中　澤　敏　明（黒橋教授）
「Fully Syntactic Example-based Machine Translation」
（構造的言語処理に基づく用例ベース機械翻訳）
平成 22 年 3 月 23 日授与

機械翻訳の歴史は古く、1940 年代後半から始まったと言われている。それにもかかわらず、現時点ではまだ高精度の機械翻訳システムは実現されていない。翻訳とはそもそも高度に知的な行為であり、これを計算機上で行なうには大きな計算機パワーが必要であったり、大規模な対訳データや高度な言語処理技術などの言語資源・言語知識が必要である。これらの問題は、近年の計算機のめざましい発達や、インターネットなどによる対訳データの利用、さらには言語処理技術の発展などにより解決されつつある。それとともに、他言語に触れる機会も急速に増加しており、機械翻訳への期待が高まっている。

機械翻訳において最も重要かつ困難な問題の一つは、言語間の違いを克服することである。ここで「言語間の違い」というのは各言語の性質の違いのことであり、言語の性質とは意味のある文を構成するために用いられる要素、表現、語順などのことである。英語とフランス語間のように似た言語対では、逐次的な単語の置き換えと局所的な単語の入れ替えだけでも、かなり高精度な翻訳が行なえるが、日本語（主語 - 目的語 - 動詞）と英語（主語 - 動詞 - 目的語）のように語順が大きく異なり、また表現の自由度も異なる言語対では翻訳は非常に難しくなり、実用的な翻訳システムは未だに開発されていない。このように違いの大きい言語対間の翻訳では、文の構造的な情報を利用することが非常に重要であり、その方法の一つとして文を単語列ではなく木構造として扱うことが考えられる。既存の手法の多くは文を単語列として扱っており、木構造を利用した手法であっても、対訳文内の単語対応の推定、単語対応結果からの翻訳知識の自動獲得、および獲得された知識を利用した翻訳の各ステップが分離しており、木構造を利用するメリットが最大限有効に利用されているものはほとんどない。本論文ではこれらの全てのステップにおいて文を単語依存構造木として扱う枠組を提案し、これが高精度な翻訳の実現に有効であることを示した。

翻訳知識の自動獲得ステップでは、1. 依存構造木上での距離を利用した、2. 方向性のない、3. 句対応獲得手法を提案し、既存の、単語列距離を用いた方向性のある単語対応推定手法による結果からヒューリスティクスにより句対応を獲得する手法に比べて 13% 程度誤りを軽減することができた。また、この対応結果から木構造で表現された翻訳用例を獲得し、これを翻訳で利用してその精度を人手により評価したところ、人手によりチューニングされた、ルールベースの商用翻訳システムと同等の翻訳精度を達成し、既存の単語列による翻訳手法よりも有意に高精度であることを示した。また構築した翻訳システムは日本語イーコマースサイトの商品説明文の英語への自動翻訳に利用されるなど、実社会への貢献も大きい。

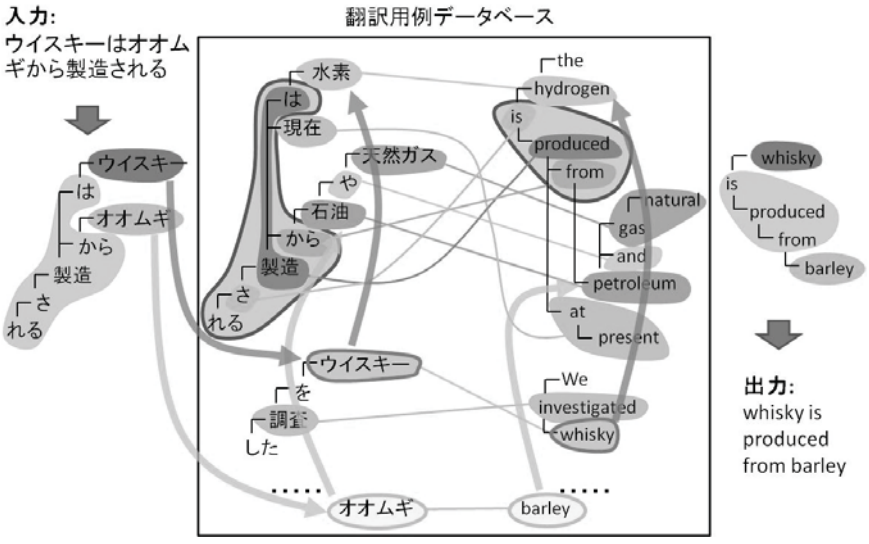

図 1：用例ベース機械翻訳の概要

今後は他の言語対での大規模な実験を行ない、提案した手法が言語に依存しないロバストな手法であることを示し、様々な場面での機械翻訳の利用の拡大に貢献したい。