

国際ワークショップ「翻訳テクノロジー評価」について

立見みどり

(立教大学異文化コミュニケーション研究科兼任講師)

As various translation technologies are increasingly employed in the translation workflow, especially in the field of technical translation, localisation, and other non-literary translation, it is increasingly important for the practitioners, educators, and researchers of translation to know how to properly evaluate the effectiveness of such technologies in order to use them wisely. This paper reports on an experimental workshop that introduced the basic concepts of the evaluation of translation technologies to a group of translation students and educators who had had limited exposure to such technologies. The workshop was held online, connecting three universities in Japan, USA, and Taiwan. In this paper, we will outline the background, the method, and the execution of the workshop as well as presenting a brief explanation of the evaluation methods introduced in the workshop, and finally we will discuss the benefit and the future potential of similar types of workshops.

1. はじめに

1.1 ワークショップ実施の背景

立教大学 SFR (学術推進特別重点資金) 翻訳研究プロジェクトチームは 2012-2014 年度、「翻訳『革命』期における翻訳者養成の新たなコンテンツと方法に関する学際的研究」に取り組み、公開シンポジウム、研究会、新科目の開発・実施などを行った。その目的は、翻訳環境に現在生じている大きな変化の実体を把握し、そこで求められる翻訳者のコンピタンスを見極め、それを満たす翻訳者を養成するために必要なコンテンツとデリバリー方法を提案することにあつた。その一環として、2013 年、世界における日本語関連の翻訳研究の促進のため、また日本から世界へ向けて翻訳教育の情報を発信する試みとして、複数の会場をインターネットで結ぶワークショップが企画され、以下の 2 回にわたって実施された。いずれも、立教大学をメインの会場とし、モントレイ国際大学と台湾輔仁大学を Skype で繋ぎ、2 時間にわたって行われた。なお、メイン会場には立教大学および東京大学の学生と教員が参加した。

TATSUMI Midori, "International Workshop: Evaluation of Translation Technologies," *Interpreting and Translation Studies*, No.14, 2014. pages 285-300. © by the Japan Association for Interpreting and Translation Studies

第1回:「翻訳テクノロジー評価」ワークショップ(2013年11月16日(土))

第2回:「文芸翻訳」ワークショップ(2014年1月26日(日))

第2回のワークショップは、作家であり文芸翻訳者であり現在は早稲田大学文学部で教鞭をとる辛島デイヴィッド氏を講師に迎えて行われた。筆者は参加者のひとりであったが、文芸翻訳の難しさと面白さを体感することのできる興味深いワークショップであった。こちらについてもその成果を広く共有することが期待されるが、本レポートでは、筆者が担当した第1回目のワークショップについてのみ報告する。

1.2 ワークショップテーマ決定の背景

今回のワークショップで講師を務めた筆者は、長らくIT関連のローカライゼーションにかかわってきた現役の実務翻訳者であり、機械翻訳関連の応用言語学研究者でもあるため、ワークショップの中心的な題材として翻訳テクノロジーを選んだ。中でも「翻訳テクノロジー『評価』」に焦点を当てたのは以下の理由による。

翻訳はさまざまな分野のテキストを対象に行われるが、その中でも広く「実務翻訳」と呼ばれる分野では、翻訳に各種テクノロジーを利用するケースが多い。一口に「翻訳テクノロジー」といっても、用語管理システムから、翻訳メモリ、ブルーフリーディング用プラットフォーム、翻訳プロジェクト管理、機械翻訳システム、後編集プラットフォーム、コラボレーション翻訳プラットフォーム、総合翻訳支援環境を含め、非常に多岐にわたる。これらのテクノロジーを効果的に利用するためには、まずその機能や利点などを適切に評価し、利用価値を見極め、さらに最適な利用方法を定める必要がある。適切に評価するには、コスト、使いやすさ、翻訳品質向上への貢献度や翻訳ワークフロー向上への貢献度を含め、さまざまな角度から測定・検討し、総合的な有用性を判断する必要がある。そのための評価基準はこれまでに複数提案されているが、中でも最も内容が充実しているもののひとつに、欧州委員会のイニシアチブにより定められた評価規格であるEAGLES(Expert Advisory Group on Language Engineering Standards)¹がある。EAGLESでは、翻訳メモリ、用語管理システム、機械翻訳システムなどのツールごとに、その設計、機能性、対応ファイル形式、オンライン・オフライン機能、セグメント化、アラインメントなど、さまざまなチェックリストと評価方法が細かく定められている。翻訳テクノロジーの利用者にとって、翻訳テクノロジーの有用性をどのように評価するかは重要な問いであり、その最適な解答を出すためには、現存する評価法を評価・検討するところから始める必要がある。このため、新しい時代の翻訳者養成を目指すワークショップには適した題材であると判断した。

