

自分と同レベルの訳文を出力する翻訳システムの探求

A Machine Translation System Generating *Natural* Natural Language Sentences

武舎広幸
MUSHA Hiroyuki

マーリンアームズ株式会社 (<http://www.marlin-arms.co.jp/>)

ABSTRACT. The author has been developing machine translation software for more than twenty years, while he has translated and published more than twenty books and has authored textbooks for training of technical translators. As a professional translator and programmer, the author has analyzed the translation process of human translator, and developed a prototype system that can generate natural (un-awkward) sentences in target languages.

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（たとえば英語）からターゲット言語（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。

本開発の第 1 段階においては、すでに筆者が翻訳した文書と翻訳者の教育用に作成したテキストを詳細に検討し、人間による翻訳中にどのような操作（処理）を行っているかを分析することによって、翻訳システムにおいてどのような機械処理を行う必要があるかを検討した。

本開発の第 2 段階においては、上記の検討の結果得られたもののうち、現在の技術レベルで実現可能な処理を洗い出し、それを実現する具体的な仕組みを考案し、（そのうち所定の期間に実現可能なものを）筆者が既に開発していた小規模なプロトタイプに加える作業を行った。

3. システムの概要

開発したプロトタイプシステムの基本的な処理手順は従来の機械翻訳システム（[4]など参照）と大きな差異はない。しかし、本システムにおいては、翻訳者が行うような柔軟な処理を実現するために、とくに以下のような機構を実現した。

- ・局所的な共起関係だけでなく、任意の範囲を見渡し、必要な変形操作を行える
- ・原文中の単語の品詞に縛られずに、自由に品詞の変換が行える

現在、本開発に関して特許出願準備中であるため、詳細は、公開可能になり次第、筆者のホームページ (<http://www.musha.com/>) 等で公開予定である。

4. 結果

本プロトタイプによる翻訳結果のサンプルを示す。

＜原文＞

(<http://maccentral.macworld.com/news/2002/10/21/ipod/> より一部引用)

One strange thing about MusicMatch and the iPod is the software versions the device will work with. The iPod comes with version 7.1 of the MusicMatch software. When you boot up the computer, MusicMatch asks if you want to upgrade to the newest version, 7.2 -- while most people would do the upgrade to stay current with the latest software, you won't be able to use your iPod if you do.

This problem should be remedied with a new point release. MusicMatch spokesperson Jennifer Roberts said that MusicMatch Jukebox 7.5 -- due out sometime in the next several weeks -- would have iPod support that should be maintained by future releases as well.

＜訳例＞

(http://www.zdnet.co.jp/macwire/0210/22/re00_ipodj.html で公開された訳文)

MusicMatch と iPod に関して1つの奇妙なことは、デバイスが動作するソフトウェアバージョンだ。iPod には、MusicMatch バージョン 7.1 が付属している。コンピュータ上でブートすると、MusicMatch は最新バージョンの 7.2 にアップグレードしたいかどうか確認してくる。大部分のユーザーは最新のソフトウェアにアップグレードしたいと思うだろうが、そうすると iPod を使用できない。

この問題は、新しいポイントリリースで修正されるべきだ。MusicMatch スポークスマンの Jennifer Roberts 氏によれば、今後数週間で発表される MusicMatch Jukebox 7.5 は、将来のリリースによっても維持される iPod サポートを得るといふ。

＜本プロトタイプによる訳文＞

MusicMatch と iPod に関してひとつ奇妙なのはデバイスが動作するソフトウェアのバージョンである。iPod には MusicMatch ソフトウェアのバージョン 7.1 が付属している。コンピュータのブート時に、MusicMatch は新しいバージョン、すなわち 7.2 にアッ

プグレードするかどうか尋ねてくる。ほとんどの人はいつも最新のソフトウェアを使うためにアップグレードをしたいと思うだろうが、そうすると iPod を利用できなくなってしまう。

この問題は次のマイナーバージョンアップで修正されるはずである。MusicMatch 担当者の Jennifer Roberts は（今後数週間のうちに出る予定の）MusicMatch Jukebox 7.5 が iPod をサポートするだろう、そしてそれは今後のリリースによっても変わらないはずであると語った。

この例からわかるように、少なくとも一部の文章に対しては必要なチューニングを行えば、公に提供されているのと遜色ないレベルの自然な訳文を生成することが可能になった。

5. 今後の展望

本プロトタイプの辞書には、700 語程度の単語しか登録されておらず、実践的な試験はまだ行えていない。今後は、実用レベルに最低限必要な数万語レベルの語数を、もつ辞書を構築して、試験を行う予定である。

6. 参加企業及び機関

- ・ 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・パーキンズ 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）