ホームページの多言語化に向けた
機械翻訳とコミュニティによる後編集の活用

相川孝子（マイクロソフトリサーチ)、 井佐原均（豊橋技科大）

1. はじめに

国際化が進む今日、多言語による情報共有の必要性が高まってきている。最近では、自治体、企業をはじめ、さまざまな組織がホームページを情報の発信元とし、その多言語化を進める動きも高まっている。そうした情報の多言語化が迫られる一方、どのように多言語化の実現をはかったらいいのかが大きな問題となっている。Web上の莫大で、かつ絶えず更新されていく情報を全て人間の翻訳者たちに依頼し、翻訳するというのは、時間的、コスト的に非現実的である。「情報の多言語化」という社会的需要を満たすための、適切な手段を見つけなければいけない状況にある。

その手段の一つとして、機械翻訳を導入している、あるいは導入を検討している組織もあるが、機械による自動翻訳では、どこでどんな間違いが起こるか分からないために、機械翻訳の導入は、危険が高すぎると懸念する組織も多いであろう。情報の信憑性を問われる自治体、企業組織などでは、機械翻訳による翻訳が「誤訳ゼロ」という状態にならない限り、「機械翻訳による情報の多言語化」へ踏み込むのは、立場上なかなかできないこともあろう。ここに大きな需要と供給のギャップがあるように思われる。

本稿では、このギャップを埋める試みとして、共同翻訳フレームワーク（CTF: Collaborative Translation Framework）を紹介し、実際にこのフレームワークを使って、大学からの多言語での情報発信を進めている例をしめすことにより、共同翻訳という考え方によって、多くの組織で情報の多言語発信が可能になることを示す。

2. Microsoft Translator

共同翻訳フレームワーク（CTF）はMicrosoft Translator[1]の翻訳システムを、Web上でWidgetとして走らせ、その上にユーザーからのフィードバックを受け入れるユーザーインタフェースを付加したものである。

マイクロソフトが機械翻訳の研究に取り組み始めたのは1999年ごろであるが、当初の対応言語は、5言語であった。それぞれの言語にパーサーと辞書を備えた、ルールベースのシステムを開発していたため、対応言語が増やせないというスケーラビリティーの問題に直面し、2005年に統計ベースのシステムへと切り替えた。これにより、現在では、対応言語は35言語以上に拡大した。

現在、パブリックAPIを提供するとともに、Office、Bing、Internet Explorer 8といったソフトウェアの中にも積極的に翻訳

[1] http://www.microsofttranslator.com/

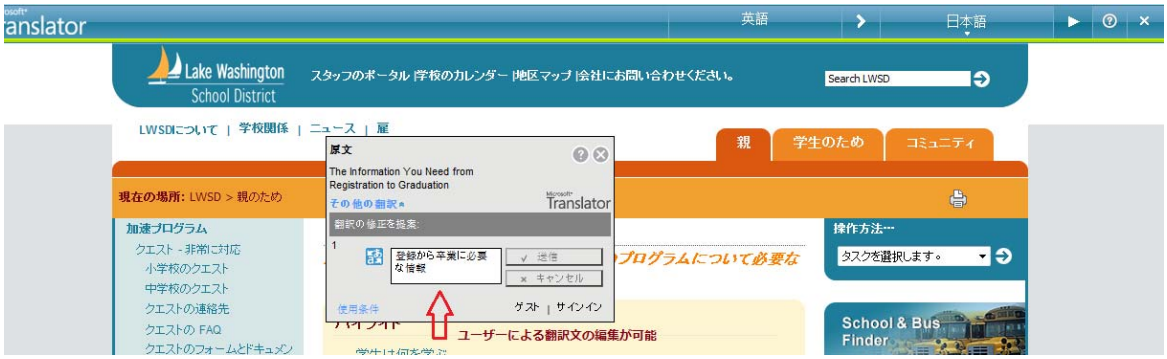

図 1　Lake Washington School District のホームページ

機能を取りいれてきている。また、無料自動翻訳サービスを Web 上にも提供して、一日平均五百万以上のビジターがホームページに訪れている[2]。

また、Microsoft Translator は統計翻訳システムであり、大規模なバイリンガルコーパスデータを必要とするので、いかに多くの、そしていかにきれいなバイリンガルコーパスデータを Web 上から自動抽出できるかなどという研究にも取り組みながら、機械翻訳精度の向上に努めている。

