

学 位 論 文 題 名

意味属性の対応構造にもとづく日中機械翻訳に関する研究

学位論文内容の要旨

近年、世界各地で機械翻訳の研究が盛んに行われ、いくつかの機械翻訳システムが実用化されている。しかし、構文構造の複雑さ、表層形と意味の対応の複雑さ、原言語と目的言語の表現方法の隔たりなど、良質の翻訳を妨げる多くの課題が残されている。また、これまで機械翻訳の研究開発は英語を主要な対象として行われることが多く、特に本論文が対象とする日中両言語間の機械翻訳に関しては本格的研究が開始されたばかりであり、未開拓の部分がきわめて多い。

一般に、日本語から中国語への翻訳において、自立語については容易に対応する中国語表現が得られる場合が多い。しかし、付属語については中国語表層として独立かつ適切な単語表現がない場合や、付属語単独でなくそれを含む文全体として中国語表現に対応する場合が多く、文全体の解析を行わなければ中国語訳文を得ることができない。このような表現形態を中国では“慣用句型(guanyong-juxing)”と呼んでいる。ここで、“慣用句型”は日本語における“慣用句”とは完全には一致しないので、本論文では“慣用句型”の日本語訳として「常用文型」という語句を用いる。一般に中国では、日本語の学習に際してまず日本語のいくつかの格助詞を学び、その後は前述のような常用文型を中心に学習する。そして、これら常用文型については詳細な文法解析を行わず、直接中国語訳を得ている。また、このようにして得られた中国語訳は常用文型に対応して選択される補助語によりアスペクトとモダリティを満足される。

本論文では、このような常用文型の性質を利用して、入力文を基本文と常用文型に分けて、意味属性を利用して多義性を解消する日中機械翻訳手法を提案している。また、提案する手法に基づく実験システムを構築し、本手法の有効性を確認している。

本論文は7章より構成されている。以下に各章毎の概要を示す。

第1章は本研究の背景と現状の概要、本論文の目的および主要なポイント、そして、論文の構成について述べている。

第2章では日中両言語の特徴を分析し、機械翻訳における主要な問題点を検討する。さらに本研究における日中機械翻訳手法の基本構想について述べている。

第3章では本論文で提案する翻訳アルゴリズムを組み込む基盤であるコード方式日中機械翻訳システムの設計について述べている。日本語は述語文節を中心とし、他の文節の役割をその文節の格助詞で指定することによって文全体の意味を表現する。一方、中国語の場合に語と語との文法関係は語順によって示される。それゆえ、日本語の格関係と中国語の語順との間には対応関係が存在すると考えられる。本章では、この両言語表現の特徴の検討にもとづき、コード方式日中機械翻訳手法を提案している。この手法では、まず、翻訳しようとする日本語文を文の意味的な基本単位に分解し、これを格助詞と活用形を中心に解析し、上記基本単位（これをコード元素という）に構文及び意味属性を表す記号を付加した2つ組（これをコードという）を生成する。なお、コード元素すなわち文の意味的な基本単位としては文節を採用している。以上の処理により文はコードの並びで表され、これをコード列という。次に、得られたコード列を中国のコード列に変換する。ここでは中国語の構文則にしたがって語順の変換が行われ、また、各コード元素は対応する中国語コード元素に変換される。最後に、このコード列から中国語訳文を生成する。このとき、コードの性質にしたがって補助語の付加などの操作を行う。この方式において、コードを生成するとき、文法的情報と意味的情報の双方が考慮され、訳語の選択の際にコードの性質が参照されるので、不要な多義語の増大が防止される。上記にもとづき翻訳システムを作成して実験を行った結果、正翻訳率87%という性能が得られ、本方式の有効性が示された。

第4章では常用文型の翻訳アルゴリズム、特に意味属性による多義性を解消する手法について述べている。この手法は大別して、①日本語文の解析ならびに常用文型の抽出、②日本語常用文型に対応する中国語表現の多義性と多訳性の解消、③中国語表現に現れる補助語の語順及び修飾対象の確定、の3部分よりなる。本手法の要点は、①各々の常用文型についてその訳文条件を調べ、訳文関数を求めること、②訳文関数を用いて直接訳文を生成すること、③意味属性によって常用文型の多義性と多訳性を解消すること、の3点にまとめられる。また、本章では、日本語文を基本文と常用文型に分ける手法、同類常用文型の合併および多重常用文型の処理などを述べる。さらに、この手法に基づく実験システムを作成し、常用文型63種類を登録し、常用文型をもつ約530文について翻訳実験を行った結果、正翻訳率は約96%であった。この結果により、本論文で提案した日中機械翻訳における常用文型の処理手法の有効性が確かめられた。

第5章では日本語の格助詞に対応する中国語の補助語の推定方法について述べている。日本語の格助詞に対応する中国語補助語は一般に1対多であるので、複数の補助語候補から正しい補助語を推定する必要がある。本章では、まず、実際の教科書、文献など約1200文から格助詞およびその関連情報を抽出して意味属性対応関係の分析を行い、その結果にもとづいて日本語格助詞に

対応する中国語補助語を推定する手法を提案している。そして、格助詞の多義性を解消する情報としている日本語格助詞と中国語補助語との関連表（KAHOT）を求め、格助詞を含む1200文の実験を行った。その結果、推定した補助語の正解率が約95%であり、本手法の有効性を確認することができた。

