

ISSN 1883-1818
No.74
June 2021

AAMT Journal

Asia-Pacific Association for Machine Translation

機 械 翻 訳



AAMT創立30周年

機械翻訳

AAMT Journal

Asia-Pacific Association for Machine Translation 機 械 翻 訳

No.74

目 次

巻頭言		
「時は来たり」	隅田 英一郎	3
追悼		
長尾 真 初代AAMT会長を偲ぶ	隅田 英一郎	4
解説記事		
人間の翻訳と機械の翻訳(4):翻訳論の翻訳とMT研究の翻訳	影浦 峯	6
ニューラル機械翻訳のための目的言語側の文脈の効果的な利用	美野 秀弥	11
ニューラル機械翻訳を対象とした試行錯誤的プリエディット事例の探索的分析	宮田 玲・藤田 篤	15
書評		
「自動翻訳大全:終わらない英語の仕事が5分で片づく超英語術」	中澤 敏明	23
機械翻訳はどこへ行くのか?『機械翻訳:歴史・技術・産業』ティエリー・ポイボー[著]、高橋聡[訳]、中澤敏明[解説]	山田 優	27
イベント報告		
AAMT 2020, Online ~機械翻訳最前線~	出内 将夫	33
第7回アジア翻訳ワークショップ(WAT2020)開催報告	中澤 敏明	46
第6回特許情報シンポジウム	須藤 克仁	52
編集後記	早川 威士	54

C O N T E N T S

Foreword		
"The time has come."	Eiichiro Sumita	3
Obituary		
Remembrance of Makoto Nagao, the First Chairperson	Eiichiro Sumita	4
Commentary		
What do translation researchers and MT researchers say about translation?	Kyo Kageura	6
Effective Use of Target-side Context for Neural Machine Translation	Hideya Mino	11
Exploratory Analyses of Instances of Trial-and-Error-Based Pre-Editing for Neural Machine Translation	Rei Miyata / Atsushi Fujita	15
Book Review		
English Hacker's Handbook	Toshiaki Nakazawa	23
"Machine Translation" by Thierry Poibeau (trans, Akira Takahashi; comments, Toshiaki Nakazawa)	Masaru Yamada	27
Event Report		
AAMT 2020, Online --- Forefront of machine translation ---	Masao Ideuchi	33
Report of 7 th Workshop on Asian Translation (WAT2020)	Toshiaki Nakazawa	46
The 6 th Symposium on Patent Information	Katsuhito Sudoh	52
Editor's Note	Takeshi Hayakawa	54

「時は来たり」

隅田 英一郎

一般社団法人アジア太平洋機械翻訳協会（AAMT）会長

第2次世界大戦直後の1946年から約70年間にわたり、機械翻訳は、利用側において膨張する **FAHQMT**（「完全自動高品質機械翻訳」の英語の頭字語）に対する期待からの予算重点化（アクセル）と、得られた翻訳精度の低さに対する落胆からの予算カット（ブレーキ）を交互に繰り返してきた。

まず、日本における機械翻訳の歴史を簡単に振り返る。AAMT 創立前の1980年代に、機械翻訳システムの開発ブームが起こり、富士通、東芝、NTT等をはじめとして10社を超える有力企業がアクセルを踏みこんで熾烈な競争を繰り広げた。

AAMTはこの状況を踏まえて、1991年4月17日に「日本機械翻訳協会」の名称で機械翻訳システムの開発・利用を促進する目的をもって創立された。翌年の1992年6月17日に「アジア太平洋機械翻訳協会」に名称変更しスコープを拡大した。さらに、2020年に8月24日に名称を変更せず一般社団法人に移行し、2021年4月17日に創立30周年を迎えた。

AAMT 創立当初の機械翻訳は対象分野が限定されており翻訳精度も高くはなかった。それにもかかわらず当時は過当競争の中、勢いで使った感が否めない。機械翻訳による低品質の訳文を翻訳者が修正する発注（ポスト・エディットと呼ぶ）が頻繁に行われた。これは大きな禍根を残す不幸であった。翻訳者は「最初から自分で翻訳した方がずっと効率的だ」とその苦役に対して怒った。翻訳者も翻訳会社も疲れ切り後進に機械翻訳の恐ろしい害毒について伝承続けた。コーパスに基づく手法、統計的な手法など機械翻訳技術には格段の進歩があっても、各企業の研究開発も大規模予算投入があだとなった後遺症から再参入にブレーキがかかり続けた。

文法も語彙も大きく異なり、最も翻訳が難しい日本語と英語の間では機械翻訳は広く普及はしなかった。

2014年にニューラル機械翻訳の芽が出てからは大きく変わった。実用化を阻んできた翻訳精度の天井は軽やかに破られ、技術進歩は加速し、たった数年間にニューラル機械翻訳は第三世代が上市された。ニューラル機械翻訳の進化を遮るものは今は無い。HUMAN PARITY(人間並み)の翻訳力とか、TOEICで高得点とか等と喧伝されるが、実際に非常に高精度になった。もっとも、出来たのは **FAHQMT** ではなく **FA** の付かない **HQMT** であって、ツールとして使いこなす人間が必須である。翻訳の主体はあくまでも人間だ。

ニューラル機械翻訳の勃興とAAMTの創立30周年とが同期した。まさに「時は来たり」である。

過去のトラウマを乗り越えて、ニューラル機械翻訳の普及を促進するのがAAMTの使命である。そのため、AAMTは今年度からポスト・エディットに注力する。具体的には、ポスト・エディットのガイドラインを作成したり、人材募集をサイトで公開したりする。研究開発側と利用側で一体となり、効率的な多言語コミュニケーションが可能な世界を創出する。

もう一つの大きな波が新型コロナウイルスによってもたらされた。人類はこのウイルスに翻弄されながらも、自らの生活・社会・経済を不可逆的に変化させている。イノベーションを阻むのは変化を嫌う心理だが、コロナ・ショックですべてが変わりつつある。特にコミュニケーションはリモート中心になり、多言語コミュニケーションの要である翻訳は機械翻訳が重用されていくだろう。

AAMTは、翻訳の未来を創ることをここに宣言する。

“The time has come.”

Eiichiro SUMITA

Asia-Pacific Association for Machine Translation (AAMT)

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

長尾 真 初代 AAMT 会長を偲ぶ

隅田 英一郎

一般社団法人アジア太平洋機械翻訳協会

AAMT の創設者の長尾真先生が、2021 年 5 月 23 日永眠されました。心よりお悔やみ申し上げます。

先生は、コンピューターによる画像処理、機械翻訳、電子図書館等で大きな功績を残された「知の巨人」です。巨大組織運営で発揮された「豪快さ」、一人一人の個人への対応で発露された「優しさ」が印象的です。



長尾先生は 1970 年代半ばから本格的に機械翻訳の研究を開始され、パイオニアの一人として機械翻訳の研究開発の基盤を築かれました。当時、10 年前に提出された ALPAC 報告書で機械翻訳の将来性が否定された米国では研究が停滞していましたが、欧州や日本では研究が進んでいました。先生は 1978 年に「論文表題」の英日翻訳システム（TITRAN）を公開し研究・開発の大規模化に向けた狼煙を上げられました。1982 年から 4 年間の国家プロジェクト¹で「論文抄録」の日英機械翻訳システム（Mu）の開発を指揮し、さらに、富士通、東芝、NEC、等 10 社を越える日本企業の事業化競争を加速しました。この大きな盛り上がり背景に、1987 年に先生は世界中から様々な立場のステークホルダーを箱根に集め、機械翻訳について議論する第一回機械翻訳サミット（MT Summit）を主催しま

した。1991 年には International Association for Machine Translation (IAMT) を設立（IAMT の初代会長も長尾先生です。）し、IAMT の配下に、世界を 3 つ領域に分けて担当する姉妹団体 AAMT、AMTA、EAMT を設立し、MT Summit を 3 団体で持ち回り開催することとしました。長尾先生は一気呵成に世界の機械翻訳を引っ張るキーパーソンとなりました。

TITRAN や Mu をはじめとする当時公開された全ての機械翻訳システムで用いられた方式はルールに基づく翻訳（RBMT、Rule-Based Machine Translation）です。RBMT では、原文を解析して単語や語順を変換し訳文を生成します。解析のルールと、単語の変換辞書と語順の変換ルールと、語形変化等の生成のルールを作成し、それを参照して翻訳するのです。

驚愕すべきことに、Mu が始まる 1 年前の 1981 年に、長尾先生は『RBMT と 180 度方向性が違う』方式、EBMT を発表されていきました。先生は、1978 年の TITRAN 以降、RBMT 方式のシステムを繰り返し試作する中で、ルールに内在する問題に気づかれてルールに捉われない方式が必要と考えられました。人間の外国語学習を観察して、対訳を記憶し類似対訳を模倣することをコンピューターにさせる方式、EBMT を提案されました。ルールは使わず対訳データを基盤とする「ノー・モア・ルールズ」の方式です。長尾先生は人工知能に関するワークショップで発表されましたが、良い反応はなかったそうです。また、本論文を含む会議録²が出版されたのは 1984 年で、当時、世界は RBMT による製品化の只中にあり、EBMT に注目し発展させる研究者もありませんでした。長尾先生の自

¹ “The Japanese government project for machine translation,” in Computational Linguistics.

² “A framework of a mechanical translation between Japanese and English by analogy principle,” in Proc. of the Int'l NATO symp.

伝『情報を読む力、学問する心』によれば、Mu 開始前にいずれの方式を採用するべきかで「大いに迷った。」とあります。EBMT は優れたアイデアでしたが、研究に必要な対訳データや計算機性能が不十分だったので先生は Mu に EBMT を採用されませんでした。早すぎた EBMT は環境変化を待つことになりました。

もう一つ驚愕すべきことがあります。長尾先生は、1990 年代に入ると機械翻訳から次のテーマの電子図書館に軸足を移されました。先生は一つの研究テーマに長居はしないという方針をお持ちだったのです。あるテーマの基盤ができれば、そのテーマをやりつくすのではなく他の研究者や若い人に任せて、未開拓のテーマに挑戦することを大事になさったのです。実際、1990 年代に入ると EBMT の研究も進みはじめ、同時に、「ノー・モア・ルールズ」の別方式である IBM 提案の SMT も登場しました。両者は切磋琢磨し精度向上を競いました。しかし、2010 年ごろには両者とも頭打ちになってアルゴリズム的に行き詰まりました。この 2010 年ごろまでに「ノー・モア・ルールズ」の手法に好都合な環境整備は大きく進みました。①対訳データの大規模化、②計算機の高性能化、③自動評価の発明、④同じデータで複数アルゴリズムを評価するコンペの定着等です。このタイミングで、深層学習の高速実行を可能にする GPGPU が出現し画像認識で成功しました。次に、隣接分野である音声認識で成功しました。2014 年に翻訳へ応用されました（この方式を NMT と呼びます）ところ、従前の手法ではいかんともしがたい天井と思われていた精度限界を越えました（第 1 世代）。天井に穴が開いた途端、研究者は雪崩のように NMT に転向し一気に研究は加速し 2017 年には更に精度をあげました（第 2 世代）。NMT の勢いは衰えず 2020 年にも大幅な精度向上を実現しました（第 3 世代）。3 年に一度、革新が起こり、システムが上市されています。研究から実用化への軽やかなこの展開は機械翻訳の長い歴史の中で誰も経験しなかったものです。第 3 世代では大規模対訳に基づく汎用の NMT と小規模高品質対訳に基づく EBMT が見事に融合しました。

長尾先生は目を見張る精度向上に満足され、英語が達者な長尾先生ですが、「英作文に役立つという理由で先生ご自身が NMT を活用されている」ことを、ニコニコと楽しそうに語っておられました。

東奔西走で活躍される元気なお姿を日頃拝見していたので喪失感は大きいです。しかし、『情報を読む力、学問する心』は我々の励みになる言葉に満ちていますし、先生の息遣いを感じることができますから、同書を手元に置けば、長尾先生は永遠です。

略歴

- 1936 年 三重県に誕生。
- 1955 年 大津東（現、膳所）高等学校卒業。
- 1959 年 京都大学工学部電子工学科卒業。
- 1961 年 同大工学研究科修了。
- 1966 年 工学博士（京都大学）。
- 1973 年 京都大学教授に就任。
- 1978 年 論文表題の英日翻訳システムを公開。
- 1982 年 科技厅より論文抄録の機械翻訳システム構築を受託。
- 1987 年 第 1 回 MT Summit を主催。
- 1991 年 International Association for Machine Translation (IAMT) を設立。
- 1994 年 言語処理学会を設立し初代会長に就任。
- 1997 年 京都大学総長に就任。
- 2004 年 情報通信研究機構初代理事長に就任。
- 2007 年 国立国会図書館長に就任。
- 2015 年 国際高等研究所所長に就任。

受賞

- 1993 年 IEEE Emanuel R. Piore Award。
- 1997 年 紫綬褒章。
- 2003 年 ACL Lifetime Achievement Award。
- 2005 年 日本国際賞とフランス共和国レジオンドヌール勲章シュヴァリエ章。
- 2018 年 文化勲章。

人間の翻訳と機械の翻訳（４）：翻訳論の翻訳と MT 研究の翻訳

影浦 峯

東京大学・大学院教育学研究科

1. これまでのまとめと今回の話題

本連載では、これまでに、翻訳者は何を翻訳しているか、翻訳者が対象としている文書とはどのようなものか、そして、翻訳のプロセスはどのような要素から構成されるか、を論じてきました。前回の最後に、今回は、メタ言語についてか、あるいは機械翻訳研究と翻訳研究のメタ科学的な対比か、いずれかを扱う予定と言いました。これは、いわゆる産業翻訳を中心とした人間の翻訳について前回までで一応の姿を示したと考えたからなのですが、作業を行う人間の側のコンピテンスやスキルがどのようなものかについても、どこかでまとめて論ずる必要はありそうです。

とはいえ、一応、対象に対する操作として捉える限りでの翻訳の大枠はこれまでで示したので、今回は、少し趣向を変えて、翻訳論で語られている翻訳と、機械翻訳研究で語られている翻訳とがどう違うのかを、計量情報学的な観点からみてみたいと思います^[1]。これまでに発表した簡単な分析に基づき、その基本的な部分を紹介するものです^[2]。翻訳論と MT 研究では分野の性質も大きく違うので、対応するコーパスを使って両者の相対的な差異を検討するのではなく、翻訳論の基本文献を参照して MT 研究は其中でどのような範囲を扱っているかというかたちで考えます。

2. 分析の対象

MT 研究の側では、The ACL Anthology Reference Corpus^[3]の 2010 年から 2015 年のデータのうち、表題に“MT”あるいは“machine translation”を含む 566

論文を分析の対象とします(以下 MT と略記します)。

2015 年までと少し古いので、NMT 関係の論文はまだ少ないのですが、その後の流れを見ても、技術的な展開は別として「翻訳」をめぐる問いはそれほど変わっていないのではないかと思います(本来ここも計量書誌学的な検討対象ですが)。

翻訳論の側では、教科書と辞書を、少し別のかたちの分析に利用します。教科書として、Newmark (1988) と Munday (2001) (それぞれ NW と MN と略記) の二つを使いました^[4]。いずれも少し古いのですが、定評のあるもので、翻訳論の範囲を俯瞰するには便利ですし、また、翻訳論の発展を考えるとそう大きな問題にはならないと思います。辞書として Shuttleworth & Cowie (1997) と Laver & Mason (2018) (それぞれ SC と LM と略記) を使いました^[5]。

計数にあたって、MT、NW、MN は名詞列を抽出し、基本形を使いました^[6]。SC と LM は見出し語のうち英語以外のものを除去しました。こちらは見出し語の主要部のみをステミングし、MT コーパスとのマッチングを取りました。MT、NW、MN の基本的な数量は以下の通り、SC は 422 見出し、LM は 1,022 見出しです。

データ	文	名詞延べ	名詞異なり
MT	88,736	681,486	43,193
NW	6,472	38,004	9,694
MN	5,067	28,068	4,653

MT、NW、MN はかなり汚いデータをほぼそのまま使っているため、数値は大まかな傾向に過ぎません。

3. 翻訳論の教科書と MT コーパス

ここでは、NW と MN を参照して MT の性格を考え

What do translation researchers and MT researchers say about translation?