3. 共同翻訳フレームワーク (CTF: Collaborative Translation Framework)

Microsoft Translator を開発する過程で、翻訳システムの精度向上、対応言語の拡大などに注力すると同時に、「どうやったら人間と機械が共同して翻訳の質を高め、情報の多言語化に努めることができるのか」という観点からの検討も進めてきた。そこで、できあがったのが共同翻訳フレームワーク（CTF）である。

図 1 は、現在実際にこの CTF を自分たちの Web サイトに取り入れているアメリカのある学区のホームページの例である[3]。この学区は、学生の多くの親が移民者であるため、Web 上で発表する学校のイベント情報、緊急事項などをどのように効果的に、いち早く（英語を母国語としない）親に連絡できるかという問題をかかえてきていた。CTF 装備の Widget を彼らの Web サイトに組み入れることにより、英語が分からない学校地区の親たちにも彼らの言語で情報共有がいち早くできるような態勢を構築しつつある。

4. CTF の機能とその特徴

本節では、CTF の機能のうちで、特徴的な機能について簡単に説明する。

「編集機能」は、文字通り機械翻訳結果を人間が確認し、訂正・編集を加えることができるという機能で、編集された結果は、

[2] http://www.microsofttranslator.com/user/

[3] この学区は、アメリカ合衆国ワシントン州レッドモンド市にある Lake Washington School District という学区で、小、中、高５０校ほどの公立学校が所属している。詳細は、http://www.lwsd.org/Pages/default.aspx を参照。

Microsoft Translator のデータベースに返され、以降の同一ページの翻訳に利用されるとともに、翻訳精度の向上に利用される。この機能を用いることにより、自分が直した翻訳文がそのウェブサイトの翻訳に反映されるばかりでなく、今後の機械翻訳精度の向上にも貢献できるという一石二鳥の効果が得られるわけである。また、機械翻訳を開発する立場からすれば、ユーザーが使えば使うほど、翻訳精度がよくなるという、いわゆるオーガニックなエコーシステムを築き上げることで、「人間と機械が手に手をとってWeb上の情報の多言語化を進める」という野心的ゴールを達成させることができる。

「権威ユーザー指定機能」は、Webマスターが特定のユーザーを選び、特別の編集資格を与える機能である。これにより、Webマスターが信頼できるユーザーを選び、この選ばれた権威ユーザー（authoritative users）によって編集された翻訳を、「信頼できる翻訳」として自分のサイトに優先して使うことができる。権威ユーザー指定機能は、自治体や企業のような、情報の正確さが問われる組織にとっては、大切な機能である。こうした組織の場合、一番問題になるのが一般ユーザーによる悪意のある翻訳編集である。ユーザーからの編集がWeb上で可能である以上、どのような翻訳訂正がされてしまうか分からない。ユーザーからの故意的な、あるいは悪意のある編集を防ぎつつ、より正確で信憑性のある翻訳を得るために実装された機能の一つである。上であげたLake Washington School Districtの場合は、修正作業にボランティアで関わる、生徒の親を権威ユーザーに指定することにより、信頼性の高い翻訳編集が行われている。

このほか、CTFには、どの翻訳が一番良いかを投票できる機能や、その投票数を基にして、Webマスターが最適な翻訳を指定できる機能なども備えている。例えば、図2では、“About Us”という英語原文に対して、その文の機械翻訳結果を修正したものがリストされており、LWSDと修正するという案に対しては投票数3であることが示されている。このような投票機能によって、Webサイトのオーナーは、翻訳目標言語が分からなくても、安心してどの翻訳が一番信頼できる翻訳なのか決めることができる。

図2　翻訳出力の修正案のリスト

5. 豊橋技術科学大学でのCTFプロジェクト

豊橋技術科学大学は、海外協定大学との交流や海外研究機関との共同研究を通し活発な国際交流活動を行っており、現在、200名を越す留学生（正規生・研究生等）を受け入れている。留学生の比率が1割に達し、

