

研究最前線からの報告 一日英両言語の機械翻訳の現状と将来の展望一

日本電信電話株式会社 NTT サイバーソリューション研究所

奥 雅博、古瀬 蔵

1. はじめに

インターネットが普及した現在、世界中の情報へのアクセス、国内外における外国の人々との交流など、外国語に接する機会が増大しており、言葉の違い、いわゆる「言語の壁」を克服することが大きな課題となっている。この課題解決に向けて、ある言語（原言語）を別の言語（目的言語）に自動的に変換する機械翻訳技術の研究開発が進められている。しかし、市販されている機械翻訳製品でさえ、翻訳作業の下訳づくりや書かれていることの概要把握など、利用者の作業を支援するレベルにとどまっており、完全自動翻訳による情報発信や異言語間コミュニケーションを可能とする本格的な実用レベルには到達していないのが現状である。

本稿では、機械翻訳研究の歴史を概観し、機械翻訳方式の概要、現状の技術レベルと課題について述べる。

2. 機械翻訳研究の歴史

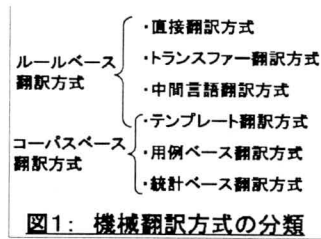
1950年代から60年代にかけて米国では、東西冷戦を背景に露英翻訳の研究が政府の資金助成によって行われたが、研究投資に比べて翻訳品質はさほど向上しなかった。1966年には、米国政府の自動言語処理諮問委員会（ALPAC: Automatic Language Processing Advisory Committee）から、「機械翻訳の効率は人手翻訳に比べて大きく劣る。基礎的な研究を推奨する」といった趣旨の「ALPAC報告」が出された。これを機に、米国における機械翻訳の研究は下火となった。

欧州では、1970年代に多言語間機械翻訳システム SYSTRAN が実用化され、現在でも欧州で使用されている言語を中心に整備が続けられている。

日本でも1950年代から機械翻訳の研究が行われていた。日本で特筆すべきは、1982年から4年間にわたる「Muプロジェクト」（科学技術庁機械翻訳プロジェクト）である。Muプロジェクトは科学技術文献を対象にした日英／英日翻訳システムの研究開発を行うことを目的としていた。Muプロジェクトは日本の機械翻訳研究を世界一流のレベルに押し上げただけでなく、日本国内で販売されている翻訳ソフトウェアに多大な影響を与えた点から見て、日本における機械翻訳研究のエポックを画するプロジェクトであったといえる。

3. 機械翻訳方式の概要

機械翻訳方式は、図1に示すようにルールベース翻訳方式とコーパスベース翻訳方式に大別することができる。



●ルールベース翻訳方式

ルールベース翻訳方式の主流は、トランスファー翻訳方式と中間言語翻訳方式である。前者は、原言語を解析した中間表現を、目的言語を生成するための中間表現に変換する過程（トランスファー過程）を有する。各言語に特有の中間表現を用いることにより、解析ルールや生成ルールを翻訳対象に合わせて調整しやすく、市販翻訳ソフトウェアのほとんどで採用されている実用性の高い方式である。

中間言語方式は、言語に依存しない中間言語を介することで翻訳を行う方式である。多言語間翻訳に有効な方式ではあるが、言語に依存しない中間言語の設計が非常に困難である。

●コーパスベース翻訳方式

1990年代以降、計算機の性能向上を背景に、大量のテキストデータを知識として利用するコーパスベース翻訳方式の研究が盛んに行われるようになった。用例ベース翻訳方式は、あらかじめ大量の翻訳用例を用意しておき、翻訳対象文と似通った翻訳用例を利用して翻訳する方式である。英借文のように例文を模倣して翻訳するのに似た方式である。

統計ベース翻訳方式では翻訳過程を雑音のある通信路とみなす。雑音のある通信路を経て観測された信号から元の信号を復元する情報理論を適用し、「観測された信号＝原言語」から「元の信号＝目的言語」を推定することで翻訳を行う方式である。雑音の性質（変換過程の特徴）は、原言語文と目的言語文の対を集めた対訳コーパスから学習する。

4. 機械翻訳の現状と課題

機械翻訳は、綿密な耐用試験により、多くの表現に対してそれなりの翻訳ができるようになってきたが、翻訳結果は直訳調であり、何とか意味は分かるという域を脱していない。すなわち、機械翻訳の現状は、「ぎこちない部分もあるが、内容はつかめるレベルであり、個人ユースなどの閉じた範囲においては支援ツールとして十分に役立つ」レベルである。

機械翻訳研究の最終的な目標は、「ネイティブが自然であると感じ、修正の必要がないレベルであり、業務用や公的な文書にもそのまま利用可能な訳質の実現」であろう。この目標を達成すること自体が不可能に近いのかもしれないが、少しでも訳質を向上させるために多くの課題への取り組みが行われている。特に重要な課題として言語の持つ曖昧性解消の問題が挙げられる。この問題は、語義の曖昧性解消と構文的な曖昧性解消の2つに大別することができる。前者は、原言語文の単語が持つ語義の曖昧性解消の問題である。

曖昧性の解消、および訳語選択時の語義の曖昧性の解消などであり、後者の例としては、英語における前置詞句の係り先の曖昧性の解消や、日本語における「～の」の係り先の曖昧性の解消などである。これらについては、単語間の意味的共起性や、階層的意味属性体系を利用した意味的制約による曖昧性解消手法が提案され、成果を挙げている。曖昧性解消の問題のほかにも、言語間の構文の違いに起因する問題、言語間の発想の違いをいかに吸収するかといった問題、省略補完の問題（日本語におけるゼロ代名詞補完を含む）などが解決すべき重要な課題として挙げられる。

5. まとめ

翻訳という知的な活動を支援するツールとしての機械翻訳は現状でも十分に役に立つと考えられるが、完全自動翻訳への道は遠いのが現状である。当然のことではあるが、現状の機械翻訳研究では、原言語を忠実に目的言語に変換することに主眼が置かれている。異言語間のコミュニケーションを考えた場合、「何を本当に伝えたいのか」という点も忠実な翻訳と同様に非常に重要である。翻訳に際して、どの部分の翻訳をさぼってもよく、何を正しく翻訳して伝えなければならないのかが理解できれば、コミュニケーションという観点から見た翻訳品質は大幅に向上するに違いない。このためには、コミュニケーションの本質は何かを見つめることが重要である。日本英語コミュニケーション学会における研究の進展が日英間の機械翻訳によるコミュニケーションの実現に大きな役割を果たしてくれることを期待している。