

自然な訳文を生成する翻訳システムの開発

1. 背景

インターネットの発展に伴い、国境を超えた情報のやりとりが爆発的に増大した。しかし、情報をやりとりする際の「言語の壁」は依然として残っており、新たな情報格差の原因ともなっている。この「言語の壁」をソフトウェアを使って取り払うための研究 ― 機械翻訳の研究 ― はすでに数十年にわたって行われており、パソコン用のアプリケーションも数多く販売されている。

機械翻訳システムは、おもに構文解析のノウハウの蓄積や辞書データの蓄積により、ここ 20 年間で着実な進歩を遂げてきた。このため、これを利用することにより、とくに技術的な文書の翻訳においては、辞書を引く時間が減ることによる翻訳作業の効率化される、ソース言語（翻訳元の言語、たとえば英語）からターゲット言語（翻訳後の言語、たとえば日本語）に置き換える作業が部分的にであれ不要になるため翻訳者の心理的な負担が軽減する、といった効果がある。

このように現状でも翻訳システムは翻訳作業の効率化に貢献はしているが、翻訳ソフトウェアの出力する訳文は、優秀な翻訳者に比べればはるかに劣っており、いわゆる「機械翻訳臭い」不自然な訳文が数多く生成される。このため、最終的な訳文を完成するためには、かなりの「後編集」を行う必要がある。翻訳の専門家にとっては「翻訳」ソフトウェアというよりも、「自動辞書引き」ソフトウェアとしての使われ方が主になっているわけである。

筆者は機械翻訳システムの開発に 20 年間携わりつつ、[1][2]をはじめとする数多くの翻訳書を出版し、翻訳者教育用のテキスト[3]の執筆も行ってきた。翻訳作業は自分が感覚的に納得した訳ができればそれで完結するが、それを「教育」する場合、なぜその訳でなければ納得しないのか、どのようにすれば納得する訳が思いつくのかを、翻訳者を目指す立場の人々にわかるよう説明する必要がある。「感覚的に」では受講者が納得しないので、善し悪しの理由を具体的に示す必要がある。この作業は、自分の翻訳処理プロセスを別の角度から見つめ直す絶好の機会となり、本開発の大きなきっかけとなった。

2. 目的

本開発の目的は、第一線の翻訳者と同レベルの訳文を出す機械翻訳システムに必要な要素を洗い出し、そのうち実現可能な要素を枠組みとして組み込んだプロトタイプシステムを完成することである。

3. 開発の内容

前年度に引き続いて行った本年度の開発では、前年度の開発者をリーダーとして、リーダー同様に翻訳の経験も豊富な開発者を新たにもう 1 人加え、昨年度開発したプロトタイプを利用して、前年度のシステムをさらに拡張するとともに、実践的なシステムのプロトタイプの開発を行った。

本年度の開発においては、自然な訳文を出力し、しかもそれが誤りのない正しい翻訳結果であることを保証するシステムを目指した。現状においては、このようなシステムの開発には、対象の領域を制限し、原文をかなり制限したものに行なければならない。

このため「海外旅行用の携帯翻訳装置」の開発をターゲットとし、利用者ができるだけ不便を感じないようにインタフェースを工夫しつつ、正しい訳文を生成できる範囲に対象文を制限して実現した。前年度提案した、対象言語に依存しない翻訳システムの枠組みを実証するため、開発リーダーがある程度の知識を持つ、日本語、英語、韓国語、中国語、フランス語の5カ国語相互間の翻訳および辞書検索を可能にするシステムのプロトタイプを実現した。

4. 従来技術との相違

図1に従来型の機械翻訳システムと本システムとの相違を示す。従来型の翻訳システムにおいては、辞書としてソース言語（翻訳元の言語）の単語に対してターゲット言語（翻訳後の言語）ではどのような訳語が対応するかという、「対訳辞書」のみが用いられている。これに対して、本システムにおいては、翻訳者が対訳辞書の情報のみならず、ターゲット言語内部の語の関連を利用して、読者が違和感を抱かない自然な訳文を生成するという事実に着目し、ターゲット言語内部の情報を記述する「ターゲット言語辞書」とソース言語内部の情報を記述する「ソース言語辞書」を用いる新しい枠組みを提案した。

