論文内容の要旨

博士論文題目　Generating and Exploiting Language Resources for Indonesian Preposition Error Correction
(インドネシア語前置詞誤り訂正のための言語リソースの生成と利用)

氏名　Budi Irmawati

(論文内容の要旨)

文法誤りの自動検出・訂正システムは，第二言語学習者にとって貴重である．学習者が書いた文の誤り箇所と誤りタイプを教えてくれるシステムを開発するには，大規模な学習者データとそれにアノテートされた誤り情報が必要となる．誤りのアノテーションとは別に，様々な言語情報，例えば，形態素，構文，意味などが学習のための素性として必要である．言語資源が豊かでない言語については，そのようなデータを得ることが難しい．さらに，学習者のデータは紙媒体のままであることが多く，それらを利用するには機械可読になるように電子化する必要があるが，それには高いコストが発生する．したがって，言語資源が豊かでない言語，例えばインドネシア語に対して，大規模なアノテーション済みデータを自動的に得る方法を探ることは挑戦的な問題である．

本論文では，インドネシア語の学習者の誤り訂正コーパスの構築と前置詞の誤り訂正に関する研究を報告する．コーパス構築については，言語学習者の相互添削サイトLang-8上のインドネシア語の作文とその訂正文のアラインメントを行うことで誤り訂正済みのコーパスを構築した．誤り訂正の自動化に向けて，対象とするインドネシア語の特性を考え，依存構造に基づくアノテーション法の提案を行った．インドネシア語の依存構造アノテーション法を提案するとともに，そのアノテーションを施したデータを作成し，機械学習に基づく依存構造解析器を構成した．構築されたコーパスは小規模ながら，前置詞の誤り訂正のモデルを学習する目的では，十分な性能を達成できることを示した．

誤り訂正タスクにおいては，母語話者による作文に人工的に誤りを挿入することで人工的な学習データを構築することが可能であるが，我々が構築した真の誤り訂正コーパスを利用することにより，大規模な人工データよりも誤り訂正能力が高い学習器を構成することができることを示した．

次に，データ量の不足を補うため，有用な人工的学習データを構築するための新しい２つの手法を提案した．いずれも言語には依存しない手法である．一つ目は，学習者データから収集した前置詞誤りの候補集合に基づいて人工的な誤りを生成し，さらに開発セットを利用して有効でない事例を削除する方法である．二つ目は，母語話者が書いた文の前置詞を，それに近い文に現れる別の前置詞で置き換えることによって誤りデータを生成する方法である．これらの人工的な学習データを用いて機械学習による前置詞の誤り訂正実験を行うことにより，提案する人工データ構築手法が，少ない学習データでも精度の高い誤り訂正能力をもつことを示した．

氏 名	Budi Irmawati

（論文審査結果の要旨）

平成28年8月2日に開催した公聴会の結果を参考に平成29年5月31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，研究活動を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．

Budi Irmawatiは，本博士論文において，インドネシア語学習者の前置詞誤りの自動訂正を目指し，言語学習者の誤り訂正コーパスの構築と，機械学習に効果的な人工的な学習データの構築方法の提案を行った．本論文の貢献は以下のようにまとめることができる．

1. 言語学習者の相互添削サイトLang-8のインドネシア語のデータの学習者の作文とその訂正文の自動アラインメントを行うことにより，機械学習に利用可能な学習者誤り訂正コーパスを構築し，一般公開した．

2. インドネシア語の依存構造解析済みコーパスを構築し，機械学習によってインドネシア語の依存構造解析システムを開発し，それを利用することによって前置詞誤りに有効な素性抽出が可能であることを示した．

3. 母語話者のデータにランダムに誤りを発生させることによって人工的な誤りデータを構成することが可能であるが，単純な人工誤りデータを用いた学習システムの能力に限界があることを示した．

4. 人工的な誤りデータを構成するための新しい２つの手法を提案した．特に，学習者の誤り傾向に基づく人工誤りの生成と，生成されたデータの有効性を客観的に確認する手法によって得られるデータが少量であっても機械学習に有効に働くことを実験により示した．

言語学習者の作文の前置詞誤りの自動訂正のための学習データ構築法法を提案した本研究は，言語に依存しない手法という意味で汎用性があり，かつ，実用的である．したがって，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．