

言語処理学会テーマセッション

『文理・産学を越えた翻訳関連研究』開催報告

山田 優 藤田 篤 影浦 峯
 関西大学 情報通信研究機構 東京大学

1. はじめに

2016年3月の言語処理学会第22回年次大会(NLP2016@東北大学)¹では、『文理・産学を越えた翻訳関連研究』と題したテーマセッションが開催された。本稿では、同セッションの提案者として、その時の模様を振り返りながら、自然言語処理(NLP)と翻訳研究(TS)を取り巻く状況をまとめ、今後の研究と活動の方向性を提示する。

テーマセッションの目的は、自然言語処理研究者だけでなく、翻訳者や翻訳ツールの開発者、翻訳研究者が文理・産学の垣根を越えてニーズとシーズを共有することであった。各分野のプレイヤーに属する技術、資産、人材、問題意識は、他分野プレイヤーにとって新たなニーズやシーズとなりうる。同セッションを通して異なる分野のプレイヤーが交流し共同研究や有機的な連携が促進されることを期待した。

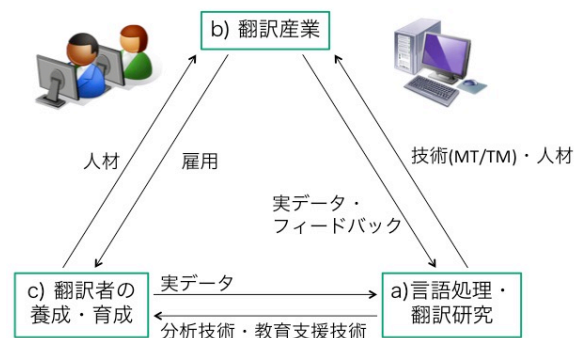
上記年次大会ウェブサイトのセッション E5、E6、E7にあるように、本テーマセッションでは合計12の発表があった。これらは以下の3つに大別できる(図1も参照されたい)。

- a) 自然言語処理・翻訳研究
NLPを中心とした情報処理・分析の研究、翻訳研究、機械翻訳(MT)や関連技術の進展に関与する研究者
- b) 翻訳産業
翻訳会社(LSP)、翻訳者などの実務者
- c) 翻訳者の養成・育成

大学のコース、職業訓練校の教育者

発表すべてが、きれいに分けられるわけではないが、このように図式化することで、プレイヤーの守備範囲と関係が把握できる。以下では、セッションでの発表内容を図1の範疇a、b、cに位置づけつつ、2者間ごとのインタラクション(すなわち三角形の各辺)に着目して、研究課題、連携の可能性、拓かれた問いとそれに対する議論を詳述する。

図1：各プレイヤーの関係図



2. テクノロジーの創出と運用：工学的な産学連携

まず図1三角形の右辺に着目する。ここには、a)自然言語処理/翻訳研究とb)翻訳産業の連携がある。平たく言えば、アカデミアと産業、いわゆる産学連携である。

a)自然言語処理/翻訳研究は、言語処理や翻訳に関する基礎研究を行うので、その技術は、少なくとも理

¹ <http://www.anlp.jp/nlp2016/>

屈の上では翻訳産業で応用できるはずである。また少し技術研究の範囲を広げて考えるならば、業務管理や翻訳準備（前処理等）への応用も考えられる。翻訳研究は、人間と機械のインタラクションやその文脈（社会的構造）を研究対象とする。つまり、a) 自然言語処理・翻訳研究にとっては、b) 翻訳産業は、技術の出口であり研究題材の場の一つでもある。これに対して、b) 翻訳産業は、翻訳を実践する現場であり、彼らにとって、a) 自然言語処理・翻訳研究はキモになる道具・資料となりうる。

産業側からの発表として、河野[4]は、MTを始めとする技術革新をいかに活用していくのが要になると主張した。例えば、翻訳メモリ（TM）を使った作業のうち、70%以上のセグメントではMT訳を下訳として利用しているという実情がある。ただし、業界内ではMTを活用して得られた効率性向上とその報酬に関して翻訳会社側とプロ翻訳者の間で十分な相互理解がなされておらず、翻訳者の間ではMTに対する否定的意見も少なくない。MTはまさに破壊的イノベーションであり、これからの時代の新しい顧客や価値観をもたらすものであるから、いま乗り越えなければならない壁も高い。一方で、ポストエディット（翻訳後編集、PE）を含めて、MT結果を人間（の翻訳者）が利用するものとみなす限り、人間支援という視点を重視したシステム開発も必要である。アカデミアにはこのような視点での研究開発が期待される。