Kyo Kageura

Graduate School of Education, The University of Tokyo

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

ます。それぞれから上位 100 語を抽出し、共通する部分と異なる部分を観察します^[7]。以下に、重なりを示します。

ALL	NW/NM	NW/MT	MN/MT	NW	MN	MT
31	22	4	7	43	40	58

3 データに共通する名詞が 3 分の 1 弱、翻訳論の教科書間では半分以上共通していて、MT は少し離れていることが伺えます。3 データに共通する 31 語を見ると、まず“translation,” “English,” “language”があって、さらに言語要素を示す名詞が 6 つ (“sentence,” “phrase,” “word,” “context,” “term”) 観察されました。残りは “work,” “type,” “example”などの一般的な名詞で占められています。翻訳論の教科書 2 つに共通し MT にはない名詞として以下は注目に値します。

人間 : “translator,” “reader,” “author.”

言語関連 : “style,” “meaning,” “concept,” “sense.”

文書関連 : “book,” “writing.”

これに対して、MT と NW あるいは MN の 2 データに共通するものは、“verb,” “grammar”といった文法レベルの言語要素を表す言葉と、一般的な名詞です。MT にしか現れないものとして“data,” “corpus,” “sequence”など、記号関係を情報学的観点から捉えた言葉があり、NWやMNでは “metaphor,” “discourse,” “poetry,” “literature”など、文法以外の言語関連概念や文書関係概念を表す言葉が目につきます。少し論点先取りの過誤になってしまう危惧はありますが^[8]、“text”が 3 データの上位 100 語に共通するのに対して、“book”と“writing”は MT では 100 語に入りません^[9]。

以上から、少し論証手続きは粗いのですが、大きく、以下の、二つの点がわかります。

1. 翻訳論では著者や読者、翻訳者といった人を含め様々な文書を明示的な検討の対象とするのに対し、MT 研究ではテキストの処理を扱うこと。
2. 翻訳論では言語学的な対象のテキストを意味論や語用論も含めて考えるのに対し、MT 研究では文法的・記号处理的な観点が中心であること。そもそも MT 研究はそのようなものだというのはそ

の通りで、MT そのものについては当たり前のことをわざわざ面倒くさく確認しただけですが、翻訳論の範囲との差異は重要です。そして、MT でカバーされていなくて翻訳論ではカバーされている領域は、本連載でこれまで行ってきた翻訳に対する解説とそれなりに対応していることがわかります。連載第一回で検討した言語／テキストと文書の区別：

- 文／段落／テキストを翻訳する。
- 図書／小説／修理書を翻訳する。

の中で、MT は前者を、翻訳論は後者を扱っており、それらは質的に異なるので、MT が処理対象とする基本セグメントを文からより大きな単位に拡張したとしても、翻訳論が対象とした実際に人間が行っている翻訳の中でカバーできない領域は依然として残ることになります^[10]。

4. 翻訳論の辞書と MT コーパス

SC と LM の共通の見出しは 202 で、そのうち 87 が MT に出現しています。SC のみの見出しは 220 でそのうち 69 が、LM の見出しは 820 でそのうち 227 が MT に出現しています。ここでは、前節で翻訳の捉え方について見たことを受け、見出し語すべてではなく、“translation”に関する範囲に着目し、この語を主要部に持つ見出しを観察することにします。SC には 168 見出し、LM には 204 見出しあります。重なりと MT の出現状況に関する基本データを以下に示します。

	SC/LM	SC	LM
ALL	61	107	143
MT	16	28	31

このうち、SC と LM に共通する 61 見出し（そのうち一つは“translation”なので実際には 60 見出しで、MT に出現するのはそのうち 15 見出しとなります）を、少し詳しく見てみることにします。“translation”を主要部とする複合語のうち、MT に現れるものを、以下に示します。修飾部のみを、一応の目安として頻度とともに、示すことにします。

machine (487), automatic (47), un (27), partial (20), full (17), direct (17), literal (13), free (7), re (5), word-for-word (3), close (2), idiomatic (2), indirect (2), reverse (1), parallel (1)

緩やかな目安ですが、概念的には、これらはいずれも、テキスト変換の方法や様態を形容するものと考えることができ、(a) “machine,” “automatic”、(b) “partial,” “full,” “un”、(c) “literal,” “free,” “word-for-word,” “close,” “idiomatic”、(d) “direct,” “re,” “indirect,” “inverse,” “parallel”のように（あくまで参考程度ですが）グループ化できそうです。MT や言語処理の研究者は、これらが何を意味するか、改めて辞書を確認しなくてもそれなりに理解できるのではないのでしょうか。

SC と LM に共通し、MT には現れない 45 見出しを、上の緩やかなグループとやはり緩やかに対応づけて示すと、以下のようになります。(a) “machine-aided”、(b) “gist,” “over,” “restricted,” “under”、(c) “abusive,” “exegetic,” “gloss,” “homophonic,” “interlinear,” “oblique,” “phonemic,” “phonological,” “rank-bound,” “semantic,” “sense-for-sense,” “thick,” “unbounded”、(d) “interlingual,” “intermediate,” “second-hand,” “serial”。

これらの他に、“aesthetic-poetic,” “blank verse,” “communicative,” “covert,” “cultural,” “documentary,” “ethnographic,” “fictitious,” “horizontal,” “instrumental,” “intersemiotic,” “intertemporal,” “intralingual,” “linguistic,” “linguistically creative,” “loan,” “metrical,” “overt,” “pragmatic,” “radical,” “rhymed,” “service,” “sight,” “vertical”があります。このうち、“aesthetic-poetic,” “blank verse,” “communicative,” “linguistic,” “linguistically creative,” “loan,” “metrical,” “oblique,” “pragmatic,” “rhymed”は(c)のグループに近そうです。ところで、これらの形容詞を有する用語のうち、どのくらいを本誌の読者の方々にご存知でしょうか。翻訳論を特に意識して学んだことがない限り、このうちいくつかは、どのような概念を表しているのかわからないのではないかと思います^[11]。

このリストを眺めていくつか興味深い点に気づ

きます。一つは“linguistic”が明示的な形容の言葉としてあることで、これは翻訳が無標の概念としては言語（学）的なものではない（に限らない）ことを示しています。これに対して、翻訳論の見出しの形容詞からは、二つの方向性が伺えます。第一は、“intersemiotic”のように、翻訳が扱う記号の系列を拡大して捉えるもので、これは言語処理におけるマルチモーダルに対応するでしょう^[12]。第二は“aesthetic-poetic,” “ethnographic,” “documentary,” “rhymed”のように、一般に記号的対象の外にある、文書が置かれたより広い文脈（言語学的な意味ではない）を考慮する方向に拡大するものです。この一つ、“rhymed translation”は、例えば歌詞を翻訳するときに目標言語側でも歌えるようにしなくてはならないといった制約のもとでどうするかという話で、こちらは、問題を適切に定義すれば、end-to-end の学習でそれなりに扱うことができそうです^[13]。

もう一つは、敢えて(c)に近いのではないかと挙げた形容詞の一部が、“abusive”や“oblique”のように、今のところ、人間が割と全体を先取りして理解しているという前提で何かするものと捉える以外にどのようなものであるかがわかりにくいという点です^[14]。このことは、MT 研究が翻訳の一部しか扱っていないということだけでなく、MT 研究で扱われていない翻訳の領域の一定程度は、そもそもそこで何が行われているのか、翻訳論の領域において、その機序が十分な解像度で解明されていないことを示しているようです。

5. おわりに

翻訳論の教科書と辞書を参照し、MT が扱う「翻訳」の範囲をメタ科学的に観察してみました。ここまでの検討で直接主題化していないことも含め、少し大きくまとめると、MT では、

1. 起点言語テキストを入力とし目標言語テキストを出力とするという問いが与えられ、一般に、起点言語テキストが文書として存在しているときのより広い文脈における属性やコミュニケー

- ション上の位置づけは考慮されないこと
2. その上で、処理を一般に言語学的なプロセスとして捉えることが観察されます^[15]。

翻訳論で扱われていて MT で扱われていない領域のうち、一部は、そのような問いが存在することが共有されれば、制約に対応した損失関数を定義してそれなりに扱うことができるようになるのではないかという程度に、翻訳論の領域での認識の解像度が高いあるいは高くするための条件は整っていると考えられます。それに対して、今のところ人間にはできるけれども、そのことができるがそもそものようなものが形式的な知識として定式化されておらず、また、手続き上の制約としても捉えられていないために、今のところ、MT が扱う課題として定式化するために必要な人間の側の認識が決定的に欠落している領域があります（例えば「文化」的差異を考慮した翻訳といったものがその典型です）。これを人間の創造性のすばらしさを示すものとしてではなく、人間知性の不足を示すものとして捉え、科学的に明確にしていくことが翻訳と MT 双方にとって大きな課題であるという認識が共有されるとよいと思います。

謝辞

本記事の一部は、科研研究費基盤(S)「翻訳規範とコンピテンスの可操作化を通した翻訳プロセス・モデルと統合環境の構築」(19H05660)に関わっています。

注・参考文献

- [1] この文章の著者は図書館情報学研究室に所属しており、計量情報学の一つの源流である計量書誌学(本や論文を対象とする)の一部は図書館情報学の中で展開されてきました。確か『詩の構造についての覚え書一ぼくの「詩作品入門」』(思潮社)でだったと思いますが、入沢康夫が本棚が詩にならないのはどうしてだろうといったようなことを言っていたと記

憶しています。詩というものの一般的な印象のもとでの特殊性をとりあえず無視すると、これは、連載の第1回で述べた、他でもありうる形式的に妥当な範囲ではなく、結局のところこうでしかなかったものとしての言語表現を、入沢康夫の言葉は言葉の側から、図書館情報学は文書の側から、捉えようとしているものと理解することもできます。なお「計量情報学」という言葉を使いましたが、以下では、あまり厳密ではないものの、質的な分析もします。

- [2] Kageura, K. (2019) “What does MT research say about translation? An empirical sketch,” *Proceedings of the 2019 ERITS Conference*, Seoul, November 9, 2019, pp. 12-18; Kageura, K. (2020) “The gap between MT research and translation studies: Observations on the difference in how ‘translation’ is viewed in the two fields,” *IATIS Regional Workshop*, Online, September 3, 2020.
- [3] Bird, S., Dale, R., Dorr, B., Gibson, B., Joseph, M., Kan, M.-Y., Lee, D., Powley, B., Radev, D. and Tan, Y-F. (2008) “The ACL anthology reference corpus: A reference dataset for bibliographic research in computational linguistics,” *LREC 2008*, pp. 1755-1759.
- [4] Newmark, P. (1988) *A Textbook of Translation*. New York: Prentice Hall.; Munday, J. (2001) *Introducing Translation Studies*. London: Routledge. 前者は実践より、後者は理論よりです。
- [5] Shuttleworth, M. and Cowie, M. (1997) *Dictionary of Translation Studies*. Manchester: St. Jerome.; Laver, J. and Mason, I. (2018) *A Dictionary of Translation and Interpretation*. <http://fit-europe-rc.org/wp-content/uploads/2019/05/Dictionary-of-translation-and-interpreting-Mason-Laver.pdf>
- [6] Genia Tagger を使いました。Tsuruoka Y. et al. (2005) “Developing a Robust Part-of-Speech Tagger for Biomedical Text,” Bozanis P. and Houstis E. N. (eds) *Advances in Informatics - 10th*

Panhellenic Conference on Informatics, LNCS 3746, Berlin: Springer, pp. 382-392.

[7] MT と NW・MN でそもそもサイズが一桁違います
が Zipf 則を踏まえ異なり数を見ると上位 100 語は安定していると考えてもそう悪くないでしょう。

[8] 一応、質的分析の手続きは参考にしています。大谷尚 (2019)『質的研究の考え方—研究方法論から SCAT による分析まで—』(名古屋大学出版会)等。

[9] MT では “book” は 1220 位、“writing” は 2859 位。

[10] 翻訳に関心がある人に「翻訳者になったとして、主に何を翻訳したいですか？」と聞いて、「私が翻訳したいのは文／段落／テキストです」という答えが返ってくることはほとんどありません。「小説です」「特許文書です」「ニュース記事です」といった答えが返ってきます。ちなみに、「小説」等とはレベルがずれますが)「図書」や「文書」は、図書館情報学の中核的な研究対象です。翻訳が第一義的に扱うのは言語／テキストではなく文書であるという点は、長尾真先生が図書館情報学に持続的な関心を持たれていたことに関係があるのかもしれませんが。

[11] 筆者も、わからないものが複数ありました。

[12] 興味深いことに Jakobson が “intralingual,” “interlingual,” “intersemiotic” の区別を導入した論文 (1959 年) のタイトルは “On linguistic aspects of translation” です。

[13] 「カチューシャ」の日本語歌詞 (関鑑子訳) は、
りんごの花ほころび

川面に霞たち

君なき里にも

春はしのびよりぬ

ですが、元のロシア語は

Расцветали яблони и груши,

Поплыли туманы над рекой;

Выходила на берег Катюша,

На высокий берег, на крутой.

