

氏 名	あま の しん や 天 野 真 家
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	論 情 博 第 41 号
学位授与の日付	平 成 15 年 3 月 24 日
学位授与の要件	学 位 規 則 第 4 条 第 2 項 該 当
学位論文題目	Studies on Natural Language Processing for Kana-to-Kanji Conversion and Machine Translation (仮名漢字変換及び機械翻訳のための自然言語処理の研究)
論文調査委員	(主 査) 教 授 奥 乃 博 教 授 松 山 隆 司 教 授 田 中 克 己

論 文 内 容 の 要 旨

本論文は、仮名漢字変換と機械翻訳における自然言語処理システムの大規模化、実用化に関する研究の成果をまとめたものであり、5章から構成されている。

第1章は序論であり、本研究の背景、目的について概観している。本研究の主たる問題の所在は、実用的な自然言語システムを構築する上で、「理論と実践との間のギャップの解消」および「組み合わせ的爆発の回避」にあることを指摘している。

第2章では、仮名漢字変換における理論と実践とのギャップを解消するために、二層型仮名漢字変換法を提案している。従来の仮名漢字変換研究では、主として、形態素解析と同音異義語の解決のための意味解析の理論的側面に焦点が当てられており、数値データや複合語などの辞書には含まれないデータの処理の重要性は認識されていなかった。それに対して、二層型仮名漢字変換法では、まず入力データを辞書に含まれるものと含まれないものに分解し、次に、辞書に含まれるデータは理論に従った仮名漢字変換規則を適用し、それ以外は特別のルールで対応している。また、同音異義語の候補選択では一度選択された語をキャッシュすることにより、利用者の利便性を高めている。さらに、動的な意味解析を可能にするために、単語をノードとし、単語間の連想の強さをアークの重みとする連想ネットワークでの解決を論じている。ノードの興奮度のスペクトラムが文脈を意味すると解釈できるようなニューラルネットワークを構成し、興奮度が最も大きいノードの単語をその文脈における正しい同音異義語であると規定する方法を考案し、動的同音異義語選択の問題を解決している。これらの成果は、日本初の日本語ワードプロセッサ JW-10 用の仮名漢字変換として実用化されている。

第3章では、機械翻訳の第一段階の処理である統語解析における組み合わせ的爆発を回避するために、語彙遷移網文法を提案している。統語解析では、統語的な曖昧性のために解の探索空間が爆発的に増大し、その結果、主記憶と処理時間を圧迫し、統語解析に失敗するという現象がよく生じる。このような失敗は翻訳の失敗につながるので、頑健な統語解析が機械翻訳のために不可欠である。従来、この組み合わせ的爆発を避けるために、統語規則の中に意味規則を導入し、統語的曖昧性を解決しつつ解析を進める方法が一般的であった。しかし、この方法は、統語規則と意味規則を同時に一つの規則の中に記述しなければならないので、規則全体の見通しが悪くなり、大規模システムの開発には不向きであった。それに対して、語彙遷移網文法では、統語規則と文法規則を分離し、前者で構築された構文木の持つ意味的曖昧性を後方で解消している。すなわち、意味規則を語彙規則として、辞書中の各単語に木-木変換規則の形態で添付し、統語解析の結果の構文木に対して、組み合わせ的爆発を起さない木-木変換により意味解析を行うことにより、意味的曖昧性を解決している。これらの成果は、世界初のエンジニアリング・ワークステーション型機械翻訳システムとして実用化されている。

第4章では、機械翻訳システムを実社会に適用する上で不可欠な機能について述べている。すなわち、前章で論じられた機械翻訳システム用いて、日英双方向の機械通訳機を構成し、ジュネーブ・川崎間を通信衛星を介して結び、二ヶ国語会話の実験を行っている。そのときの会話記録から言語現象等を分析し、翻訳誤りが会話をメタ会話へと引き込んでしまうこと、あるいは、長距離通信における時間遅れが会話のタイミングをずらせてお互いに違和感を感じさせること、などの知見

を得ている。これらの知見を基に、音声通訳機に必要な機能を論じ、翻訳誤りの提示方法を改良することにより、メタ会話への移行が減少し、本来の会話に集中できるなどの成果を得ている。本論文で報告されているような一般人を被験者に使った8日間にも及ぶ大規模な二ヶ国語会話実験は過去にも現在にも例がない。現在、ブロードバンド・インターネット、IMT-2000規格による携帯電話のような海外との通信を容易に可能にする環境が構築され普及しつつあり、そのような環境に機械通訳機を使用した場合に必要な機能要件について考察を行っている。

第5章では結論を述べている。

論文審査の結果の要旨

本論文は、仮名漢字変換と機械翻訳における自然言語処理システムの大規模化、実用化に関する研究の成果をまとめたものであり、得られた主な成果は次のとおりである。

1. 仮名漢字変換において、二層型仮名漢字変換法を考案し、辞書に含まれる語彙とそうでないものに分解し、次にそれぞれに対応した処理を行うことで、多種多様な入力に対する仮名漢字変換法を確立した。また、同音異義語の候補選択では一度選択された語のキャッシュ化や文脈に基づいた選択を可能にするニューラルネットワークの構築を通じて、利用者の利便性を高めている。これらの成果は、日本初の日本語ワードプロセッサとして実用化された。
2. 機械翻訳において、語彙遷移網文法を提案し、統語解析で構築された構文木の意味的曖昧性を意味規則により解消することにより、頑健な統語解析を実現した。また、語彙遷移網文法は、統語規則と意味規則が分離されているので、大規模システムの開発に適している。これらの成果は、世界初のエンジニアリング・ワークステーション型機械翻訳システムとして実用化された。
3. 開発した機械翻訳システムを用いて構築した日英双方向機械通訳システムをスイス日本間で8日間多数の被験者により試験している。その結果、翻訳失敗によるメタ会話や、長距離通信における時間遅れの影響を指摘しており、ユビキタス通信での翻訳通信に必要な機能要件が見出された。

本論文の成果は、日本語処理研究が理論から実用化へと新たな一步を踏み出す契機となっただけでなく、昨今のCPUパワーやメモリ容量の飛躍的な拡大にもかかわらず、依然としてそれを大きく凌駕する仮名漢字変換システムは登場しておらず、本論文は、学術上、實際上寄与するところが少なくない。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成15年2月6日実施した論文内容とそれに関連した試問の結果合格と認めた。