

氏 名 ^{すみ}隅 ^た田 ^{えい いち ろう}英 一 郎
 学位(専攻分野) 博 士 (工 学)
 学 位 記 番 号 論 工 博 第 3437 号
 学位授与の日付 平 成 11 年 3 月 23 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 An Example-Based Approach to Transfer and Structural Disambiguation within
 Machine Translation

(言語間変換と構造的曖昧性解消のための用例に基づく言語翻訳手法)

(主査)

論文調査委員 教 授 池 田 克 夫 教 授 松 山 隆 司 教 授 上 林 彌 彦

論 文 内 容 の 要 旨

現在多数の言語翻訳システムが実用化されているが、それらの訳文の質は必ずしも満足できるものではなく、その根本的な改良には、要素技術である言語間変換と構造的曖昧性解消の精度改善が欠かせない。両課題に対して、本論文では『用例に基づく手法』を発展させて適用している。提案の手法は、入力表現に最適照合する用例に基づいて処理を行うもので、言語間変換、構造的曖昧性解消、さらには、文の翻訳にも適用でき、言語翻訳の訳文の質を向上できることを示した。加えて、用例を並列に検索するアルゴリズムを考案し、その高速化実験の結果に基づいて『用例に基づく手法』を用いた言語翻訳システムの大規模化について良い見通しを示している。

論文は、5章から構成されている。

1章は序論で本研究の目的と論文の構成について述べている。

2章は用例に基づく言語間変換について述べている。言語間変換とは原言語の表現を、目的言語の同義の表現に変換することであり、どのような言語対にも共通な言語翻訳の本質的な課題である。原言語表現が多義であるほど、目的言語表現の可能性も多様であり、言語間変換がより困難になる。機能語は多義性が高いと同時に高頻度で出現するため、その言語間変換は困難であるがその働きは極めて重要である。ここでは代表的な例である日本語の助詞「の」について論じている。『用例に基づく言語間変換』は、入力表現に最適照合する用例に基づいて言語間変換を行う。本手法では、原言語表現と目的言語表現の対を集めた用例データベースと単語の意味概念の階層関係を木構造で表現した辞書を用いる。提案手法により、低頻度語でも適切に扱え、高頻度語についてもデータベースの大きさを小さくすることができた。

3章は用例に基づく構造的曖昧性の解消について述べている。一般に文の構造は文法的観点からは複数成立する場合(構造的曖昧性)があるが、意味的な観点からより妥当な構造を選択することが自然言語処理の応用にとっては重要である。本論文では代表的な構造的曖昧性である英語の前置詞句の係り先の曖昧性を取り上げている。『用例に基づく構造的曖昧性解消』は、言語間変換の場合と同様に入力表現に最適照合する用例に基づいて行うが、用例データベースは、表現対ではなく、正しい原言語構造から構成される。ここでは、前置詞句について始めてその有効性を示し、少ないデータ量で高い精度を実現できることを示した。さらに、様々な用例を用意し、入力文全体をカバーする照合を行い、上記の言語間変換と構造的曖昧性解消を同時に行うことによって、文の翻訳にも同手法が適用できることを示した。

4章は用例検索の並列化による高速化について述べている。『用例に基づく手法』の処理時間について検討すると、検索時間が支配的である。検索時間はデータベースのサイズに比例し、用例データベースを大規模化した場合に検索時間及び全体の処理時間が非常に大きくなるという問題がある。そこで、本論文では、用例を連想メモリに載せて連想メモリの並列照合機能を使って並列に検索するアルゴリズムを提案し、連想プロセッサを木構造に結合した実機で検索の高速化を実現している。このように構成したシステムにおいて、大規模化した場合の処理時間は、用例データベースのサイズには比例せず、各連想プロセッサ上での検索時間と連想プロセッサ全体を結合した木の深さに比例するプロセッサ間通信コストの和となる

が、通信コストは元々各連想プロセッサ上での検索時間より2桁小さいので、処理時間の増加は無視できる程小さい。

5章は結論で、本論文の内容をまとめ、将来の課題について述べている。

論文審査の結果の要旨

本論文は、言語翻訳の精度向上を目指して「入力表現に最適適合する用例に基づいて処理する手法」を発展させ、言語翻訳の要素技術である言語間変換と構造的曖昧性解消に適用し、さらに、用例検索の高速化のため、並列に検索するアルゴリズムを考案し、その有効性を示したものであり、得られた成果は次の通りである。

1. 言語間変換に関して「用例に基づく手法」を適用し手法の有効性を示した。原言語表現が多義であるほど、目的言語表現の可能性も多様であり、言語間変換が困難になる。また、機能語は多義性が高い上に高頻度で出現するため、その言語間変換は困難であるが、その役割は重要である。ここでは代表的な多義語である助詞「の」を対象として、用例データベースとシソーラスを用い、最適照合用例を求めて言語間変換を行い、小容量の用例でも提案手法が極めて有効であることを示した。

2. 文法的観点からは解釈が複数成立する構造的曖昧性を解消するために、意味的な観点からより妥当な構造を選択することとし、前置詞句については始めて、係り先の構造的曖昧性解消が用例に基づいて効果的に行われることを示した。

3. 文の翻訳にも、入力文全体に対して照合し、言語間変換と構造的曖昧性解消を同時に行うことによって、同じ枠組みで、訳文の質を向上できることを示した。

4. 提案手法の処理時間では検索時間が支配的である。そこで、用例を連想メモリに載せ、これを並列に検索するアルゴリズムを考案し、用例データベースを大規模化した場合の高速化を実現した。

以上要するに本論文は、用例に基づく言語間変換、構造的曖昧性解消が言語翻訳の質の向上に極めて有効であることを示し、それを効果的に適用できるようにする手法を考案したもので、学術上、実際に寄与するところが少なくない。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。また、平成11年1月6日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。