プロ翻訳者の多くがMTに否定的な意見を持つ状況において、プロの翻訳者ではない人向けのCATツールを提案するのがヤラクゼン[7]である。主にポストエディットに最適化したMTとユーザインタフェースを開発し、漠然とした「翻訳品質」よりも「ユーザー満足度」を重視したツールの開発に務める。すでに1万人以上のユーザーを抱える同サービスは翻訳コーパスも蓄積しており、研究者との連携も期待される。

宮田ら[6]は、プリエディット（翻訳前編集）、すなわちMTによる翻訳を前提としたオーサリング支援を検討している。この研究では、起点言語の話者に、満

足のいくMT訳が得られるまで、原文を試行錯誤にプリエディットさせて、その工程を記録し、修正履歴の分析を通じてプリエディットの全体像と有用性を見出すことを試みている。

人間翻訳者の振舞いは、作業モード（例えば入力メソッド）により異なる。Carlら[2]は、キーボード入力による人手翻訳（HT from scratch）、ポストエディット（PEMT）、音声認識入力（dictation）の3つのモードを比較し、作業効率の観点から、音声認識入力での翻訳がPEMTに迫ることが示唆された。この他にも、PEにおける人間の負荷の指標として作業中の平均ポーズ時間が有用である（相関が高い）ことが示された[5]。この指標を、逆にMT品質の測定に応用できるのではないかという意見も交わされた。

発表が示すように、a) 自然言語処理・翻訳研究とb) 翻訳産業の連携は拡大しているが、まだまだ発展の余地がある。今後は翻訳産業界とアカデミアの更なる連携に期待したい。実データを用いて実世界の問題を解決するというスタイルの連携は、アカデミアにとっても、(若い) 研究者を呼びこむ意味で有益である。また、翻訳産業界のテクノロジーの高度化に伴い、SEあるいは社内開発のために理系人材の活用が活発化することも予想される。

3. 翻訳産業が求める人材と翻訳者の育成： 文系人材の市場

次に左辺のb) 翻訳産業とc) 実務翻訳者の養成・育成の関係を見る。上述したようにa) 自然言語処理・翻訳研究を前提としたテクノロジーの発展により、その進化に対応すべく、b) 翻訳業界で求められるスキル、能力、人材も変化している。

MTを始めとする技術の展開は、翻訳産業の現場だけでなく、翻訳教育およびその背景にある、翻訳者に求められる力とは何かという理論的な問題にまで影響を与えている。とりわけ、大学院レベルでの翻訳教育はあまりなされておらず、翻訳産業との結びつきもあ

まり強くない日本の大学では、翻訳教育カリキュラム全体の設計と同時にその中に翻訳関連技術をどう位置づけるかについても検討しなくてはならない状況に置かれている。これは、翻訳関連技術も考慮した見通しのよいカリキュラムを構成する観点からチャンスでもあるが、一方で、翻訳実務を考慮した基本的な翻訳教育実践がそれほど進んでいない状況でそうしたカリキュラムが仮に作られたとしても砂上の楼閣になりかねない危険性が存在しているということでもある。

また、翻訳者として生計を立てていくのは、理系のSE等が翻訳産業で働くのとは異なり、様々な意味でハードルが高い。そもそも職人芸的な要素が強い仕事であるのに加え、近年多様化が進む業務形態への対応も求められる。例えば、PEMTへの対応を考えた場合、そのための訓練をどのように教育カリキュラムに折り込むべきかは白紙の状態である。それ以前に、PEMTで生計を立てることができるという見通しを作ることが、まず翻訳業界に求められているとも言える。

教育側はこうした業界の実情をふまえた人材育成に迫られている。近年のISOの翻訳サービス規格（ISO 17100）の発行を契機に、国内でも専門職教育に対する必要性の意識が高まっている。立教大学を中心に「翻訳通訳教育研究会」が立ち上がり、専門職訓練を意図した学科・コースもようやく開設しつつある。また、翻訳教育実践を統合的にサポートするシステムの開発と利用を目的とするプロジェクトなども進められている。

