

# AAMT

Asia-Pacific Association for Machine Translation

# Journal



March 2018

No.67

アジア太平洋機械翻訳協会

# 目 次

巻頭言：	AIは通訳者にとって代わるのか？ .....	武田 珂代子.....1
シンポジウム報告 MT Summit XVI：全体総括 .....		中岩 浩巳.....3
シンポジウム報告 MT Summit XVI：Research Track .....		黒橋 禎夫.....5
シンポジウム報告 MT Summit XVI：Users Track .....		山田 優.....7
シンポジウム報告 MT Summit XVI：Technology Showcase .....		長瀬 友樹.....9
シンポジウム報告 MT Summit XVI：JTF Workshop .....		菊田 直歩.....11
シンポジウム参加報告：	MT Summit 2017 での参加・発表報告 .....	山本 ゆうじ.....13
シンポジウム参加報告：	MT Summit XVI 報告 .....	工藤 竜広.....14
シンポジウム報告：	MT Summit XVI:Technology Showcase XTM as a perfect platform for post editing Machine Translation.....	目次 由美子.....16
ワークショップ報告：	The 7th Workshop on Patent and Scientific Literature Translation (PSLT 2017) .....	須藤 克仁.....18
研究報告：	Baidu Machine Translation.....	Zhongjun HE ,Hua WU ...20
シンポジウム報告 MT Summit XVI：招待講演“Neural translation technologies and futures” .....		中岩 浩巳.....27
シンポジウム参加報告：	アジアでの機械翻訳サミットに参加して.....	井佐原 均.....29
シンポジウム参加報告：	第27回 JTF 翻訳セッション「ニューラル機械翻訳は翻訳プロセスをどう変えていくか ー最近の機械翻訳技術と利用に関する動向ー」 .....	島津 美和子.....31
シンポジウム参加報告：	Women in Localization Japan 第13回イベント報告.....	目次 由美子.....34
会員の広場（法人会員）	.....	株式会社ウィズウィグ.....38
会員の広場（個人会員）	.....	河目 志津加.....40
事務局からのお知らせ：協会活動報告（2017年9月～2018年2月） .....		AAMT 事務局.....42
編集後記	.....	宇津呂 武仁.....46

## CONTENTS

Foreword:	Is AI going to replace professional interpreters? .....	<i>K.Takeda</i> ..... 1
Symposium Report MT Summit XVI: .....		<i>H.Nakaiwa</i> ..... 3
Symposium Report MT Summit XVI: Research Track Report .....		<i>S.Kurohashi</i> ..... 5
Symposium Report MT Summit XVI: Users Track Report.....		<i>M.Yamada</i> ..... 7
Symposium Report MT Summit XVI: Technology Showcase Report.....		<i>T.Nagase</i> ..... 9
Symposium Report MT Summit XVI: JTF Workshop Report.....		<i>N.Kikuta</i> ..... 11
Symposium Report:	MT Summit 2017 presentations .....	<i>Y.Yuji</i> ..... 13
Symposium Report:	Report on MT Summit XVI .....	<i>T.Kudo</i> ..... 14
Report:	MT Summit XVI: Technology Showcase XTM as a perfect platform for post editing Machine Translation.....	<i>Y.Metugi</i> ..... 16
Workshop Report:	The 7th Workshop on Patent and Scientific Literature Translation (PSLT 2017) ...	<i>K.Sudoh</i> ..... 18
Research Report:	Baidu Machine Translation.....	<i>Z.HE, H.WU</i> ..... 20
Symposium Report:	"Neural translation technologies and futures" .....	<i>H.Nakaiwa</i> ..... 27
Symposium Report:	My participation in Machine Translation Summit in Asia .....	<i>H.Isahara</i> ..... 29
Symposium Report:	Report on the 27th JTF Translation Festival: "How neural machine translation (NMT) systems transform translation processes: Recent trends in MT technologies and their uses" (AAMT session) .....	<i>M.Shimazu</i> ..... 31
Symposium Report:	Report on the 13th Women in Localization Japan event.....	<i>Y.Metugi</i> ..... 34
AAMT Members:	.....	<i>WysiWyg Co., Ltd.</i> ..... 38
AAMT Members:	.....	<i>S.Kawame</i> ..... 40
AAMT Activities:	AAMT Activities (from September 2017 to February 2018).....	42
Editor's Note:	.....	<i>T. Utsuro</i> ..... 46

## AI は通訳者にとって代わるのか？

武田 珂代子

立教大学異文化コミュニケーション学部

### 1. コミュニケーションの特徴と通訳

この1年間、「AI によって通訳者の仕事はなくなりますか？」という質問を何度となく投げかけられた。グーグルやマイクロソフトなどの巨大企業が AI ベースの機械翻訳・自動通訳サービスの開発と展開に力を入れているというニュースが次々と流れる中、「今さら通訳者を養成してどうするんですか？」とでも言いたげな態度を感じとってしまう。

私はテクノロジーの専門家ではない。しかし、米国の翻訳通訳大学院で学び、フリーランス通訳者として四半世紀働いてきた。「シンギュラリティ」を予言したレイ・カーツワイルの通訳も何度かしたことがある。また、米国および日本の大学院で通訳者の養成に 20 年近く携わってきた。こうした実践経験および研究活動に基づき、上記の質問に対しては、今のところ次のような返事をしている。

「短い答は『ノー』。買い物をしたり、道を聞いたりといった短いトランザクショナルな通訳では AI が役立つかもしれないが、そういうものはもともとプロの通訳者が対応してきた領域ではない。話し言葉によるコミュニケーションの特徴や通訳者の複雑なタスクを考えると、近い将来 AI がプロの通訳者にとって代わることはないだろう。」

音声言語の通訳は、話し言葉によるコミュニケーションを仲介する行為だ。話し言葉によるコミュニケーションは、発話者間の相互行為によって共同構築される動的なプロセスであり、状況的行為である。つまり、発話者間のやりとりは自発的で即時性を伴う。また、プロソディや強調など話し方に関わる準言語的要素や、ジェスチャーや表情などの非言語要素、さらに、話す

スピード、ポーズ、言い淀み、言い直しなど時間的要素も関わる。これらの要素が全て絡み合って、発話の意味が形成されるのだ。また、話し言葉には、書き言葉とは異なる口語的表現、アクセント、方言、敬語、ニュアンスなども関わってくる。

こうした話し言葉によるコミュニケーションの特徴に AI ベースの自動通訳 (machine interpreting) はうまく対応できるのだろうか。まず、原発言の音声認識をする段階で、アクセント、発音の誤り、準言語・非言語要素などへの対応という難題がある。次に、音声認識で産出されたテキストを機械翻訳する際、口語表現やさまざまなレジスター、敬語を適切に訳すことができるのかという課題がある。最後に、機械翻訳が産出したテキストを音声として合成するとき、原発言のトーンを再現することは可能だろうか。また、ユーザーはモノトーンのロボットの発話に長時間耐えられるのだろうか。

以上のような課題はこれまでも多くの人たちが議論してきたと想像するが、何よりも大事なこととして、倫理的な課題がある。自動通訳の質に対する説明責任や賠償責任は誰が負うのだろうか。医療や法廷などの場で自動通訳の誤訳が悲惨な結果につながることもあるかもしれない。テキストを扱う機械翻訳であれば、誤訳や形式の問題などは人手によるポストエディティングで修正することもできるが、自動通訳では品質保証をおこなう責任の所在という倫理的問題も考慮していかなければならないだろう。

通訳の実践に関するこれまでの実証研究で、通訳者が実際には訳出行為以上のことを行っていることは証明されている。通訳者が発言の順番やタイミングを調整したり、コミュニケーションに齟齬が生じればその

解決に向けて、あるいは参加者の「フェイス」を守るために「介入」したりすることは実際にあるのだ。通訳者はまた、コミュニケーションの参加者間の力関係やオーディエンスが誰なのかについて常に注意を向け、それに適した敬語やレジスターを選びながら通訳をしている。

通訳行為におけるこうしたプラグマティックな側面に自動通訳がどれほど対応できるのか。それが、私たち通訳者、通訳教育者、通訳研究者が自動通訳に対して抱いている疑問（と懐疑心）の最たるものだといえよう。

## 2. AI と通訳の今後

以上の議論から、AI を応用した自動通訳はノンクリティカルな短いトランザクションベースのコミュニケーションでは役立つ可能性が大いにあると考える。特に、学習者が少ない言語への対応には社会的意義もあるだろう。

今後、プロ通訳者の領域に自動通訳が食い込む可能性があると考えられるのは、あらかじめ用意されたスクリプトを忠実に読むだけのスピーチの通訳だ。スピーカーがスクリプトを部分的に飛ばしたりアドリブで話したりしない限り、また、オーディエンスがモノトーン的な合成音声に長い時間耐えられる用意があるのであれば、通訳者が自動通訳にとって代わられる可能性はあるのかもしれない。

また、米国の司法の場で、デポジション（証言録取）や法廷での証言でよく使われるリアルタイム速記録をベースに機械翻訳を利用する可能性もゼロではないだろう。ただ、証拠能力や説明責任に関する厄介な議論がともなうのであまり魅力的とは言えない。

もう一つ考慮すべき領域は、あらたな通訳の形態として欧州を中心に研究されているリスピーキング（re-speaking）である。テレビ番組やビデオで、聴覚障がい者のための字幕を作成する際、原発言を字幕に適したものに言い直す（リスピーク）作業を「言語内

通訳」として捉え、実証研究が行われている。このアプローチを適用するとモノリンガルの言語内通訳者が原発言を音声認識や機械翻訳がしやすいものに言い直し、自動通訳の精度が上がる可能性も出てくるだろう。

東京オリンピックを2年後にひかえ、訪日外国人が増加し続ける中、日本政府は「言語の壁」を乗り越えるコミュニケーション技術の開発に大きな期待を寄せている。そうした国からの期待や支援を背景に、技術者たちは観光場面に対応するレベルを超え、スピーチの自動同時通訳の開発にも取り組んでいると聞く。通訳の実践者、教育者、研究者として、自動通訳開発の今後の動きに注目し続けたい。

## 全体総括

中岩 浩巳

アジア太平洋機械翻訳協会 会長、JTF 監事、名古屋大学

MT Summit は、機械翻訳に関する様々な最新の成果の発表や、招待講演などが行われる国際会議である。この会議の特徴は、機械翻訳の研究開発者に加え、翻訳者、翻訳会社、クライアント、翻訳支援ツールメーカー等の機械翻訳に関連する関係者が一堂に会し、機械翻訳の研究開発に関する最新の成果に関する研究開発者による発表だけではなく、機械翻訳技術を翻訳業務で実際に活用する方法など、翻訳会社や翻訳者などの利用者による発表も行われる点である。その意味では、本会議に参加すると、機械翻訳に関するあらゆる観点からの動向を把握することができるといえる。

本会議は、アジア太平洋機械翻訳協会（AAMT）の上位組織である機械翻訳国際連盟（IAMT）が主催で、欧州機械翻訳協会（EAMT）、アメリカ機械翻訳協会（AMTA）、及び、AAMT が交代で、欧州、北米、アジア・太平洋と開催場所を変えて、隔年で開催している。今まで AAMT が運営した会議は、日本・箱根（第 1 回：1987）、日本・神戸（第 4 回：1993）、シンガポール（第 7 回：1999）、タイ・プーケット（第 10 回：2005）、中国・アモイ（第 13 回：2011）だったが、2017 年に行われた第 16 回の MT Summit は、24 年ぶりに日本・名古屋市の名古屋大学のシンボリック建物である豊田講堂にて 2017 年 9 月 18 日から 9 月 22 日まで開催された。筆者は、運営組織である AAMT 会長として、また、同会議の組織委員長及び現地委員長とし同会議の運営全般に携わった。

日本は、2020 年の東京オリンピック・パラリンピック開催を契機とした、来日観光客の急増などに伴うインバウンド需要の増大、円安傾向などを契機とする様々な国際ビジネスの増大、また 2016 年 11 月にニュー

ラル機械翻訳技術が Google 翻訳に導入されたことに伴う機械翻訳による訳文品質の向上等、今まで以上に機械翻訳が注目されているタイミングであったこともあり、今回の MT Summit は盛況のうちに終了した。



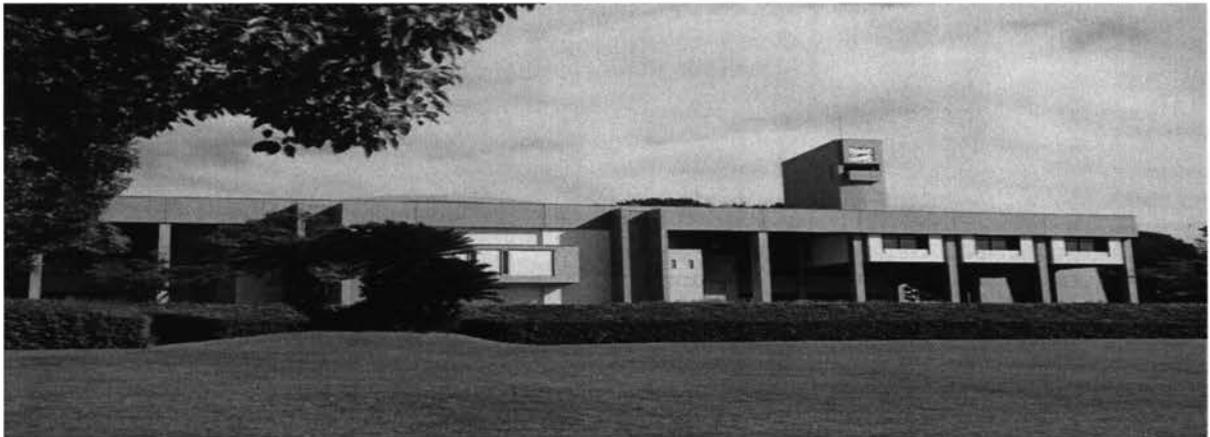
本会議には、約 300 名の機械翻訳関係者（研究者、機械翻訳ソフト業者、翻訳業者、翻訳者等）が参加した。海外からの参加者が 100 人弱、日本の参加者が 200 人強であった。特に今回は、経済産業省及び総務省に加え、翻訳業界を代表する団体である日本翻訳連盟（JTF）及び製品やサービスの使用説明を扱う専門家の団体であるテクニカルコミュニケーター協会（TC協会）の後援を頂いたこともあり、機械翻訳の研究開発者とともに、機械翻訳を利用している方もしくは今後利用を検討している方の参加が多くあった。

本会議は 2 つの一般セッション（Research Track 及び Commercial MT Users and Translators Track）を中心とし、他に招待講演、パネルディスカッション、技術展示を行う Technology Showcase で構成されていた。ニューラル機械翻訳技術が多くの実サービスに適用された最初の MT Summit であることもあり、様々なセッションにおいて、研究開発面だけではなく、利用技術に関しても最新の成果が多数報告された。一

般セッションの2つTrack及びTechnology Showcaseが企画したセッションに関しては、個々の委員長の記事に譲るとして、ここではそれ以外のセッションについて概説する。