で、りんご (яблони) だけではなく梨 (груши) の花もほころんでいますし (秩父では梨の花はりんごの花よりだいぶ遅く咲きます。地域の問題なのか

洋梨と和梨の違いなのかは調べていません)、後半は、ロシア語ではカチューシャの行動を記述して思う人の不在を示唆しているのに対して日本語ではカチューシャからの光景になっています。ロス・マクドナルドの *The Way Some People Die* の日本語表題 (中田耕治) は『人の死にゆく道』で、日本語表題の方が哲学的ですが、内容と記述のレベルを考えるとやはり原題にある “some” は非常に重要です。もちろん、訳者がそれを気づかなかったわけでは恐らくなく、(a) 日本語で “some” をうまく表示しにくい、(b) 小説の表題としてはなおさら難しい、という要因があったのだと思います。Google 翻訳では『一部の人々が死ぬ方法』となりました。表題としては難しいだろうという点に加え、今度は記述の対象が「方法」側に偏り過ぎていて『人の死にゆく道』よりも大きく原題とずれてしまっている感じがします。

歌詞の自動生成は複数論文がありますが、歌詞の自動翻訳は (私が知らないだけかもしれませんが) あまりないようです。Nishimura, A., Itoh, T. (2020) “Implementation and experiments for interactive lyrics transcreation system,” *Visual Computing and Industry, Biomedicine, and Art* 3(17) ではルールベースの手法が使われています。

[14] 例えば LM では、“oblique translation” は “any translation where the translator judges that a direct translation respecting the lexical or grammatical sequence of the source text would yield neither an adequate nor an acceptable target text, and therefore changes the latter’s linguistic structure, style or meaning.” と定義されています。

ニューラル機械翻訳のための目的言語側の文脈の効果的な利用

美野 秀弥

NHK 放送技術研究所

1. はじめに

本記事では、The 28th International Conference on Computational Linguistics (COLING2020) に採択された論文である“Effective Use of Target-side Context for Neural Machine Translation” (Mino et al. 2020) の主要な部分の解説を行う。COLING は 2 年に 1 度開催されている計算言語学に関する国際会議であり、2020 年は、併設ワークショップ・チュートリアルを含め、12 月 8 日から 13 日までの 6 日間で開催された。スペインのバルセロナで開催される予定だったが、コロナウイルスの影響によりオンラインでの開催となった。解説する論文は文脈を考慮した機械翻訳に関するものである。従来の機械翻訳システムの多くは文単位で独立に翻訳を行うことが多いが、単文が持つ情報だけでは適切な翻訳を行うことができない場合がある。その一つに、訳語の統一の問題がある。例えば、同一文書内で、“source language” の日本語訳として“起点言語”、“原言語”、“ソース言語”を混在して用いることは一般的に好ましくない。また、文末表現において「ですます調」と「である調」が混在した文書も良い文書とは言えない。単文が持つ情報だけではこれらの表現を統一して翻訳することは難しい。また、語義曖昧性解消も単文翻訳の問題となる。例えば、“I go to the bank.” を日本語に翻訳する場合、“bank” が“銀行”なのか“土手”なのかをこの文だけで判別して翻訳することは難しい。他にも、日本語などに代表されるゼロ代名詞が含まれる言語から主語の省略を許さない英語などへの翻訳ではゼロ照応解析が必要となる場合が多いが、単文翻訳では文を跨いだゼロ照応

解析は行えない問題もある。これらの問題は、翻訳対象の前後の文などの文脈情報を使うことで解決できる可能性があり、多くの先行研究が存在する。文脈を考慮した機械翻訳に関する先行研究を知りたい方は Maruf ら (2021) の論文を参照されたい。本記事が解説する論文では、既存手法の中から、目的言語側の文脈情報を用いた機械翻訳 (Bawden et al. 2018) に着目している。Bawden ら (2018) は目的言語側の前文を機械翻訳器の入力に追加するモデルを提案したが、翻訳精度の向上が見られなかったと報告している。しかしながら、例えば訳語の統一のためには前文でどのように翻訳されていたかが重要な情報となり、目的言語側の文脈情報が有益な場合もあると考えられる。本記事では、従来研究の問題点の洗い出し、目的言語側の文脈情報を用いた翻訳精度を向上させる提案手法、提案手法の効果を確認するために行った実験について解説する。

2. 従来手法の問題点

目的言語側の文脈情報を用いた機械翻訳の従来手法では、以下のように学習、翻訳を行う。

【学習】

目的言語側の文脈情報の参照訳と原言語側の翻訳対象の文を入力とし、翻訳対象の文の参照訳を出力するように翻訳モデルを学習する。

【翻訳】

目的言語側の文脈情報の機械翻訳結果と原言語側の翻訳対象の文を学習済みの機械翻訳器に入力し、機械翻訳結果を得る。

このように、目的言語側の文脈情報として、学習時は参照訳を使うのに対して翻訳時には機械翻訳結果を使う。著者らは、この違いが問題点となっていると考えた。参照訳は目的言語側の文脈情報を正しく翻訳しているのに対して、機械翻訳結果は訳抜け、過剰訳、誤訳などの機械翻訳特有の誤りが含まれる。仮に機械翻訳結果に誤りがない場合でも、機械翻訳の出力には Translationese (翻訳調) と呼ばれる偏りが含まれることが知られており、Toral (2019) は、機械翻訳は人手翻訳よりもシンプルかつ標準化された結果になる傾向にあると示している。従って、参照訳と機械翻訳結果との間には、機械翻訳特有の誤り、および表現の多様性の面において特徴の違いがあると考えられる。

学習時と翻訳時で特徴の異なるデータを用いることで生じるモデルの偏りは exposure bias (Bengio et al. 2015) と呼ばれ、翻訳品質の低下に繋がることが報告されている (Zhang et al. 2019)。特徴の異なる参照訳と機械翻訳結果を学習時と翻訳時で別々に用いる従来手法では、文脈情報を考慮した翻訳モデルが効果的に学習されずに精度低下を引き起こす可能性がある。そこで、学習時と翻訳時に用いる文脈情報の特徴の違い (ギャップ) の問題を低減させる手法を検討した。

3. 提案手法

学習時と翻訳時の文脈情報のギャップをなくす最も簡単な手法は同じ情報を使うことである。すなわち、学習時と翻訳時ともに参照訳を使うか、機械翻訳結果を使うか、の二択である。効果的に文脈を利用するという観点では翻訳誤りのない参照訳を使うことが好ましいと考えられるが、翻訳時には参照訳が存在しないために機械翻訳結果を用いることしかできず、結果的に学習時と翻訳時の両方で機械翻訳結果を用いる選択肢のみが残る。しかし、翻訳誤りが含まれ、翻訳調となった機械翻訳結果を用いて学習することで、文脈情報から適切な情報を抽出することができなくなる可能性がある。すなわち、翻訳時と同じ特徴を有する目的

言語側の文脈情報を使うという観点では機械翻訳結果を用いることが好ましいが、効果的な学習という観点では参照訳を用いることが好ましい。そこで、学習時に、翻訳時と同じ機械翻訳結果のみを用いるのではなく、参照訳と機械翻訳結果の両方を用いる手法を提案した。この手法は、カリキュラム学習 (Bengio et al. 2009) とスケジュールサンプリング (Bengio et al. 2015) のアプローチを参考にしたサンプリング制御法である。提案手法は、下記の手順により、学習エポック (訓練データ全体の学習の単位) ごとに、参照訳を文脈情報として用いるデータの割合 (サンプリングレート p) と機械翻訳結果を文脈情報として用いるデータの割合 (サンプリングレート $1-p$) を変更して学習データを再構成し、翻訳モデルを学習する。

- (1) 1 エポック目の学習には、目的言語側の文脈情報が全て参照訳 (サンプリングレート $p=1$) である学習データを用いる。
- (2) 2 エポック目以降の学習には、変更したサンプリングレート p に従い、参照訳の割合と機械学習結果の割合を変えて再構成した学習データを用いる。

サンプリングレート p は、逆シグモイド関数を用いて計算し、学習エポックごとに変更する。学習が進むにつれて p の値は小さくなり、段階的に参照訳から機械翻訳結果へと学習データ内の文脈情報を変更することで、翻訳誤りに頑健で翻訳調の特徴に対応した翻訳モデルの学習を実現する。

4. 実験

提案手法の効果を、時事通信社ニュースをベースにしたニュースコーパスを用いた英日・日英機械翻訳実験と、IWSLT2017 (Cettolo et al. 2012) で公開されている TED トークコーパスを用いた英日・日英、および英独・独英機械翻訳実験で確認した。

比較手法として、下記の3種類の手法を用いた。

手法	ニュースコーパス		TED トーク コーパス			
	EN-JA	JA-EN	EN-JA	JA-EN	EN-DE	DE-EN
Single-Sent.	41.99	24.23	15.69	8.48	24.04	28.46
Trg-Context GT	42.40	24.80	16.15	8.98	25.96	30.25
Trg-Context Pred.	42.45	24.31	16.27	9.15	25.98	30.23
提案手法	42.79	24.86	16.37	9.66	26.01	30.60

表 1：実験結果

手法1. 文脈情報を利用しないニューラル機械翻訳手法 (Single-Sent.)

手法2. 学習時の文脈情報として目的言語側の参照訳のみを用いたニューラル機械翻訳手法 (Trg-Context GT)

手法3. 学習時の文脈情報として目的言語側の機械翻訳結果のみを用いたニューラル機械翻訳手法 (Trg-Context Pred.)

手法 1 は Transformer モデル (Vaswani et al. 2017) のエンコーダとデコーダを用いて実装した。手法 2, 3 は文脈情報を用いたニューラル機械翻訳手法であり、結合ベース文脈考慮型ニューラル機械翻訳とマルチエンコーダ文脈考慮型ニューラル機械翻訳の 2 種類での実装が可能である。どちらもエンコーダとデコーダにより構成されるが、文脈情報の入力方法が異なる。結合ベース文脈考慮型ニューラル機械翻訳は、文脈情報と翻訳対象の文とを特殊記号で結合して 1 つのデータとみなし、単独のエンコーダに入力するアーキテクチャである。マルチエンコーダ文脈考慮型ニューラル機械翻訳は、2 つのエンコーダと 1 つのデコーダで構成され、文脈情報と翻訳対象の文を別々のエンコーダに入力するアーキテクチャである。原論文では上記 2 種類のアキテクチャを用いた翻訳実験を報告しているが、本記事では特に注目したいマルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた実験結果を紹介する。文脈考慮型ニューラル機械翻訳についても Transformer モデルを用いて実装した。翻訳性能の比較には、シードの異なる 5 つのモデルを学習してそれ

ぞれのモデルが出力した翻訳結果の BLEU スコア (Papineni et al. 2002) の中央値を用いた。BLEU スコアの計算には、case-insensitive BLEU-4 を用い、multi-bleu.perl¹で計算した。

実験結果を表 1 に示す。まず、文脈情報を用いない Single-Sent.の手法と比較して、文脈情報を用いた手法は全てのタスクにおいて BLEU スコアが改善しており、文脈を利用することで翻訳精度が向上していることを確認した。次に、学習時と翻訳時で異なる文脈情報（学習時は参照訳、翻訳時は機械翻訳結果）を用いた Trg-Context GT と、学習時と翻訳時ともに機械翻訳結果を文脈情報として用いた Trg-Context Pred.との比較では、ニュースコーパスの JA-EN タスク、TED トーク コーパスの DE-EN タスクを除いて、Trg-Context Pred.の翻訳精度が高かった。翻訳タスクによっては、学習時と翻訳時の両方の文脈情報に機械翻訳結果を使ってギャップを解消することで翻訳精度が向上する可能性がある。最後に、Trg-Context GT、及び Trg-Context Pred. と提案手法との比較では、全タスクにおいて提案手法が BLEU スコアで上回っており、参照訳と機械翻訳結果の両方のデータを制御して用いる本提案手法の効果を確認した。

5. おわりに

本記事で解説した論文は翻訳対象の目的言語側の前

¹ <https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu-detok.perl>

文を文脈情報として用いた手法であったが、近年、機械翻訳の精度向上のために翻訳対象以外の様々な外部情報を使う研究が盛んに行われている。マルチモーダル翻訳でよく用いられる画像情報（マルチモーダル翻訳とは視覚や聴覚、触覚など複数のモダリティを用いた機械翻訳を指す）や、BERT を始めとする事前学習された言語モデルなども外部情報に含まれる。扱う情報の種類が増えれば翻訳精度向上の可能性は高まるが、コンピュータの計算処理能力の向上が必須であり、闇雲に情報を追加すれば良いわけではない。また、それぞれの情報が独立したものでなければ、追加した情報のどの部分が精度向上に寄与したかなどの分析が困難となる場合もある。文脈情報を用いた機械翻訳はまだ発展途上であり、引き続き研究を進めていきたい。

参考文献

- Bawden, R., Sennrich, R., Birch, A., and Haddow, B. (2018). Evaluating Discourse Phenomena in Neural Machine Translation. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 1304-1313, New Orleans, Louisiana. Association for Computational Linguistics.
- Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. (2015). Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 28*, pp. 1171-1179. Curran Associates, Inc.
- Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). Curriculum learning. In Danyluk, A. P., Bottou, L., and Littman, M. L. (Eds.), *ICML, Volume 382 of ACM International Conference Proceeding Series*, pp. 41-48. ACM.
- Cettolo, M., Girardi, C., and Federico, M. (2012). WIT3: Web Inventory of Transcribed and Translated Talks. In Proceedings of the 16th Conference of the European Association for Machine Translation, pp. 261-268, Trento, Italy.
- Maruf, S., Saleh, F., and Haffari, G. (2021). Survey on Document-level Neural Machine Translation: Methods and Evaluation. *ACM computing surveys*, Volume 54, No. 2.
- Mino, H., Ito, H., Goto, I., Yamada, I., and Tokunaga, T. (2020). Effective Use of Target-side Context for Neural Machine Translation. In Proceedings of the 28th International Conference on Computational Linguistics, pp. 4483-4494, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: a Method for Automatic Evaluation of Machine Translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311-318.
- Toral, A. (2019). Post-editsese: an Exacerbated Translationese. In Proceedings of Machine Translation Summit XVII Volume 1: Research Track, pp. 273-281, Dublin, Ireland. European Association for Machine Translation.
- Zhang, W., Feng, Y., Meng, F., You, D., and Liu, Q. (2019). Bridging the Gap between Training and Inference for Neural Machine Translation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4334-4343, Florence, Italy. Association for Computational Linguistics.