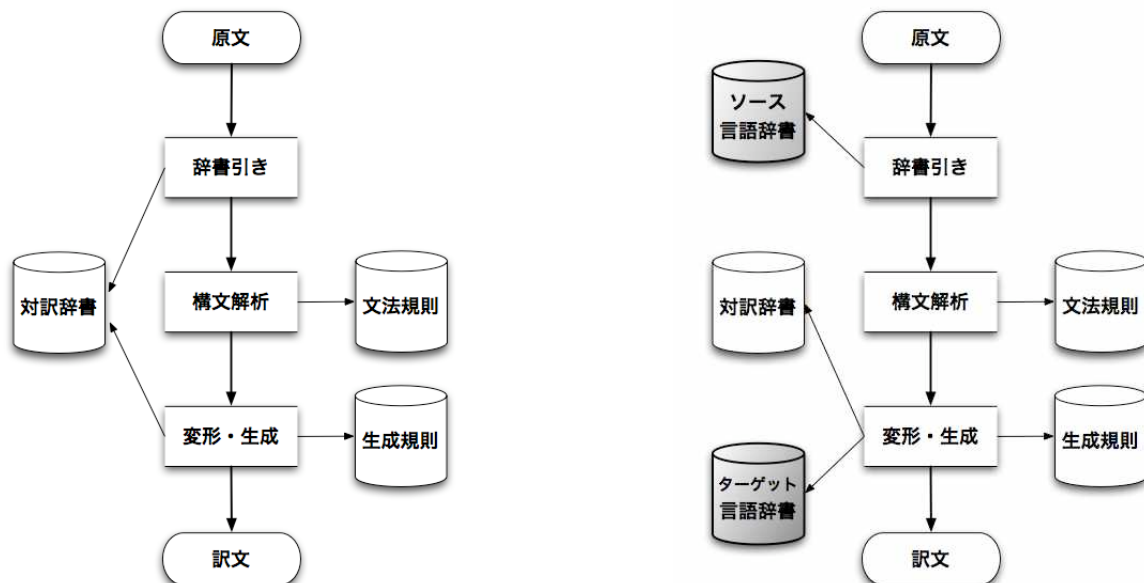


図1 従来の機械翻訳システムの構成（左）と本システムの構成（右）

この枠組みのうち、ターゲット言語辞書の有効性を、次の例で説明する。

【原文1】

The Mixer object gives you direct control over the volume.

【従来の訳例】

ミキサーオブジェクトはあなたにボリュームについての直接のコントロールを与える。

【本システムの訳】

Mixer オブジェクトにより、ボリュームを直接制御できる。

従来の機械翻訳では原文で形容詞である語句は訳文でも必ず形容詞として訳される。このため、"direct"は日本語で「直接の」あるいは「直接的な」などとなる。一方、本システムでこれをは副詞として「直接」と訳している。翻訳者が翻訳する時には訳文を

自然にするために頭の中で品詞を変換することを、多くの場合無意識に行なっている。この例に対して翻訳者は、形容詞を副詞に変えて「直接制御できる」などと訳すが、本システムではこれが可能になってる。このような品詞の変換などのためにターゲット辞書が用いられる。この様子を図2に示す。