本会議初日には、運営委員会が企画した招待講演セッション“NMT Technologies and Their Futures - Baidu, Google and Microsoft”が行われた。ここでは、ニューラル機械翻訳のサービスを世界的に展開しているバイドゥ、グーグル、マイクロソフトの翻訳サービス展開や新技術についての紹介があった。特にバイドゥは、本講演初公開の自動翻訳機能を搭載したモバイルWi-Fiルーター「百度智能Wi-Fi翻訳機」のデモが

あった。またこのセッションの前には機械翻訳に関する長年の貢献をたたえるIAMT Award of Honorの表彰式があり、今回はIAMT、AAMT元会長であり現産業技術総合研究所人工知能研究センター長の辻井潤一氏が受賞した。また、会議初日9/18及び最終日9/22には、機械翻訳に関連するワークショップ3件及びチュートリアル3件が行われた。そのうち9/22午後のワークショップはJTF関西セミナー委員会が企画運営し同時通訳付きで実施したもので、国内外の専門家が講師となり、ニューラル機械翻訳の最新の活用法等活発な議論が行われた。本ワークショップの詳細は別稿を参照のこと。



会場となった名古屋大学豊田講堂



## Research Track

黒橋 禎夫

京都大学 大学院情報学研究科 教授

MT Summit の本会議 3 日間のうち技術に関するプログラムを Research Track とよび、私と香港科技大学 (HKUST) の Pascale Fung 教授が Program Co-chair を務めた。

Program Committee は MT の最先端の研究開発を担っている各国の研究者・技術者約 50 名に依頼した (一覧は会議のホームページに記載されているのでここでは省略)。

Research Track では論文の一般募集を行い、Program Committee 3 名による査読に基づき 26 件の論文を採択した。26 件の論文発表は 16 件の口頭発表と 10 件のポスター発表とし、口頭発表は以下の 7 セッションのプログラム構成とした。

- Enhancement of NMT I
- Enhancement of NMT II
- Enhancement of SMT
- Hybrid Method
- Domain Adaptation
- Data Construction
- Evaluation and Cognitive Model

なお、口頭発表セッションは Users Track と合わせて常時 3 並列で、3 並列のうち 1 セッションまたは 2 セッションが Research Track という構成であった。

これらのセッションのタイトルを見て頂ければ分かりますとおり、ニューラルネットワークによる翻訳 (Neural Machine Translation; NMT) の発表が多く見られた。

また、NMT と SMT のハイブリッド方式や、SMT の改良、ドメイン適応、データ構築、評価、さらには認知的モデルなど、機械翻訳に関する多様な発表が行われた。



さらに、Research Track では 2 件の招待講演を行った。まず、本会議 2 日目朝一のプレナリーセッションとして、University of Maryland の Marine Carpuat 教授に「Semantic and Stylistic Divergences in Machine Translation」というタイトルで講演して頂いた。機械翻訳では、SMT にしても NMT にしても対訳コーパスを用いて学習を行うが、対訳コーパスそのものの質についてはほとんど議論されることがない。Carpuat 教授の講演では、TED コーパスを対象として、対訳コーパスの中に必ずしも意味的に等価といえない対訳文が少なからず存在することを具体例をまじえて紹介し、それらを多言語テキストの含意関係認識器を用いて取り除くことにより、対訳コーパスのサイズが半分になっても翻訳の質が向上することを示した。さらに、文体についても対訳コーパス中に多様なものが混在しており、形式性の観点でコーパスを整理することにより、やはり翻訳精度が向上することを示した。現状では多くの研究者がそのまま利用している対訳コ

ーパスの中身に対して議論を深めた発表であり、会議参加者から大変好評であった。なお、前者は ACL2017 のワークショップ、後者は EMNLP2017 で論文発表されている内容で、興味を持たれた方はそちらを参照して頂きたい。

2 件目の招待講演は、本会議 3 日目朝一のプレナリーセッションで、NICT の隅田英一郎博士に

「Social innovation based on speech-to-speech translation technology targeting the 2020 Tokyo Olympic/Paralympic Games」と題する講演を頂いた。NICT では、2020 年の東京オリンピック・パラリンピックへ向けて活発な研究開発が行われており、その包括的な紹介を中心とする講演であった。具体的には、英語の Wikipedia 記事をアジア各国の言語に翻訳し構文タグ付与を行う Asian Language Treebank の構築、JST/京都大学と共同での WAT (Workshop on Asian Translation) の運営、Translation Bank の構想等について紹介があった。また、NICT の翻訳エンジンに基づく実証実験として、富士通による医療現場での実験や、Logbar によるウェアラブル音声翻訳デバイスの実験が impressive なデモビデオによって紹介され、聴衆を魅了した。最後に「No MT, no life」という強いメッセージで講演が締めくくられ、NICT の気合いが感じられる講演であった。



全体を通して、ニューラル翻訳によって劇的に進展した機械翻訳に対するさらなる技術向上や実利用への熱い期待、熱気が感じられる会議であった。と技術の観点からは感じるのであるが、翻訳者がどう感じておられたのか、Users Track の報告を読ませて頂くのが楽しみである。

## Users Track

山田 優

関西大学 外国語学部・外国語教育学研究科 教授

Users Track は、MT の応用に関する研究発表の場で、翻訳研究者、翻訳会社、実務翻訳者による口頭発表が中心になる。Mark Seligman 氏 (Spoken Translation) と私が Program Co-chair を努め、他 23 名の Program Committee に査読を依頼した。

MT Summit XVI は色々な意味でひと味違った。特にユーザトラックは、それを強く感じた。

まず、日本開催ということで、ローカル色が強くできたことだ。ユーザトラックの口頭発表は 20 件であったが、そのうち日本からの発表は 4 件あった。国際会議において、日本の研究者が発表することは珍しくないが、4 件ともが日本の翻訳業界で実践に携わる方々による発表であったことは、非常に意義がある。これを契機に、業界や実務者から世界に発信する研究に期待したい。

次に、今回の MT Summit がニューラル機械翻訳 (NMT) の本格的な実用化以後、はじめてであったことである。そのため、NMT への関心はとても高かった。口頭発表でも NMT に関するものは半数近くを占めていた。MT 利用が効率化につながるの、もはや自明な事になりつつある一方で、NMT の台頭により、機械翻訳を如何にしてコントロールするのかが、新たな課題として議論されていた。NMT の出力には、予測不能な訳出が比較的頻繁に含まれることが知られている。これらをユーザ側でどのように見極めていくのかが今後の MT の応用と活用に大きく左右するに違いない。

このような流れから、NMT (および AI) の影響を、少し幅広く捉え、我々人類にとってどのような恩恵や脅威をもたらすのかを真剣に考えようという目的で、

本会議の初日にはユーザトラックの特別企画のパネル・ディスカッションが催された。『MT と AI』と題したこの企画では、「アシロマの原則」を基にして、NMT をとりまくステークホルダー達が意見を交わした。パネリストには、自然言語処理の研究者 Marine Carpuat 教授 (University of Maryland)、開発者の Chris Wendt 氏 (マイクロソフト)、翻訳教育現場から Tony Hartley 特任教授と武田珂代子教授 (立教大学)、そして業界を代表して Olga Beregovaya 氏 (AMTA 会長、Welocalize) が登壇した。NMT の時代だからこそ、このような分野を超えた人たちによる議論とコラボレーションがどれだけ重要であるかを再確認できた企画であった。



また、本会議 2 日目には SAP 社の Chris Pyne 氏を招いて、同社内での MT ワークフローについて、特にプロ翻訳者の適合性の観点から現実に即した現場の実態の話聞くことができた。

このように MT の応用研究分野は、特に NMT の実用化を契機とした今回の名古屋大学での MT Summit で、確立した研究分野に成長したという印象を受けた。筆者は今回の準備の過程で、これまでの MT Summit で数多くのポストエディットや機械翻訳の応用研究に

関するワークショップを組織してきた Sharon O' Brien 氏(Dublin City University)と Michel Simmard 氏(National Research Council Canada)と頻繁に連絡し合う機会を得た。彼等は、自然言語処理の研究者と翻訳研究者、それと実務者研究者と一緒に発表や議論できる場を提供しようと、過去の MT Summit 等で多くのワークショップを開催してきた。しかし、今回はワークショップを行わなかった。理由は次の通りだ。

「自分たちの役目は終えた。我々が目指していた『MT 応用研究』は、翻訳研究および自然言語処理において、独立した研究分野として認知されるようになった。これまでのように別立てでワークショップを開催しなくても、MT Summit のユーザトラックがその役割を果たしてくれる。」これが、先述したパネル・ディスカッションを彼等が企画した動機にもなっていたわけで、今回の MT Summit がひと味違うものであった理由である。そして、この事実と、次回の開催地がアイルランドの Dublin City University (Sharon O' Brien 氏の所属機関)である行われることは、決して無関係ではない。翻訳研究と MT 開発と実践は、今後、これまで以上の協力関係を築いていくことを、今回の会議を通して強く感じた。

## Technology Showcase

長瀬 友樹

株式会社富士通研究所 人工知能研究所 主管研究員

Technology Showcase は、一般セッション、ポスターセッションとともに本会議を構成するプログラムのひとつであり、プロトタイプやデモシステムを実際に動作させながら機械翻訳技術やツールの説明を行う展示主体のセッションである。本会議二日目の 9 月 20 日の午後 2 時半から 5 時まで、豊田講堂の中二階のギャラリーにて開催された。



### 1. 出展企業、団体

国内外の企業や団体から 16 件の展示が行われた。出展者の内訳をカテゴリー別でみると、企業が 8 件と最も多く、大学が 3 件、国の研究機関などその他の団体が 5 件であった。比較的実用化に近いテーマが多いこともあり、一般セッションやポスターセッションに比べて企業の存在感が強くなっている。国別には、イギリス、中国、アメリカ等海外からの出展もあったが、

約半数が日本の企業または団体による展示であった。マイアミで開催された前回の MT Summit で 2 件あった政府系団体からの出展は、今回は見られなかった。

### 2. 出展内容

16 件の出展内容は、プロダクトの機能紹介から学術的内容まで多岐にわたる。出展件数が最も多かったのは、翻訳会社や企業内翻訳者のためのシステムやツールの展示であった。XTM、memoQ、Memsources、八楽、Edinburgh 大学等から、機械翻訳と連携可能な最新の翻訳管理システムやプラットフォームの説明とデモが行われた。連携対象として特定の翻訳エンジンに固定せず、翻訳対象や条件に合わせて翻訳エンジンを選択して使える仕組みを取り入れているシステムが主流となっている。

会場を回って強く印象に残ったのは、機械翻訳と音声技術とをつなぎ、人と人のリアルタイムの情報伝達やコミュニケーションを支援するシステムの展示である。NICT は英語のプレゼン内容をリアルタイムに日本語字幕にしてスクリーンに表示するシステムを、Microsoft はプレゼン内容を利用者の母国語に翻訳して個人のスマホにリアルタイムに配信するサービスを実演していた。音声認識の精度は良好で技術的に実用化に近いところまで来ていると感じた。また富士通はハンズフリーで使える医療機関向けの音声翻訳端末のデモを行った。

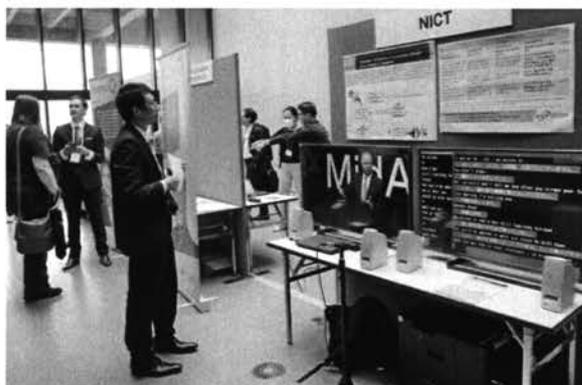
言語リソースの収集や共有については、今後さらにホットなトピックになることが予想される。これに関連して、東芝、豊橋技科大から、複数の利用者や組織の間で翻訳リソースを共有するための枠組みについて

展示が行なわれた。

学術的な展示では、Institute of Scientific and Technical Information of China (ISTIC)から NMT (ニューラル機械翻訳)の未知語の問題を PBSMT (フレーズベース統計的機械翻訳) 技術との組合せによって解決したシステムについて、また、JST からは科学技術論文データベースを用いて構築した高精度な NMT エンジンについての出展があった。NHK は機械翻訳技術を用いて通常の日本語表現を簡単な表現に自動的に書き換える技術のデモを行った。このほか、日本の翻訳会社である知財コーポレーションは、実際に導入効果のあった自社の機械翻訳のデモを行い注目を集めていた。

### 3. 会場と参加者の様子

Technology Showcase の開催中は他のセッションは行われなため、MT Summit 参加者のほぼ全員が Technology Showcase に足を運んだと思われる。特に前半は、切れ目なくデモ対応しなければならないほどの盛況を呈していた。会場スペースに余裕があり、隣の展示ブースとの間隔も十分とれていたことから、人の滞留が起きることなく快適な環境で展示を行うことができた。



### 4. おわりに

今回の Technology Showcase が応用の視点から機械翻訳の現状を理解するための一助となれば、筆者ら

Technology Showcase Committee として幸甚である。

今回展示された 16 件の展示が、次回ダブリンで開催される MT Summit までにどのような発展を遂げるのか、2 年後の Technology Showcase がとても楽しみである。

## JTF Workshop

菊田 直歩

一般社団法人日本翻訳連盟 広報委員会

規模の大きな学会では、本会議の前後にチュートリアルやワークショップを併設することが多い。チュートリアルは、技術やトレンドについて分かりやすく講義するもので全体的な動向把握に役立つ。一方、ワークショップでは第一線で活躍する専門家が集まり、ある特定のテーマについて掘り下げた視点から発表や議論を行う。

JTF ワークショップでは、言語産業において機械翻訳がどのように受容されているかをテーマに、機械翻訳の技術がすでに普及している欧米からベストプラクティスを学ぶべく 3 名のパネリストを招いた。また、日本でも NMT の登場で機械翻訳の導入に向けた機運が高まっていることを受け、日本のパネリスト 2 名が最新の現状を報告した。

さらには、ワークショップ冒頭に機械翻訳の導入について技術的な側面から解説する講座を設け、ワークショップ全体を通して機械翻訳の活用を積極的に検討する開かれた場とした。

ワークショップの口火を切ったのは、NICT の内山将夫氏。機械翻訳エンジンとしては NMT が現在最良の選択肢であること、また自動翻訳の品質向上には、アルゴリズムの改良、ドメイン適応とカスタマイゼーション、大規模対訳コーパスの構築、さらには機械翻訳エンジンと人間との間で訳文評価をフィードバックさせていくエコシステムの構築が重要との認識を示した。



続いてパネリスト 5 名による発表に移り、LinkedIn の Dillinger 博士による招待講演が行われた。

博士は一貫して、AI が翻訳者を駆逐するという単純な話ではないこと、そうした視点から脱却し発想の転換が必要なこと、人間の知と機械の知能とを融合させ二者間でどう共通知を作りあげていくかが問われていることを説いた。特に、ポストエディットは、機械翻訳を人間が編集して訳文を完成させるため、結局は機械と人間との分業であり、本当の意味でのハイブリッド翻訳ではないと鋭く切り込んだ。翻訳の未来は、機械と人間とが互いに得意・不得意を補完し合うだけでなく、人間が機械を制御し、インタラクティブに機械に学習させていく適応型の機械翻訳が切り拓くとした。