しかし翻訳テクノロジー全般をワークショップのテーマにすると対象が広くなりすぎることで、また参加予定者の多くは翻訳研究者および翻訳実践者であるが、機械翻訳を含む翻訳テクノロジーの利用者は少数派であることから、2時間のワークショップで取り扱える範囲に対象を限定し、翻訳テクノロジーにあまりなじみのない参加者にも興味を持っても

らえそうな話題として、「機械翻訳システム」を対象を絞った。しかし一口に「機械翻訳システムの評価」といってもやはりその範囲は広く、関与者の立場によって評価のポイントは異なる。翻訳業務の発注者にとってはコストが重要であり、翻訳ベンダーにとってはワークフローのスピードも大きなポイントなる。翻訳者や編集者にとってはシステムの操作性や、システム導入による作業の軽減の度合いが気になるところであるし、翻訳システム管理者なら拡張性や他のシステムとの統合性が問題となる。しかし翻訳物の読者を合わせたすべての関与者にとって共通の関心事は、「機械翻訳システムによる翻訳文の品質」である。そこで、今回のワークショップでは、「翻訳テクノロジー評価」の中でも、「機械翻訳文の品質評価」に焦点を絞った。

2. 機械翻訳文の品質評価

機械翻訳文の品質評価は機械翻訳研究の歴史の中で常に重要なテーマであり続けているが、近年、機械翻訳が広く実用されるに従い、機械翻訳の研究と実践の場において改めて関心が高まっている。2012年、欧州委員会の出資により複数の大学と研究機関が参加するプロジェクト QTLaunchPad² が開始された。このプロジェクトでは、翻訳テクノロジーを活用するため、人手翻訳と機械翻訳の品質のギャップを埋めるべく、文書のタイプ、目的、対象読者などに応じて適切な評価基準を選択する枠組みを提案している。また、翻訳業界で機械翻訳を有効活用するための調査研究、データ、および各種サービスを提供する組織である TAUS (Translation Automation User Society)³ でも、同年、コンテンツプロファイリングに基づいて最適な評価モデルを構築する試みを開始している (O'Brien, 2012)。これら、翻訳品質評価の枠組み作りに向けたイニシアチブの始動が、機械翻訳品質評価に対する関心の高さを示している。

また、人間が訳文を読んで評価する方法では時間がかかる上に評価者間のばらつきが大きいことから、自動評価手法が各種開発されている (Pepineni et al. 2002, Snover et al. 2006, Turian et al. 2003)。しかし、「最も信頼できるのは人手による評価である」という考えは、機械翻訳システムの研究者の間でも根強い (Bojar et al. 2013)。品質が重要であるのは、人手による翻訳でも機械翻訳でも変わりはないため、翻訳の研究者・実践者にとっては興味深いテーマである。人手翻訳の評価方法については、翻訳教育の現場で工夫されたり、翻訳サービス提供の場において提案されたりし、活用されている。しかし機械翻訳文の品質評価についてもこれまでさまざまな手法が研究され実践されていることは、今回の参加者の間では知られていない可能性が高い。このことを踏まえ、各種評価手法 (人手、自動、タスク指向) についてスライドを使った解説を行うとともに、人手評価については全員で実践しその結果を共有するという、ハンズオン中心のワークショップを計画した。また、ハンズオン実施のため、参加者には、各自でパソコンを持ち込み、インターネット接続を確立するよう依頼した。

今回、ハンズオンを実施する人手評価手法は以下の 4 種類である。ハンズオンの例も合わせて紹介する。

2.1 エラー分類による評価

訳文に含まれるエラーを分類しカウントする方法。機械翻訳の普及以前より、人手翻訳の品質測定基準として使われていた。ローカリゼーションの分野では、ローカリゼーションに関する標準規格の作成・推進団体である LISA (Localisation Industry Standards Association, 2011年に活動終了) が提案した以下のエラー分類モデルが、人手翻訳、機械翻訳の両方を対象に広く使われている。

