	水本智也
----	------

(論文審査結果の要旨)

平成27年4月27日に開催した公聴会の結果を参考に平成27年5月28日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動が続けていくための十分な素養を備えていることを示すものと認める。

水本智也は、本博士論文において、語学学習SNSから大規模な文法誤り事例と誤り修正文の対応データを自動的に抽出できることを示し、こうして得られた学習者データに統計的機械翻訳手法を適用する様々な手法を提案した。本学位論文の貢献は次のようにまとめることができる。

1. 言語学習者の作文に見られる文法誤りを母語話者が修正する相互添削サイトLang-8のデータから、学習者の文法誤りが存在する文とその修正結果の対応を取った大規模なコーパスを整理し、公開した。
2. 文法誤りを含む日本語学習者の文が通常の状態素解析等で正しく解析できない場合が多いことから、学習者の文を単語単位だけではなく文字単位で扱い、修正後の文を単語単位で扱う方法を提案し、その有効性や特徴について確認し、学習者の母語による影響などを明らかにした。
3. 学習者コーパスのサイズと文法誤り訂正の精度の関係について調査し、データサイズを増すことで精度が向上する誤りのタイプとそうでないタイプがあることを確認し、文法誤りタイプに応じた素性の設計の重要性を明らかにした。
4. 複単語表現を考慮する文法誤り訂正法を提案し、その効果を検証した。
5. 文中の広い前後文脈を見なければ訂正が困難な文法誤りの存在に対応するため、誤り訂正の複数の候補を生成した上でそれらをリランキングする手法を提案し、誤り訂正の精度が向上することを示した。

機械翻訳技術を利用した文法誤り訂正手法を提案した本研究は、独創性が高く、また、大規模な学習者データを提供するという視点からも自然言語処理の分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。