続いて欧州の機械翻訳エンジン開発ベンダーである Pangeanic の Herranz 氏と KantanMT の Shterionov 氏が登壇した。Pangeanic、KantanMT はともに、世界でも最大規模の翻訳需要を抱える欧州委員会と手を組み、2019 年 2 月までに行政を機械翻訳エンジンと連結させる巨大プロジェクトに参画している。

Herranz 氏は、ニューラルネットワークにおける深層学習の原理からエンコーダ・デコーダモデルにおけるアテンションの役割までを分かりやすく解説し、出力の改善には繰り返し学習させることが重要であると示した。

Shterionov 氏は、機械翻訳を使って産業翻訳の生産性をあげる 3 要素について述べた。1 つめはドメイン適応などによる機械翻訳エンジンの制御性、2 つめはエンジン学習やポストエディットの効率性、3 つめは翻訳環境のセキュリティ面をあげた。なかでも、従来の SMT に比べ NMT では制御性が低下していると指摘し、用語設定や前処理・後処理ルールの適用によるエンジン制御が商用ニーズで鍵を握るとした。

続いて登壇したポストエディット東京の河野弘毅氏 (JTF 理事) は、機械翻訳導入コンサルタントとして翻訳会社側の目線に立ち、機械翻訳を使って生産性をあげるには、翻訳者の積極的な関与を引き出すポストエディターの採用、機械翻訳に適した文書の選定、ポストエディットに対応する品質保証プロセスの適用、この 3 要素が肝要だと語った。特に、品質評価の枠組みが 3 要素に共通して重要とし、評価指標および評価基準をいかに設定するか、評価プロセスをどのように構築するかについて具体的に示唆した。

SDL ジャパンの佐藤弦氏は、翻訳業界で衝撃をもって受け止められた NMT の登場だが、セキュリティ面や価格設定、評価方法、用語指定など、依然として課題が多いことを指摘し、文書によって機械翻訳を適用できる場合もあれば、人間翻訳が適している場合もあることを強調した。その上で、発注者側は用途によって翻訳方法を振り分ける視点が必要とし、翻訳会社や翻訳者にはサービスや対応力の多様化が求められると述べた。



パネリストによる発表後は、パネルディスカッションの時間が設けられた。ワークショップの開始から終了まで 4 時間半に及ぶ長丁場だったが、会場は終始熱気に包まれ、開発者、研究者だけでなく、翻訳者、翻訳会社と多様な参加者を得た。

なお、パネリストの発表資料は一部を除いて下記 URL からダウンロードできる。ご興味をお持ちの方はぜひアクセスしてほしい。

<https://sites.google.com/view/jtfmt/english>

## MT Summit 2017 での参加・発表報告

山本ゆうじ

AAMT WG3 UTX チーム リーダー

AAMT 主催により、機械翻訳に関する国際会議 MT Summit 2017 <<http://aamt.info/app-def/S-102/mtsummit/2017/>>が 2017 年 9 月 18 日から 22 日の間、名古屋大学で開催されました。日本での MT Summit の開催は 24 年ぶりとなります。近年の機械翻訳への関心の高まりを反映して、各国からの参加者数は、当初の想定を上回る盛況となりました。期間中は、多数のチュートリアル、ワークショップ、発表が行われました。技術ショーケースでは最近注目を集める音声翻訳機器のデモも行われました。また興正寺で開かれたバンケットも、各国の参加者に好評でした。

報告者は、商業機械翻訳ユーザーおよび翻訳者トラック委員会の一員として参加し、また 2 件の発表を行いました。1 件は単著、もう 1 件は共著です。

1 件めは "Terminology-Based Post-Editing of Neural MT Using the Structured Glossary Data Format, UTX" (「構造化用語データ形式 UTX を使用した、ニューラル機械翻訳の用語ポスト エディット」) です。ニューラル機械翻訳は、訳文の自然さには優れていますが、訳抜けが発生することも多く、用語レベルでの一貫性には課題があります。本発表では、特許庁から提供された用語データを AAMT が策定した UTX 用語集形式 <<http://www.aamt.info/japanese/utx/>>に変換して、特許文書を Google のニューラル機械翻訳で翻訳した結果を、翻訳支援ツール Trados および Memsource、また用語ツール Xbench を用いて用語チェックしました。このように、共有・再利用しやすい用語データを活用することで、ニューラル機械翻訳の欠点を補うことができます。

2 件めは、Honda R&D Americas の CAT ツールリーダー、Saemi Hirayama 氏との共著で "Rule-based

MT and UTX Glossary Management – Honda's Case Dealing with Thousands of Technical Terms" (「ルールベース機械翻訳と UTX 用語管理—ホンダでの多数の専門用語を扱うケース」) です。

ルールベース機械翻訳はニューラル機械翻訳に押されていますが、訳文の流暢さより、専門用語の正確性や機密保持を重視する企業では現在も活用されています。本発表では、実際に UTX 用語集形式を活用して、八万語の専門用語を新旧の機械翻訳システム間で移行した事例が紹介されました。

また、一つの原因語に対して複数の訳語が対応している用語を、UTX 用語集形式で管理することも紹介されました。

例:「整合会」が correlation meeting や coordination meeting と訳される。

本発表では、用語データが人間翻訳者と機械翻訳の両方に役立つこと、UTX 用語管理が Honda R&D で有効であったことが述べられました。また企業での機械翻訳の活用ではニューラル機械翻訳が唯一の選択肢ではなく、ルールベース機械翻訳でもユーザーの満足度が得られていることが報告されました。企業でのこのような事例紹介は稀であり、非常に貴重な事例です。

上記の発表スライドはどなたでも以下から無料でダウンロードできます。

"Proceedings, vol. 2: Commercial MT Users and Translators Track"

<<https://goo.gl/oDLoHH>>

## MT Summit XVI 報告

工藤 竜広

株式会社サン・フレア 商品保証部

## 1. カンファレンスの概要

2017年9月18日(月)～9月22日(金)、機械翻訳国際連盟(IAMT)、アジア太平洋機械翻訳協会(AAMT)、名古屋大学大学院情報科学研究科主催で第16回MTサミット(MT Summit XVI)が開催された。1993年に第4回MTサミットが神戸で開催されてから24年ぶりに日本で開催された。同サミットは機械翻訳に関する研究発表、ワークショップ、招待講演、パネルディスカッションなどが行われる国際カンファレンスである。2016年末にGoogle社がGoogle Neural Machine Translation(Googleニューラル機械翻訳:GNMT)を採用したことにより、機械翻訳の品質が急激に向上し、ニューラル機械翻訳(以下、NMT)が各界から注目を集めるようになった。今回はNMTを筆頭に機械翻訳の技術、運用、翻訳通訳技術の教育を含めて、熱心な研究発表、ワークショップ、ディスカッションが行われた。参加者数は300名、参加国は20か国に及んだ。

9月18日は、KantanMTのDimitar Shterionov氏によるワークショップと、Microsoft社のChris Wendit氏によるワークショップが行われた。それぞれ自社システムを利用して、NMTの実装(Shterionov氏)とMicrosoft Translation Hubを使った音声翻訳の実演デモ(Wendit氏)に興味深かった。9月18日には豊田講堂において“MT and AI: Probing Near and Medium-Term Impacts”と題してパネルディスカッションが行われた。パネリストのCh. Wendt(Microsoft)、O. Beregovaya(Welocalize)、W. Hua(Baidu)、T. Hartley、武田珂代子(立教大学)の諸氏が、NMTや人工知能(artificial intelligence:AI)を

使った翻訳・通訳産業の可能性と問題点について議論をかわした。NMTとAIの発展に期待を寄せるIT企業各社と機械翻訳の品質そのものに懸念を抱くアカデミズムとの、NMTに関する姿勢の違いが興味深い内容だった。同日午後には、“NMT Technologies and Their Futures – Baidu, Google and Microsoft – Practical Machine Translation”と題して、H. Kazawa(Google)、W. Lewis(Microsoft)、W. Hua(Baidu)の諸氏がGoogle、Microsoft、中国の百度(Baidu)のNMT技術の解説や音声機械翻訳の技術に関する招待講演を行った。注目を集めたのは百度とGoogleのNMT技術の解説と開発中の新技術の概要説明だった。今回のMTサミットでは話題の中心がNMTであっただけに、会場からも多くの質問が寄せられた。

20日午前には、“Semantic and Stylistic Divergences in Machine Translation”と題してM. Carpuat氏(メリーランド大学)による招待講演が行われた。本講演は特に機械翻訳が抱える問題点について論じたものである。同日午後には、長年にわたって機械翻訳を実案件で利用してきたSAP社CEOのCh. Pyne氏による招待講演が“Introduction of MT into Industrial-scale Translation Workflows with Translator Acceptance”と題して行われた。

21日には“Social Innovation based on Speech-to-speech Translation Technology Targeting the 2020 Tokyo Olympic/Paralympic Games”と題して、隅田英一郎(NICT:情報通信研究機構/UCRI:ユニバーサルコミュニケーション研究所)の招待講演が行われた。NICTの音声翻訳技術を活用して東京オリンピック/パラリンピックのインバウンド事業拡大に努める官民を含めた取り組みを本講演では紹介した。

招待講演の他にも、数多くの口頭発表（oral presentation）や、テクノロジーショーケース、ポスター発表が行われた。口頭発表は、“Research Track”と“Commercial MT Users and Translators Track”の2トラックに別れ、機械翻訳の技術分野に関する研究発表がメインとなる前者と、機械翻訳の運用メソッドに関する研究および翻訳者通訳者の教育に関する研究を扱う後者とで多くの聴衆を集めた。研究者、企業、個人翻訳者通訳者の垣根を越えて、発表者と会場参加者とが熱心に質疑応答する姿が印象的だった。NMTの研究発表が多いのは当然だが、ルールベース機械翻訳や統計ベース機械翻訳や、これらで蓄積されてきた技術や知見をNMTで活かす工夫をこらした研究も多いことが印象的だった。また機械翻訳技術を教育現場で活かそうとする研究もみられた。

## 2. 日本における機械翻訳

最終日 9 月 22 日には、“Machine Translation Acceptance among the Language Industry”と題して、日本翻訳連盟（JTF）のワークショップが行われた。M. Dillinger 氏（LinkedIn）、M. Herranz 氏（Pangeanic）、D. Shterionov 氏（KantanMT）、内山将夫氏（NICT）、河野弘毅氏（ポストエディット東京）、佐藤弦氏（SDL ジャパン）の諸氏が登壇し、NMTの技術向上と生産性向上に関わる提言や、NMTが抱える問題点とその解決手段について発表した。引き続き登壇者によるパネルディスカッションが行われ、会場を巻き込んで議論がかわされた。本ワークショップの目的は、機械翻訳導入が日本よりも進んでいる欧米翻訳企業から第一人者を招いて、機械翻訳利用の最先端の事情を聞くとともに、機械翻訳導入にあたっての業務に関する知見を深めるというものだった。ワークショップの参加者は、欧米の機械翻訳利用の最前線について生の声を聞くとともに、実際の機械翻訳導入にあたって取り組むべき課題について有意義な手がかりを得られた。

## 3. 補足

なお、最終日には、JTFのワークショップのほかに、“Controlled Language Helps You Improve Comprehensibility and Translation Quality”と題して中村哲三氏（エレクトロスイスジャパン）による制限言語を利用した翻訳品質向上に関する紹介と、AAMT/Japio 特許翻訳研究会による「第七回特許・技術文書翻訳ワークショップ（Patent and Scientific Literature Translation）」が開催された。日本語と英語の制限言語を活用して翻訳品質を向上させるために中村氏が提唱する手法が興味深い。また「特許・技術文書翻訳ワークショップ」では、特許翻訳を専門とする翻訳会社 MK 翻訳事務所の梶木正紀氏が登壇し、事業の方針として機械翻訳に注力し、今後はポストエディットに対応したビジネスモデルを構築すると話されたことが印象的だった。実案件を扱う翻訳会社が今後どのようなかたちでビジネスを展開してゆくのかこれからも見守ってゆきたい。

なお、9 月 20 日夜には名古屋市興正寺の大書院にて晩餐会が開催された。優雅な日本庭園を備えた寺院の大寺院で、各国の参加者が酒食をともにして、楽しく歓談した。

今回の MT サミットは、2019 年にアイルランドのダブリンで開催が予定されている。

以上

## MT Summit XVI: Technology Showcase XTM as a perfect platform for post editing Machine Translation

目次 由美子

XTM International Ltd.

2017年9月に名古屋大学にて盛大に開催された MT Summit XVI では、「Technology Showcase」と題した展示会もプログラムされ、弊社 XTM International Ltd. も出展させていただいた。9月20日（水）の午後はこの展示会のみが執り行われ、展示会場は多くの参加者の熱気で溢れた。

弊社は「XTM Cloud」という名前の TMS (Translation Management System、翻訳管理システム) を展示した。日本の産業翻訳界で十分に浸透していると考えられる「翻訳支援ツール」の代表的な機能である翻訳メモリや用語集管理、さらに DTP 作業を軽減させるための機能を有し、なおかつワークフロー管理、ユーザ管理、コスト管理、メッセージング管理などを兼ね備えている。世界規模を意識した Web ベースの Cloud ソリューションであり、大量の翻訳を効率良く展開し、迅速なターンアラウンドを実現する機能性や、作業を委託する機能などにおける拡張性の高さから「企業向けシステム」とも呼ばれている。

この XTM Cloud では、もちろん機械翻訳エンジン（以下、MT エンジン）との連携も可能である。2017年12月現在、12種類の MT エンジンとの連携に対応している。ユーザ側で入手済みの API キーや独自にビルドした MT エンジンがある場合、連携に必要な情報を登録するのみで XTM での利用が可能となる。また、連携対象の MT エンジンの種類は今後も増加する予定であり、詳細は弊社 (sales@xtm-intl.com) までお問い合わせいただきたい。

今回の Technology Showcase では、特に CrossLang の MT エンジンと Google Translate との連携を実演した。CrossLang (<http://www.crosslang.com>) は弊社のテクニカル パートナー企業であり、翻訳オートメーションを専門として 2002 年にベルギーで設立された。MT エンジン開発も業務の 1 つとしているが、システム インテグレーションやコンサルティングも請け負っており、Adobe 社のコンテンツ管理プラットフォームソリューションである Adobe Experience Manager と XTM を連携させるための Connector も共同開発している。

XTM Cloud のインスタンスで使用する MT エンジンはあらかじめシステム コンフィギュレーションとして設定しておく。翻訳プロジェクトを作成する際、プロジェクトごとに使用するエンジンやオプションを調整することもできる。翻訳プロジェクトのテンプレートを用意しておけば、プロジェクトを作成するたびに同じオプションを選択する必要はなくなる。

たとえば、CrossLang の MT エンジンを使用する場合、翻訳対象のソース文書を XTM に取り込む際、MT エンジンに訳を自動生成させてターゲットセルに挿入しておくこともできる。この場合、XTM Editor で必要とされる作業は「翻訳」ではなく、「ポストエディット」となり、MT エンジンが生成・挿入した訳が適正であるかを確認し、必要に応じて修正するという作業を実施することになる。