カリキュラムの作成に際しては、すでに大学での翻訳者養成が社会に浸透している国・地域における取り組みが参考になる。例えば、モントレイ国際大学（現、ミドルベリー国際大学）における翻訳訓練プログラムでは、これまでは語学力とドメイン知識が中心であったが、現在はそれらに加えて翻訳管理能力やTMなどのツールなどのトレーニングが必須となっている。また翻訳産業界では、ローカリゼーション翻訳管理（マネージメント能力）に特化したプログラムの修了生に対するニーズが高まっているという。

現時点では、教員不足の問題、特に現状では b) 翻訳業界の実務翻訳者が c) 大学教育現場で教える人材になるためのコンピテンシーの整理、という問題が未解決である。欧米では両者が固く結びついている。日本でも同じような連携の実現が急がれる。

本セッションでは、業界の現場と教育についての問題が確認できたが、実際に b) 翻訳産業と c) 実務翻訳者養成が連携した発表はなかった。このことは、実践レベルで両者の連携があまり進んでいないという大きな問題を示している。

4. 翻訳教育におけるテクノロジーの活用： 文理融合・教育工学的アプローチ

3節の状況を想定して、最後に三角形の下辺である a) 自然言語処理研究・翻訳研究と c) 実務翻訳者養成の関係を見る。今回のテーマセッションでは、翻訳者の養成と翻訳関連技術、翻訳者に求められる力の関係をめぐる共同研究に関する発表も行われた。

影浦ら[3]は、産業翻訳のプロセスで標準的に採用されており、ISO17100でも前提とされているプロジェクト型翻訳のプロセスをエミュレートしつつ、翻訳の力を総合的に高めるための翻訳教育／学習システム「みんなの翻訳実習」の背景と理念を紹介した。ここでは、翻訳者の力として、狭義の「翻訳力」とプロジェクト管理やプロジェクト参加者間のコミュニケーションも含めた「翻訳者力」を区別し、システムが「メタ言語」の修得を通じた翻訳プロセスに対する反省的認識を促す「足場」を提供することで、翻訳力と翻訳者力を同時に涵養する見通しが示された。

上述の「みんなの翻訳実習」では、欧州の翻訳教育プログラムを通して開発された校閲カテゴリ（MeLLANGE）をベースに、新たに英日・日英翻訳への適用を想定した校閲カテゴリが設計された。豊島ら[8]は、実際の翻訳教育現場において校閲カテゴリを用いるための方法論的検討を行った。翻訳の学習者が作成した英日翻訳における校閲箇所のアノテーション

および校閲カテゴリ分類を通じて、校閲カテゴリを一貫して付与すること、ひいては校閲カテゴリが教育者と学習者の間の意思疎通のための「メタ言語」として利用しうることが示された。豊島らの調査をふまえ、山本ら[10]は、翻訳の初学者（学部学生）を対象とした翻訳演習講義において校閲カテゴリを用いた添削を繰り返し実施し、学習者の習熟状況の変化を分析した。その結果、校閲カテゴリを用いることによって反省的な認識が促進され、用語集に従うことや直訳調の翻訳を避けるべきであることは効果的に習得できたこと、一方で、原文の内容を正しく理解して過不足なく適切に目標言語に移す（誤訳をしない）ことや想定されるレジスタに合わせて訳すことなどは短期間では習得できないことが明らかにされた。これらは、翻訳実務段階での OJT や問題改善においても有用な知見と言える。

翻訳者を取り巻く環境や翻訳者に求められる要件の変化にいち早く対応すべく、山田ら[9]は、大学・大学院における翻訳テクノロジーに関する教育を促進するためのオンライン教材「翻訳テクノロジー」を開発している。本発表では、機械翻訳(MT)、MT後編集(PE)、翻訳メモリ(TM)を題材として取り上げ、動画およびクイズ形式の確認問題からなるeラーニング教材を提供していることが紹介された。山田らは、実際の講義においてこの教材を用いるとともに、翻訳テクノロジーに関する学習者の意識についても調査した。そして、MTを単体で用いる際の期待値の低さに対して、欧州で標準的な翻訳フローになりつつあるPEを学ぶことにより、道具としてのMTの可能性に気づく者も多いことを明らかにした。