- Mistranslation
- Accuracy
- Terminology
- Language
- Style
- Country
- Consistency

今回のハンズオンでは、この LISA のモデルを基にし、ただしスタイルガイドを使用しないため、Style を除いた 6 項目を用いた。以下、ハンズオンに使用したスクリーンを示す。

図 1. ハンズオンスクリーン「エラー分類による評価」

エラー分類による評価

エラー分類評価 1/5

<原文>
What is a travel authorisation?

<機械翻訳文>
渡航認証とは何ですか?

<参照翻訳文>
渡航認証とは何ですか?

Mistranslation

Accuracy

Terminology

Language

Country

Consistency

戻る 前行

エラー分類による評価

機械翻訳文に含まれるエラーの数を、カテゴリごとにカウントします。
原文と機械翻訳文を比較し、各機械翻訳文に含まれるエラーを、規定のエラー分類法に基づいてカウントし、数を記入してください。
(なお、このハンズオンでは、用語などの確認のため、「参照翻訳文」が提供されています。)
全形でも文あります。

前行

2.2 Fluency & Adequacy 評価

訳文の Fluency (訳文の、文としての完成度) と Adequacy (訳文における、原文の情報の再現度) をそれぞれ評価する方法。Fluency は訳文だけを見て評価する。Adequacy は、原文と訳文を比較して評価する場合と、あらかじめ「参照翻訳文」(「正解」とされる翻訳

の例)を用意し、それと訳文を比較して行う場合がある。この評価方法は、1966年のALPACのレポート(ALPAC, 1966)での使用がよく知られており、同様の尺度は、Fluencyの代わりにIntelligibilityやClarityなど、またAdequacyの代わりにFidelityやAccuracyなどの呼び名を使って、今日でも機械翻訳の研究や実践に使われている。今回のハンズオンでは、それぞれ5段階で評価し、Adequacyの評価には参照訳文を使用した。

図 2. ハンズオンスクリーン「Fluency & Adequacy 評価」



2.3 複数の訳文の比較による評価 (ランキング)

通常は複数の機械翻訳システム間で性能を比較するために行われる。同じ原文を複数のシステムで翻訳させ、評価者は複数の訳文を品質の順にランク付けする。Fluency と Adequacy を使った評価法よりも評価者間でのばらつきが小さいなどの利点により、機械翻訳システムの研究の分野などで採用されている(Bojar et al. 2013)。ランキングであるため2文以上ならいくつでも比較することができるが、今回のワークショップでは簡単のため二者択一方式を採用した。

図 3. ハンズオンスクリーン「複数の訳文の比較による評価」



2.4 モノリンガル・ポストエディティング

この評価は2段階で行い、通常は各段階を別の評価者が担当する。第1段階では、原文を見ずに、機械翻訳結果である訳文を、意味が通るように編集する。編集しなくても意

味が通る場合は編集せず、文の意図がまったく読み取れない場合は「編集不可」とする。第2段階では、編集された訳文と原文を比較し、Adequacyを5段階で評価する。これにより、機械翻訳文を読んだ人が実際にどの程度原文の意味を理解できたかを判断することができる。「機械翻訳文に不自然さや文法的な誤りがあったとしても、読んだ人が正しく理解できるなら翻訳文としての役割は果たしている」という考えに基づく評価法である。