ソース文書内に翻訳メモリ (TM) のマッチが検出された場合、TM を優先させることもできる。完全一致のみは TM を優先させるのか、あいまいマッチや文書内の重複も TM マッチを優先させるかを選択できる。

または、MT エンジンに生成させた訳例をあらかじめ解析時にターゲットセルに挿入はせず、XTM Editor での編集時に TM マッチとともに訳例として提示させることも可能だ。この場合、翻訳者は原文を見ながら、提示される TM マッチや MT マッチのなかから最善と考えられる訳例を流用することができる。もちろん用語集の機能も併用できるため、登録されている訳語の使用を確実にできる。

Google Translate を使用する場合、言語コンビネーションが対応していれば、XTM でもニューラル機械翻訳を利用できる。XTM Editor 上でターゲットセルごとに自動的に MT マッチを挿入させることも可能。MT エンジンや API によって、動作には相違が見られるので、MT エンジンを選択する際に考慮したい。

CrossLang と Google Translate の MT エンジン両方をプロジェクトに設定して、併用することも可能である。CrossLang の MT 訳例と、Google Translate の訳例を一緒に表示させることができる。最適な MT エンジンを選ぶために一時的にそのような利用をすることは可能である。しかしながら、ソース文書やドメインに合わせて最適な MT エンジンを選択するのであれば、翻訳プロジェクトごとに異なる MT エンジンを選択し、TAUS DQF プラグインを利用して、プロジェクトごとの生産性（時間単位での翻訳処理分量）や、修正の度合いなどを比較することを推奨したい。

No.	英語(米語)	日本語	ステータス
	Sample file for Machine Translation	Sample file for Machine Translation	1 MT
		あいまいマッチ 93% - 承認済み	
1	Sample file for Machine Translation	機械翻訳のためのサンプル	✓
		CrossLang MT 機械翻訳のサンプルファイル	✓
		Google NMT 機械翻訳のサンプルファイル	✓

## The 7th Workshop on Patent and Scientific Literature Translation (PSLT 2017)

須藤 克仁

奈良先端科学技術大学院大学

### 1. 開催概要

AAMT/Japio 特許翻訳研究会の活動の一環として、2005 年の第 1 回から数えて 7 回目となるワークショップを第 16 回機械翻訳サミットの会議最終日である 9 月 22 日（金）に開催した。前回 2015 年に引き続き特許・技術文書翻訳ワークショップ（Workshop on Patent and Scientific Literature Translation）と題し、特許・論文等技術文書全般の翻訳について広く扱うこととした。Workshop Co-chair は宇津呂武仁先生（筑波大）・須藤の両名が務め、網川隆司先生（静岡大）に Publication Chair をご担当いただいた。プログラム委員として特研究会委員及び 6 名の国内外の研究者にご協力いただき、技術講演論文の査読を依頼した。

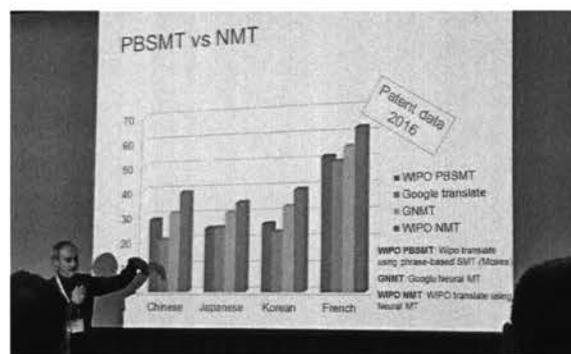
また、知財関係公的機関、翻訳者／翻訳会社関係者、機械翻訳研究者に招待講演を依頼した。技術論文講演に関しては、本会議研究論文トラックの投稿期限及び採否発表スケジュールが延長されたこともあったが、投稿は 4 件にとどまり、査読の結果 3 件を採択した。

また、ワークショップ当日の参加者は 30 名程度で、前回 2015 年マイアミで開催されたときと同程度であったが、半数以上が日本からの参加者であったと見受けられた。

### 2. ワorkshop 報告

午前のセッションは招待講演 2 件であった。

1 件目の招待講演は Bruno Pouliquen 氏（World Intellectual Property Organization: WIPO）で、前回の PSLT 2015 に引き続いて WIPO の機械翻訳シス



Pouliquen 氏の講演の様様

テムについてご紹介いただいた。今回はフレーズベース統計的機械翻訳（PBMT）を基本として、データ選択、多言語化、事前並べ替え等の前処理等についてのものであったが、今回はその後のニューラル機械翻訳（NMT）への技術動向の変化を捉えた、特許翻訳向け NMT についての講演であった。NMT は PBMT に比べモデルのサイズが非常に小さくできること、非常に流暢な訳文が生成できること等特許翻訳においても大きなアドバンテージがあり、実際 WIPO の新しい特許向け NMT システムにおいても従来の PBMT システムを大きく上回る翻訳精度を達成しており、同じく NMT を採用する Google 翻訳よりも優れていること等が示された。

2 件目の招待講演は梶木正紀氏（MK 翻訳事務所）で、2017 年 2 月の JTF ジャーナルの Google NMT 特集において翻訳者の視点から見た NMT について寄稿をされている等、我々にとって有益なご見解が伺えると考えて講演を依頼した。梶木氏はすでに機械翻訳を使った翻訳ワークフローの転換を近々に進めようとしていて、そのために Google や Microsoft の NMT を用いた翻訳後編集作業のテストを行い、担当した翻訳

者からも 20-30%の作業効率の改善への効果が期待できるといった反応を得たことについてご紹介いただいた。

午後のセッションは招待講演 1 件と技術講演 3 件であった。

招待講演は Andrew Finch 氏 (Apple) で、前職の NICT 在籍時の統計的翻字に関する技術の進展についてご講演いただいた。翻字は音訳とも言い、日本語では“computer”という英語を「コンピュータ」という表記に変換するような過程を指す。機械翻訳を簡略化した問題と捉えることもできるため、統計的機械翻訳と同様の技術によって進化している。近年のニューラルネットワーク翻訳の技術は翻字においても同様に有効であり、従来の統計的機械翻訳を応用したものを上回る精度が達成できていることが示された。

技術講演 1 件目は木村氏 (筑波大) らによる、ニューラル機械翻訳における訳抜けの影響に関する研究の発表であった。著者らの研究グループで取り組んでいる専門用語等を統計翻訳のフレーズテーブルを利用して翻訳した結果をニューラル機械翻訳結果の該当箇所に埋め込む翻訳方式の利点として訳抜けが少ないことが確認できたことが報告された。

技術講演 2 件目は熊野氏ら (東芝ソリューションズ/静岡大) による、対訳辞書エントリの派生語を自動的に生成して対訳辞書を拡張する研究の発表であった。既存の対訳エントリから派生変化パターンを抽出し、品詞や接辞辞書の情報から付与した素性と合わせた派生変化規則の構築と、構築された派生変化規則を適用して得られる派生候補を Web テキストデータでの検証を行い、動詞的名詞のエントリ拡張で有効性を確認したことが報告された。

技術講演 3 件目は李氏ら (北京師範大学) による、特許に関する言語リソースを活用した中英特許機械翻訳の改善に関する研究の発表であった。著者らが SIPO の特許コーパスを利用して構築している CTKB (中国特許知識ベース) や対訳辞書、翻訳規則等を活用して NTCIR9 共通タスクにおけるルールベース機械翻訳の性能を上回ったことが報告された。

### 3. 所感

今回ワークショップの企画当初より特に注目したかった内容は、近年急速な発展を続けるニューラル機械翻訳技術が特許・技術文書翻訳に与える影響や、専門用語辞書のように、従来技術文書翻訳のために構築・利用されてきた言語リソースの活用可能性であった。Pouliquen 氏の講演は近年のニューラル機械翻訳による大きな変化を如実に感じさせるものであり、梶木氏の講演ではそうした機械翻訳技術の変化が人手の翻訳にまで明確に波及しつつあることが語られた。Finch 氏の講演はその変化を翻字というタスクの視点から捉えたものであったし、技術講演でもニューラル機械翻訳と対訳言語リソースの関係について議論があった。

技術文書特有の問題のうち、多数の固有名詞の問題と非常に長い文の翻訳については広く活発に研究されていると言えるが、実際に難解な技術文書での実験評価を学術サイドで行っているケースはあまり多くない。以前は情報検索の国際ワークショップ NTCIR で特許翻訳の共通タスクが何度か開催されてきたが、現在では JST を中心に 2014 年から開催されているアジア言語ワークショップ (WAT) が、科学技術論文や特許翻訳の共通タスクを行っている国際的にも現在唯一のものである。

多様なパターン認識・機械学習技術の進化が加速している反面、訳語の統一や翻訳メモリの活用、人手翻訳との連携等、翻訳産業界の需要に機械翻訳技術がまだ十分応えられていないのではないかと感じることもある。機械翻訳技術に注目が集まっている今、人手の翻訳との関係や実際の産業応用等、昨今の技術進展に合わせた展開を見据えた取り組みが重要であると強く感じる。

## Baidu Machine Translation

Zhongjun HE and Hua WU

Baidu Inc.

---

### 1. Introduction

Machine Translation (MT) has been developing for more than 70 years. Thanks to the evolution of techniques, e.g. rule-based method, example-based method, statistical method, as well as the explosive growth of data, translation quality is greatly improved.

Recently, Neural Machine Translation (NMT) has been widely studied [1][2]. NMT works by using a large neural network which directly translates source text to target text, outperforming conventional MT methods in many language pairs. However, as a newly emerged approach, NMT has some limitations that may jeopardize its practicality.

We have applied a group of novel methods to improve the robustness and translation quality of our online NMT system, for example, advancing speed of NMT with constrained softmax, handling the OOV problem by incorporating SMT features, and improving multilingual translation with multi-task learning.

It was on May 20<sup>th</sup>, 2015 that we successfully launched the first Internet NMT system [3]. Till now, Baidu Translate supports translations between any two of 28 languages. Every day, it translates over 1 billion sentences. In addition to the web

interface of our translation service, we have an APP for mobile access. We also provide API for commercial users. Over 50,000 developers are using Baidu translate API in their products for localization, international trades, language learning and other purposes.

In this article, we will discuss the challenges and methods of building a practical MT system in Section 2. In section 3, we will introduce Baidu's translation products, which integrate MT with other AI technologies including speech recognition and OCR. We conclude the article in Section 4.

### 2. Challenges and Methods for Practical MT System

Our online MT system receives billions of translation requests every day. The requests are turned into the form of all sorts of texts, from various domains, e.g. poems, news articles, patent documents, colloquial expressions. The first challenge is to provide reliable translation for all these requests. Secondly, these requests require immediate responses, hence the need for a real-time translation system. The third challenge is the data sparseness problem. For some language pairs, only limited bilingual corpora are available. Even in resource-rich language pairs, data sparseness problem could exist in

some specific domains.

In the following, we will introduce our methods of building a practical MT system.

### 2.1. Improving NMT with SMT Features

Among the problems confronting NMT are the out-of-vocabulary (OOV) problem, caused by limiting the vocabulary to a certain size to avoid the time consuming training and decoding, and the under-translation problem, i.e. missing words in translation, caused by the mechanism that prefers short translation.

In order to tackle these problems, we incorporated three SMT features into the log-linear framework [4]: a phrase table, an n-gram language model, and a word reward feature (Figure 1). The phrase table was trained on word-aligned bilingual corpora in the conventional phrase-based SMT approach, and employed to score word pairs and alleviate the OOV problem. The n-gram language model was trained on monolingual sentences in target language to enhance the local fluency. The word reward feature encourages the model to produce longer sentences and thus to avoid missing words in translation.

Experiments show that our method obtains significant improvements on Chinese to English translation. To the best of our knowledge, this is the first work to integrate SMT components into NMT, which has motivated many studies on this research line, for example, to improve NMT with a coverage model [5], a reordering model [6], or a phrase table [7].

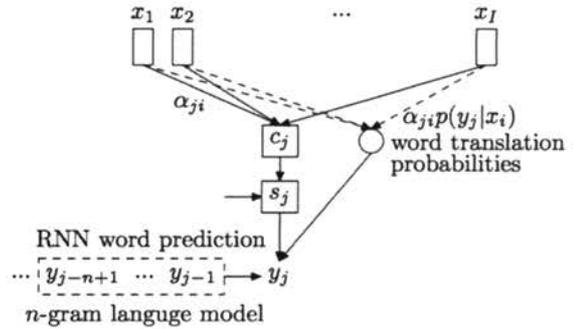


Figure 1: Illustration of the Log-linear NMT. To predict the target word, we introduce SMT features, such as the word translation probabilities, the n-gram language model, word length reward, together with the probabilities estimated by RNNs.

### 2.2. Optimized Beam Search with Constrained Softmax for NMT

Due to the computational complexity of the large neural network, NMT suffers a problem of time-consuming decoding. The basic version of the open source NMT system GroundHog translates at a speed of only 10 words/seconds, far from being able to provide real-time translation service.

To improve the decoding speed of our NMT system, we proposed and applied an improved beam search decoding algorithm with constrained softmax operation over the output layer [8]. Firstly, rather than extend all hypotheses for each stack, we utilized a priority queue to store partial hypotheses and chose the best to extend (Figure 2). We found that the unexpanded hypotheses are safely pruned without hurting translation quality. Secondly, to reduce the computation complexity in softmax layer, we used a phrase table to generate translation candidates for source sentences. Therefore, when predicting a target word, the network only perform softmax on the candidates rather than all vocabulary words.

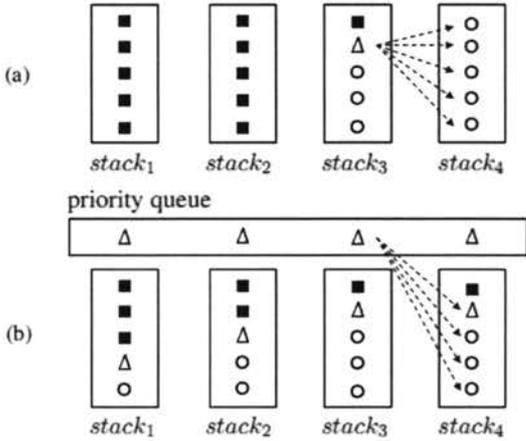


Figure 2: (a) standard beam search (b) improved beam search;  $\square$  extended hypothesis,  $\Delta$  best hypothesis to be extended,  $\circ$  unextended hypothesis

Experimental results show that the optimized algorithm achieves a speed about 3.5 times faster than that of the well optimized baseline system, without hurting translation quality. Specifically, the improved beam search with priority queue helps to reduce about 37.6% hypothesis extension operations and brings a 1.7 times speedup. The constrained softmax strategy reduces the output layer size from 30,000 to about 300 and contributes to another 2.1 times speedup.

### 2.3. Multi-task Learning for Multiple Language Translation

Conventional MT models perform one-to-one translation, that is, translate a source language to a target language at a time. We employed a unified framework that performs one-to-many translation, using multi-task learning [9]. The model is based on standard Encoder-Decoder neural network with a shared encoder on the source side and multiple decoders, each for a target language.