ニューラル機械翻訳を対象とした試行錯誤的プリエディット事例の探索的分析

宮田 玲¹ 藤田 篤²

¹名古屋大学 ²情報通信研究機構

1. はじめに

機械翻訳 (MT) を実務で活用する方法として、ドメイン適応や用語集登録などの MT システム自体のカスタマイズに加えて、入力文を編集するプリエディット (pre-edit) と出力結果を修正するポストエディット (post-edit) があります。このうち、ポストエディットについては、産業翻訳において取り入れられることも増えており、また、例えば ISO 18587^[1]ではポストエディットのプロセスやポストエディターの技能要件も明文化されています。それに対して、プリエディットについては、長年、研究・実践で取り組まれてきたものの、産業翻訳での活用はポストエディットほど進んでいません。

その理由として、翻訳会社がクライアントから提供された原文書に手を加えることは越権行為であるという規範意識も根強い、という話を聞いたこともあります。また、プリエディット単体では MT 訳の品質を保証することは困難であり、産業翻訳においては、いずれにせよポストエディットなどの出口側の品質保証が求められるため、なかなか工数削減に結びつかないという側面もあるかもしれません。

上記に加えて、筆者らは、プリエディットとはどういう行為・プロセスであり、どのような効果が見込めるのか、といったことが十分明らかになっていないことが問題であると考えています。これまでも文の分割や主語の補完など、有効なプリエディット方法に関する個別の知見は提示されていますが、それらの全体像は明らかではありません。また、入出力を end-to-end で学習するニューラル MT (NMT) は、一般にルール

ベース MT (RBMT) や統計的 MT (SMT) よりも予測可能性・制御可能性は低いと言えます^[2]。これまで制限言語ルール^[3]の形で蓄積されてきた知見がそのまま有効に使えるかどうかは定かではありません。

とはいえ、プリエディットは、ここ数年ますます注目されているように思います。筆者らが参加した 2019 年 3 月の言語処理学会第 25 回年次大会のテーマセッション「産業翻訳に役立つ自然言語処理技術」では、参加者を交えた総合討論の中で、少なからぬプリエディットの話が出され、実際の活用事例の紹介もありました。AAMT のイベントにおいてもプリエディットの事例や研究の報告が複数なされているように^[4, 5]、産業翻訳での活用も徐々に進みつつあるようです。まだ数は多くないものの、NMT 向けのプリエディットの研究もなされつつあります^[6, 7]。

このような背景と問題意識のもとで、筆者らは NMT 向けプリエディットのベストプラクティスを探求する一歩手前で、プリエディットの全体像の理解を目指した研究を進めてきました。本稿では以下、2021 年 4 月開催の国際会議 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021) に採択された筆者らの論文“Understanding Pre-Editing for Black-Box Neural Machine Translation”^[8]の内容を解説します。理解の解像度を上げること自体がこの論文の目的ですので、本解説においても、論文全体を一般論として要約するのではなく (もちろん適宜要約しますが)、なるべく具体性を失わない形で結果を紹介します。

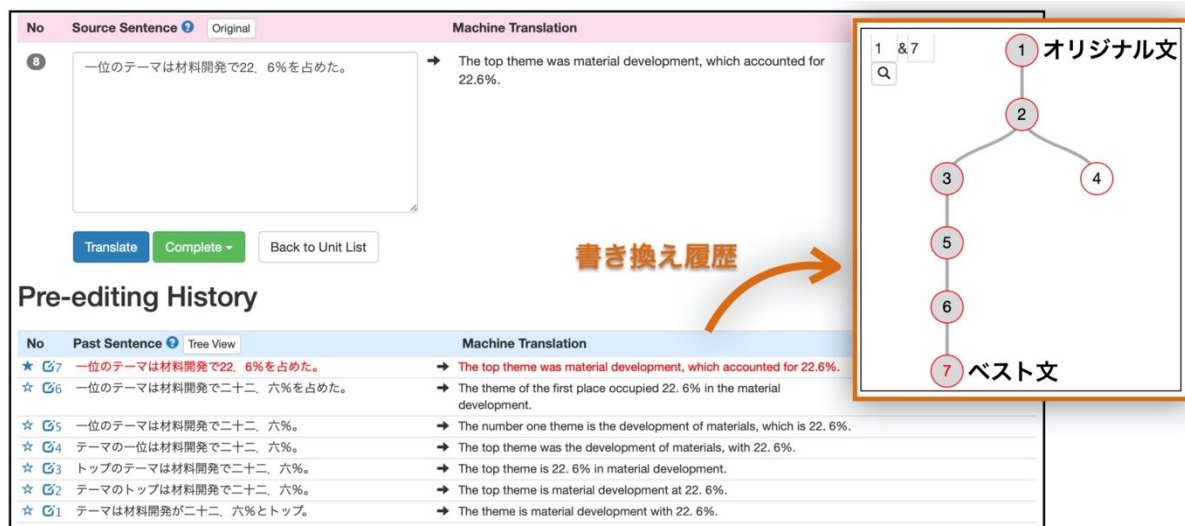


図 1: 書き換え事例収集プラットフォーム

2. 研究の枠組み

本研究は、人手による試行錯誤的なプリエディットの内容・操作とその結果を探索的に分析することを通じて、NMT 向けのプリエディットに関する広範かつ詳細な理解を得ることを目的としています。具体的には、高解像度のプリエディット事例を収集し（3 節）、得られた事例を次の 3 つの観点から分析しました。

- プリエディット後の「機械翻訳しやすい」原文の特徴（4 節）
- 原文書き換え操作の多様性（5 節）
- 各書き換えが NMT に与える影響（6 節）

なお、本研究は、有効なプリエディット手法の提案・検証を目的としていませんが、分析の過程で、どのような書き換え操作が有効そうか、という予備的な知見も一部得られています。

3. 人手書き換え事例の収集

図 1 のプラットフォームを用いて、MT 訳が十分な品質に達するまで、最小単位の原文書き換えを繰り返す、というプロトコルでプリエディットの事例を収集しました^[9]。ここでは手順の詳細は省きますが、書き換え作業の主なポイントは以下の通りです。

- 具体的な書き換えの指針は与えられず、MT 訳を見ながら、効果のありそうな原文書き換えを試行錯誤的に繰り返す。
- 過去の書き換え事例は全て記録され、作業者は随時、任意の事例に戻り、別の書き換えを試すことができる。
- 様々な書き換えを試しても MT 訳が十分な品質に達する見込みがないと判断した場合は、作業を途中で終了する。

図 1 に示すように、書き換え履歴は木構造で表すことができます。以下では、書き換え対象の日本語原文（オリジナル文とします）とそれに対する書き換え文、および MT 訳の履歴をまとめて、**ユニット**と呼びます。また、ユニット内で最も品質の高い MT 訳が得られた日本語書き換え文を**ベスト文**と呼び、オリジナル文からベスト文に至るパス（図の例では、1→2→3→5→6→7）を**ベストパス**と呼びます。このベストパス上の書き換えには、MT 訳の品質向上に寄与する書き換えが必ず含まれていますので、次節以降の分析でも主な対象としています。

事例収集の対象として、以下の条件を幅広く扱いました。

- **翻訳方向**：日英、日中、日韓
- **MT システム**：Google 翻訳（Google）^[10]、みんなの自動翻訳（TexTra）^[11]

表 1: 収集した書き換え事例の基本統計

翻訳方向	MT システム	書き換え数		ユニット数	
		合計	平均	オリジナル文＝ベスト文	十分な品質を達成
日英	Google	1332	13.32	27/100	98/100
	TexTra	1260	12.60	23/100	96/100
日中	Google	1371	13.71	2/100	91/100
	TexTra	812	8.12	8/100	94/100
日韓	Google	950	9.50	1/100	96/100
	TexTra	927	9.27	8/100	96/100
合計		6652	11.09	69/600	571/600

表 2: オリジナル文とベスト文の日本語原文の言語学的特徴

	オリジナル文	ベスト文					
		日英		日中		日韓	
		Google	Textra	Google	Textra	Google	Textra
文長	25.4	27.8	26.9	28.6	27.1	27.8	26.9
係り受け距離	1.95	1.97	1.99	1.99	1.99	2.00	1.98
係り受け最大深度	3.57	3.73	3.68	3.73	3.77	3.78	3.76
延べ語数	2538	2779	2685	2861	2709	2780	2693
異なり語数	1010	1074	1060	1106	1061	1068	1055
延べ語数/異なり語数	2.513	2.588	2.533	2.587	2.553	2.603	2.553
単語の頻度ランク(中央値)	170	143	154	143	155	143	169.5

- テキストドメイン：病院内会話、自治体生活情報、日本語で書かれた新聞記事(BCCWJ)、英語で書かれ日本語に翻訳された新聞記事(Reuters)

各テキストドメインから 25 文ずつ、計 100 文を対象に、翻訳方向、MT システムの全 6 通りの組み合わせについて、それぞれ 1 名の作業者（翻訳および MT 評価の経験有する）を割り当てました。事例収集は、2019 年 12 月から 2020 年 2 月にかけて実施しました。

収集した事例の統計を表 1 に示します。計 600 ユニットに対して、6652 事例が収集されました。特筆すべきは、約 95% のユニットにおいて、十分な品質の MT 訳が得られたことです。NMT に対するプリエディットの高い有用性を示す結果だと言えます。

ちなみに、日英の SMT システム (TexTra) を対象とした先行研究^[12]においても、プリエディットの潜在的な有効性は示されていますが、書き換え回数の平均は約

30 回/ユニットと、今回の結果を大幅に上回ります。両方の事例収集に協力してくださった作業者から、今回はだいぶ作業がやりやすかった、という所感もありました。実験条件の違い (MT の訓練データは不統一) には注意が必要ですが、現在の NMT を用いたプリエディットの効率は、以前よりも高いと言えます。

4. 機械翻訳しやすい原文の特徴とは？

機械翻訳しやすい原文の特徴を観察するために、翻訳方向、MT システムの各組み合わせ 100 ユニットずつについて、オリジナル文とベスト文の日本語文を比較しました。構文・語彙に関する定量的な指標を計算した結果の一部を表 2 に示します。

構文的な特徴の指標としては、文長、係り受け距離、係り受け最大深度を用いました（いずれの指標も 100 ユニットの平均値を使用しています）。興味深いことに、

いずれの指標においても、構文的な複雑さは僅かに増えています。これまでプリエディットの原則として、なるべく文長を短くすることや文構造を簡単にすることが提示されてきましたが、その原則を支持する結果とはなりません。当然、今回の結果から「文を短くすることは有効ではない」という結論は導けませんが、少なくとも「ほとんどの文は、短くしなくても機械翻訳でうまく訳せる表現に書き換えられる」ということが明らかになりました。この点については、次節でも改めて触れます。

語彙的な特徴の指標としては、延べ語数、異なり語数、延べ語数と異なり語数の比率、Wikipedia における単語の頻度ランク（値が小さいほど Wikipedia で頻出することを示す）を用いました。プリエディットを経ることで、全体として延べ語数、異なり語数ともに若干増えたことが分かります。また、延べ語数/異なり語数の値も上昇していることから、新しい語が多く導入されたというよりも、特定の語への偏りが増えたという傾向が見て取れます。また Wikipedia における単語の頻度ランクの中央値が上位に移動しています。本稿では詳述しませんが、オリジナル文とベスト文で変化が起きた語の分布を調べると、Wikipedia における頻度が低い語が回避され、より高い語が使用されるようになる傾向が観察されました。使用した NMT の訓練データにはアクセスできないため、低頻度語の定義は Wikipedia を用いた擬似的なものに過ぎない点に注意が必要ですが、先行研究^[13]においても NMT の弱点の一つとして指摘されている低頻度語を避ける形で書き換えが進んだことが分かります。

5. どのような書き換えが可能か？

続いて、機械翻訳しやすい表現に至るまでに使用された書き換え操作のタイプについて整理します。翻訳方向、MT システム、テキストドメインの計 24 通りの組み合わせから 10 ユニットずつ無作為に選んだ合計 240 ユニットの対象に、ベストパス中の書き換え事例 935 件を、書き換え方法の観点からボトムアップに類

型化しました。

その結果、構造、句、内容語、機能語、表記、エラーの 6 つのカテゴリからなる 39 タイプの書き換え操作を同定しました（表 3）。最も頻繁に用いられたのは、**C01「類義語の使用」**でした。MT 訳において適切な訳語を得ることを目指したものであり、様々な語の置き換えが試されていました。次に頻度の大きい**C07「内容変更」**も基本的には単語の置き換えですが、こちらは文脈知識や外界知識に基づくより高度な書き換えを指します。例えば、「根室署」に対する「根室警察署」はいわゆる辞書的な類義語ではなく、文意をふまえて常識を補完することで導けるものです。

S01「文分割」が 14 件（1.5%）しかない点は、4 節の構文的特徴に関する知見を裏付けます。ちなみに、分割された 14 文の内、9 文は文長が 60–67 語とかなり長いものでした。NMT にとっては依然として 60 語を超えるような長文の処理は難しいことが、先行研究においても実証されています^[13]。この知見が、プリエディットの観点からも補強されました。

さらに、翻訳学の知見^[14, 15]を参考にしながら、情報方略の観点からも事例を分類しました。その結果、書き換えによってテキストの意味・含意が明確に表現される**明示化**は 329 件、その逆の**暗示化**が 88 件、いずれでもない**情報保持**が 480 件という結果になりました。類義語に置き換えるような情報保持が一番多いものの、明示化が多く観察された点は重要です。表 4 は、明示化の事例を類型化した結果です。曖昧な表現を避けたり、暗黙の前提とされる情報を言語化したりすることは、機械翻訳しやすさにつながる方略と言えそうです。

6. NMT の出力結果はどれだけ変わるか？

最後に、プリエディットが NMT の出力結果に与える影響を見ていきます。図 2 は、表 3 の書き換えタイプごとに、原文の書き換えがもたらした NMT 出力文字列の変化量の分布を示しています。変化量は、TER（Translation Edit Rate）^[16]という編集距離の指標を用いました。

表 3: 書き換えタイプの類型と事例頻度

ID	頻度	構造	ID	頻度	内容語
S01	14	文分割	C01	118	類義語の使用
S02	27	文構造の変更	C02	21	省略語の回避・使用
S03	19	主題構造の使用・回避	C03	14	照応表現の使用・回避
S04	14	主語・目的語の追加	C04	11	強調表現の使用・回避
S05	12	体言止めの回避・使用	C05	30	カテゴリの表示・非表示
S06	4	態の変更	C06	10	辞書的説明
S07	5	その他構造	C07	94	内容変更
ID	頻度	句	ID	頻度	機能語
P01	71	読点の追加・削除	F01	47	助詞の変更
P02	28	チャンキングの使用・回避	F02	31	複合辞の変更
P03	31	語句の順序変更	F03	12	アスペクトの変更
P04	7	修飾の仕方の変更	F04	4	テンスの変更
P05	40	接続表現の変更	F05	11	モダリティの変更
P06	36	並列表現の変更	F06	10	敬語の使用・回避
P07	16	同格表現の変更	ID	頻度	表記
P08	13	名詞句・動詞句の交替	O01	61	日本語表記の変更
P09	28	複合語の使用・展開	O02	16	全角・半角の変更
P10	17	接辞の使用・解除	O03	2	記号の追加・削除・置換
P11	5	サ変名詞表現の変更	O04	4	その他表記
P12	9	形式名詞表現の変更	ID	頻度	エラー
P13	5	その他語句	E01	22	表記誤り、文法誤り
			E02	16	内容的な誤り

表 4: 明示化方略の下位分類・頻度・事例

明示化方略	頻度	書き換え例	MT 出力結果
情報の追加	142	12 日は台湾の休日のため休場。	The twelfth is a holiday in Taiwan.
		→ 12 日は台湾の休日のため 株式市場 は休場。	The stock market was closed on the twelfth due to a holiday in Taiwan.
関係の明示	103	来院しなくても 10 日前後で登録のクレジットカードから 引き落とし を行います。	Withdraw from your registered credit card in about 10 days without visiting the hospital.
		来院しなくても 10 日前後で登録のクレジットカードから 引き落とし が行われます。	Even if you do not visit the hospital, your credit card will be debited in about 10 days.
意味の限定	54	採尿と採便を 出 してください。	Please collect urine and feces.
		→ 採尿と採便を 提出 してください。	Please submit urine and stool samples.
正規化	30	単位は億円。	Figures are in billions of yen.
		→ 単位は億円 です 。	The unit is 100 million yen.