第1段階：

機械翻訳文を正しい日本語に修正して、「ポストエディティング後の文」の欄に記入してください。

図4. ハンズオン資料「モノリンガル・ポストエディティング」第1段階

1	機械翻訳文	どのように申請者は、アプリケーション上の間違いを修正することができますか？
	ポストエディティング後の文	

第2段階：

図5. ハンズオンスクリーン「モノリンガル・ポストエディティング」第2段階

3. 実施に向けた準備

参加者に関する予備知識を得るため、また参加者に翻訳テクノロジー評価に関する予備知識を得ってもらうため、さらにハンズオンの効果的な実施のため、以下の準備を行った。

3.1 事前アンケート

機械翻訳およびその評価に関する参加者の経験を知るため、簡単な事前アンケートを行った。質問と回答については5.1で詳しく述べる。

3.2 リーディングリスト

ワークショップではなるべく、講義時間は短く、ハンズオンの時間を長く取ることが望ましい。しかしハンズオンの効果を最大にするには、ある程度の予備知識が必要となる。このため、機械翻訳システムのさまざまな側面からの評価についてまとめられた文献（Hutchins & Somers, 1992）をなるべく事前に読んできてもらうよう通知した。またこれに加え、ハンズオンで使用する評価法について詳しく知りたい人や、機械翻訳品質評価における最近の動向を知りたい人向けに、リーディングリストを作成して配布した。

3.3 環境の準備

今回のワークショップは、Skype を使用して日本、米国、台湾の 3 会場を繋ぎ、お互いの様子を画面で見ながら、またハンズオンの結果を即時集計して画面で共有しながら進める予定であるため、安定した接続の確立や画面の見やすさなどが成功の鍵となる。テクノロジー面での準備については、青山学院大学他兼任講師の山田優氏に一任し、環境を整えた上で接続のテストなども実行してもらった。

3.4 Google フォームを利用したハンズオンインターフェイス

今回のワークショップにおけるハンズオンでは、単に参加者がそれぞれ自分で評価作業を体験するだけでなく、全員の評価結果を共有した上で評価に関するディスカッションを行うことを目的としている。それには 3 つの会場にいる参加者のハンズオン結果をスムーズに集計するシステムが必須である。このため各参加者に、インターネットに接続できるパソコンを用意してもらう必要があった。加えて、参加者が特別なプログラムをインストールする必要なく、またどのプラットフォームからも参加できることが望ましい。このため、無料で使用でき使い方が簡単なウェブベースのアプリケーション、Google ドライブの Google フォームを使用することにした。Google フォームでは択一式や自由記入形式のフォームを簡単に作成・共有でき、また各参加者の入力内容は自動的に集計して表にまとめられるため、全員が記入し終わった時点で即時にデータを共有できる。すべてのハンズオン課題をこの方式で実行できるよう、フォームを準備した。フォームの例は 2.1、2.2、2.3、2.4 に示したとおり。

4. 実施の概要

以下、ワークショップ実施の概要を示す。

日時：2013 年 11 月 16 日 10:00 am 開始（日本時間）

場所：立教大学 池袋キャンパス 16 号館 3 階 第 1 会議室

講師：立見みどり

参加者（全 24 名）：

立教大学 大学院 異文化コミュニケーション研究科学生 7 名＋教員 4 名

東京大学(大学院・学部)2名

台湾 輔仁大学通訳通訳大学院日本語科学生2名+スタッフ1名

モントレー国際大学通訳・翻訳学科 学生6名+教員2名

参加者の母語は、日本語、英語、中国語のいずれかであったが、参加者全員が日本語と英語の両言語を理解するとの前提に基づき、また参加者の半数以上が日本語母語話者であったため、ワークショップはすべて日本語で行った。評価に使用したテキストは、米国への旅行者向けの情報を掲載したウェブサイトから抽出した。これは文芸翻訳などと異なり意識が求められない分野であり、しかし翻訳や評価の際に専門的な知識を必要としない一般向けの内容であることから、機械翻訳の対象として、また品質評価の題材として適切であると判断したためである。すべて英文から和文に翻訳し、評価対象は和文のみとした。

内容：

機械翻訳の各種評価手法(人手評価、自動評価、タスク指向評価)を概観し、人手評価手法については演習を行う。

実際に評価を行ってみた上で、それぞれの手法で何がわかるのか、わからないのか、利点は何か、問題点は何かについて考える。

4.1 講師によるプレゼンテーション(1)

会場間のインターネット接続(Skype)の確認、立教、モントレー、台湾それぞれからの参加者の簡単な紹介、各自のパソコンからのインターネット接続の確認の後、講師によるプレゼンテーションを行った。内容は、翻訳テクノロジーや評価に関する全般的な基礎知識およびハンズオンのための説明である。