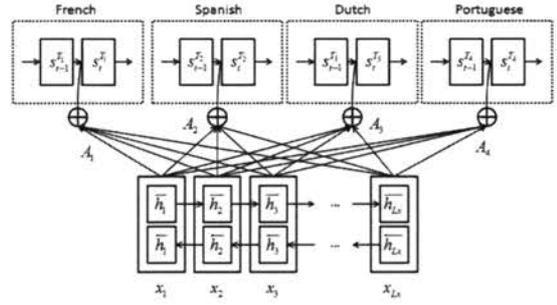


Figure 3: Multi-task learning framework for multiple-target language translation

This method is able to address the data sparseness problem for resource-poor language pairs by sharing information from other resource-rich language pairs on the source side. So far as we know, we are the first to work on multiple target language NMT and has inspired many work following this research line [10][11].

### 2.4. Domain Adaptation

As mentioned, in some specific domains there are data sparseness problem, even in resource rich languages. To further improve domain-specific translation quality, we used a simple and effective domain adaptation method for NMT.

Suppose we have two parallel corpora, a large scale out-of-domain dataset and a small scale in-domain dataset. We first pre-trained a general model on the large bilingual dataset, then continued to train, or say fine-tune the model on the small one. However, the in-domain dataset is so small that we found the translation quality decreased largely after 1-2 iterations, indicating the model was over-fitted.

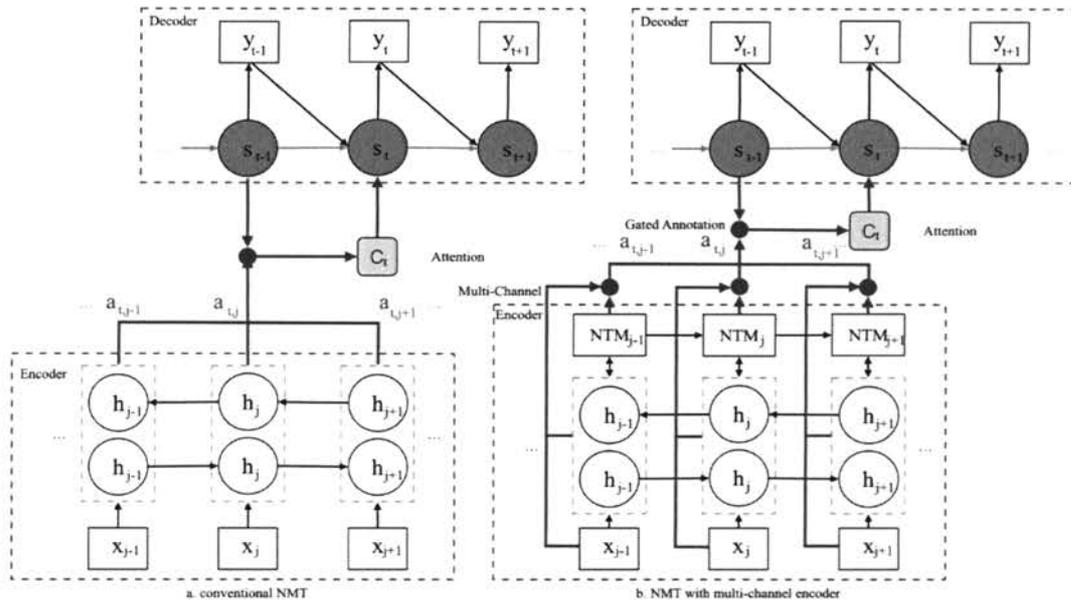


Figure 4 : Left is the conventional attention-based NMT which consists of encoder, decoder and an attention mechanism. Right is our novel NMT model with multiple channel encoder, which consists of hidden state from biRNN, external memory in the NTM and embeddings directly from the input sequence. A gated annotation is designed to automatically learn the weights for different encoding components.

To solve the problem, we incorporated the general model with the fine-tuned model. As a result, the incorporated model get significant improvement on in-domain test set.

### 2.5. Multi-channel Encoder for NMT

The standard Encoder-Decoder architect for NMT abstracts information from a source sentence through hidden states, yielding a sequence of vectors. The information flow is a kind of one-channel flow. For translation, we often want the model to take pieces of the source sentence at varying compositional levels suiting its own linguistic structure. For example, we may want to take an entity name in the source in its raw form while taking an idiom as a densely composed unit. Thus we proposed a novel multi-channel encoder model taking encoding components at different levels of composition [12].

More specifically, in addition to the hidden state of encoding RNN, MCE takes the original word embedding for raw encoding with no composition, and a particular design of external memory in NTM for more complex composition, in a way analogous to visual channels with different frequency. All three encoding strategies are properly blended during decoding. More specifically, we design a gate that can automatically tune the weights of different encoding channels. Empirical study on Chinese-English translation shows that our model can improve by 6.52 BLEU points upon a strong open source NMT system: DL4MT. On the WMT14 English- French task, our single shallow system achieves a BLEU of 38.8, comparable with the state-of-the-art deep models.



Figure 5: Simultaneous Translation System

### 3. Multi Modal Translation and Innovative Products

Traditional MT systems perform text translation. Recently, with the progress of other AI technologies, such as speech recognition, vision understanding, multi-modal translation has been widely studied. Integrating these techniques together, we have launched several innovative products.

#### 3.1. Voice Translation

Voice translation takes voice as input and produces text/voice as output. We have two voice translation products: a simultaneous translation (ST) system for translating speaker's speeches in conferences in real-time, and a smart Wi-Fi translator for travelers' conversation.

The ST system takes speakers' voice as input and shows translation results on screen in real-time. At Baidu World Conference 2017, our ST system successfully translated from Chinese to English for five speakers (Figure 5).

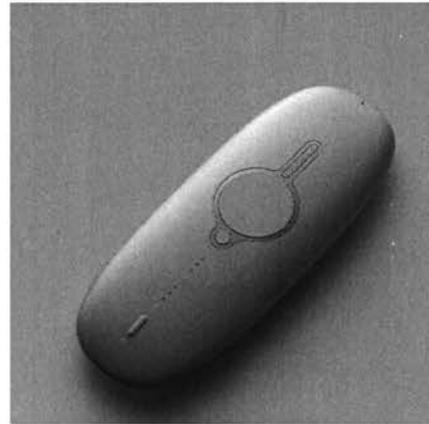


Figure 6: Smart Translator

Our smart translator (Figure 6) was firstly introduced at MT SUMMIT 2017. It provides automatic Wi-Fi connection in over 80 countries and now supports translation in three languages, including Chinese, English and Japanese. More languages will be supported in coming version. Another interesting feature is that it automatically recognizes what language the user is speaking. So users do not need to manually choose translation directions. We are cooperating with some travel agencies. Now some cruise ships are equipped with the devices.

#### 3.2. Image Translation

Image translation takes images as input. If one cannot or do not want to read or write, image is another medium for him/her to communicate.

If the input is an image of an object, the system will recognize the object and provide its description in both source and target languages, which is very helpful in language study (Figure 7).



Figure 7: Object Translation

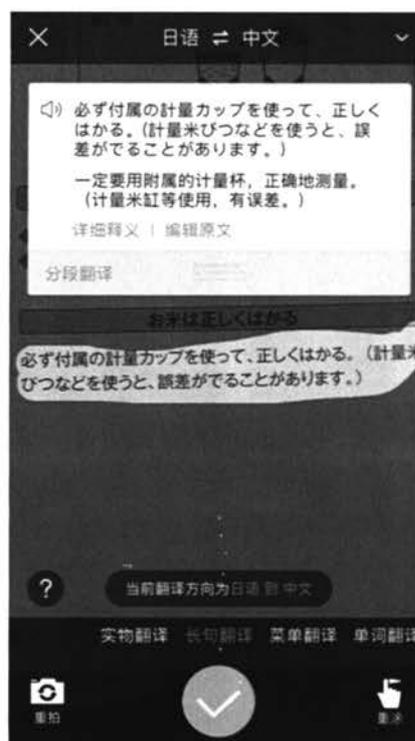


Figure 8: OCR Translation

If the input is an image of a passage of text, we use OCR to recognize the text and perform translation. This function is useful for translating menus, product descriptions (Figure 8), and etc.

#### 4. Conclusion

In this article, based on the practice of Baidu machine translation, we discussed challenges for building practical machine translation system and introduced novel methods we used to address these challenges.

In the future, we will continue focusing on improving user experience with innovative technology, which brings better translation quality, and innovative products combing MT with other AI technologies.

#### Acknowledgements

This research is supported by the National Basic Research Program of China (973 program No. 2014CB340505).

#### References

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. In arXiv:1409.0473 [cs.CL].
- [2] Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems, 3104–3112.
- [3] Zhongjun HE. 2015. Baidu Translate: Research and Products. In proceedings of ACL-IJCNLP 2015 (2015): 61.
- [4] Wei He, Zhongjun He, and Hua Wu Haifeng Wang. 2016. Improved neural machine translation with smt features. In

Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), Phoenix, USA.

[5] Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, Hang Li. 2016. Modeling Coverage for Neural Machine Translation. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 76–85, Berlin, Germany, August 7-12, 2016.

[6] Jinchao Zhang Mingxuan Wang Qun Liu Jie Zhou. 2017. Incorporating Word Reordering Knowledge into Attention-based Neural Machine Translation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 1524–1534 Vancouver, Canada, July 30 - August 4, 2017.

[7] Xing Wang, Zhaopeng Tu, Deyi Xiong and Min Zhang. 2017. Translating Phrases in Neural Machine Translation. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1432–1442 Copenhagen, Denmark, September 7–11, 2017

[8] Xiaoguang Hu, Wei Li, Xiang Lan, Hua Wu and Haifeng Wang. 2015. Optimized beam search with constrained softmax for NMT. In Proceedings of MT Summit XV.

[9] Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, Haifeng Wang. 2015. Multi-task learning for multiple language translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (2015), pp. 1723–1732.

[10] Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, Lukasz Kaiser. 2016. Multi-task sequence to sequence learning. In proceedings of ICLR2016.

[11] Orhan Firat, Kyunghyun Cho and Yoshua Bengio. 2016. Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism. Proceedings of NAACL-HLT 2016, pages 866–875.

[12] Hao Xiong, Zhongjun He, Xiaoguang Hu and Hua Wu. 2018. Multi-channel Encoder for Neural Machine Translation. AAAI2018, to appear.

## 招待講演 “Neural translation technologies and futures”

中岩 浩巳

アジア太平洋機械翻訳協会・名古屋大学

## 1. はじめに

2016年11月にGoogle翻訳への深層学習技術に基づくニューラル機械翻訳(NMT)技術が導入され、日本語に関連する機械翻訳品質が大幅な改善を実現したことから、機械翻訳の利用に関する期待が急激に高まっている。ニューラル機械翻訳の基本方式であるニューラルネットワーク技術は、1980年代にブームがあったニューラルネットワークが2010年代に、計算機能力の増大や新たな処理方法の考案により再度ブームになったものである。ニューラルネットワーク技術の導入は画像処理(画像認識・理解等)や、音声処理(音声認識・合成等)から始まり、その性能が飛躍的性能向上した。機械翻訳を含む自然言語処理での効果は当初限定的であったが、様々な方式の改良(RNN, アテンション等)により数年前より効果が表れるようになってきた。機械翻訳にもニューラルネットワーク技術が導入され、従来主流であった統計翻訳を凌駕する性能を達成するようになった。

本稿では、2017年9月にアジア太平洋機械翻訳協会(AAMT)主催で24年ぶりに日本で開催(開催地:名古屋大学)された機械翻訳の代表的国際会議第16回機械翻訳サミット(MT Summit XVI)で行われた招待講演“Neural translation technologies and futures”において紹介された、代表的翻訳プロバイダーであるバイドゥ、グーグル、マイクロソフトにおける最近機械翻訳技術、特に、ニューラル機械翻訳(NMT)についての講演内容を概説する。

## 2. バイドゥの機械翻訳技術

バイドゥは、世界で初めてニューラル翻訳技術を実

サービスに導入した会社として知られている。実用的システムの構築に向けては、翻訳対象の多様性、データが少ない言語対や翻訳対象分野のデータが少ない場合の翻訳技術、スピードとメモリ、に関する研究開発に注力している。例えば、ニューラル翻訳の欠点としてよく知られている未知語の訳出に関しては、従来型の統計翻訳技術を併用することで、解決しようとしている。また、データの少ない言語対に関しては、各言語対で別々に学習したモデルを統合する(マルチタスク学習)ことにより克服しようとしている。この技術を利用することで単独の翻訳対象言語対データだけより翻訳性能(翻訳品質及び学修速度)が向上することが示されている。なお、第16回機械翻訳サミットの講演では、音声翻訳機能にモバイルWifiを統合したハンディ翻訳機を世界で初めてお披露目した。

## 3. グーグルの機械翻訳技術

前述のとおり2016年秋にグーグル翻訳の方式が従来型の統計翻訳からニューラル翻訳に変わったのを契機に、日本語関連の翻訳品質が劇的に向上したことから、機械翻訳に対する期待が急速に高まった。このような品質向上は、①グーグルが持つWebを中心としたデータ(対訳コーパス)が膨大であること、②世界中の優秀な研究開発者が集まり最新技術を実サービスに展開していること、③強力なデータセンターを持ち他社を凌駕する規模の計算能力を持つこと、等の強みにより実現できたと考えられる。この豊富な計算機資源を活用し、グーグル翻訳では、他社よりも多層(8層)のニューラル翻訳方式を導入している(図1)。ニューラル翻訳では扱える語彙数が学習段階の計算量の多さ

から限界があるため、未知語の問題が発生するが、これに対しては、Wordpiece という、単語中の部分文字列（例えば英単語 “Brexit” では、“Br”, “ex”, “it” と分割）を活用することで、扱える語彙数を増やすことで翻訳精度の向上と、翻訳時間の短縮を実現している。また、グーグル翻訳では現在 103 言語を対象としています。これを実現するために、前述のマルチタスク学習に加え、すべての言語対のデータを活用して 1 つのモデルを作るゼロショット翻訳という技術も導入している。

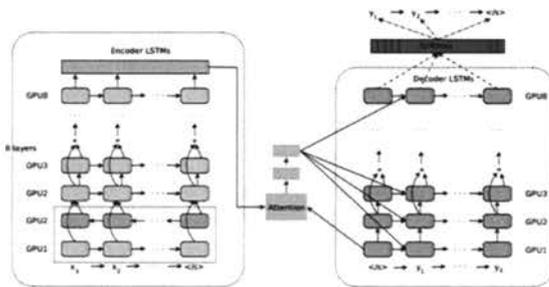


図 1：グーグル翻訳におけるニューラル機械翻訳の構成図

#### 4. マイクロソフトの機械翻訳技術

第 16 回機械翻訳サミットの講演では、音声自動翻訳技術が紹介され、実演を通じてその品質の高さが示された。これは、①音声認識の技術的躍進、②データの大規模化、③頑健で高速な学習インフラが整った、④ニューラル翻訳技術が導入された、というに様々な条件がそろったことによると報告されている。音声翻訳の課題は、①話し言葉は書き言葉と異なること、②人の発話は驚くこと勝手であること（考えと発話の違い、黙ったと思ったら話し出したり言い換えたり等）、③句読点がなく部の区切りがわからない、等といったディスフルエンシ（disfluency）を含んだテキストに対する翻訳になる。これに対してマイクロソフトでは、音声認識結果に対してこれらのディスフルエンシを削除し、完全な文章に変換する処理を実装することでこれらの課題を解している。