5. 研究発表と討論を通じて拓かれた問い、暫定的な解の案

以上、言語処理学会第22回年次大会のテーマセッション『文理・産学を越えた翻訳関連研究』の模様を、各プレイヤー間の連携と問題・可能性について整理・

言及しながら俯瞰した。翻訳産業がテクノロジーの高度化とともに高効率化を目指す上で、文理・産学の連携が今後も重要であることが再確認できた。

本論考でプレイヤー別に整理して見えたこととして、特筆すべきは、b) 翻訳産業と c) 実務者養成の連携、すなわち翻訳教育における産学の連携が、実際にはかなり少ないということである。一方、文理・産学の共同研究自体は実施されてきていること、「教育」分野においても文理が連携した活動は存在することが参加者間で共有された。

今後、例えば「深層学習によって取って代わられる業務もあるのか」という議論や課題も出てくると予想される中で、翻訳教育における産学連携は急務と思われる。

6. 今後の我々の活動予定

我々は、今回の言語処理学会テーマセッションのような企画を、今後も継続的に実施したいと考えている。まずは、2016年9月に日本通訳翻訳学会(JAITS)第17回年次大会(@同志社大学)で本活動の近況と今後の課題について発表する。2017年3月に筑波大学で開催予定の言語処理学会第23回年次大会でも2度目のテーマセッションを開催したい。2017年の秋には名古屋でMT Summitが開催されるので、ここでもワークショップなどを行って、産学の連携強化等を図りたい。

いずれにしても、状況としては海外先行、日本後発であるので、世界の最先端での新規性を求めるよりは国内の意識向上がねらいである。

謝辞

同テーマセッションの提案者である武田珂代子先生・立見みどり先生(立教大学)にも、本稿の内容を確認いただいた。

参考文献

- [1] M. Carl, I. Lacruz, M. Yamada, and A. Aizawa. Measuring the translation process. 言語処理学会第

- 22 回年次大会, 発表論文集, pp. 957–960, 2016.
- [2] M. Carl, I. Lacruz, M. Yamada, and A. Aizawa. Comparing spoken and written translation with post-editing in the ENJA15 English → Japanese translation corpus. 言語処理学会第 22 回年次大会発表論文集, pp. 1209–1212, 2016.
- [3] 影浦峯, M. Thomas, A. Hartley, 内山将夫. みんなの翻訳実習における「足場」と翻訳力・翻訳者力～みんなの翻訳第 6 報～. 言語処理学会第 22 回年次大会発表論文集, pp. 857–860, 2016.
- [4] 河野弘毅. 機械翻訳ビジネスの成功要件あるいは MT で金持ちになる方法. 言語処理学会第 22 回年次大会発表論文集, pp. 965–968, 2016.
- [5] I. Lacruz, M. Carl, M. Yamada, and A. Aizawa. Pause metrics and machine translation utility. 言語処理学会第 22 回年次大会発表論文集, pp. 1213–1216, 2016.
- [6] 宮田玲, 藤田篤, 内山将夫, 隅田英一郎. 機械翻訳向け前編集の事例収集と類型化. 言語処理学会第 22 回年次大会発表論文集, pp. 869–872, 2016.
- [7] 坂西優. 非翻訳者向け CAT ツール「ヤラクゼン」の仕組みと課題. 言語処理学会第 22 回年次大会発表論文集, pp. 961–964, 2016.
- [8] 豊島知穂, 田辺希久子, 藤田篤, 影浦峯. 翻訳教育での利用を意識した翻訳エラー分類体系の再構築. 言語処理学会第 22 回年次大会発表論文集, pp. 861–864, 2016.
- [9] 山田優, 立見みどり, 武田珂代子. 「翻訳テクノロジーを学ぶ」: 教材オンライン化の現状と展望. 言語処理学会第 22 回年次大会発表論文集, pp. 953–956, 2016.
- [10] 山本真佑花, 田辺希久子, 藤田篤. 翻訳学習者の学習過程におけるエラーの傾向の変化. 言語処理学会第 22 回年次大会発表論文集, pp. 865–868, 2016.