I. 各種翻訳テクノロジーの紹介

II. 機械翻訳システムの総合的な評価についての概説

III. 機械翻訳システムの翻訳品質評価の紹介

a. 人手評価

- ・ エラー分類
- ・ Fluency と Adequacy
- ・ ランキング
- ・ モノリンガル・ポストエディティング

b. 自動評価

c. タスク指向評価

d. ハンズオンの導入

4.2 機械翻訳品質評価ハンズオン

各評価法について、まず講師がスライドを使用してその概要を紹介し、実行方法を解説した。その後参加者が各自のコンピューターからハンズオン用のサイトを開き、評価を実行。各ハンズオンの演習が終わるごとに講師のコンピューターで結果を即時集計し、参加者と共有し、評価方法についての意見交換を行った。

4.3 講師によるプレゼンテーション (2)

全ハンズオンの終了後、機械翻訳評価に関する補足知識のためのプレゼンテーションを行った。内容は、コンピュータープログラムを使用した機械翻訳品質の「自動評価法」と、機械翻訳の出力物の品質を直接評価するのではなく、機械翻訳システムによる翻訳ワークフローの向上や機械翻訳されたテキストの実用性などを評価する、「タスク指向評価法」である。

IV. 自動評価法

- a. BLEU (BiLingual Evaluation Understudy) (Papineni et al. 2002)
- b. TER (Translation Edit/Error Rate) (Snover et al, 2006)

V. タスク指向評価法

- a. 後編集の時間や労力を計測
- b. 訳文の理解度を質問により判断
- c. 訳文に記述されたタスクを実行してもらい、その成功度を計測

終了

5. 結果・反響・反省点

5.1 事前アンケート

3.1 で触れた「事前アンケート」への回答者は 12 名で、質問と回答は以下のとおり。

質問 1. これまでに機械翻訳システムを利用したことがありますか？

回答

- ある : 9 名
- ない : 3 名

質問 2. 「質問 1」に「ある」と回答した方にお尋ねします。機械翻訳システムは役に立ちますか？ (ひとつ選択)

回答

- 常にある程度は役に立つ : 3 名
- 役に立つときと立たないときがある : 5 名
- これまで役に立ったことはない : 1 名

質問 3. すべての方にお尋ねします。機械翻訳品質の評価手法についてどの程度知っていますか？（複数選択可）

回答（重複あり）

- 人手による評価手法について、ある程度の知識がある：2名
- 人手による評価手法を使ったことがある：1名
- 自動評価手法について、ある程度の知識がある：3名
- 自動評価手法を使ったことがある：1名

参加者 24 名中回答者が 12 名であるため参加者の半数のデータしか集められていないが、回答者に限れば、その 3 分の 2 が機械翻訳システムを使用したことがあり、そのほとんどが機械翻訳システムはある程度は役に立つとしている。しかし機械翻訳品質の評価については、人手、自動とも、ある程度の知識があったのは回答者の 4 分の 1 以下であり、実際に評価を行った経験については、人手、自動とも回答者のほとんどが持っていなかった。

5.2 インターネット接続とディスプレイ環境

Skype 接続はワークショップ中に一度切断されてしまったが、それ以外は安定していてもお互いによく届いた。ただ、メイン会場である立教大学の設備では、講師のコンピューターの表示（スライドや集計結果など）、モニターの画像、台湾の画像がすべて 1 つのディスプレイに表示されるため、ディスプレイが混み合い、ハンズオンの集計などがしにくいことがあった。また他の会場でも同様に、他の 2 大学の様子がわかりにくかったようである。

5.3 ハンズオン

ハンズオンに予想以上に時間がかかり、やや時間が押してしまったため、ハンズオン後の集計結果の検討や意見交換に十分な時間を確保できなかった。事前に何度かシミュレーションをしたが、やはり機械翻訳品質評価は、慣れているといたないとで時間のかかり方が非常に異なることが改めて明らかになった。ハンズオンで実際に評価を行ってもらうことも重要なポイントだが、実践後に、各自が経験に基づく感想を述べたり意見を交換したりすることで、より理解を深め、今後役に立つ見も得られるため、ディスカッション時間を短縮せざるを得なかったのは残念であった。

評価方法の数を減らすこともできたが、今回のワークショップの目的のひとつは、機械翻訳品質評価を広く概観してもらうことであったため、4 種類を設定した。また、各評価方法で使用する文の数を減らすこともできたが、実際に機械翻訳システムを使用していれば日常的に出会うような品質のバリエーションを体験してもらうには、ある程度の分量は必要となる。将来、同様の方式のワークショップを行う際には、より目的を絞り込むことが望ましい。