#### 5. 終わりに

前述のとおり、NMT の進化は大変早く、現時点での NMT に対する評価が数か月で大幅に変わってしまうこともあり得る。AAMT ではこれからも NMT の最新の動向を調査し、タイムリーに情報発信をする予定でいるので、ご期待いただきたい。



図 2 招待講演の様子

（左からバイドゥ Wu Hua 氏、グーグル賀沢秀人氏、マイクロソフト William Lewis 氏）

（参考文献）

- Neural translation technologies and futures - Baidu, Google, Microsoft -, MT Summit XVI, 2017
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, et.al, Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, arXiv:1609.08144, 2016

## アジアでの機械翻訳サミットに参加して

井佐原 均

豊橋技術科学大学 情報メディア基盤センター

### 1. アジアでの機械翻訳サミット

アジアで始まり、ヨーロッパ、アメリカと隔年で開催されてきた機械翻訳サミットの第 16 回が名古屋で開催された。これを機会にこれまでの私の機械翻訳サミットとの関わりを振り返ってみたい。

第 1 回の機械翻訳サミットは 1987 年に箱根で開催された。この時には私は参加していない。初めて参加したのは 1991 年にワシントン DC で開催された第 3 回である。企業や大学の方々とともに日本からの視察団の一員として参加した。会期中に米国の独立記念日があり、視察団を担当した旅行会社の人に、危険ですから独立記念日の夜は出歩かないようにと言われたことが思い出される。私も含め、誰も従わなかったようであるが、International Association for Machine Translation (IAMT) は、このサミットで設立が合意され、翌 1992 年に設立された。1991 年に「日本機械翻訳協会」として設立されていた AAMT も、これに合わせて 1992 年に「アジア太平洋機械翻訳協会」と名称変更し、IAMT の下にアジア、ヨーロッパ、アメリカの各協会が存在するという現在の体制となった。

1993 年の第 3 回は再び日本での開催となった。私はチュートリアル講師として参加した。会場となったホテルオークラ神戸のカーテンが電動だったことが参加者の間で話題になっていたが、あの当時は珍しかったのだろうか。1999 年のシンガポールでの第 7 回時には、現地の準備が遅れ気味となり、急ぎよ、私と企業の方々数名のチームを作って対応したことが思い出される。この時が私が AAMT に深く関わることになったきっかけである。この時のチームの方々には、その後も AAMT の活動に深く関わっていただいた。

2005 年のタイ・プーケットでの第 10 回は、当時私が前職でタイにラボを持っていたため、ラボのメンバーであった Virach さん (AAMT 理事) とともに現地手配を担当させていただいた。欧米ではまだまだ知られていないアジアでの機械翻訳研究を広く知ってもらうために、アジア圏の機械翻訳研究を紹介した AAMT ジャーナルの特別号を発行した。サミット前年の 2004 年 12 月にスマトラ島沖地震が発生し、プーケットも被害を受けた。状況確認のために現地に派遣されたことも思い出される。

2011 年の中国・アモイでの第 13 回は私が AAMT 会長および IAMT 会長として担当した。会議の実務面はアモイ大学や中国中文信息学会の方々をお願いしたため、開催まで詳細が分からないことも多く、戸惑ったが、実際に開催してみると特に問題もなくスムーズに開催できたと思う。この年には東日本大震災が発生したが、海外から今回はヨーロッパでの開催にしないかとの打診が来たのには驚かされた。我々のアメリカやヨーロッパの地理に関する感覚も同じようなものではあろうけれど。

### 2. 名古屋での機械翻訳サミット

アモイでの第 13 回から 6 年が経ち、機械翻訳サミットが再びアジアに回ってきた。名古屋大学の中岩先生が AAMT 会長、IAMT 会長として担当され、名古屋での開催となった。私はこれまで、シンガポール、タイ、中国と機械翻訳サミットの開催に深く関わってきたが、今回は Steering Committee のメンバーとして、時折コメントする程度で、個人的には「気楽に」関わる機械翻訳サミットであった。神戸以来の日本開

催となった今回は中岩先生を中心に本号の他の報告にあるように多くの方々の努力で大成功となった。関わった皆様のご尽力に敬意を表します。

今回の機械翻訳サミットはニューラル機械翻訳が実サービスに入って最初のサミットであり、百度、グーグル、マイクロソフトによる紹介のセッションなど非常に興味深いものであった。Technology Showcaseでも機械翻訳の応用に向けた多くの展示があった。機械翻訳を実際の翻訳プロセスに組み込むという立場からは、日本翻訳連盟と協調してサミットを開催できたことは大きな意味を持つ。今回のサミットは機械翻訳システムが社会実装に向けた大きな転換点を迎えた今に

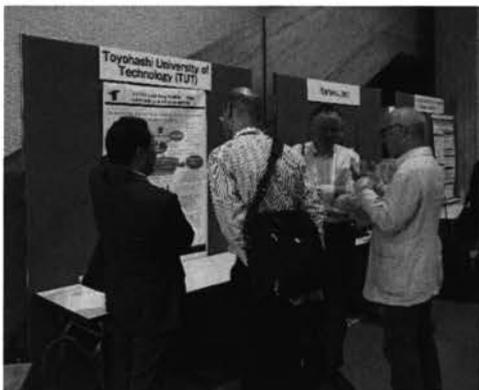
ふさわしいサミットであった。

会場はコンパクトにまとまっており、地下鉄の駅からも近く、参加者には好評であった。コーヒーブレイクの行われた1階のスペースに行くと、いつでも誰かしら知人に会えて好都合だった。懇親会も非常に日本的な場所と食事であり、海外からの参加者だけでなく日本人参加者にとっても印象深いものになった。

次回の機械翻訳サミットは2019年にアイルランドで開催される。機械翻訳の新しい社会実装がどこまで進んでいるか楽しみである。そして次のアジア開催時には機械翻訳がワープロのように我々の生活に溶け込んでいることを期待するものである。



懇親会で笑顔の参加者



Technology Showcase



懇親会での日本伝統の鏡開き

第27回 JTF 翻訳セッション  
「ニューラル機械翻訳は翻訳プロセスをどう変えていくか  
—最近の機械翻訳技術と利用に関する動向—」

(アジア太平洋機械翻訳協会 AAMT 企画セッション)

島津 美和子

東芝デジタルソリューションズ株式会社

## 1. はじめに

2017年11月29日、アルカディア市ヶ谷において恒例のJTF翻訳祭(日本翻訳連盟主催、AAMT後援)が開催された。昨年好評であったパネルディスカッション「いまさら聞けない機械翻訳の基礎」に引き続き、AAMTは1セッションを持ち、中岩会長とAAMT機械翻訳課題調査委員会の長瀬委員長が掲記のタイトルで講演を行った。今年も収容人数120名を優に超え、150名ほどの熱心に聞き入る聴衆を得た。

本講演では、機械翻訳(MT)の研究開発に長年携わってきた2名が昨今の急速な環境の変化の下でのMTの最新技術動向とMTの性能比較結果について、今年開催された国際会議(MT Summit XVI, EAMT2017)の発表内容も交え、話題のニューラル機械翻訳(NMT)を中心に、分かりやすく解説した。本稿ではその概要を報告する<sup>1</sup>。

## 2. 機械翻訳(MT)を取り巻く状況(中岩会長)

日本におけるここ数年の変化は、翻訳需要に拍車を



満員の会場(写真撮影・提供:中村ヒロユキ氏)

<sup>1</sup> 本稿は『日本翻訳ジャーナル』(2018年1/2月号)掲載の筆者による報告をもとにしている。

かけ、MT技術に対する注目度を高めている。背景には、急速な少子高齢化とグローバル化の進展により、日本産業の海外依存度が増し、国際ビジネスにおける翻訳場面が増大したことがある。訪日外国人数の増大と2020年の東京オリンピック・パラリンピック開催により、今や外国人とのコミュニケーションは不可欠になっている。一方、Google翻訳をはじめとする深層学習技術に基づくニューラル機械翻訳(NMT)技術の導入により、従来のMTでは難しいとされてきた言語対の翻訳品質が大幅に改善し<sup>2</sup>、MTの利用に対する期待が世界的に高まっている<sup>3</sup>。

## 3. ニューラル機械翻訳(NMT)とは(中岩会長)

MT方式は、専門家が作成したルールに基づいて翻訳を行うルールベース方式(RBMT)、蓄積された対訳データを分析・学習した結果からモデルを作成し、このモデルに基づいて翻訳を行うコーパスベース方式に大



講演者の中岩会長・長瀬委員長(写真提供:JTF事務局)

<sup>2</sup>

<https://research.googleblog.com/2016/09/a-neural-network-for-machine.htm>

<sup>3</sup> Castilho, S., et al. (2017). Is neural machine translation the new state of the art? *EAMT2017*  
[https://ufal.mff.cuni.cz/eamt2017/slides/Wednesday/Research\\_presentations/Castilho\\_et\\_al.pdf](https://ufal.mff.cuni.cz/eamt2017/slides/Wednesday/Research_presentations/Castilho_et_al.pdf), p. 4

別される。後者は統計的機械翻訳(SMT)とニューラル機械翻訳(NMT)に代表される。RBMTは一定品質の翻訳を保ち、辞書・規則によるカスタマイズが容易であるが、対象言語ごとに専門家が必要である。また、訳抜けや訳余りが起きにくい反面、訳文はぎごちない。コーパスベース方式は、モデルを自動で学習し、多言語化が容易であるが、最低100万文の大量データを必要とする。このモデルは権利関係が不明確という難点もある。

NMT以前のSMTでは、対訳データに基づいて句レベルの対訳辞書を自動的に学習し、最も確率の高いものを選択する。語順が異なる言語対では句の並び替えを行うが、この失敗がSMTの問題となっていた。そこで登場したのがNMTである。

NMTの背景には1980年代にブームとなったニューラルネットワーク(NN)が2010年代復活したことがある。計算機能力の急激な普及とアルゴリズムの改良により、大規模なNNを使った技術が実装可能になった。NNの自然言語処理での効果は当初限定的であったが、数年前からMTにも導入され、SMTを凌駕する性能を達成した。2016年11月にはGoogleがNMTのサービスを開始し、一般の人々がその性能を身近に体感できるようになり、大きなインパクトをもたらした。

ではNMTとSMTの違いは何か。NMTは句ではなく文全体の情報を用いて翻訳する。まず、NNを通じて原言語を意味解析し、数値的な意味表現に置き換える。そしてNNを使ってその意味表現から単語を生成し、目的言語を出力する。この中間的な意味表現は二次元ベクトルの形である。NMTはリカレントニューラルネットワーク(RNN)を使う。RNNは、入力、出力、入力ベクトルに対し重みづきの演算を行った結果得られる複数の中間層に加え、一つ前の時間の情報を入力としてフィードバックし、時系列のベクトルの情報を継承する点に特徴がある<sup>4</sup>。処理は次のように進む。最初に原文の各単語がNNに入り、それをあるベクトルにエンコードした後、ベクトルをもとに隠れ層(中間

層)というニューラルの活性情報に変換する。次に、RNNにより、新しい単語の情報と一つ前の単語の情報を使い、新しい意味の情報を得る。最終段階で原文に対する意味表現を保持し、この意味をもつ目的言語の単語を次々と生成(デコード)していく。

Google翻訳の技術はアテンション機構を持つ。アテンションに基づくNMTでは、入力文の情報をベクトルに変換した後、どの単語に注目(アテンション)するかを一出力ごとに計算し、最も高い注目度の単語の情報を使う<sup>5</sup>。アテンション機構はNMTで標準的に使用されている。

MTの人手評価の例としてATRの基準は訳文の自然さを見る流暢さ(fluency)と原文の意味を訳文に伝えているかを見る適切性(adequacy)の2軸で5段階評価する。自動評価には世界的に使われているBLEU(機械訳と参照訳のn-gramの一致度を両者の文長の違いも考慮し、計算する手法)の他RIBES(機械訳と参照訳の語順の違いを重視する手法)がある。

#### 4. 機械翻訳技術及び活用方法の動向(中岩会長・長瀬委員長)

第20回ヨーロッパ機械翻訳協会年次大会(EAMT2017)(2017年5月)<sup>6</sup>ではMTの研究開発及び活用の最先端の事例が紹介された。ここではNMTの活用と評価を概観する。性能評価では、新聞記事、e-commerce、MOOCsにおいてNMTが従来方法を上回る結果が報告された。MTのエラー分析結果に基づいて流暢さと適切性を評価するMultidimensional Quality Metrics (MQM)の報告では、NMTは適切性のうち訳漏れの数は最大だが、誤訳の数はSMTの一種である従来のフレーズベース機械翻訳(PBMT)よりも少なく、流暢さは全項目で上回ることが確認された。SMTでは効果のあった事前並び替えをNMTに適用した結果、効果がみられなかったという指摘もあった。カスタマイ

<sup>4</sup> <https://www.slideshare.net/YusukeOda1/encoderdecoder-tis>, p. 7

<sup>5</sup> <https://www.slideshare.net/ToshiakiNakazawa/attentionbased-nmt-description>, p. 1, p. 3

<sup>6</sup> <https://ufal.mff.cuni.cz/eamt2017/>

ズなしのNMTとカスタマイズしたPBMTの比較では、単語については後者が上回ったものの、全体的にはNMTの方がよいとの結果も示された。

MT Summit XVI (2017年9月、於：名古屋大学)<sup>7</sup>に関しては、NMTの主要3社が現状と今後の方向性について語った招待講演"Neural translation technologies and futures—Baidu, Google, Microsoft"に限定して述べる。Baiduは、SMTの特徴を使ってNMTの精度向上を図るとともに、Multi-Task Learningにより原言語の単語表現の学習効果及び運用効率を改善した。GoogleはWeb上の文を英語に機械翻訳し、一致する文を抽出して大規模パラレルコーパスを作成している。これを用いて、複数言語対を同時に学習させ(Multi-Task Learning)一つのモデルを構築することにより、データが存在しない言語対の翻訳(zero-shot translation)を可能にした。音声自動翻訳にフォーカスしたMicrosoftの発表では、課題として話し言葉は書き言葉と異なること、人の発話は秩序だっていないこと、句読点が発話されないこと、会話データが少ないことを挙げた。

## 5. NMTの訳文評価(長瀬委員長)

AAMT機械翻訳課題調査委員会が日英・英日・中日方向でNMT、RBMT、SMTの訳文品質を評価したところ、全言語方向の定量評価においてNMTが最も優れていた。定性評価ではNMTの長所・短所、すなわち、訳語選択が正確で、訳文が自然で、並列句の認識に優れ、主語なし文もうまく翻訳する一方、訳抜け、過剰訳出、低頻度語の誤訳が起きやすいことを確認した。文書別にみるとNMTは特許の翻訳精度が突出してよかった。

一方、運用面のNMTの問題として、従来方式と比較して多くの学習時間や計算資源が必要であることを挙げた。また、原文と訳文の単語対応が取りにくいこ

<sup>7</sup> <http://aamt.info/app-def/S-102/mtsummit/2017/>  
 なお、日本語の詳細な報告は『日本翻訳ジャーナル』2017年11/12月号の巻頭特集「MT Summit XVI」を参照されたい。

ともNMTのデメリットである。

## 6. 問題提起(長瀬委員長)

締めくくりに、次のようなNMTとRBMTの英語原文と日本語訳文を示し、直訳のRBMTと意識のNMTのどちらが有用と考えるかをフロアに問うた。

	例文
原文	Pedestrians are still allowed to talk on their phones while crossing the streets, as long as they look at their surroundings.
NMT	歩行者は、周囲を見る限り、通りを横切って電話で話すことができます。
RBMT	彼らが環境を見る限り、通りを横断している間、歩行者はまだ彼らの電話について話すのを許されます。

RBMTとNMTの訳、どちらがお好みですか？  
 直訳のRBMT：  
 英語の構造のまま翻訳。ごこちないが、ほぼ確実に対応する訳が出る。  
 意識のNMT：  
 "while crossing the streets"→「通りを横切って」  
 読みやすく意味もわかるが、原文の意味が訳文に明示的に表れないことも。

すると、NMT支持とRBMT支持がほぼ同数という興味深い結果が出た。これについては今後議論を深めていく必要がある。

## 7. さいごに

本講演では、講演者が演題に掲げる問いに直接回答することはしなかったが、各人がその答えを模索するなかで必要となる情報を提供できたのではないと思われる。なお、同じく機械翻訳をテーマにした翻訳祭セッションにはMLV(マルチランゲージベンダー)3社によるパネルディスカッション「機械翻訳は翻訳業界を壊すのか？」があった。ディスカッションを通じて、パネリストらは、機械翻訳の導入の是非について決断を下す際には、よりどころとすべき情報源をどこにするかこそが重要であるとの共通見解に達した。司会の渡邊氏は信頼できる情報源の例としてAAMTを挙げていた。AAMT会員による日頃の地道な啓蒙・普及活動が功を奏したものと思われる。

AAMTは本講演と並行して展示ブースで出展し、ユーザアンケートと委員会の活動紹介を行った。MTに関心を持つ多くの方々が足を運んでくださった。翻訳者・翻訳会社と機械翻訳研究者・開発者との距離が徐々に縮まってきたことを感じさせる一日であった。

## Women in Localization Japan 第13回イベント報告

目次 由美子

XTM International Ltd.

