主　論　文　要　旨

報告番号	甲 乙 第　　号	氏　名	佐 野　智 久

主 論 文 題 目：

表層情報に基づく固有表現の属性推定に関する研究

（内容の要旨）

本研究は，自然言語処理の課題である未知語処理の積極的な解決を目的とし，固有表現の属性推定手法の提案およびその有効性の評価を行うものである．

固有名詞等の外国語句を含む文書は数多く存在するが，単一言語を前提として処理されるため文書中に埋め込まれた外国語句は未知語として扱われてきた．外国語句全てを辞書に登録することは現実的ではなく，辞書ベースの手法には限界がある．文書に埋め込まれた外国語句は固有名詞の場合が多く，言語推定よりもエリア推定が必要とされる．エリア推定とは，ある固有表現が所属するエリアを推定する処理である．ここでのエリアとは，国や地域などを指す．従来のエリア推定は，辞書引きによる検索処理と文脈情報を用いた曖昧性解消処理の組合せであり，辞書やヒューリスティクスへの依存度が高く，頑健性や汎用性の面で問題がある．本研究は，固有表現を構成する文字の並びに着目し，その表層的な情報のみを用いることで，文書の言語や文脈に依存しない，頑健性と汎用性の高いエリア推定の手法を提案するものである．

本研究では，第一に，地名の表層的な情報を用いたエリア推定の基本的な手法を提案した．この手法では，10 のエリアの 100,000 の地名を対象とした実験で，94.05%の再現率とＦ値 0.74 を実現し，表層的な情報の有用性を確認した．第二に，効果的な表層情報の利用を目的として，言語的な特徴を有しつつ十分な量の情報を抽出できるブロックと呼ぶ新たな単位を導入し，低い適合率を改善することに成功した．第三に，類似した地名を有するエリアが複数存在するとの地名特有の問題に取り組み，類似エリアを含むエリア推定においても再現率を下げることなく適合率を向上させることに成功した．最終的に，これらの手法を効果的に組み合せることで，辞書情報や文脈情報に依存しない頑健性とエリア数の変化や類似エリアの有無に対応できる汎用性を兼ね備えたエリア推定手法を実現した．類似エリアを含む 10 のエリアでの実験で適合率 88.30%，再現率 94.03%，Ｆ値 0.91 を達成し，エリア数を 20 に増やした場合にもＦ値0.82 を得た．語句の短さや複数エリアでの出現等の地名という対象の特殊性を考慮すると，この結果は，辞書やヒューリスティクスに頼らない手法としては十分な結果と考えられると同時に，有効性も十分に有するものである．

本研究は，外国語句からの有用な情報の抽出を可能にするものであり，属性情報が十分でない固有表現から自動的に属性を引き出すことで，機械翻訳や情報抽出等への適用や，自然言語理解への応用などが期待できる．