ただし全体としては、少なくとも参加者が興味を持続しつつある程度の知識と経験を得る機会にはなったため、ハンズオンという手法自体は成功と言えるだろう。

5.4 ディスカッション

各ハンズオン後のディスカッションでは、各評価方法の長所、短所、その他の問題点や評価法改善の可能性について意見を求めた。評価結果をすぐに見ることができるため、評価者間でいかに判定が分かれるかなどを実感しながら評価方法について検討することができた。また、評価基準の解釈のぶれや、それぞれの評価法がどのような場合に有用であるのかなどについての意見交換があった。全般的な視点に加え、この場合はどう判断すればよいのか、などの各論としての質問も出された。特筆すべきは、モノリンガル・ポストエディティングの第1段階を比較的難なく実行できた参加者と、非常に困難に感じた参加者がいたことであった。機械翻訳の出力結果には、人間による翻訳物にはまず含まれないような、基本的な助詞の誤りや、不自然な表現などが現れることがある。それらは当然読む物に違和感を与えるが、特に機械翻訳に慣れない場合には、そのような文を読んだ上でさらに修正することは、大きな苦痛となり得ることが示された。

5.5 事後アンケート

ワークショップ後、参加者に任意でアンケートに回答してもらった。回答者は12名で、質問と主な回答は以下のとおり。

質問 1. 今回のワークショップは、役に立ちましたか？

回答

- 非常に役に立った：3名
- 役に立った：8名
- どちらとも言えない：0名
- あまり役に立たなかった：0名
- 役に立たなかった：1名

役に立った点、悪い点を具体的にお聞かせください。

- 評価方法を初めて経験したことが刺激的だった
- 評価方法の種類を知ることができたのがよかった
- 自分なりに機械翻訳で役に立ちそうなものを選びやすくなったのではないかと思う
- 機械翻訳のポストエディットをする際に、どれだけ時間がかかるのか、目安にできそう
- ハンズオンのスタイルが効果的だった。
- エラー評価では、人による翻訳でも何をもってして一つの「エラー」とするのか考えるいい機会になった

- 人的翻訳でも評価は難しい点なので、参考にしたい
- 評価方法の難しさと共に、良い翻訳についても考える良い機会になった
- 機械翻訳にとって良い翻訳とは何か、ということが一般的にあるいは業界内にどのような常識があるのかわかっていれば評価がしやすくなると思った
- 評価する時間をたいへん短く感じたが、時間がもっとあれば評価にブレがなかったのだろうか、という疑問が残った
- 参加者の英語ネイティブ・日本語ネイティブの識別をした上でハンズオンを行ない、データにどんな差が出るか知りたかった
- 実際にプロとして評価を行っている人の評価などを見ることができると、さらに役に立ったかもしれない

質問 2. 技術的環境（スクリーンの見え方、画像・音声品質など）は、どうでしたか？

回答

- 非常に良かった：0名
- 良かった：9名
- どちらともいえない：2名
- 悪かった：1名
- 非常に悪かった：0名

改善点などあれば、お聞かせください。

(回答欄記入なし)

質問 3. 三か所（東京、台湾、モントレール）を繋いで実施した環境について、どのように感じましたか。

回答

- 非常に良かった：3名
- 良かった：3名
- どちらともいえない：5名
- あまり意味がなかった：1名
- 意味がなかった：0名

良かった点、改善点などあればお聞かせください。

- 画期的な試みで、わくわくした
- 画像も音声もほとんど問題なかった
- 一部、聴きづらいところがあったが、3元コミュニケーションとしてはよくできていた

- ときどき画面に問題があったが、講義の進行には差し支えなかった。音声もクリアだった
- 諸事情が許せば、Ustream などを使って中継し、もっと幅広く、誰でも参加できるような形で実施できれば素晴らしい
- せっかく 3 箇所を結びお互いの顔も見える環境だったので、参加者がそれぞれ自己紹介をすればさらによかった。
- 席の作り方を工夫すればもう少しスクリーンが見やすかった
- 最初の設定に時間を取られなければ、全体的に良かった
- もう少し、他の地域とつながってる感があればよかった
- 繋いだ意味や利点がよくわからなかった

