論文内容の要旨

博士論文題目　統計的機械学習を用いた日本語歴史コーパス構築時の表記整理作業の自動化

氏名　　岡照晃

(論文内容の要旨)

近年，コーパスを利用した日本語研究が増えつつある．しかしながら日本語研究の大きな位置を占めるのは日本語の歴史的研究であり，そこで扱うような日本語の歴史的資料（古い時代の文献資料）はコーパスとしての整備が現代語ほど進んでいない．歴史コーパスの整備が進まない原因の一つとして，コーパス整備時の表記整理にかかるコストが高いことが挙げられる．表記整理作業は専門家にしか行えず，作業人員を大量に確保することが難しい．またその反面，作業対象が膨大であるため，作業完了までには大変な時間を要する．

そこで本研究では，歴史的資料の表記整理作業を自動化することを最終的な目的とする．これにより，誰でも簡単に低コストかつ大規模な表記整理作業が可能になる．本論文では，まず第1段階として，統計的機械学習を用い，濁点付与作業の自動化に取り組んだ．その後，濁点付与の自動化で得られた知見を基に，その他の表記整理項目も合わせた自動化に取り組んだ．

本論文における貢献は以下の通りである．

1. 表記整理の作業項目の一つである濁点付与の自動化に取り組み，識別学習を用いて近代文語論説文に対しF1値で96を超える実用的精度での自動濁点付与手法を開発した．

2. 1で開発した手法を実装した自動濁点付与アプリケーションを作成・公開した．

3. 表記整理と形態素解析を同時に処理する手法を開発した．この手法により濁点付与に限らず，すべての表記整理項目を一斉に自動化できるようになった．また自動濁点付与の性能向上，その他の表記整理項目でも94～96%の適合率を実現した．

以上により，これまでに多大な時間と労力を要していた表記整理のコスト削減が可能になった．

氏　名	岡照晃

（論文審査結果の要旨）

平成27年1月23日に開催した公聴会の結果を参考に平成27年2月18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．

岡照晃は，本博士論文において，日本語の歴史資料の表記整理を行う手法を提案し，実装した解析システムの性能評価を行った．

1. 歴史的資料の表記整理として一番の問題である濁点付与を識別学習に基づいて行う方法を提案し，実用的な精度で自動濁点付与が可能であることを示した．

2. 中古和文や近代日本語文に対して濁点自動付与を行うアプリケーションを開発し，インターネット上で公開した．

3. 歴史的資料の表記整理と形態素解析を同時の行う手法を開発した．これにより，表記整理に必要なほとんどの処理を高い精度で達成することができた．

日本語歴的資料の表記整理を自動的に行う手法を提案した本研究は，独創性が高く，実用的であり，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．