第6章では日中機械翻訳における可保留曖昧関係について述べている。一般に、非同族言語間の機械翻訳では、原言語の係り受け構造の曖昧性を解消しなければならないとされている。しかし、日本語と中国語は非同族の言語であり、全体の構造は異なるが、文中の部分的な構造が同じである場合が多い。この部分に対して、日本語の係り受け構造の曖昧性を残したまま中国語に翻訳しても日本語の意味を復元できると考えられる。本章では、並列助詞「と」と連体助詞「の」と名詞からなる名詞句及び用言連体形からなる文の係り受け構造の可保留曖昧性について検討し、これを利用した翻訳手法について述べている。可保留曖昧関係は原言語の係り受け構造の曖昧性を解消しなくてもその訳文を生成できる曖昧関係である。さらに、技術論文2919文から可保留曖昧関係パターンをもつ806文を抽出し、翻訳実験を行った結果、正解率は97.6%であった。これにより、本論文で提案した手法の有効性を確認した。

第7章は全体の結論で、上記実験結果および翻訳アルゴリズムに関する考察を行い、提案した手法の有効性の確認および今後の課題について述べている。

学位論文審査の要旨

主査	教授	梶内香次
副査	教授	伊藤精彦
副査	教授	小川吉彦
副査	教授	宮本衛市

機械翻訳の研究開発は近年世界各地で活発に行われており、すでにいくつかの実用システムも出現している。しかしながら、これらの多くは英語と他言語間の機械翻訳を目的としており、日中両言語間の機械翻訳に関しては未開拓の部分がきわめて多い。

日本語から中国語への翻訳において、自立語については対応する中国語表現が容易に得られる場合が多い。しかし、助詞、助動詞などの付属語については、独立かつ適切な中国語の単語が存

在しない場合や、個別単語ではなく付属語を含む文全体として中国語文に対応する 경우가多く、一般にきわめて複雑な解析が必要とされてきた。

本論文は、このような日中機械翻訳の問題点を解決し、実用性のある翻訳アルゴリズムの構築を目的として著者が行った一連の研究をまとめたもので、全体は7章より構成されている。

第1章は序章であって、本研究の背景と現状の概要、本論文の目的、主要なポイント、および論文の構成について述べている。

第2章では日中両言語の特徴を分析し、機械翻訳における主要な問題点の検討を行うとともに、本研究における日中機械翻訳手法の基本構想について述べている。

第3章では本論文で著者が提案した翻訳アルゴリズムを組込む基盤であるコード方式日中機械翻訳システムの設計について述べている。この方式は日本語の格関係と中国語の語順との間の対応関係を利用したもので、まず、翻訳しようとする日本語文を格関係にもとづく基本単位に分解し、これに構文及び意味属性を表す記号を付加して2つ組（コード）を生成して文をその並びで表し、次いでこれを同様な中国語のコード列に変換し、このコード列から中国語訳文を生成するものである。この方式は、コード生成の際に文法的情報と意味的情報の双方が考慮されるため、訳語選択において不要な多義語の増大が防止されるという特長を有する。

第4章では常用文型、すなわち日本語文に頻出する慣用的な文型の翻訳アルゴリズム、ならびにその際に問題となる多義性の解消について述べている。ここで著者は、助詞、助動詞等の複雑な結合からなる日本語常用文型に対し、その詳細な解析を行うことなく一体として対応する中国語の文型に変換する翻訳方式を提案し、学術論文など多数の文書から63種類の常用文型を抽出して各々の訳文条件を求め、さらに意味属性を用いて多義性と多訳性を解消する手法を提案している。さらに、これを上記実験システムに組み込み、常用文型をもつ約530文について翻訳実験を行い、正翻訳率約96%という結果を得ている。

第5章では日本語の格助詞に対応する中国語補助語の推定法について述べている。日本語の格助詞に対応する中国語補助語は一般に1対多であるので、複数の候補の中から正しい補助語を推定する必要がある。著者は、教科書、学術文献中の多数の文から格助詞とそれに関連する情報を抽出して、日本語－中国語間の意味属性対応関係を詳細に分析し、その結果にもとづいて日本語格助詞に対応する中国語補助語を推定する方法を提案している。そして、これにもとづき格助詞を含む1200文について翻訳実験を行い、正翻訳率約95%という結果を得ている。

第6章では日中両言語間の翻訳における可保留曖昧関係について述べている。一般に機械翻訳において、原言語文の係り受け構造に曖昧さが存在する場合は予めそれを解消しなければならな

いとされてきた。これに対し著者は、日本語と中国語を比較すると、文全体の構造は異なるが個別単語の意味や文の部分的な構造が同じである場合が多いために、日本語文に係り受け構造に曖昧さが存在しても、それを残したまま中国語に翻訳できる場合があることを見出した。具体例としては、助詞「と」、「の」と名詞からなる名詞句、および用言連体形からなる文の係り受け構造を検討し、技術論文2919文について、このような保留可能な曖昧さをもつ806文を抽出して翻訳実験を行い、正翻訳率97.6%という結果を得ている。

第7章は全体の結論で、上記実験結果および翻訳アルゴリズムに関する考察を行い、提案した手法の有効性の確認および今後の課題について述べている。

これを要するに本論文は、多数の実文献の詳細な検討にもとづいて日本語と中国語の間に存在する意味属性の対応構造を抽出し、それを利用した高性能な翻訳手法を提案し、さらに実験によりその有効性を確認したもので、自然言語処理工学の進展に寄与するところが大である。よって著者は博士の学位を授与される資格あるものと認める。