質問 4. 今後の要望（トピック、講師、講義フォーマット等）をお聞かせください。

- 今回は既に翻訳されたものに対して評価をしたが、翻訳テクノロジーを使って実際に翻訳を試してみるという作業もやってみたい
- 実際に翻訳現場（翻訳者やエージェント、開発者）でどのように評価方法が利用されているかという具体例を聴きたい
- 翻訳支援ツール関連の講義にも参加したい
- 講師も講義フォーマットもよかったが、ハンズオンで使用したテキストがあまり良くなかった
- 次回文芸翻訳ワークショップも楽しみにしている
- 映画字幕、翻訳と自己検閲について興味があるので、そのようなテーマの話が聴きたい

6. 考察と今後の展望

6.1 複数の国や地域を繋いで行うワークショップについて

アンケート結果の「画期的な試みで、わくわくした」に代表されるように、複数の国や地域をインターネットで接続して行うワークショップには、それだけでどこか心躍るものを感じる参加者も多いのではないかと推察される。翻訳に携わる人間にとって、異文化交流は常に興味深いものであり、インターネットを活用することで、居ながらにして複数の国や地域の参加者と意見交換できる機会が提供されることは喜ばしい。また、自国以外での翻訳研究や翻訳教育の場でどのようなことが行われているか、どのような点に関心が集まっているかを知る良い機会にもなる。しかし「もう少し、他の地域とつながってる感があればよかった」に示されるように、3か所の参加者がお互いに顔を見て議論するといった状況を作ることはできなかった。実際に別の会場の参加者と面と向かって話しているような臨場感を作り出すには、画面の大きさや数、映し方、着席のレイアウトなどを工夫する必要がある。また時差の関係から、参加する国や地域の組み合わせによっては時間設定が難しくなる。しかしこれらの問題が克服できれば、複数の国や地域を繋ぐ翻訳関連ワークショップは有意義なものであろう。

6.2 翻訳テクノロジーワークショップの活用

今回のワークショップの主な目的は、機械翻訳において品質というものがどのように考えられているのか、どのように評価されているのかについて、ハンズオンを通じて実際に体験しながら知ってもらうことと、各種評価法についてその問題点や改善点について議論することであった。そして、これらの点について、また人手翻訳の評価との類似点や相違点などについて、参加者から有意義な意見が多数出された。機械翻訳品質の最適な人手評価手法として確立された唯一の手法はない。「機械翻訳にとって良い翻訳とは何か、ということが一般的にあるいは業界内にどのような常識があるのかわかっていれば評価がしやすくなると思った。」とのアンケート回答にもあるとおり、一般に、人手による翻訳と機械翻訳では、期待される品質の種類やレベルが異なり、その違いを意識した評価が必要となる。今後、このようなワークショップを通じて、さまざまな立場から翻訳に携わる研究者・実践者の間で意見交換することが、より優れた評価基準や評価手法の確立につながるのではないかと。

しかしアンケートの質問 1 のさまざまな回答に見られるように、機械翻訳になじみのない参加者にとってこのワークショップは、評価法を検討する以前に、評価を通じて、機械翻訳についての知見を得たり、人手翻訳との違いについて考えたりする機会ともなった。ワークショップ中の意見交換でも、機械翻訳でできることとできないことについて理解が深まったといったような感想もあった。これらのことから、「評価」という作業は「学ぶ」ための道具としても有用なのではないかと感じた。今後機械翻訳結果の評価を実践する機会の有無にかかわらず、また評価方法の改善や洗練への貢献度とも関係なく、「訳文の評価」という作業に取り組むことで、システムティックな視点で機械翻訳に触れ、機械翻訳に関する理解を深められる可能性がある。今日、学生であれ教育者であれ実践者であれ、翻訳に携わっていれば、実際に自らが使用するかどうかは別として、機械翻訳に多少なりとも興味のある人は多いだろう。そのような人に、機械翻訳について知ってもらう方法としても、「評価」を主眼としたワークショップが有用である可能性が示された。