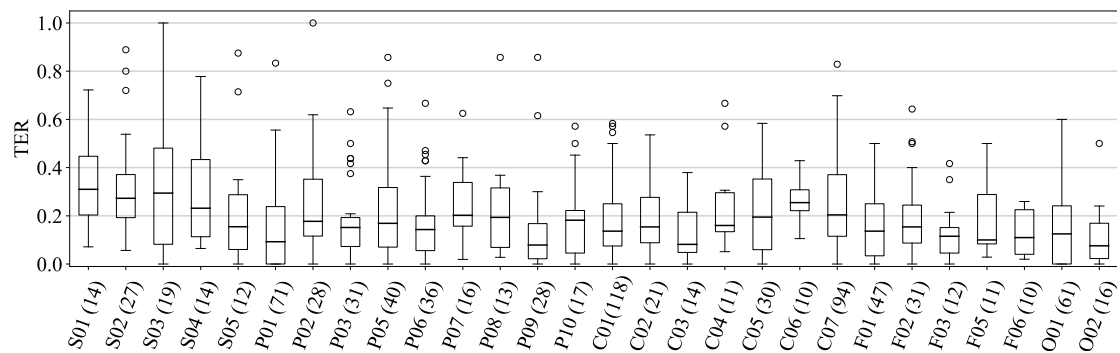


図 2: 書き換えタイプごとの MT 変化量 (TER) の分布 (括弧内の数字は書き換え事例数)

それぞれの書き換えタイプに注目すると、変化量の分布には幅があることが分かります。最小単位の原文書き換えで、がらりと出力が変わることもあれば、それほど変わらないこともある、という NMT の出力を予測することの難しさが具体的に可視化されたといえるでしょう。一方で、書き換えタイプ横断的に見てみると、出力の変わりやすさに一定の傾向はありそうです。例えば、構造レベルの書き換え（特に S01–S04）は MT 訳に大きめの変化をもたらしている一方、機能語の変更が NMT に与える影響は小さいということが観察できます。

また、P03「語句の順序変更」はそれほど NMT の結果に影響を及ぼしていません。これまで SMT においては、自動的な語順の事前並び替え手法が翻訳精度の向上に有効であるとされてきましたが、NMT においてはあまり効果がない（場合によっては逆効果）という報告もなされています^[17, 18]。今回の結果から、最新の NMT は人間が行うような語順の変化に対しても頑健であることが示されており、プリエディットの方法として語順変更は必ずしも有用であるとは限らないことが示唆されます。

く分かっていたこと/行われてきたことに回収されてしまいます。本研究の意義は、まさにこのような何となく考えられてきたことを、具体的なレベルで愚直に調査したことにあります。もちろん、その過程で、プリエディットを経ても必ずしも文構造が簡単になるわけではない、という意外な発見もありましたが、あえていえば、それは派生的なものです。ここで重要なことは、平均 10 回程度の細かい原文書き換えにより 95% の事例で十分な MT 訳品質が達成されたことや 6 カテゴリ 39 タイプの書き換え方法が同定されたことなどの具体的な記述です。

本稿で示したような解像度・範囲で記述することで初めて、これをベースとした具体的な議論が可能になります。「他にもこのような書き換え方法があるのでは」、「内容変更の方法としては具体的にはどういうものがあるのか」、「別の条件で試したらどうなるのか」といった疑問も湧いてくるでしょう。筆者としては、本研究の知見を起点として何かしらの議論（反論も含めて）が芽生えることを期待しつつ、解説を終えたいと思います。

謝辞

本研究の一部は、科研費若手研究（課題番号：19K20628）、科研費基盤研究（S）（課題番号：19H05660）および KDDI 財団調査研究助成の支援を受けました。データの一部は、総務省の情報通信技術の研究開発「グローバルコミュニケーション計画の推進 - 多言語音声翻訳技術の研究開発及び社会実証 - I.

7. おわりに

ここまで解説してきた内容を、「プリエディットはうまく活用できれば NMT の品質を十分に高められる」、「多様な書き換え方法が存在する」、「明示化は大事」、「NMT の挙動は読めないけれど、一定の傾向はある」といった抽象度でまとめてしまうと、これまで何と

多言語音声翻訳技術の研究開発」から提供を受けました。

注・参考文献

- [1] ISO (2017). ISO 18587:2017 Translation services – Post-editing of machine translation output – Requirements.
- [2] もちろん、MT システムがブラックボックスである限りは、RBMT でも SMT でも、出力そのものを予測することは困難ですが、NMT とは異なり、翻訳前後の語や句は陽に対応付けられているため、「この部分を変えれば、この部分が変わりそうだ」という予測は立てやすいといえます。しかしながら、この予測可能性に関する想定自体が、抽象的に仮定されたものに過ぎず、データに基づく検証が十分なされていないことを、ここでは強調したいと思います。
- [3] 例えば、Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1): 121–170. において様々な英語の制限言語が提示されています。日本語の制限言語については、Miyata, R. (2020). *Controlled Document Authoring in a Machine Translation Age*, Routledge, London. で整理されています。
- [4] 星井智 (2019). 句点挿入と倒置：日英 MT における長文リエディットの二つの手法. AAMT 2019, Tokyo.
- [5] 早川威士 (2019). 機械翻訳に対するリエディットのタスク分類に関する考察. AAMT 2019, Tokyo.
- [6] Hiraoka, Y. & Yamada, M. (2019). Pre-editing plus neural machine translation for subtitling: Effective pre-editing rules for subtitling of TED talks. *Proceedings of MT Summit*, pp. 64–72.
- [7] Mehta, S., Azarnoush, B., Chen, B., Saluja, A., Misra, V., Bihani, B. & Kumar, R. (2020). Simplify-then-translate: Automatic preprocessing for black-box translation. *Proceedings of AAAI*, pp. 8488–8495.
- [8] Miyata, R. & Fujita, A. (2021). Understanding pre-editing for black-box neural machine translation. *Proceedings of EACL*, pp. 1539–1550. <https://www.aclweb.org/anthology/2021.eacl-main.132/>
- [9] 本研究において、「十分な品質」は「多少流暢さに欠けても、情報の過不足がなく、文法的にも正しい訳文」としました。また「最小単位の書き換え」とは「文法性を保ちながら、それ以上細かくは分割をできない最も単純な書き換え」としました。その他プロトコルの詳細は、元の論文もしくは以下の文献をご参照ください。
- 宮田玲・藤田篤 (2017). 機械翻訳向けリエディットの有効性と多様性の調査. 通訳翻訳研究への招待, 18: 53–72.
- [10] Google 翻訳. <https://translate.google.com>
- [11] みんなの自動翻訳 @ TexTra. <https://mt-auto-minhon-mlt.ucri.jgn-x.jp>
- [12] 宮田玲・藤田篤 (2017). 注 9 番参照.
- [13] Koehn, P. & Knowles, R. (2017). Six challenges for neural machine translation. *Proceedings of NMT*, pp. 28–39.
- [14] Vinay, J.-P. & Darbelnet, J. (1958). *Stylistique Comparée du français et de l'anglais*. Didier, Paris, trans. and ed. by J. C. Sager & M.-J. Hamel (1995) as *Comparative Stylistics of French and English: A Methodology for Translation*. John Benjamins, Amsterdam.
- [15] Chesterman, A. (1997). *Memes of Translation*. John Benjamins, Amsterdam.
- [16] Snover, M., Dorr, B., Schwartz, R., Micciulla, L. & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. *Proceedings of AMTA*, pp. 223–231.
- [17] Zhu, Z. (2015). Evaluating neural machine translation in English-Japanese task. *Proceedings of WAT*, pp. 61–68.

- [18] Du, J. & Way, A. (2017). Pre-reordering for neural machine translation: Helpful or harmful? *The Prague Bulletin of Mathematical Linguistics*, 108: 171–182.

「自動翻訳大全：終わらない英語の仕事が 5 分で片づく超英語術」

中澤 敏明

東京大学

1. はじめに

本書は八楽株式会社代表取締役である坂西優氏と 2021 年 4 月より関西大学から立教大学に移られた山田優教授の 2 名によって書かれた本で、いかにして自動翻訳をうまく使いこなすか、という点に集中した内容となっている。1 章では近年進歩が目覚ましい自動翻訳の概要と、これをうまく使いこなすには入力する文を自動翻訳しやすいように事前に編集しておく「プリエディット」が重要であるということを述べている。2 章では英語の文章を自動で日本語に翻訳して「読む」際のポイントを解説し、3 章では逆に日本語の文章を英語に直して「書く」際のポイントを解説している。4 章では音声翻訳アプリやデバイスに注目し、外国人との対話において相手が話している英語を日本語で「聞く」ためのノウハウを解説している。また対話の場面に応じた最適なアプリも紹介されている。5 章では逆に英語を「話す」ためにはどうすれば良いかが解説されているが、5 章だけは他の章とは毛色が異なり、英語の授業を受けているような感覚を覚える。6 章ではまとめとして、自動翻訳をうまく使いこなすことで、仕事の効率化だけでなくさまざまなメリットがあると説明している。

以降では 2 章以降の各章についての簡単なまとめと、筆者の感想を述べたいと思う。

2. 秒速で「読む」

2 章の前半では英語の文章の内容を自動翻訳を駆使していかに早く、正確に把握するかというノウハウが

紹介されており、章の後半ではノウハウを活用することで様々な文章をうまく読むことができるという具体的な例を示している。ノウハウは入力文に対してプリエディットを適用することで、自動翻訳の誤りを減らすことを目的としており、長い文をカンマや接続詞などを手がかりとして短く区切る、各文にはピリオドをつけるなどといったものが紹介されている。また訳語に違和感を感じた際に、語や句単位での訳語のチェック方法も紹介されている。

この章で紹介されているノウハウは普遍的なものであり、どの翻訳エンジンに対しても有効であると考えられる。長い文を入力してしまうと、自動翻訳が文の解釈（係り受け関係など）を誤ってしまい結果として誤訳につながる可能性があるが、入力文を短くすることでこの問題をある程度回避することができる。

ただし一点だけ注意が必要なのは、本書では自動翻訳は基本的に 1 文単位だと書かれているが、2021 年 3 月時点で、DeepL 翻訳も Google 翻訳も 1 文単位ではなく、入力された複数の文をまとめて翻訳している（ここではこれを文書翻訳と呼ぶ）。文書翻訳のメリットとしては、前の文にあって後の文で省略されている情報なども正しく補って翻訳できるようになったり、前後の文脈に沿って適切な訳語が選択できるようになったりする可能性がある点にある。一方でデメリットとしては、一度に多くの文を翻訳することになるため、結果として長い文を入力しているのと同じ状況になってしまう、訳抜けや文解釈誤りが発生する可能性があるという点である。

3. 完璧に「書く」

3 章は自動翻訳を駆使して英文を書くことに主眼を置いている。やはり前半でノウハウを紹介し、後半で実例を挙げるという構成になっている。ノウハウの内容は前章と同様、プリエディットによりいかにして自動翻訳の誤りを減らすかという観点でまとめられている。文長に関するものなどは言語によらず共通だが、英日翻訳と日英翻訳では注目すべきポイントが異なる。

日本語には代名詞の省略、漢数字の使用、常体と敬体、オノマトペの多用、略語使用などの独特の特徴があり、これらは全て自動翻訳精度に悪影響を及ぼす。プリエディットによりこれらを極力減らすことで、自動翻訳精度の悪化を防ぐというのが本章のポイントである。つまりは入力文の解釈に曖昧さを残さないということが重要なのである。

本章で紹介されているノウハウは筆者からみてもどれも有効であると考えられるが、ここでも一つだけ注意していただきたい点がある。ノウハウの中で固有名詞や専門用語など定訳が存在し、間違えてはならないものは英語表記にしてから日英翻訳をすると良いと書かれているが、この対処法は確実に英語側に同じ表記が出現するとは限らないという点である。これは現在のニューラル機械翻訳の仕組みをある程度理解している方、サブワードとは何かを理解している方であれば想像はつくであろう（詳細は本稿では割愛する）。同様に、アラビア数字も必ずしも入力と出力で一致するとは限らない。本書でもこれらはポストエディットで誤りがないかをチェックすることが重要だと述べられている。

4. あらゆる状況で「聞く」

本章では自動音声翻訳に注目し、音声翻訳アプリを駆使することで外国語の発話を理解するためのポイントを解説している。また様々な音声翻訳アプリやデバイスも紹介されており、一度に認識できる音声の長さ（および長さごとの翻訳精度）と認識した結果が履歴

として保存できるかどうかという 2 軸で各アプリの比較がされている。この 2 軸による比較はどのような場面で自動音声翻訳を利用するかと関係しており、旅行先でのコミュニケーションなのか、ビジネスシーンでの利用かなどの利用シーンごとにアプリを使い分けるのが良いと述べられている。

また音声翻訳アプリの使いこなしテクニックとして、音声認識精度を向上させるための方法や音声認識アプリを利用したコミュニケーションを円滑に進めるための方法なども紹介されている。翻訳エンジン自体の精度はもちろん重要だが、本章で書かれている通り音声認識に誤りがあると、結果として翻訳精度が低いように見えてしまう。最近では音声認識の精度は非常に向上しており、多少の雑音ならあまり影響を受けなくなってきたはいるが、可能な限り悪影響を与えるような要因は除外しておくべきであることは間違いない。

5. 今すぐ「話す」

本章は英語を「話す」ことをテーマにしている。英語を話すには自動音声翻訳を利用して話す場合と自力で英文を作り出して話す場合の 2 種類がある。自動音声翻訳を利用してうまく英語を話すコツは、3 章でも説明されているプリエディットを行うことだと述べられている。音声なので頭の中でプリエディットしておいてから発話する必要があるが、もしうまく翻訳できていない（相手に言いたいことが伝わっていない）場合でも、言い方を変えるなどして再度翻訳すればプリエディットしたことになるであろう。

問題は自力で英文を作り出して話す場合である。本書では自分で話す場合であっても、プリエディットを応用すれば良いとされている。話したい内容を省略や曖昧性のない形で整理し、これを英語の形に変換することで英語が話せるようになるというのだ。英語の形にする上で最も重要なポイントは語順であると説明されている。日本語には格助詞があるため文の構成要素をある程度自由に並べ替えることができるが、英語では単語の並び順が重要であり、文を構成する単語の意

味や役割がその位置（他の単語との関係）によって定義される。このため英語文を生成するには語順を意識することが大切だというのだ。

英語を初めて学校で学んだ時には5文型というのを覚えた記憶があるが、本書では3つの語順さえ覚えておけばどんな表現であっても英語に変換できるとしている。この3つの語順についてあえて5文型との対応を考えてみると、1つ目の語順は第2文型、2つ目の語順は第5文型、3つ目の語順は第4文型となるであろう。最初のうちはこの3つの語順さえ覚えれば、言いたいことのほとんどは英語で表現できるというのだ。3つの語順に絞るメリットとして、使う動詞が少なく済むという点が挙げられている（本書では全部で30弱の動詞が紹介されている）。例えば「彼の報告は信じられない」という日本語を英語にする場合、第3文型を使って「I can't believe his report.」と直訳することも可能である。しかしこのように直訳をしていると例えば使うべき動詞が出てこなくて英語に変換できなくなってしまうということがいつか起こりうるとしており、そうではなく使う動詞を絞って、その範囲でうまく表現するようにする方が汎用性が高いと説明されている。また言いたいことを換言したり、組み換えたりしてうまく整理すれば、どの語順でも同じ内容を表現できると説明されている。

