

氏名(本籍)	<sup>すず</sup> <sup>き</sup> <sup>えみこ</sup> 鈴木 恵美子 (茨城県)
学位の種類	博士(工学)
学位記番号	博乙第1552号
学位授与年月日	平成11年7月23日
学位授与の要件	学位規則第4条第2項該当
審査研究科	工学研究科
学位論文題目	日本語処理における頑健な文節区切り手法とその応用
主査	筑波大学教授 工学博士 大田 友一
副査	筑波大学教授 工学博士 西原 清一
副査	筑波大学教授 工学博士 宮本 定明
副査	筑波大学教授 工学博士 鬼沢 武久
副査	筑波大学教授 工学博士 池田 克夫

### 論文の内容の要旨

単語間が空白で区切られている欧米諸言語と異なり、日本語を計算機によって処理する場合には、単語間を区切る処理が必要となる。本論文は、この単語間を区切る手法、すなわち、文節区切り手法とその応用に関するものである。特に、膨大な辞書を用意して形態素解析を行う通常的手法と異なり、辞書に含まれていない未知語に影響されにくい、言語の統計情報に基づく手法を扱っていることに特徴がある。

論文は7章から構成されている。1章の序論に引き続き、2章では本論文の主張である簡便な日本語処理手法の必要性について論じている。

3章では、辞書や文法を用いずに、漢字とその後に続くひらがなの文字組に関する統計情報によって文節区切りを行う手法を述べている。JICST文献抄録および新聞1ヶ月分のデータから文字組の出現頻度を求めた結果を用いて、2～3文字の組み合わせを考慮することによって、精度のよい文節区切りが可能であることが示されている。

4章では、3章で提案された文節区切り手法の応用として、ワードプロセッサによって作成された技術文書に出現しやすい文書構造上の誤りを自動検出し、ユーザによる校正を支援するシステムについて述べている。文書を構造化しておくことにより、単語・文節・文・段落といった様々なレベルでの処理を可能としている。

5章では、従来の形態素解析で問題となっている解析結果の曖昧性を減らすために、統計的な情報を用いて形態素の多義性を解消する手法を提案している。

6章では、点字翻訳における自動分かち書きする問題において、既存の辞書から自動的に構築した見出し語テーブルを用いて、字種の情報と特定の助詞のみにして着目して分かち書きする手法を提案し、実験システムを構築している。

7章は、結論である。

### 審査の結果の要旨

日本語処理の最も基本的な問題である形態素解析に用いる辞書の構築と維持には、膨大な人的労力と時間を必

要とする。本論文は、大量の日本語文書を統計処理することによって自動的に得られる情報に基づいて、日本語の文節区切りを行う手法を提案し、様々な観点からのプロトタイプシステムの構築によってその有効性を検証している。論文における議論の展開が、やや、表層に流れすぎる感は否めないが、実際に役に立つ手法を着実に積み重ねていく姿勢は、工学の研究として評価できる。

よって、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。