

# AAMTセミナー開催報告

Report of AAMT Seminar



東京大学 大学院情報理工学系研究科特任研究員

**中澤 敏明**

2010年京都大学大学院情報学研究所知能情報学専攻博士課程修了。博士（情報学）。現在は東京大学大学院情報理工学系研究科客員研究員。機械翻訳の研究に従事。

## 1 はじめに

一般社団法人アジア太平洋機械翻訳協会（Asia-Pacific Association for Machine Translation、以下 AAMT）では機械翻訳の発展・啓蒙の一環として、MT に関するトピックの情報配信を目的として 2022 年度より定期的にセミナーを開催している<sup>1</sup>。本稿の著者は AAMT の理事であり、AAMT セミナー委員として AAMT セミナーの企画運営を行なっている。本稿では 2022 年度に開催された 3 回と、2023 年 5 月に開催された第 4 回について開催報告を行う。

## 2 第 1 回：第 17 回 AAMT 長尾賞 / 第 9 回 AAMT 長尾賞学生奨励賞 受賞記念講演<sup>2</sup>

第 1 回 AAMT セミナーは、AAMT 長尾賞および AAMT 長尾賞学生奨励賞受賞者の記念講演として参加費無料で開催した。第 17 回長尾賞は「ニュースを対象とした日英機械翻訳システムの研究開発」というタイトルで NHK の日英機械翻訳開発プロジェクトに贈られた。このプロジェクトは英語ニュースの制作を支援する日英機械翻訳の開発を目的としており、総合テレビ特設ニュースにて AI 翻訳の英語字幕をつけてインターネットでライブ配信を行うなどの実用化も達成している。

第 9 回長尾賞学生奨励賞は 2 名に贈られた。一人目の受賞者は“Expanding the Applicability of Machine Translation”というタイトルで東北大学（現

NTT コミュニケーション科学基礎研究所）の森下睦氏であった。ヨーロッパ言語と英語間では ParaCrawl と呼ばれる大規模な対訳コーパスを作成するプロジェクトが EU の助成を受けて行われているが、日本語は含まれておらず、一般に無償公開されている大規模対訳コーパスは存在しなかった。そこで森下氏は Web から大量のデータをクロールし、クリーニングなどを行うことで 2100 万文対を超える日英対訳コーパス J Para Crawl を構築し、公開した。また構築した対訳コーパスを用いて、世界的に有名な機械翻訳の評価型ワークショップである WMT において日英の機械翻訳タスクを創設し、日英翻訳をよりメジャーな言語対とすることに貢献した。

二人目の受賞者は「学習時と推論時における入力データの特徴の違いを考慮したニューラル機械翻訳モデルの学習手法」というタイトルで東京工業大学（現 NHK 放送技術研究所）の美野秀弥氏に贈られた。この研究は翻訳モデルの学習時と翻訳時のデータの特徴の違いに起因する翻訳精度の低下を低減することを目的としたもので、2 つの手法を提案している。一つ目は原言語側・目的言語側それぞれのデータの特徴と、それらの関係に着目した複数のドメインタグを用いたニューラル機械翻訳である。もう一つは文脈情報として目的言語側の前の文を利用したニューラル機械翻訳において、学習時に参照訳と機械翻訳結果を適度に混ぜ合わせる手法である。どちらも翻訳精度の向上に寄与することが示されている。

1 <https://www.aamt.info/event/seminar>

2 <https://www.aamt.info/event/seminar/20220928>

### 3 第2回：理工系の英語論文執筆における課題とAI自動翻訳ツールの活用<sup>3</sup>

第2回 AAMT セミナーはタシケント国立工科大学副学長・教授であり工学博士の西山聖久氏より「理工系の英語論文執筆における課題とAI自動翻訳ツールの活用」というタイトルでご講演いただいた。西山氏は数百名を超える工学部学生の英語論文執筆の指導を行ってきた経験を通じ、機械翻訳を積極的に活用した英語論文執筆の新手法を紹介した。

西山氏は英語論文を執筆する上で行き詰まってしまうことの原因は、まずそもそも書きたいことの概要が頭の中で整理されていないことが問題であると指摘し、論文の概要を作成するためのテンプレートを提案している。また理工系の英文ライティングにおいては3C（Clear：明確、Correct：正しい、Concise：簡潔）であることが求められるが、これを実現するためのポイントとして（1）誤りがない英文を書く、（2）できるだけ少ない語数にする、（3）名詞を意識して扱う、の3つを挙げ、機械翻訳を利用しながらそれぞれのポイントを実現するための方法を紹介した。

### 4 第3回：海外投資家を引き付ける英文IRコストや時間をかけない自動翻訳の活用<sup>4</sup>

第3回 AAMT セミナーはジャパンエレベーターサービスホールディングス株式会社 常務執行役員 広報・IR室長の西村賢治氏より「海外投資家を引き付ける英文IRコストや時間をかけない自動翻訳の活用～日本語のみの開示で公平・公正・適時という開示の原則に沿っていると言えるか～」というタイトルでご講演いただいた。IR活動は企業の価値を公正に評価してもらうために非常に重要なものであり、投資家はこの情報を元に判断を行う。特に海外の投資家向けに、日本語だけでなく英語でも情報発信を行うことが重要である。しかしながら、現時点でまだ英訳されていない情報がたくさん存在するとのことであった。

西村氏は機械翻訳を利用することで、まだ英訳されていない情報も迅速に、効率的に英訳できると指摘した。機械翻訳が間違えたり、うまい訳が出てこないのは、原文の日本語が不完全（主語が曖昧であるなど）であるこ

3 <https://www.aamt.info/event/seminar/20230111>

4 <https://www.aamt.info/event/seminar/20230309>

とが大きな要因であると述べ、「誰が」「何をやる」のかを明確にすることで機械翻訳の出力の質を向上させることができると提案した。

### 5 第4回：ポストエディットの実態<sup>5</sup>

第4回 AAMT セミナーは AAMT 会長であり、情報通信研究機構フェローの隅田 英一郎氏より「ポストエディットの実態～英日 PE に従事した翻訳者のアンケート結果より」というタイトルでご講演いただいた。ポストエディット作業は翻訳者にとって重労働である、効率性が下がるなどの意見もある中で、講演では大規模なポストエディットプロジェクトに携わった翻訳者 277 名にポストエディットに関するアンケートを行い、回答を得られた 178 名の結果を紹介した。

アンケート結果では、PE で作業効率性が下がると答えた人は全体の 3% しかなく、83% の人は良くなると感じたことや、平均して生産性が 50% 程度向上すること、PE をやるにつれて徐々に作業効率が向上することなどが報告された。また PE が成功するためには受注側と発注側で目標を一致させることが重要との指摘があった。

### 6 おわりに

本稿では過去 4 回の AAMT セミナーの開催報告を行った。今後も定期的に AAMT セミナーを開催する予定で、機械翻訳とその周辺のトピックを幅広く取り扱う予定である。今後の AAMT セミナーにご期待いただき、是非とも参加をご検討いただければ幸甚である。

5 <https://www.aamt.info/event/seminar/20230531>