### Women in Localization Japan 第13回イベント報告

日時：2018年02月02日（金）19:00～21:30

開催場所：ネットアップ株式会社 会議室

#### セッション I

テーマ：中国人のものの考え方やコミュニケーション

登壇者：金子 行宏 ブライトンヒューマンキャピタル  
コンサルティング会社

#### セッション II

テーマ：いま MT をどう使うか？ MT にどう立ち向かうか？

登壇者：三笠 綱郎 株式会社十印

#### セッション III

テーマ：翻訳とポストエディットの本質的な違い

— 国際標準規格から考える —

登壇者：森口 功造 株式会社川村インターナショナル

Women in Localization Japan 事務局メンバー：

森 みゆき、上田 有佳子、千葉 容子、澤村 雅以

ローカリゼーションに携わる女性のための非営利団体として 2008 年に米国加州で設立された「Women in Localization」の日本支部は 2015 年に発足し、3 ヶ月ごとに勉強会兼ネットワーキングのイベントを開催している。発足当初は約 10 名であった会員は今や 100 名を越え、クライアント企業のみでなく、翻訳会社や個人翻訳者、翻訳ツールベンダーを含む。ローカリゼーション関連の新しい技術や共通の問題をトピックとしたセッションを行っており、この 2 月のイベントでは機械翻訳が大きな話題として取り上げられた。

### セッション I

『中国人のものの考え方やコミュニケーション』

金子 行宏

「発話された言葉が相手にどのように受け取られているのか。そこには中国人と日本人の違いが存在する」

中国と日本に研修・コンサル会社を設立して 15 年の金子氏は、中国に所在する日系企業を顧客に擁している。中国人を部下に持つ上司の不満の一つとして、中国人は受けた仕事上の指示に対して、「できる」は気軽に言うが、「できない、分からない」とは決して言わないことがあげられると指摘し、次のように述べた。

日本人が考える「できる」と中国人の「できる」では、その言葉の意味することが異なる。日本人が「できる」と言えば、それは完遂できることを意味し、その発言にも責任や義務がまわりつく。中国人のそれは、私ならできそうだ、やってみますという程度の、いわば仕事の受領宣言に近い意味がある。なので、その仕事を完遂する能力が有るかどうかは別の話だから、実際に受けた仕事が不十分な成果になることもしばしば起こる。日本人上司は、「できると言ったのに！できないなら先に言っておいてよ」という不満を持つことになる。他方「できない、分からない」と発言する部下への反応も違う。日本人の上司は育成的な態度を示してフォローする傾向にあるが、中国人の上司は私が機会を与えているのになぜやろうとしないの？と、部下のやる気に疑いの目を向ける。このように同じ言葉であっても受け取られる意味にずれが生じてしまう。なので、仕事で大きな問題を生じさせることがある。結局、日本人上司が中国人部下に対してできることは、期待値を現実的に設定し、遂行プロセスに介入する、

できないことよりもできることを中心に考える、刻々と変化する状況においてはその時々に対応するということになる。

そもそも、このような違いが生まれる背景には社会とその成り立ちに関係がある。中国社会において、自分の能力を多少過剰に発信することは極めて一般的な行動である。その理由として、中国が厳しい個人の競争社会であることや、人の流動性が高いことなどが挙げられる。逆に日本人は、自らの能力をやや過小に申告してしまう傾向がある。

それぞれの社会にはそれぞれに合理的な行動があると、私たちは相互に理解し合いながら仕事をすることは可能であると言える。

以上、金子氏の講演に対し、参加者から「日本人は自らの長所を過小申告しているとするが、米国在住の日本人の提示はさらに低く発信しているように思われる」との発言もあった。講演内容は翻訳自体に直接関連する話題ではなかったが、私たちは多言語翻訳プロジェクトなどで、ターゲット言語ごとに翻訳者や該当国の特色が伺えることがある。それはその国の国民性であると簡単に片付けてしまいがちだ。しかし、それは取りも直さず私たちの翻訳業務の難しさそのものを示しているのかもしれない。「日本語への翻訳は、なにがそんなに難しいのか、なぜチェックにそんなに時間が必要なのか」と問われることがあるが、それは金子氏が指摘したことと同じように関連しているようにも思われた。

## セッション II

『いま MT をどう使うか？ MT にどう立ち向かうか？』

三笠 綱郎氏

株式会社十印にて品質管理業務に従事する三笠氏は、現在、機械翻訳 (Machine Translation : MT) の実務への導入に取り組んでいるそうだ。ルールベース、統計

ベース、そしてニューラルネットワークといった機械翻訳の代表的な手法についても、それぞれの特徴をざっと紹介してくれた。

特に、2016年11月にリリースされた Google 翻訳のニューラル機械翻訳を含め、ニューラルネットワークを搭載した MT エンジンから生成される和訳の流暢さを提示した。なかには、英作文が得意ではないことが顕著であり、文法の誤りが含まれる英文であっても、MT にかけてと一見して流暢に見受けられる和文が生成される例を示してくれた。

また、MT の弱みとして訳語の「揺れ」が指摘された。deploy (デプロイする、配備するなど)、configuration (コンフィギュレーション、構成、設定など) の訳に不統一が見られ、ソフトウェアのローカリゼーション業務においては必須とされる用語統一が難しい様子であることが言及された。

MT を利用する形態としては、MT を利用して訳文を生成するのみのこともあれば、「ポストエディット」(Post-Edit : PE) と呼ばれる、MT が生成した訳文を人手によって修正する「MT+PE」という作業があるとのこと。

代表的な MT エンジンとして Google 社の「Google Translate」、情報通信研究機構 (National Institute of Information and Communications Technology : NICT) の「みんなの機械翻訳」、Microsoft 社の「Microsoft Translator」などが紹介された上で、実施している試験利用と結果を紹介してくれた。

たとえば、あるユーザガイドを対象に Google と NICT のエンジンを利用して PE を実施した場合、人手翻訳 (Human Translation : HT) よりも作業効率が向上するという結果が出たそうだ。ただし、4人の担当者に対して PE を急いで行うようにという指示は出さなかったとのこと。翻訳品質の評価は確認してはいないものの4人の被験者には同じ傾向が見られ、他の MT エンジンに比較して、Google Translate を利用した場合は最速で PE 作業が完了したそうだ。

さらに、三笠氏はテクニカル製品のマーケティング

資料の英和翻訳においても、MT の利用が有効であるかを試験してみたとのこと。ノートパソコンなど実際の Web サイト上の製品紹介ページを提示した上で、MT 結果と HT 結果を混在して提示してくれた。MT 結果には良いところとそうでないところの「ムラ」が見られ、「Choose yours!」というような英語ならではの表現ともなると、それらしい和訳の生成は難しいことが見て取れた。

### セッション III

『翻訳とポストエディットの本質的な違い - 国際標準規格から考える -』

森口 功造氏

ISO TC 37 の国内委員としても活躍する森口氏は、まず、「ポストエディット」(PE) とはなにかを参加者たちに質問した。その場では、「機械翻訳後の修正作業」という表現が挙げられた。

そして、翻訳に関連する ISO (International Organization for Standardization、国際標準化機構) の規格がどのように策定され、施行されるかをざっと解説してくれた。たとえば ISO17100: 翻訳サービスに関する国際規格が策定された背景として、翻訳提供者間に真に国際的な基準を設ける需要があったことや、副次的な効果として翻訳業界でプロセスが共有されたこと、ISO を取得したことによって企業の信頼性が向上したことが紹介された。ISO17100 は、日本ではすでに 32 の組織が取得済みとのこと。

本セミナーでは、2017 年 4 月に正式版がリリースされた ISO18587: ポストエディットに関する国際規格を ISO17100 と比較しながら解説してくれた。この規格については、日本では認証機関が存在しないために「自己適合宣言」しか選択肢がなく、信頼性が低くなるという点と、しかしながらこの規格のおかげで「ポストエディット」という言葉が普及したように思われるという利点が紹介された。

ISO17100 は品質が重視されており、翻訳の成果物に特化しているとのこと。一方、ISO18587 は納期短縮と予算低減が重視されているようだ。しかしながら、HT と同レベルの品質は求められているとのこと。

さらに、ISO18587 でいうポストエディットは「フルポストエディット」であるとも解説があった。「製品としてリリース可能な品質」のためにはフル PE が必要であるが、最終的に公開することを目的とはしておらず、主な概念やポイント、情報を収集する目的であればライト PE で十分という表現を聞くこともある。そして、フル PE には、翻訳者による翻訳と同水準が求められており、つまりは ISO17100 で規定される翻訳者と同水準のスキルセットや資格が必要であるとのこと。

ISO17100 で規定されている HT では、翻訳をした後、セルフチェックをし、バイリンガルチェックをすることが求められており、チェック担当にも翻訳の学位や経歴が求められている。ISO18587 では、「MT+PE」、つまり機械翻訳に訳文を生成させて、ポストエディットをするというみのプロセスが定められており、それ以降のチェック作業のプロセスは明確には定義されていないようだ。しかしながら、「ポストエディター」には ISO17100 に規定されている翻訳者と同様の力量が要求されており、翻訳に関連する学位や資格、検定などの力量評価、さらには専門知識や能力まで求められているとのこと。

MT や PE を提供している翻訳会社は、ISO18587 に関わらず、発注者と仕様を協議して必要なプロセスの項目を増やすことも可能であるとの言及もあった。また、フル PE 作業に関して発注者と合意を形成するポイントとして、以下が挙げられた。

1. 用途は正しいか (重視するのは品質か、価格・納期か)
  2. ターゲット品質の違い
  3. チェック作業の有無
  4. チェック項目の明確化
- MT が向上している今だからこそ、こういった定義

をする必要があると森口氏は力説された。

#### 登壇者プロフィール

金子 行宏

1964 年生まれ。

株式会社リクルートにて人材育成コンサルティングおよび研修企画、プログラム開発を担当。退社後、2003 年に日本と上海にブライトンヒューマンキャピタルコンサルティングを設立し、現在に至る。

中国日本商会、JETRO、海外職業訓練協会、金融機関などでの講演多数。

上海外国語大学（非常勤）講師

OVTA 国際アドバイザー、厚生労働省キャリア・コンサルティング手法等調査研究委員、経営行動科学学会、国際ビジネス研究学会、アジア・国際経営戦略学会所属

著書『中国進出企業における成功／失敗事例集』（共著）  
技術情報協会

『老成時報』（労務行政）Web 版連載

三笠 綱郎

大手電機メーカーでシステム開発に携わった後、1990 年代半ばローカリゼーション業界に入り、翻訳サービスのプロセス標準化や品質管理、機械翻訳をはじめとした言語関連の各種テクノロジーの導入、翻訳チームのマネジメントに従事。

途中フリーランス翻訳者も経験し、IT を中心としたさまざまな分野の翻訳に携わる。

現在は品質管理プロセスの強化・改善とともに最新の MT 技術の導入に取り組んでいる。

森口 功造

品質管理担当として株式会社川村インターナショナルに入社後、翻訳、プロジェクトマネジメントなどの制作業務全般を経験。

2007 年に品質工学の考え方を取り入れた独自の統計的品質管理手法の社内導入を推進。2011 年からは営業グループを含めた業務全般の統括として社内の管理に携わっている。

現在は、業務プロセスの標準化推進にも関わり、ISO TC 37 国内委員として、主に ISO17100、ISO18587、ISO21999 の策定に関わる。

# AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第二十二回を迎えることができました。今号では、個人会員一名、法人会員一社からのご寄稿をいただいております。

独自のお取り組みのご紹介、機械翻訳研究への提言、AAMT の活動へのご要望など、今回も貴重なご意見をお寄せいただきました。

AAMT Journal では今後も引き続き、会員の皆さまからのご寄稿を心よりお待ちしております。

ご寄稿・お問い合わせは AAMT 事務局(E-mail: AAMT-info@AAMT.info)まで宜しくお願いいたします。

---

## 法人会員（敬称略・50 音順）

---

### 会員名

株式会社ウィズウィグ/WysiWyg Co., Ltd.

