

研究課題別中間評価結果

1. 研究課題名: セマンティック・タイポロジーによる言語の等価変換と生成技術

2. 研究代表者名: 池原 悟 (鳥取大学工学部 教授)

3. 研究概要

従来の要素合成法を基本とする自然言語処理技術の技術的限界を突破するため、新しい原理に基づく言語処理の基礎を確立しようとするものである。日英対訳コーパス(100 万文対)を構築し、その中から重文・複文の対訳標本文(15 万件)を抽出した。引き続き、対訳標本文から意味的に非線形な表現構造を取り出して文型パターン化を行い、単語レベル、句レベル、節レベル合わせて 22.1 万件の文型パターン辞書を試作した。現在、その改良と意味類型化の作業を実施中である。

4. 中間評価結果

4 - 1. 研究の進捗状況と今後の見込み

100 万件という膨大な対訳コーパスの収集・分析、重文・複文の対訳標本文の抽出、非線形な表現構造の文型パターン化など、大きな労力により文型パターン辞書が構築され当初計画通りの進捗がみられる。

文型パターン等についてほぼ整理されたが、意味要素についてさらに改良されたものを作り、再度整理し直している段階である。また、意味のシステムなどを従来のものから新しいシステムに作りかえて全体の精度をあげつつあり、今後の進展に期待したい。

4 - 2. 研究成果の現状と今後の見込み

膨大な言語データをコンピュータで分析し、網羅的に作成された文型パターンは広く自然言語処理研究の基盤となる貴重なデータであり、直ちに外部の研究者が活用できる資料が調い始めている。自然言語への極めて体系的なアプローチであり、世界的にみても文型パターンをこれだけ網羅的に整備したものは他になく、その科学的・技術的重要性は高い。

今後、十分な評価・検証を行い、言語処理の新しい技術体系として完成することを期待する。

4 - 3. 今後の研究に向けて

文型パターン辞書の検証・評価が今後のポイントの一つになると思われるが、あまり精度の向上を意識すると堂々巡りに陥る恐れもある。このため一度現時点でのシステム概念で最後までまとめ上げることも必要であろう。

4 - 4. 戦略目標に向けての展望

新しいパラダイムに基づく実用的な言語処理システムの実現により、質の高い言語翻訳システム機能が一般国民にも手軽に利用できるようになり、言語的バリアーから解放されるなど、まさに「高度メディア社会の生活情報技術」の中核となり得る研究と云えよう。

4 - 5 . 総合的評価

文型パターン 100 万件の収集、整理、分類など、全体的に地味で力業(ちからわざ)も必要な研究であるが、非常に堅実に進められてトップレベルの価値ある成果が出つつある点は高く評価できる。これまでに高い精度の文型パターン等は学問的に貴重であるだけでなく、企業等に対する技術的なインパクトは大きく、日英機械翻訳システムが精度の高いものとなってさらに広く利用されるようになることが期待できる。

今後評価実験を加速して残りの期間で何をどこまで達成できるか早期に見通しをつけ、新しい言語処理体系としてまとめられることを期待する。