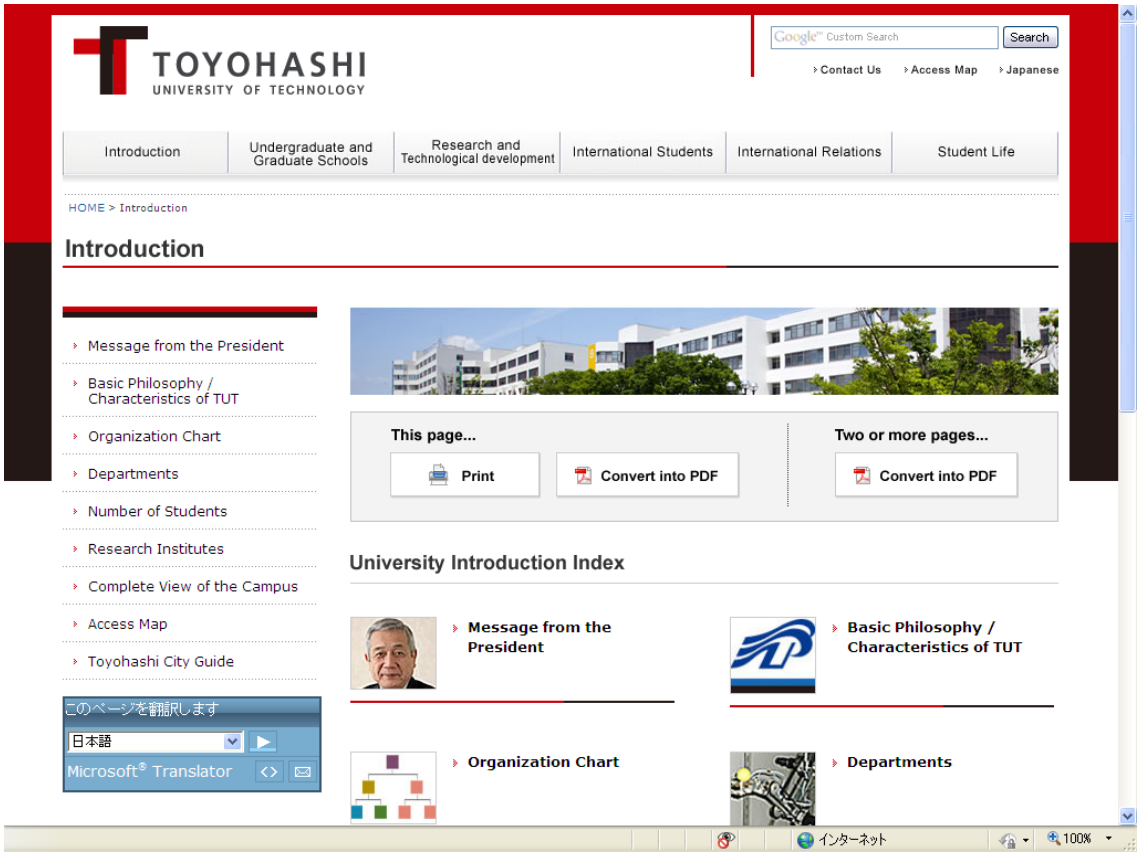

図３　豊橋技術科学大学の英語版ホームページ

特に東南アジア諸国からの留学生が多い。このような状況から海外への情報発信には力を入れており、平成２２年度には英語でのホームページを全面的に改訂した。

全面改訂に合わせて、英語だけではなく、より多くの言語での情報発信を目指して、英語でのホームページに Microsoft Translator と CTF を組み込むこととなった（図３）[4]。

英語版のホームページに CTF 付の機械翻訳のボタンを付けることには以下のようなポイントがある。

１）英語から他言語への翻訳は、日本語から他言語への翻訳と比べて、翻訳精度が高いことが期待できる。

２）利用者は英語のページが機械翻訳される過程を目にした後、母語への翻訳を読むために、それが機械翻訳の結果であり、保証された訳文ではないことを実感しつつ、訳文を読む。

３）留学生等を使って、大学の実態に沿った翻訳修正を行うことにより、その修正結果は以後の大学のホームページの翻訳に反映される。

現在、実際に翻訳修正作業を実施し、ホームページの訳質の向上、より良い対訳の獲得、翻訳処理へのフィードバックを進めている。

[4] http://www.microsoft.com/japan/presspass/detail.aspx?newsid=3878