筆者も英語が得意というわけではないので、確かに言いたいことを言うための最適な動詞が出てこない（もしくは知らない）ということはよくある。最適な動詞さえ知っていれば、少ない単語数でスッキリと言いたいことが言えるのであらうと思いつつ、同じ内容に近くなるように手持ちの語彙の中から遠回りして話したりしている。自動翻訳というトピックからは多少外れてはいるが、本章の内容は英語を話す必要に迫られているがあまり得意ではないという人にとっては非常に役立つ内容になっていると思う。

6. 自動翻訳を使いこなして英語力を上げる

本章では全体のまとめとしてプリエディットの重要

性を再度主張している。またプリエディットをすることで脳の言語を司る部分が刺激され、自然と多言語を扱う能力が向上すると説明されている。

英語学習を行う上で大事なことはアウトプットをすることだと言われており、その理由として「気づき（伝えたいことと伝わったことのギャップに気づく）」「仮説検証（相手に伝えるにはどうしたら良いかを試行錯誤）」「意識化（自分のやっていることを認識し、別の言葉で説明）」の3つができるからだと言われている。本書では自動翻訳を使うこともこのアウトプットのひとつと考えることができるとされている。つまりプリエディットをすることでこの3つが自然と達成されているというのである。プリエディットをしている時点で、伝えたいことと機械が認識できることのギャップに気づき、翻訳結果を確認して再度入力を修正することで仮説検証を行い、プリエディットしたルールそのものが意識化に相当するのである。

このように自動翻訳を使うだけでも自然と英語力が向上する可能性があり、「英語を好きになろう！」というモチベーションを持つ必要も「自動翻訳を使っている人は英語力は伸びない」という批判に耳を傾ける必要はないと本章では述べられている。この点に関しては、反対意見を持つ人も多くいるのではないだろうか。筆者の見解は、自動翻訳を使ってもおそらくある程度までは英語力は伸びると思う。このある程度というのは必要最低限の英語力、つまり簡単な文章を書いたりコミュニケーションを取ったりといったレベルを想定している。実際多くの人にとって、このある程度のレベルで十分なのではないだろうか。そうになると、自動翻訳を使うだけでも十分だという結論になるのも不思議ではない。ただしここで注意が必要なのは、自動翻訳で英語力が向上するためには、本書で何度も繰り返されているように、入力文（つまり母語）をしっかりと理解するという行為を行わなければならないという点である。

一方で「自動翻訳を使っている人は英語力は伸びない」という批判をする人の多くは、翻訳という行為自体を生業としている人たち（翻訳者・翻訳家）であらう。

これらの人は当然、機械翻訳のレベルをはるかに上回る英語力を持っており、そのレベルの英語力に達するには自動翻訳になんか頼ってはいけないという意見を持つのも理解はできる。だが、前にも述べた通り、多くの人にとってそのレベルの英語力は必要ではないのだ。「私たちのように翻訳を生業として生きていくなれば」という前提条件のもとでは「自動翻訳は使うべきではない」という主張は理解できるが、その前提条件なしに頭ごなしに自動翻訳はダメと主張するのは大きな間違いだと思う。

本章では最後に自動翻訳の限界についても触れられている。自動翻訳の研究者は限界についてよくわかっているが、一般的なユーザーはここがわかっていないため、自動翻訳の誤った使い方をしてしまっているケースがしばしば見受けられ、たまにニュースになったりもしている。この辺りの啓蒙活動は研究者や自動翻訳サービスのプロバイダーが行うべきなのだが、あまりしっかりと情報共有されているとは言い難い。本書ではこの点についてもしっかりと述べられており、非常に良いと思う。

7. 終わりに

本書は自動翻訳をいかに役立てるかというトピックについて、読む、書く、聞く、話すという各状況に応じてノウハウが提示されており、自動翻訳に関してほとんど知識がない人にとっての入門書、もしくはとにかく自動翻訳を仕事に役立てたい人にとっての虎の巻としてとても良い内容になっていると思う。ただし、途中でも述べたし本書の注意書きにもあるとおり、自動翻訳の今後の発展次第では本書の内容が通用しなくなる可能性も十分にある。自動翻訳のユーザーも、最新の状況がどうなっているか、何か変化があったのかといった情報は常に仕入れておく必要があるだろう。

機械翻訳はどこへ行くのか？

『機械翻訳：歴史・技術・産業』

ティエリー・ポイボー[著]、高橋聡[訳]、中澤敏明[解説]

山田 優

立教大学

1. はじめに

2020 年に出版された『機械翻訳』（ティエリー・ポイボー[著]、高橋聡[訳]、中澤敏明[解説]）は、これまでの機械翻訳（以下、MT）の解説書とは違う。本書は、MT の仕組みについて、数式を使わないで分かりやすく解説する書であると形容できる。しかしながら、私はこの書を「機械翻訳の歴史ドキュメンタリー」として読み、機械翻訳に携わってきた人たちのノンフィクションとして楽しんだ。さらに、本書が中心的に扱うテーマ、「自然言語の曖昧性の問題」「翻訳するのは、どういう意味なのか」に関しては、よくある技術的な解説記事に比して強く意識され、機械翻訳の開発がそもそも、本質的な翻訳という行為の一部の側面のみしか処理できていないということを出発点として始まったその歴史、その技術的説明を提供している。これらの点において、この『機械翻訳』の書は異彩を放っている。

本稿では、この書、『機械翻訳』の特徴を、特にそこに詳述された歴史的側面に焦点を合わせ説明する。本稿の後半では、自然言語の「曖昧性」について、本書でも述べられている通り機械が扱えない（人間ならばできてしまう）翻訳の側面を、翻訳支援という技術で補うという観点から、MT の可能性を論じる。これには、筆者の関心のひとつでもあるポストエディットも射程に入る。すなわち、MT はどこへ行くのかを考える時、短中期的な MT の社会実装を思慮すると、人間とのインターアクションという要素を考慮すべきでは

ないかと、主張する過去の偉人トロヤンスキーやバー＝ヒレルの提言に、我々は耳を傾けなければならない。かつて、世界的に莫大な国家予算が投じられ、バベルの塔の前の言葉の壁を復元できるように信じられた過去の盲目的な時を、本書の「歴史」の記述が紐解く。今の私たちは、果たして同じ過ちを犯してはいないだろうか。MT はどこから来て、どこへいくのか。その答えは、それ自体の歴史に刻まれている。

2. 『機械翻訳：歴史・技術・産業』

本書の章構成に沿って概要をまとめる。

第 1 章「はじめに」では本書の狙いが述べられている。「限られた紙面のなかでめざすのは、第二次世界大戦後に機械翻訳が遂げてきた発展を概観することである...詳しい技術的な内容を知らなくても、読者が主な原理を理解できるようにすること、それが本書の狙いである」(p. 3)。

第 2 章「翻訳をめぐる諸問題」では、翻訳の難しさに関する重要な議論がされている。「良い」翻訳の基準とは何か。翻訳とは、「もとの文章の主な特性、トーン（調子）やスタイル（文体）、思想の細部、全体の構成まで反映されている必要がある。ターゲット側での読みやすさと言語的な正しさの両方が求められるので、繊細な作りかえのプロセスが求められる」(p. 6)。これに対して、機械翻訳は、長い間、局所的な技術のみを利用して、逐語翻訳に相当するプロセスしか行っていないことも強調されている。そして、本書で中心的に扱

Book Review: “Machine Translation” by Thierry Poibeau (trans, Akira Takahashi; comments, Toshiaki Nakazawa)

Masaru Yamada

Rikkyo University

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License. License details: <https://creativecommons.org/licenses/by-sa/4.0/>

う問題、自然言語の曖昧性についても説明する。たとえば、“the chicken is ready to eat”は「鶏が何かを食べようとしてる」ところなのか、「人が鶏肉を食べようとしている」ところなのか、日常的に曖昧性が存在し、この決定は文脈を考慮すれば容易に解決する問題でないことを、改めて説く (p. 11)。「ネッカーの立方体」(図1)を示し、どの面が前でどの面が後ろかを特定できる手がかりがないことから、この図は「曖昧」であるのだが、人間は合理的な見方をひとつだけ自然に選択できること。機械で同じことをすることが、いかに難しいのか、そしてこの曖昧性の問題が、この後で展開する機械翻訳にとっての最大の難関となると予告する。

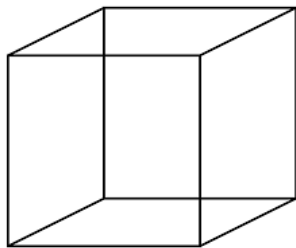


図1：ネッカーの立方体

第3章「機械翻訳の歴史の概要」では、技術的な仕組みの観点から、ルールベース翻訳の直接翻訳方式、トランスファー方式、中間言語方式と、統計的機械翻訳の簡潔な説明が提供されている。これ以降の章で展開する歴史の詳細説明を理解するための予備知識となる。

第4章は「コンピューター登場以前」の歴史が記されている。筆者が本書を「ドキュメンタリー」として楽しんだのは、この4章から6章の厚い記述である。機械翻訳をめぐるキープレイヤーたちの人間模様も垣間見られる。翻訳者の高橋氏の流暢な訳も、読者を飽きさせない。

機械翻訳が生まれた背景には、そもそもヨーロッパには古くからの伝統として、バベルの塔の逸話以前に人が話していたとされる「アダムの言語」、すなわち仮説上の普遍言語があり、中世の哲学者デカルトやライプニッツはこれに興味を示していたところからストーリーが展開する (p. 25)。デカルトの言葉が引用され

る。

「aymer, amare, ahilein (いずれも、「愛する」の意)」という語と、それぞれの同義語に対応する記号を辞書に登録すれば、その辞書さえあれば誰でも、そうした記号で書かれた本を翻訳することができる」(デカルト、メルセンヌに当てた書簡、1629年11月20日) (p. 25)。

デカルトの提案は、言葉の「曖昧性」のないコードへの置き換えを示唆し、この一節が、機械翻訳の先駆者たちを触発することとなる。こうした提案に続き、17世紀のヨーロッパでは「数値辞書」を作ろうとする試みもあった (p. 25)。数値辞書とは、言葉や概念に特定の数値(識別子)を対応させた辞書のことで、さまざまな研究者の業績がある。このような構想を機械翻訳の直系の前身と見なすことはできないとしても、多くの研究者にとって、インスピレーションの源になったのは間違いない (p. 26)。このようにコンピューター以前の時代から、機械翻訳に通ずる概念の歴史を端的かつ、書簡の引用などを見ながら追えるのは嬉しい。

第5章「機械翻訳のはじまり：初期のルールベース翻訳」では、初期(1940年代)から、黎明期(1950～60年)、幻滅の期間(1960～64年)が語られる。

機械翻訳の父ウオーレン・ウィーバーが1949年、クロード・シャノンとともに、通信の数学的モデルに関する論文を発表。その提唱はきわめて汎用的だったことから、多くの場面に応用される。1947年から、ウィーバーは機械翻訳をめぐる、サイバネティクス学者のノーバート・ウィナーと書簡を交わし、翻訳を「デコード」の問題とし考えてみたらどうかと提案した (p. 32)。

翻訳という問題は、暗号技術の問題と考えられるかもしれない。例えば、ロシア語で書かれた記事を見たとき、「これは、本当は英語で書かれているのだ。ただ、知らない記号で暗号化されている。そう考えて解説をこころみるというわけだ。(p. 33)

ウィーバーの構想は「Translation」という覚書にまとめられ、一般的にはこれが機械翻訳研究の出発点とされている。その影響力は極めて大きく、研究資金源となった組織と深いかわりがあったという理由から、影響力は政治的にも大きかった (p.34)。

ウィーバーがウィナーと交わした書簡では、ウィナーは構想そのものには慎重であった (p.33)。ウィーバーは、逐語的な翻訳に伴う基本的な誤りを避けるための原理を提言する。ここで強調されているのは、冒頭で述べた曖昧性の問題だ。ウィーバーは、曖昧性のほとんどは周囲の文脈を見れば解決できることを示す。確かに、この手法は現在でも用いられている。あらゆる種類の曖昧性を解決できるというわけではないが、ほとんどは解決できる (p. 35)。ウィーバーの意見は言語学者の精査を必要とし稚拙であるとしているが、それでも、この覚書は実際にきわめて先見性に飛んでおり、だからこそ、後世に多大な影響を残すことになったのである (p. 36)。

機械翻訳の真の明瞭期 1950 年代の前半は、冷戦という国際事情からロシア語・英語間の翻訳の必要性が高まっていた。こうした時代背景のなか、1950 年代のアメリカで機械翻訳の発展をリードする重要な役割を果たしたのが、イスラエルの研究者、バー=ヒレルだ。1954 年、ジョージタウン大学の研究チームと IBM は共同で開発したシステムに基づいて機械翻訳を実証する世界初のデモを実施した。直接翻訳方式で、ロシア語 49 文を英語に翻訳し、その効果は絶大で、機械翻訳の財政支援は増え続ける。メディアの反響も大きく、世間一般からも注目が集まり、社会は熱狂する。そして機械翻訳は学問として確立し、国際会議が定期的に開催されるようになる。また、専門学術誌『Mechanical Translation』も同年に発行されている (p. 37 – 38)。

しかし、熱狂後まもなく、直接翻訳方式の限界が露呈する。当初から問題視されていた曖昧性が解消できない。このため、構文または意味レベルで機能するトランスファー方式が模索される (p. 38)。1950 年代に形式文法という概念が、主にノーム・チョムスキーの

研究成果によって有力になってきたことと関係が深い。しかし、曖昧性の問題は解消されず、開発者の間で、むしろ問題の重大さがやっと認識され始めた本書では述べられている (p. 41)。

1953 年、バー=ヒレルは、機械翻訳の根本的な問題として考えた言語上の課題をあげている。翻訳には文章の完全な統語解析や意味的な曖昧性の解決が必要であると説く。曖昧性のある一部の単語は、たとえ文脈を考慮しても、その意味を限定することはできなし、以下のよく知られた例をあげた。

“Little john was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.” (p. 44)

この文を理解するには、“pen”という単語が「子供を遊ばせる小さい囲い」を意味しており、筆記用具ではありえないと認識する必要がある。しかし、この意味を推論できる手がかりはない。システムでこうした問題を解決するのが不可欠ことは明らかであり、しかもこれは珍しい例でもない (p. 44)。したがって、質の高い完全自動の機械翻訳 FAHQMT を、短期的または中期的に構想することは不可能である、とバー=ヒレルは結論づける (p. 44)。

このようにして、熱狂的な機械翻訳開発は、第 6 章「1966 年の ALPAC レポートと、その影響」の時代に突入する。本書では、ALPAC レポートがバイアスのかかった報告であったこともしっかりと述べられている (p. 48-49)。米国における機械翻訳開発が公的資金から研究資金の援助を受けているため、研究そのものが基礎研究というよりも、現実的な問題解決をめざしている研究の類であると指摘し、短期的に質の高い完全自動の機械翻訳 FAHQMT が実現する見通しがなくことを主張した (p. 48-49)。ALPAC レポートは、現実的な機械翻訳の問題についての評価が定まっておらず、判断基準が、それまで投じられてきたコストであった。それどころか、コストに関しては機械翻訳を使うより人間に翻訳させたほうが低く抑えられると、レ