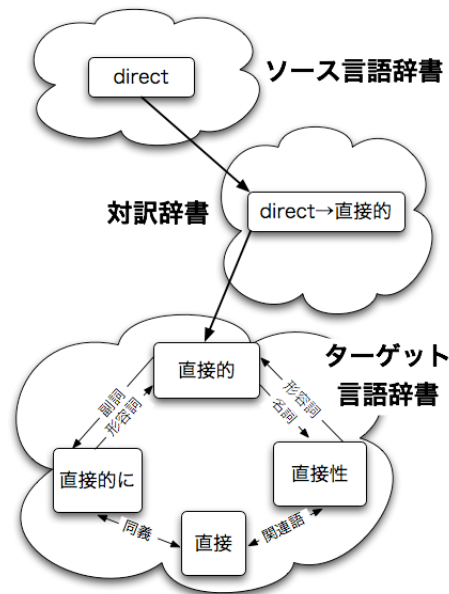


図2 ターゲット辞書の活用

ソース言語（英語）の direct は、ターゲット言語（日本語）においては、たとえば、「直接的」に対応する。"give direct control"のうち、"give control"は「制御する」などと訳するのが自然である。動詞 give と名詞 control が2語で「制御する」という動詞に対応していることになる。原文において control を修飾している direct は形容詞であるが、訳文においては、修飾先の control が他の単語と合体して動詞になってしまったため、形容詞ではなく、副詞に変換される必要がある。従来の機械翻訳システムではこういった処理が不可能であったが、本システムにおいては、「直接的」の関連語の中からその副詞形である「直接」を選択することができ、「直接制御する」という自然な訳語を出力することができる。

この仕組みにより次のような訳も可能になった。それぞれ、原文、従来の翻訳ソフトによる訳例、そして本システムの訳である（例文は[5]から引用）。

Ignorance of foreign customs can result in unexpected misunderstandings.
 外国の習慣についての不案内は、予想外の誤解に終わることがありえる。
 外国の習慣を知らないと思いがけない誤解を生ずることがある。

His application of the rule to this case was in a sense quite natural.
 このケースへの規則の彼の適用は、ある意味で全く自然だった。
 彼が今回のケースにこの規則を当てはめるのはある意味ではきわめて自然なことであった。

このほか、ここでは詳しくは説明しないが、本システムは以下のような機能を実現した。

- ・ 広範な情報の参照が可能 一プロの翻訳者は、非常に広い範囲を見渡して訳語や

訳文の構造を決定するが、従来のシステムでは近傍の単語だけしか参照することができなかった。本システムでは、文の構造の任意の範囲を自由に見渡して必要な変形を行うことができる。

- ・ 対象の言語に依存しないアーキテクチャ — 従来型の機械翻訳システムにおいては、ソース言語あるいはターゲット言語に依存した処理をプログラムで記述する必要がある場合が多いが、本システムでは、ほとんどの変換はデータによって行われる。

さらに、本年度の開発において作成した海外旅行用携帯翻訳機のプロトタイプにおいては、中間言語を媒介とした多言語インタフェースを考案した。この携帯翻訳機においては、ソース言語およびターゲット言語として、日本語、英語、韓国語、中国語、フランス語をサポートし、これらの言語間相互の翻訳ができる。この開発に関しては、特許を申請する予定なので、詳細は今後開発者のホームページなどで公開する。

5. 開発者名

武舎広幸 (マーリンアームズ株式会社 代表取締役)

河村政雄 (個人)

電子メール: <http://www.musha.com/contact.html> より

開発者 URL: <http://www.musha.com/>

開発パートナー: マーリンアームズ株式会社 <http://www.marlin-arms.co.jp/>

謝辞

本プロジェクトの実施にあたり、様々な支援をいただいたIPAの未踏ソフトウェア創造事業関係者の方々、並びにプロジェクトマネージャの紀信邦氏に深く感謝します。

参考文献

- [1] ローラ・リメイ+チャールズ・L・パーキンス著, 武舎広幸+久野禎子+久野靖訳, 『Java 言語入門』, ピアソンエデュケーション, 東京 (1996)
- [2] ローラ・リメイ+アーマン・ダニッシュ著, 武舎広幸+久野禎子+久野靖訳, 『HTML 入門 第2版』, ピアソンエデュケーション, 東京 (1998)
- [3] 武舎広幸監修, 『プロ翻訳家養成英日コンピュータコース ADVANCED』, TEXTBOOK 1~6, DHC 総合教育研究所, 東京 (2001)
- [4] 武舎広幸著, 『パソコン翻訳ソフト活用法』, pp. 37-58, ピアソンエデュケーション, 東京 (1994)
- [5] 安西徹雄著, 『翻訳英文法一訳し方のルール』, バベルプレス, 東京 (1986)