<https://www.wysiwyg.co.jp/>

【本社】〒104-0032 東京都中央区八丁堀 2-21-2 コア京橋ビル 6F

【大阪支社】〒541-0046 大阪府大阪市中央区平野町 1-7-3 吉田ビル 2F

---

### 自己紹介

当社は、医学・薬学関連における翻訳、メディカルライティング、安全性情報管理、及び文献情報、副作用情報の整備等を主な業務として行っています。

会社設立時より、一貫して科学的文書をメインとして取り扱っており、「科学的に正確」な文章であることにこだわったサービスを提供をしております。

よく、社名の由来を聞かれますが、「WysiWyg (ウィズウィグ)」という社名は「What You See Is What You Get」の頭文字をとったものです。これは「表示されたものがそのまま出力される」という意味を有する DTP 用語で、設立当時から、私たちはこの言葉を「科学的に正確な文書の作成」をするというキータームとして大切にしてきました。しかしながら、時代の変化とともに、私たちに求められることも変わり、昨年より私たちはそこに「k」を加え、「見えている」からクライアントが「望むこと」「目指しているもの」を形にできる「What you seek is What you Get!」としました。

翻訳会社として翻訳者を育てることもひとつの使命と考え、医薬品の安全性に特化した翻訳学校「アンセクレツォ」も開校しています。

---

## MT/翻訳とのかかわり

### MT および翻訳業界に期待すること

医薬品の開発は長い期間に渡りますが、必要となる翻訳のひとつひとつは短い期間で対応することが要求されます。私たちも、求められる品質を確保しつつ、短納期に応えるため日々模索し続けています。MT はその大きな解決策となり得るはずで、翻訳会社の役割も今後大きく変わっていくことは間違いありません。AAMT の活動を通して、私たちも MT の開発や今後の展開についても積極的に情報収集を行っていきたいと考えております。MT は翻訳会社のみならず翻訳者にとっても不安な部分があります。個人での情報収集にも限界があり、翻訳業界から翻訳者に向けて今後の動向などの情報発信する機会を設けられることを期待しています。

### AAMT への要望

日々進化している分野ですので、AAMT の中でもいろいろな立場の方からの情報発信と情報共有の場を提供する中心的役割を担っていただくことを願っています。

# AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第二十二回を迎えることができました。今号では、法人会員一社、個人会員一名の皆さまからのご寄稿をいただいております。

独自のお取り組みのご紹介、機械翻訳研究への提言、AAMT の活動へのご要望など、今回も貴重なご意見をお寄せいただきました。

AAMT Journal では今後も引き続き、会員の皆さまからのご寄稿を心よりお待ちしております。

ご寄稿・お問い合わせは AAMT 事務局(E-mail: AAMT-info@AAMT.info)まで宜しくお願いいたします。

---

## 個人会員（敬称略・50 音順）

---

### 会員名

河目志津加

---

### 自己紹介

特許・法律分野の翻訳、翻訳チェック、ポストエディットをしています。昨年二度目の独立開業をし、今のところ順風満帆です。

GNMT の登場で焦って、「翻訳でごはんを食べられる時間はもう長くない。早めにポストエディットのエキスパートになってしばらく稼いだらほかの業界に移ろう」という、翻訳業界の方々には知られたくない動機から翻訳会社を退職してポストエディットを始めました。

予定と違って、機械翻訳を知れば知るほど関心がわき、また、GNMT を知れば知るほど翻訳業界の寿命が長いことにも気がつきました。今は私の、あるいは翻訳業界の最期のときまで、翻訳業界と添い遂げてしまいそうな感触でいます。

---

### MT/翻訳とのかかわり

仕事で MT を使うようになったのはつい最近です。ポストエディットの依頼は積極的に引き受

けていますが、今のところ苦手です。どうしても MT 訳文をもとに「いつもの自分の訳文」を作ろうとしてしまいがちで、手が疲れます。翻訳者の訳文を生かすチェック業務のときと同じように、「MT 訳文を生かす」ポストエディットができるようになれば、と思っています。

プレエディットがほぼ趣味になっています。いろいろと言い換えをして出力を比べるのが楽しいです。プレエディターの求人にはまだお目にかかったことがありませんが、プレエディットのお仕事があればぜひ私に依頼していただきたいです。

#### MT および翻訳業界に期待すること

翻訳を始めてから 11 年になりますが、その間、外国人への日本語教育、難民支援団体の研究資料翻訳、医療通訳研修の受講など、趣味と自己満足のためにいろいろな活動をしてきました。

言語の壁があるおかげで生活できている私ですが、こうして同じ言語の壁に苦しむ人たちをたくさん見てきたので、MT の品質が向上してからは、「いずれこの人たちのためにも MT を使えるのでは」と考えています。これが私が MT に強い関心を持っている大きな理由でもあります。自分が生活できなくなったらもちろん困りますが、「何でも読める、誰とでも話せる世界って楽しそう」とも思います。

MT 技術の発展で、これからなくなる仕事も新しく生まれる仕事もあると思います。翻訳者も翻訳業界も臨機応変に対応していかなければならなりません。また、AI が人間の労働力の大部分に代わるときには、社会の仕組み自体が変わらないといけないのかもしれませんが、そんなずっと先の難しいことは私には分かりませんが、「何でも読める、誰とでも話せる」楽しい世界を目指して、ポジティブに MT と付き合い合っていける環境を一緒につくっていけると嬉しいです。

#### AAMT への要望

セミナー、講習会、研究会など、機械翻訳について知る機会を提供していただきたいです。

# 協会活動報告

(2017年9月～2018年2月)

## AAMT 交流会・MT Summit XVI 報告会を兼ねて—

13:30—17:00 講演会

17:30—19:30 懇親会

場所：関東ITソフトウェア健康保険組合（大久保）会議室

上記日程にて第二回 AAMT 交流会および MT Summit XVI 報告会を開催した。

講演会前半は、中岩浩巳会長から、「MT Summit XVI および EAMT2017」に関する報告、長瀬友樹課題調査委員長から「AAMT 課題調査委員会の活動報告」、内山将夫氏（情報通信研究機構）から「自動翻訳技術の最新動向」について広く講演をいただいた。

後半は隅田英一郎（情報通信研究機構）をオーガナイザーとして、以下の御三人に、各自15分程度のご講演の後、会場の参加者も質問者として参加し、活発な意見交換がなされた。これは、オーガナイザーの発案のもと、MT 開発者と MT ユーザーの双方の立場からの意見交換の場となり、参加者を含め、有意義な会となった。

中山雄貴（ヒューマンサイエンス）：エンジニアの立場から

三笠綱郎（十印）：翻訳者・品質管理者の立場から

森口功造（川村インターナショナル）：経営者の立場から

参加者も昨今の MT に関する関心の高まりを見せるように昨年度の倍となった。

講演会参加者：90名 懇親会参加者：46名

## 機械翻訳課題調査委員会

2017年9月29日（2017年度 第6回）

① 各WGの活動について（各WGに分かれて議論）

(WG3)

・UTX仕様について

免責事項

UTX シンプル仕様書

ISO DCR

・宿題

シンプル仕様書のたたき台を作る。

免責事項を固め、UTX 1.20 に追記する。

2017年11月10日（2017年度 第7回）

① 各WGの活動について（各WGに分かれて議論）

（WG1、WG2）

- ・ 翻訳祭の準備
  - 展示ブース（アンケートの準備）
  - 景品の確認（抽選、回答者全員）
  - 備品など
  - 配布物など
  - 講演セッション
  - 当日のシフト
- ・ 機械翻訳評価の発注について
  - 各翻訳エンジンの5段階評価ができた後での評価・分析について

2017年12月8日（2017年度 第8回）

① 全言語処理学会発表の原稿について検討

2018年1月26日（2017年度 第9回）

① 各WGの活動について（各WGに分かれて議論）

（WG1、WG2）

- ・ 会員向け交流会について
  - ・ 言語処理学会発表について
  - ・ 17年度報告、18年度計画
  - ・ その他
- ② 全体会議
- ・ 各WGの報告
  - ・ 来年度の活動計画
  - ・ 医療における音声機械翻訳実証の報告

2018年2月16日（2017年度 第10回）

- ① 会員向け交流会について(報告)
- ② AAMT/Japio 研究会の評価サイトに関する報告内容確認
- ③ 言語処理学会発表について
- ④ 2018年度計画検討
- ⑤ AMTA 発表について
- ⑥ 自動評価サイトの英語化
- ⑦ 医療における音声機械翻訳実証の報告（長瀬）

### **インターネットWG**

- ① 総会後の AAMT ホームページの更新
- ② MT Summit に関するWEB対応

### **編集委員会**

- ① No.67(2018年3月発行)の原稿依頼、編集作業を行った。

#### AAMT/Japio 特許翻訳研究会

2017年10月13日(金)(2017年度 第4回)

1. 議事録の確認
2. 「ニューラル翻訳での未知語対応論文の調査」発表(筑波大学 宇津呂武仁先生)
3. ワークショップ開催報告
4. 次回の開催について
  - ・今後の研究会・拡大評価部会開催の日時予定(場所)
  - ・主な議題

#### AAMT/Japio 特許翻訳研究会 拡大評価部会

2017年10月13日(金)(2017年度 第2回)

1. 各グループの活動発表
  - ・「MTの文法能力を評価する」(王向莉様)
  - ・人手評価グループ
  - ・テストセットグループ
2. その他

#### AAMT/Japio 特許翻訳研究会

2017年12月8日(金)(2017年度 第5回)

1. 議事録の確認
2. 発表「単語の意味を考慮した自動評価法の提案」(北海学園大学大学院 越前谷博先生)
3. 平成29年度AAMT/Japio 特許翻訳研究会報告書作成スケジュールなど
4. 次回の開催について
  - ・今後の研究会・拡大評価部会開催の日時予定(場所)
  - ・主な議題

#### AAMT/Japio 特許翻訳研究会

2018年1月19日(金)(2017年度 第6回)

1. 議事録の確認
2. 発表「Fタームによる外国特許検索手法の検討」(静岡大学大学院 綱川隆司先生)
3. 平成29年度AAMT/Japio 特許翻訳研究会報告書タイトルと概要報告など
4. 次回の開催について
  - ・今後の研究会・拡大評価部会開催の日時予定(場所)
  - ・主な議題

AAMT ジャーナル 67 号をお送りします。

今号の巻頭言は、立教大学の武田珂代子先生より、御寄稿を頂きました。

今号に先立ちまして、昨年 9 月 18-22 日に、名古屋大学におきまして、AAMT の上位組織である機械翻訳国際連盟 (IAMT) が主催する第 16 回 MT サミットが開催されました。AAMT は本サミットの運営組織として、プログラム構成ならびに現地運営全般に携わりました。今号におきましては、本サミットのプログラム構成における各企画をご担当されました先生方から、ご担当の企画についての御寄稿を頂きました。

名古屋大学の中岩浩巳先生 (AAMT 会長) からは全体総括、ならびに、招待講演 “Neural Translation Technologies and Futures” の御報告を、京都大学の黒橋禎夫先生 (AAMT 理事) からは Research Track の御報告を、関西大学の山田優先生からは Users Track の御報告を、富士通研究所の長瀬友樹様 (機械翻訳課題調査委員会委員長) からは Technology Showcase の御報告を、日本翻訳連盟菊田直歩様からは JTF Workshop の御報告を、奈良先端科学技術大学院大学の須藤克仁先生からは PSLT ワークショップの御報告を、それぞれ御寄稿頂きました。

さらに、本サミットに御参加あるいは御出展された豊橋技術科学大学の井佐原均先生 (AAMT 前会長)、XTM International Ltd の目次由美子様、サン・フレアの工藤竜広様、山本ゆうじ様 (機械翻訳課題調査委員会 WG3)、の皆様からも、それぞれの御報告を御寄稿頂きました。

また、本サミットにおいて招待講演をされた Baidu Inc. からは、Zhongjun HE 様、Hua WU 様より同社の機械翻訳についての解説を御寄稿頂きました。

その他、イベント報告として、東芝デジタルソリューションズ島津美和子様からは、今号に先立ちまして、昨年 11 月 29 日に、AAMT の後援にて開催されました JTF 翻訳祭における AAMT 企画セッション 「ニューラル機械翻訳は翻訳プロセスをどう変えていくかー最近の機械翻訳技術と利用に関する動向ー」の御報告を御寄稿頂きました。また、XTM International Ltd の目次由美子様からは、2 月 2 日に開催されました、「Women in Localization Japan 第 13 回イベント報告」を御寄稿頂きました。

その他、「AAMT 会員のひろば」の企画におきましては、法人会員 1 件、個人会員 1 件の紹介文を掲載しました。

# AAMT

Asia-Pacific Association for Machine Translation

## AAMT 入会のご案内

AAMT は、機械翻訳の発展を目的として、機械翻訳の研究者、開発者、製造者、利用者が集まった任意の組織です。委員会による定期的な調査研究をはじめ、機関誌の発行、シンポジウムの開催など活動を行っています。

機械翻訳にご関心のあるすべての方にご入会をお勧めします。

### \*\* AAMT 会員の特典 \*\*

#### 1. AAMT Journal の購読ができます。

会員には、機関誌である AAMT Journal（年 2～3 回発刊予定）が送付されます。購読料は年会費に含まれています。

#### 2. 機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機械翻訳関連の情報をお届けします。

MT 新製品、新サービスの紹介、国際会議、シンポジウムのお知らせ、WEB での MT 関連記事の紹介など盛りだくさんです。

#### 3. AAMT が組織する委員会や調査活動に参加し、機械翻訳や翻訳に関心のある方との交流を深め、知見を広めることができます。

機械翻訳に関する言語資料の調査、広報、標準化活動に参加したり、AAMT Journal や会員専用メーリングリストで、自社製品、サービスの紹介を行うことができます。

#### 4. 関連機関の主催する国際会議に参加できます。

IAMT の主催で隔年開催される MT Summitをはじめ、AAMT、AMTA\*、EAMT\*\* の主催する会議やワークショップに参加できます。

AMTA\* : Association for Machine Translation in the Americas

EAMT\*\* : European Association for Machine Translation

#### 年会費は以下の通りです。

法人会員：入会金 1 口 10,000 円 年会費 1 口 50,000 円

個人会員：入会金 1,000 円 年会費 5,000 円（学生は学生会費 1,000 円）

ご関心のある方は、事務局までお問い合わせください。

アジア太平洋機械翻訳協会（AAMT）

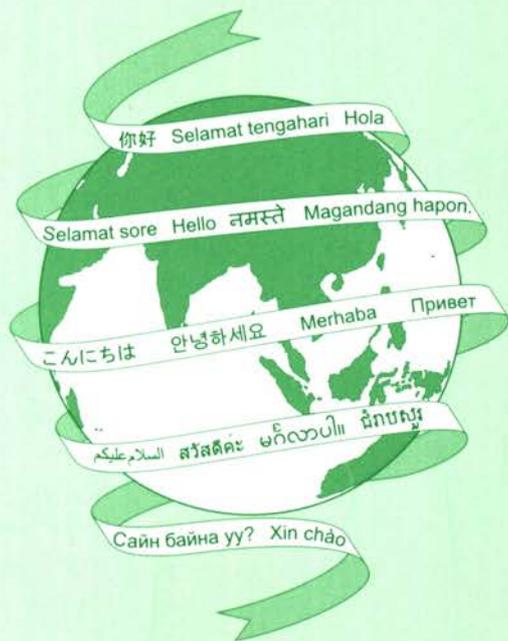
ホームページ：<http://www.aamt.info>

電子メール：[aamt-info@aamt.info](mailto:aamt-info@aamt.info)





# AAMT



## AAMTジャーナル No.67

発行：アジア太平洋機械翻訳協会 (AAMT)

ホームページ：http://www.aamt.info

住所：〒171-0014 東京都豊島区池袋2-55-2鈴木ビル3階  
(株)日本システムアプリケーション内

phone：03-5951-3961 fax：03-5951-3966

編集委員会：宇津呂 武仁 小谷 克則 大倉 清司

阿部 さつき 釜谷 聡史 河野 弘毅 目次 由美子

表紙(図部分)デザイン：阿部 さつき

事務局：神崎 享子 荻野 孝野

印刷所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)

c/o Japan System Application Co., Ltd.

Suzuki Building 3F 2-55-2, Ikebukuro, Toshima-ku Tokyo 171-0014, JAPAN

aamt-info@aamt.info

phone:+81(0)3-5951-3961 fax:+81(0)3-5951-3966