ポートは結論付ける (p. 49)。当時は、人手翻訳のほうが良質な翻訳を速くこなすことができた。機械翻訳の出力を修正するのは、熟練の翻訳者が直接翻訳するより時間がかかることも多かった (p. 49)。これは、現在の文脈でいうところの、ポストエディットは、人手翻訳よりも高くつくということであり、また機械翻訳よりも人手翻訳の方が安いということである。この報告の真偽のほどはともかくとして、機械翻訳の時代は、これをきっかけに長い停滞期へと入っていくのは、既知の通りだ。

歴史の説明に一旦区切りをつけ、第7章では、「パラレルコーパスと文アラインメント」の説明がなされる。時代的には1980年代に対応する。バイテキストの概念、自動アラインメントの手法などテクニカルな説明が展開する。

第8章では、長尾真氏が発明した「用例ベース機械翻訳」が解説される。曖昧性を解消するために、高度な構文解析が必要になる一方で、システムが複雑化の一途をたどっていた。そんな中で長尾氏が、プロ翻訳者が文章の断片を扱い、それを組み立てながら、一貫性のある文を作り出していることを洞察し、翻訳者はたいてい、目の前にある文について予め完全な解析を実行しているわけではないと主張。と同時に、パラレルコーパスには、専門性の高い辞書も含めほとんどの対訳辞書に欠落している貴重な情報が大量に含まれていると指摘した (p. 69)。したがって対象とする言語間で新しい辞書や、新しい解析規則、トランスファー規則を作るより、既存の対訳コーパスで見つかる翻訳の一部を直接利用するほうが有用であるとした (p. 69)。ここに用例ベース機械翻訳が誕生する。

第9章は「統計的機械翻訳と単語アラインメント」である。自動アラインメントと用例ベースの仕組みを理解した上で、本章ではソース言語とターゲット文の単語レベルでのアラインメントの対応を予測するという発想から、ターゲット言語のさまざまな単語列について、どれが最も高い確率でソース言語の翻訳であるのかを判定するという考え方をベイズの定理でモデル化した、という流れが容易に理解できる。

第10章「セグメントベースの機械翻訳」では、単語から句レベルでの統計的機械翻訳への挑戦が記されている。続く、第11章「統計的機械翻訳の課題と限界」では、それまでの統計的機械翻訳の限界点を、まず言語距離や訳出の方向性の観点から指摘する。例えば、日本語と英語の組合せよりも、フランス語と英語の翻訳のほうが精度が高くなることなどである。そこから、ハイブリッド機械翻訳システムの挑戦など、なんとなく「懐かしさ」をも感じる記述が続く。そして、根本的な問題「統計が多すぎる？」こと、言葉の「意味」を扱ってこなかったこと、そもそも意味はどのように表せるのか、という問題を取り上げている。明らかに、その後の深層学習をベースとするニューラル機械翻訳へのつながりになる。

第12章「ディープラーニングによる機械翻訳」では「単語の意味は、使われている文脈で分かる」という言語学者ファースの言葉が引用され、「単語埋め込み」処理から始まり、初期のニューラル機械翻訳の仕組み (RNN+注意機構型) が説明される。なお、原書出版のタイミングの関係で技術的にカバーできていない、self-attention を有する Transformer の仕組みについては中澤氏による解説「2020年時点でのニューラル機械翻訳」で補われている。

第13章「機械翻訳の評価」では、いわゆる BLEU スコアなどの自動評価の仕組みが語られ、14章「産業としての機械翻訳: 商用製品から無料サービスまで」では、機械翻訳の商用利用の実情が記されている。欧州委員会で行われる翻訳のうち自動の支援ツールを利用しているのは、5%にも満たない (ほとんどが人手翻訳) という現状は (p. 137)、今もそう変わらないのだろう。ポストエディットを行うにも、細心の注意を払わないと、取り返しのつかない結果になる恐れもある、という警告もされている (p. 151)。

第15章「結論として: 機械翻訳の未来」は締めくくりである。これまでの機械翻訳の開発者が、往々にして人間の翻訳の仕方や認知的な問題を扱おうとして来なかった点を指摘する。翻訳という作業は無限であり、そもそも知識とは何であり、何をコード化すれば

よいのか自明でない (p. 156)。最後に、繰り返すように、「bank」(川岸か銀行か?) の例を持ち出し、曖昧性の問題を訴える。それでも、人間の幼児が言語に接することで言語を習得していく過程に、単語の分布や使用頻度の高い単語から学ぶ、短文から学ぶというパターンが、それとなくニューラルネットワークの仕組みに類似するという観点から、機械翻訳の明るい未来を想像しながら本書は締めくくられる。

以上のように本書は、新しい情報を提供してくれるものではないにしても、これまで、断片的な理解しかしていなかった機械翻訳の歴史と仕組みを、あらためて整理してくれる書である。しかし、それよりも、ドキュメンタリー調かつドラマチックな感覚を読者に与えつつ「読ませる」スタイルで書かれている点が、なによりもオススメポイントである。ぜひ、一気読みしていただきたい。

3. 機械翻訳はどこへ行くのか？

さて、本稿の最後に、筆者が扱っているテーマの一つである、人間と機械翻訳の相互行為 (インタラクション)、具体的には、翻訳支援システムとしての機械翻訳の可能性について、本書から得たインスピレーションも含め述べる。

作者ポイボー氏は、言語の「曖昧性」を最大の問題と考えていた。そして、上で見たように、機械翻訳の開発の歴史とは、端的に言えば、曖昧性を解決する方法の模索であった。それは今も進行中である。またその問題が、並大抵の努力と既存の限定的な技術では解決できないという認識を持てるのかどうかであった。それに早い段階から気づいていたトロヤンスキーとバー=ヒレルは、共に「翻訳支援」ということを訴えていた。本書からの引用を交え、二人の考えをみる。

ロシアのスマルノフ=トロヤンスキーが、1930 年代に、特許を出願した装置は、単語を選択して符号化し、複数の言語間で翻訳を実現するとされていた (p. 28)。この発明が特徴的なのは、装置だけではなく作業環境に焦点を置いていたところである。翻訳者は、装置の

力を借りながら単語レベルで翻訳要素を見つけなければならないという観点で設計されていた。そのうえで、プロの編集者あるいは翻訳者が最終段階で介入し、文章を編集して文体論的な観点から修正を施すのである。完全な自動翻訳ではなく、翻訳支援の環境を思い描いていた点は注目に値する (p. 29)。しかし、この構想に注目すべき点が多いが、ほとんど忘れ去られている (p. 29)。

スマルノフ=トロヤンスキーの翻訳支援装置は、あまりに最先端すぎて当時のロシアでは、写植機の一種としかみなされていなかったようである。Hutchins and Lovtskii (2000) が、スマルノフ=トロヤンスキーの構想を再発見し、その考え方が再評価されるようになった。このあと、歴史的には、機械翻訳の開発競争に突入するのだが、トロヤンスキーが翻訳支援ツールにこだわった理由は、ある意味で、ことばの曖昧性の問題を見抜いていたのかもしれない。

機械翻訳の黎明期をリードしたバー=ヒレルもまた、翻訳支援システムの必要性を訴えた。上で記した“pen”という単語の曖昧性の問題、「子供を遊ばせる小さい囲い」の意味を推論できる手がかりはないからこそ、質の高い完全自動の機械翻訳 FAHQMT の非実現性を説いた。

彼は、自動翻訳の代わりに、研究者はコンピューター支援システムに目を向けるべきだと提唱した。黎明期に科学者の関心を引きそうにないのは明らかだったが、翻訳支援が開発されれば、翻訳に適した効率的なツールが提供されて、翻訳の生産性が、特に編集前や後編集の段階で上がるはずだと彼は考えていた。目標は翻訳者を支援することにあるので、従来の機械翻訳システムの出力とは異ならなければならない (p. 45)。例えば、翻訳者には翻訳の候補を提示するほうが一般的に望ましい (p. 45)。訳文をそのまま提示されても、修正に苦勞するだけだからである。そうすることで、いやそうすることでしか、無限の曖昧性の問題を解決する術はないと、彼もまた考えていたのかもしれない。

ちなみに、バー=ヒレルが挙げた下記の例文を、今のニューラル機械翻訳に入れてみても、残念なが

ら、“pen”に関する曖昧性は解決されていない。

Little john was looking for his toy box. Finally, he found it. The box was in the pen. John was very happy.

某 MT：小さなジョンは、自分のおもちゃ箱を探していました。ついに彼はそれを見つけました。その箱はペンの中にありました。ジョンはとても幸せでした。

ポストエディットとは、拡大解釈をすれば、機械翻訳を翻訳支援ツールとみなして利用するということである。バー=ヒレルは、翻訳支援ツールの出力は、機械翻訳のそれと同じであってはならないと主張した。それは、曖昧性を含めた適訳を決定するための選択肢の幅を翻訳者にできるだけ残しておくべきだと解釈することができる。プロ翻訳者の井口耕二は自身のブログの中で、柴田元幸氏の某著書の一節にリアクションを取りながら、以下のように述べる。

「影響されやすい人間」

自分は特に影響されやすいと柴田さんは言われているけれど、人間なんて、みんなそんなもんだと私は思っています。違いは、意識しているかないかであって。だから、MTPE で機械翻訳の出力を毎日あびるように読むのはあぶないだろうかと、ヒトゴトながら、心配しているわけです（井口, 2020）。

これは、機械翻訳の出力を人間が修正することばかりしていると、人間の翻訳者の側が、MT 出力に影響を受けてしまい、本来、適切な訳を選択できるはずなのに（曖昧性を適切に解消できる実力を持ち合わせているはずなのに）、それが（短期的にも中長期的にも）できなくなってしまうと、警告していると解釈できる。

また別の次元において、本書の解説を執筆した中澤氏は、「機械翻訳結果を細かく見れば、人間が間違えるはずのない箇所でも誤った翻訳がされていたり、一部の翻訳が抜けていたりなど、まだまだ人間の翻訳には及ばないことば明白である」と指摘する（p. 172）。これは、機械翻訳がボカミスを犯す点において、MT は人間に

劣ると言っているわけですが、であるならば、MT 出力をポストエディットするということは、単に MT の尻拭い的な作業とみなされても仕方ないということを示唆している。このように考えては、MT は翻訳支援ツールにはなりえない。全体を勘案すると、本書のあとがきで、訳者の高橋氏が、ポストエディットとは、従来の翻訳作業とは「指向性が全く異なる新しい仕事になるでしょう」（p. 185）ということの意味が、手に取るように理解できるのである。

翻訳支援に必要なことは、曖昧性を解消するための作業をより楽に行える環境を整えることが一つと、理想を言えば、人間がよりよい翻訳を発想できる環境とリソースを提供することであるが、まだまだそのような次元で、翻訳（支援）は考えられていないようである。MT は翻訳支援ツールになりえるのか。この方向に、MT は行くのか、それが問題だ。

参考文献

Hutchins, J., & Lovtskii, E. (2000). Petr Petrovich Troyanskii (1894-1950): A forgotten pioneer of mechanical translation. *Machine Translation*, 15(3), 187-221. Retrieved April 15, 2021, from <http://www.istor.org/stable/40009018>

井口耕二（2020）『ぼくは翻訳についてこう考えています』 ブログ：Buckeye the Translator（2020年2月12日）

<http://buckeye.way-nifty.com/translator/2020/02/post-878c35.html>

AAMT 2020, Online

～機械翻訳最前線～

出内 将夫

情報通信研究機構

1. はじめに

2020年12月2日に開催されたAAMT 2020, Onlineのイベント報告をします。コロナ禍の影響により、初めてのオンライン開催となる中、長尾賞受賞者紹介および受賞者講演4件、招待講演3件、パネルディスカッション1件と密度の高いイベントでした。また、Webサイト上に展示コーナーが設けられ、商品紹介も行われていました。講演資料は <https://aamt.info/aamt/2020online/handout-20201202/> からアクセスできますので、ぜひ、これらの資料をご参照ください。

2. 長尾賞受賞者紹介および受賞者講演

長尾賞受賞者紹介および受賞者講演では、第15回AAMT長尾賞2件と第7回長尾賞学生奨励賞2件の、計4件の受賞講演が行われました。

第15回AAMT長尾賞（2件）

◎コニカミノルタが取り組む多言語通訳システム：

川崎健氏／コニカミノルタ株式会社 BIC Japan

◎特許庁 機械翻訳システムの開発：

西本俊之氏／東芝デジタルソリューションズ株式会社
第7回長尾賞学生奨励賞（2件）

◎Non-autoregressive Neural Machine Translation
Based on Latent-Variable Models 潜在変数モデルに
基づいた非自己回帰型ニューラル機械翻訳：

朱中元(Raphael Shu) 氏／東京大学大学院情報理工
学系研究科（現 Naver AI Research）

◎大語彙フレーズ翻訳機能付き SMT・NMT ハイブリッド翻訳方式：

龍梓氏／筑波大学大学院システム情報工学研究科（現
深圳技術大学）

下記は、講演資料や質疑内容より、報告者がまとめたものであることを注記します。またAAMT Journal No.73に各受賞者ご自身による記事が掲載されていますので、そちらもご参照ください。

◎コニカミノルタが取り組む多言語通訳システム：川崎健氏

コニカミノルタは、人間の通訳と機械の通訳を併用したタブレット型医療用翻訳アプリ「MELON」を2016年に開発した。起動が早く短い言葉のやり取りが得意な機械通訳と、方言や重い話への確実な対応ができる人間の通訳を、シーンに応じて使い分けることで、医療現場に広く受け入れられている。コロナ禍においても、空港での検疫や自治体での相談などで利用されている。

MELONでは、機械翻訳(MT: Machine Translation)の結果が話者の発話内容の直下に表示され、時系列に会話が把握でき、逆翻訳による翻訳結果確認が行える。現場検証を通じて顧客の本音を調査し、顧客を取り巻く環境変化を把握したことが、開発の成功要因となった。また、アジャイルで素早く新機能を提供し顧客の生の反応を得たことも、顧客に受け入れられるソリューション提供に繋がった。

質疑では、学習データの集め方に対する質問があり、論文のコーパスを活用して医療分野の専門用語の対応

をしたことについて回答があった。また、外国でのビジネス展開について質問があり、データ管理に関する法令や規則や海外の母国語対応が必要なため、現状対応していない旨の説明があった。また、逆翻訳の活用について質問があり、文化や社会制度の差異により置き換える言葉が無い場合に訳が崩れてしまうという課題の説明があった。また、医療現場での致命的な翻訳誤りを防ぐ工夫についての質問に対して、どのような時に MT から人の遠隔通訳に切り替えるかの説明があった。

◎特許庁 機械翻訳システムの開発：西本俊之氏

特許庁の MT システムは、国内特許情報を海外に発信する際、中国語や韓国語の文献を国内ユーザが参照する際に、MT による訳文を提供している。特許情報プラットフォーム「J-PlatPat」のバックエンドとして動作しており、日英翻訳では、ピーク時に 20,000 アクセス/時の翻訳処理を行っている。2019 年の日英翻訳機能リリース、2020 年の中日・韓日翻訳機能リリースを、東芝デジタルソリューションズが担当した。

