

(様式 2)

学位論文の概要及び要旨

氏 名 坂田 純 印

題 目 非線形言語モデルに基づく文型パターン翻訳に関する研究

学位論文の概要及び要旨

本研究において、非線形言語モデルに基づく日英文型パターン機械翻訳方式を提案し実装する。

非線形言語モデルでは、話者の対象認識のあり方のような抽象的な意味を、文構造が保存し表現すると仮定している。またこのような文構造の持つ抽象的な意味は、単純な部分構造の組み合わせでは再現できないとしている。そこで非線形言語モデルでは、部分構造を組み合わせて文構造を組み立てるのではなく、文構造の持つ抽象的な意味を保存するように作成された文型パターンを用いて文構造を得る。文型パターンは対訳文対ごとに原言語パターンと目的言語パターンの対として作成される。そのため原言語の文型パターンの適合により、入力文全体に対する構文解析のような複雑な処理なしに文構造が抽出でき、対となる目的言語の文型パターンの使用により、文構造の抽出と目的言語への構造変換が同時に達成される。非線形言語モデルでは、文構造の持つ意味を文型パターンとして抽出するために、他の要素に置き換えても文構造の持つ意味が変化しない要素を「線形要素」、置き換えにより変化してしまう要素を「非線形要素」と定義している。この定義に従い、すでに重文複文文型パターン辞書が作成されている。また線形要素は変数として記述されており、変数化する要素の範囲に応じて、単語、句、節の 3 レベルの文型パターンが作成されている。変数は翻訳の際、入力文の任意の要素に適合する要素であり、翻訳文を得るには変数に適合した要素を翻訳する必要がある。

本研究では、非線形言語モデルに基づく文型パターン翻訳方式を提案し実装した。具体的には文型パターン辞書の記述に対応する単語レベルと句レベルの英文生成システムを構築し、単語レベル文型パターン翻訳方式と句レベル文型パターン翻訳方式を実装した。そして翻訳実験を行い、文型パターンが適合して翻訳文が得られた割合（パターン適合率）と、翻訳精度を調査した。翻訳精度の比較には、構文解析結果を利用する **tree-to-string** の SMT の結果を用いた。

単語レベル文型パターン翻訳では、単語変数に適合した要素の翻訳を、文型パターン辞書から抽出した単語辞書を用いて行った。抽出された単語辞書の日本語見出し語数は、約 5 万であった。翻訳実験より、パターン適合率は約 11%と低いことが明らかになった。しかしながら、適合する文型パターンを得られた場合は、人手評価において約 51%の文で翻訳精度が高く、比較システムの約 24%に比べ、高い翻訳精度を得ていることが明らかになった。また文構造の調査から、文構造変換の精度の高さにより、高い翻訳精度が得られていることが示された。句レベル文型パターン翻訳では、句変数に適合した要素の翻訳を、句パターンと単語レベル文型パターン翻訳を用いて翻訳した。単語レベル文型パターンを元に、約 12 万の句パターンを抽出した。翻訳実験より、パターン適合率は約 28%に改善したが、翻訳精度の高い文は 27%であり、単語レベル文型パターン翻訳に比べて翻訳精度が低下した。一方、比較システムで評価の高い文は 13%であり、比較システムよりは翻訳精度が高い結果となった。問題分析から、半数近い文において不適切な句レベル文型パターンを選択しており、これにより翻訳精度が大幅に低下していることがわかった。

上記の翻訳実験から、パターン適合率の高さは、句レベル文型パターン翻訳、単語レベル文型パターン翻訳の順に高く、翻訳精度の高さは単語レベル文型パターン翻訳、句レベル文型パターン翻訳、

比較システムの順となった。したがって入力文に対し、単語レベル文型パターン翻訳、句レベル文型パターン翻訳、比較システムの順に適用するハイブリッド方式が可能である。