6.3 特定分野に特化したハンズオン型翻訳ワークショップの発展

ワークショップ参加者からも聞かれたことであるが、翻訳の分野は多岐にわたるため、翻訳の実践や教育に長年携わっていても、自分の専門とする分野以外のことはあまり知らないことが多い。しかしアンケートの質問 3 の回答からもわかるように、さまざまな分野における翻訳の現実について興味を持っている人もまた多い。筆者自身も第 2 回の文芸翻訳ワークショップに参加して感じたことだが、このような参加型・経験型のワークショップを通じて得られる知見は多く、また振り返って自分の専門分野の翻訳について考える機会にもなる。今後、さまざまな翻訳の分野から講師を招いてこのようなワークショップを開催することができれば、学習者、教育者、実践者ともに、さまざまな分野の翻訳について、ただ情報として知るだけでなく、少しでも体験する機会が得られる。そこから視野を広げ、自分の専門分野以外の翻訳についても理解を深めることができる。さらに、分野間

の情報共有や交流、共同研究などのチャンスも生まれることが期待できる。

7. まとめ

以上、翻訳テクノロジー評価を題材に 3 つの場所をインターネットで繋いで行ったワークショップの背景、準備、内容、実施、およびその結果を報告し、同様のワークショップの潜在的有用性について考察した。今回のワークショップは、普段あまり翻訳テクノロジーになじみのない翻訳学習者や教育者に、評価という作業を通じて機械翻訳に触れてもらう機会となったとともに、現在、翻訳テクノロジーの開発、研究、実践の場で新たな関心を集めている機械翻訳品質評価の手法や品質そのものの定義についても検討する機会となった。

日本における翻訳研究は、欧州などと比較してまだ歴史も浅く範囲も限られている。しかし翻訳環境に大きな変化が生じている今、学際的な視点で翻訳者養成に取り組むため、大学などの研究機関が大きな役割を果たすことが期待されている。国内において、また海外との間で、広く活発に情報交換できる機会を作ることもそのひとつである。インターネットを活用した国際ワークショップは、そのような場を提供する効果的な手法といえる。今後もさまざまなテーマで同様のワークショップを開催することで、分野を越えて翻訳学習者、実践者、研究者を繋ぐ場を作り出すことができるだろう。

【著者紹介】

立見 みどり(TATSUMI, Midori) 立教大学兼任講師および東洋大学非常勤講師、応用言語学研究者、実務翻訳者。Dublin City University (アイルランド) にて、機械翻訳の後編集に関する研究で博士号取得。連絡先: midori.tatsumi2@mail.dcu.ie

【註】

1. 詳細は EAGLES のウェブサイト (<http://www.ilc.cnr.it/EAGLES/browse.html>) を参照。(2014年7月21日)
2. 詳細は QTLaunchPad のウェブサイト “The New Goal of Quality Translation” (<http://www.qt21.eu/launchpad/>) を参照。(2014年7月10日)
3. 詳細は TAUS Dynamic Quality Framework のウェブサイト (<https://evaluation.taus.net>) を参照。(2014年7月10日)

【参考文献】

ALPAC (1966). *Language and machines: computers in translation and linguistics*. Division of Behavioral Sciences, National Academy of Sciences, National Research Council.
Bojar, O., Buck, C., Callison-Burch, C., Federmann, C., Haddow, B., Koehn, P., Monz, C., Post, M.,

- Soricut, R. & Specia, L. (2013). Findings of the 2013 workshop on statistical machine translation. *WMT 2013: 8th Workshop on statistical machine translation, Proceedings of the workshop*, August 8-9, 2013, Sofia, Bulgaria; pp.1-44.
- Hutchins, J., & Somers, H. (1992). *An introduction to machine translation*. (Chapter 9. Evaluation of MT systems). San Diego, CA: Academic Press
- O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation, *The journal of specialised translation*, issue 17 [Online] http://www.jostrans.org/issue17/art_obrien.php (July 5, 2014)
- Papineni, K., Roukos, S., Ward, T. & Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. In: *Proceedings of the 40th annual meeting of the association for computational linguistics (ACL)*, Philadelphia, pp. 311-318.
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L. & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In: *Proceedings of the 7th conference of the association for machine translation in the americas*, Cambridge, Massachusetts, USA, pp. 223-231.
- Turian, J.P., Shen, L. & Melamed, I.D. (2003). Evaluation of machine translation and its evaluation. In: *Proceedings of MT summit IX*, New Orleans, USA, pp. 386-393.