特許庁の MT システムの翻訳対象は、主要な特許公報（請求項、明細書、要約書）だけでなく、意匠・商標・審判公報や多様な審査書類など、多種多様である。書類ごとに語彙や表現が異なるほか、一つの書類の中でも書類ヘッダ、本文、書類フッタなどの部分によって特徴が変わる。この特徴に合わせて、ニューラル翻訳（NMT: Neural Machine Translation）、統計翻訳（SMT: Statistical Machine Translation）、ルールベース翻訳（RBMT: Rule-Base Machine Translation）から、最適な翻訳エンジンを組合せた。そのほか、レイアウト解析処理、特殊表現退避処理、長文分割や誤訳判定など、翻訳品質を向上するための自然言語処理を組合せ、大規模パブリッククラウドでの並列分散処理により、数秒でのリアルタイム翻訳を実現した。

特許庁の MT システム開発にあたり、翻訳の前後処理を支える辞書整備や、1 文が 8,000 文字を超える長文への対応、1 文献の大きさが大きい場合の訳質調整

やエラー処理なども行う必要があった。今後も、アプリケーションで NMT の前後処理を工夫する NMT とアプリケーションの融合や、特許翻訳の最適解追及、他分野展開を行っていく。

質疑では、特許庁の辞書が公開されているか質問があり、現状は非公開であり、公開は特許庁の判断となる旨の回答があった。

◎Non-autoregressive Neural Machine Translation Based on Latent-Variable Models 潜在変数モデルに基づいた非自己回帰型ニューラル機械翻訳：朱中元 (Raphael Shu) 氏

NMT では、通常自己回帰型のネットワーク構造が活用されており、各単語を生成する際に、前にくる単語を元に次の単語を予測している。この方法の問題点として、逐次的に予測するため並列化ができない点や、文が長くなると翻訳速度が落ちる点がある。非自己回帰型のネットワークにより、すべての単語の予測を、一度の前向き推論で完了でき、並列化が容易となる。既存研究としては、マスキングした単語の予測を繰り返す Facebook の Conditional Masked Language Model (CMLM)^[1]がある。

そこで、潜在変数モデルを使い非自己回帰型ネットワークを実現するアプローチを提案した。入力文の各単語がどう翻訳されるかを、連続潜在変数を用いて訓練する。実験では 8 次元の潜在変数を用いた。訓練では、変分下限 ELBO (evidence lower bound)の最大化を目的関数として、入力文から参照文とその文長が予測できるよう、潜在変数を求める。予測では、潜在変数の更新を繰り返しより良い訳文を得る Delta inference^[2]と、潜在空間すべてを評価するエネルギーモデルに従い勾配に沿って最適な訳文を得る Energy-based inference^[3]を提案した。WMT14 En-De を用いた実験の結果では、自己回帰型のネットワーク構造と比較し、BLEU スコア 1~2 程度の品質劣化に抑えつつ、翻訳速度を 6~10 倍に高速化できた。

質疑では、自己回帰型の翻訳品質までもう一歩だが、

もう一步となっている要因は何か、という質問に対し、一回目の潜在変数の予測、翻訳の予測が重要だと分かっており、文が長くなると単語の重複などのミスが増えている、との回答があった。

◎大語彙フレーズ翻訳機能付き SMT・NMT ハイブリッド翻訳方式：龍梓氏

本講演は、電子情報通信学会で受賞した論文^[4]の内容に基づく。NMT は従来の SMT に比べて高い翻訳精度を実現するが、語彙制限による未知語問題が知られている。大規模語彙を扱うために事前に未知語トークンに置き換える従来手法^[5]も存在するが、フレーズに対応できない問題があった。そこで、NMT における未知語の問題に SMT のフレーズ間対応を活用することで、複合語を含むフレーズに対応する手法を提案した。

提案手法はモデル訓練と訳文生成の 2 ステップからなる。ステップ 1 では、日中対訳特許文に対して、SMT を用いてフレーズ翻訳テーブルと単語間対応を取得しフレーズ対訳対をトークン対に置き換え、置き換え後の文を NMT で訓練する。ステップ 2 では入力文に対し、NMT モデルで訳文を生成し、SMT モデルを用いてフレーズ翻訳する。

特許翻訳を対象とした実験の結果、従来の未知語処理手法より BLEU、人手評価において最も高い翻訳精度を実現し、名詞句の評価も SMT に次ぐ精度を実現した。また、訳抜けを 30% 程度低減できた。現状は Transformer^[6]の学習力や、大規模コーパスから訓練された Pretrained モデル^[7]によって、フレーズの翻訳問題も解決しつつあるが、フレーズ単位の精度評価や非構成的なフレーズの翻訳などの研究領域があると考えている。

質疑では、提案手法で辞書を利用できるかの質問があり、辞書も組み合わせて利用できる、との回答があった。

3. 招待講演

招待講演は次の 3 件でした。

◎ポストエディター教育について：

川口仁氏／インタースクール

◎機械翻訳の最新技術動向調査 ～特許翻訳への応用を見据えて：経過報告～：

須藤克仁氏¹／奈良先端科学技術大学院大学

◎機械翻訳と翻訳支援ツールの関係性について：

佐藤弦氏／SDL ジャパン株式会社

江上聡氏／シストラランジャパン合同会社

それぞれの講演について、前述の講演資料より次にまとめます。

◎ポストエディター教育について：川口仁氏

ポストエディター教育について、「なぜ教えているのか」「何を教えているのか」「何を感じたのか」「これから何を教えるのか」の観点から説明する。

ポストエディター教育を行っている理由は、市場環境の変化に対応できるしたたかな翻訳者を育てるためである。翻訳者には 2 つの面があり、職人としての翻訳者、ビジネスパーソンとしての翻訳者がある。前者は、翻訳品質を最優先に考え、自分の翻訳結果にこだわりを持つ。後者は、お客様の満足を最優先に考える。今までは人手翻訳(HT: Human Translation)と MT が対立軸で捉えられていたが、これからは効率を上げるために MT を活用していく with MT の時代となる。

ポストエディター教育は、顧客満足重視のビジネスマインドを持つ柔軟性のあるエディターが対象となる。ポストエディット(PE: Post Editing)講座の内容は、PE の重要性についてのガイダンス、翻訳支援(CAT: Computer Aided Translation)ツールを使った PE 体験、Light PE と Full PE の差異、事前課題文添削などである。基礎スキルを身に着けるため、実践的、体系的、双方向性重視の講座となっている。

PE に対する翻訳者の理解を深めるため、ポストエ

¹ AAMT/Japio 特許翻訳研究会副委員長

ディテーター教育では、まず PE におけるエディターの重要性、需要の高さ、仕事のラインアップ拡大について説明する。PE では HT とは別のスキルが必要となる。また、Light PE と Full PE について、Light はスピードとコスト重視であること、Full では単文ベースで MT の訳が正しく見えたとしても全文を見た調整が必要であることを説明する。ポストエディター養成は、業界全体の課題であるため、翻訳会社間でポストエディターの役割を揃えていくとともに、MT に向かない原文を PE 案件として引き受けない工夫が必要である。

これからは翻訳というビジネスへのスタンスを考えていく必要がある。HT の案件だけで PE の案件は受けない、という選択肢もあるし、PE と HT 双方の案件を受けていく、という選択肢もある。HT だけではなく、いろいろな技術の選択肢を選べることで幅が広がる。

質疑では、MT の利用により英語力が落ちるのでは、という質問に対し、MT 訳が参考となることもあり、最大のパフォーマンスを発揮するために MT 活用もありではないか、という回答があった。また、PE の報酬が HT より低い点について、報酬の向上に対する取り組みはあるか、という質問に対して、PE の報酬が下がるのは、ある程度やむをえないが、仕事として PE が HT より下ということはない。報酬については業界全体で取り組む課題であると考えたとの回答があった。また PE での心がける必要があることとして、特に Light PE の場合はマインドセットを変え、割り切りが必要になることについて説明があった。

◎機械翻訳の最新技術動向調査 ～特許翻訳への応用を見据えて：経過報告～：須藤克仁氏

MT の技術は急速に変化している。1950 年代の RBMT から始まり、1980 年代に用例ベース、2000 年頃から SMT、2010 年代半ばから NMT と変遷してきた。NMT 以前は構成性(compositionality)に基づいており、文を構成する語や句での対訳関係を考えていた。NMT は文全体を符号化して、それを復号化するプロ

セスとなっており、構成性は保証されない。従来型 MT は「木を見て、森を見ず」、NMT は「森を見て、木を見ず」とも言える。結果として入力的全被覆が保証されず、抜けや重複が発生することがある。現在の NMT の標準モデルとして Transformer が用いられており、層を単位とした類似の計算が複数回行われる。一般的に行われる工夫として、単語より短い単位（サブワード）への分割^{[8][9]}や、逆翻訳による単言語コーパスからの対訳コーパス生成が行われている。

AAMT/Japio 特許翻訳研究会では、特許翻訳に使えるような技術を探すために、特許翻訳の課題と関連する MT 技術の動向を調査している。特許文書の特徴として、内容が専門的、文体が特殊、公的な文書、文構造が複雑で長い、などがある。今回は、デプロイ対策、低リソース言語対策、自動評価の 3 つの課題について報告する。特許翻訳そのものを題材にした研究は少ないため、特許翻訳に役立つと思われる技術を調査している。

デプロイ対策としては、パラメータ共有や知識蒸留などのモデルの軽量化^{[10][11]}、浮動小数点の半精度(fp16)化や整数への量子化などのハードウェア特性活用^{[12][13]}、C/C++での実装や途中計算の再利用など実装の工夫が行われている。

低リソース言語対策では、NMT が低リソース言語に対して不利と言われていた^[14]が、チューニング次第で不利とは言えないことが分かってきた^{[15][16]}。高リソース言語（典型的には英語）を仲介して翻訳するピボット翻訳^[17]や、高リソース言語で事前学習した後に低リソース言語で追加学習する転移学習^{[18][19][20]}が行われている。また、複数言語を翻訳できる単一の翻訳モデルが研究されている^{[21][22]}。

自動評価は、基本的に表層を用いる言語横断的なモデルフリー手法、言語ごとにモデルを持つ汎用モデルベース手法、タスクごとにモデルを持つ専用モデルベース手法の 3 つに大別できる。モデルフリー手法には、BLEU^[23]、NIST^[24]、TER^[25]、IMPACT^[26]、RIBES^[27]、METEOR^[28]、MEANT^[29]、chrF^[30]、EED^[31]があり、

高速で利用が容易だが精度が今一つである。汎用モデル手法には、YiSi^[32]、WE_WPI^[33]、BERT_R^[34]、BERTScore^[35]、MoverScore^[36]があり、学習済みの資源を活用し表層のみの手法より精度が高いと言われていたが、専用モデルほどではない。専用モデルには、BEER^[37]、BLEND^[38]があり、精度が高いがデータが必要で計算量が多い。自動評価法の「メタ評価」も行われ、人手評価との相関の評価自体への問題提起もされており^[39]、今後さらに研究が進むと思われる。

質疑では、特許翻訳研究会での他のテーマについて質問があり、専門用語の翻訳、未知語、長文の翻訳などのテーマがある、という回答があった。また、特許翻訳会社や特許翻訳者との共同研究について質問があり、特許情報シンポジウムでは連携しているものの、特許翻訳研究会として現在連携していないという回答があった。前編集の取り組みについての質問には、特許翻訳研究会では現時点の範囲としていないが、産業日本語研究会²で行っているという回答があった。

◎機械翻訳と翻訳支援ツールの関係性について（翻訳支援ツールの観点から）：佐藤弦氏

翻訳支援ツール（CAT ツール）の概要を含め、MT の関係を説明する。Trados の歴史は 1984 年にスタートし、1997 年に Microsoft 社が選んだことでメジャーとなった。2005 年には、翻訳会社である SDL 社が Trados を買収し、業界を揺るがした。現在の最新バージョンは SDL Trados Studio 2021 となっている。CAT ツールは翻訳メモリ(TM: Translation Memory)と翻訳エディティングツールが合わさったもので、TM は過去の翻訳結果を原文と訳文を対にしてデータベースに蓄積したものである。CAT ツールでは、原文とマッチ率が高い文を参照でき、編集した訳文が自動的に TM に登録される。マニュアルなど繰り返しが多い文書や、改版がある文書で力を発揮する。

MT と TM の使い分けとして、SNS や Web のコメントなどでは、MT が良く使われており、マニュアル、

IR、法務などのビジネス文書では、TM と MT が併用されている。日本における併用の方法では、ネイティブではない言語方向への HT における MT 結果参照や、ネイティブな言語方向における MT/PE 導入がある。英日翻訳では、顧客が欧米企業であることが多く、海外で決められた標準に従うことにより、MT/PE の採用が増えている。CAT ツールを併用する理由として、MT での翻訳ミスの修正や、用語・文体の統一、XML などのタグ処理がある。Trados は様々な MT と連携できるようにになっている。

PE が抱えている課題として、用語の統一、文体の統一、流畅さをどの程度確保するのか、といった品質基準整備や、ビジネスを進める上では品質に加え、価格や手順を確立していく必要があると考えている。

◎機械翻訳と翻訳支援ツールの関係性について（機械翻訳の観点から）：江上聡氏

MT の観点から CAT ツールとの関係性について説明する。SYSTRAN は 1968 年に設立され、1969 年に世界初の商用自動翻訳システムを米国政府に納入した。2000 年代に入り、確率を組み合わせたエンジンとなり主要ポータルサイトへの導入が進んだ。2016 年にハーバード大と共同開発した NMT を無償公開した。現在は SYSTRAN Pure neural SERVER に最新技術を搭載している。また、翻訳者が専門エンジンを学習・販売できる仕組みである SYSTRAN Market place を展開している。

SYSTRAN Pure neural SERVER には、PE 向けの機能として、PE 後の翻訳文を TM にそのまま追加するほか、それを追加学習する機能を搭載している。この機能を活用して、日々の翻訳フローと MT エンジンの学習フローを連携させることで、翻訳対象に特化した MT が作成でき、PE 量を持続的に削減できる。製造業のカatalog日英翻訳の事例では、導入 1 年後で翻訳費を 56%削減、導入 2 年後で 80%削減する効果があった。

TM をさらに MT で有効活用する技術として、MT

² <https://www.tech-jpn.jp/>

でTMのファジーマッチ情報を加味して翻訳する技術である Neural Fuzzy Adaptation (NFA)を2021年前半に実装予定である。専門エンジンとNFAを組み合わせることで、複数分野で60を超えるBLEU値が出せることが確認できている。

佐藤氏、江上氏の発表への質疑では、ファジーマッチの活用について両者に質問があり、江上氏からはベクトル化してAIで活用する旨、佐藤氏からはマッチ率が高い場合に差分を自動修正する技術を活用している旨、回答があった。また、ツールの機能が増えることによるデジタルディバイドへの対策について質問があった。両氏からは、多くのユーザが使うツールとUIを合わせることで、直感的に操作できる工夫をしていると回答があった。また、MT精度向上のためにSDL社とSYSTRAN社が提携または共同研究しているか、という質問については、行っていない旨回答があった。ファジーマッチの解析技術がCATツールによって差があるか、という質問に対して、佐藤氏から日本語では文字単位か単語単位で差が出る旨、回答があった。

4. パネルディスカッション「ポストエディットのこれからを考える」

次のメンバーでパネルディスカッションが行われました。モデレーターは、安達久博氏／一般社団法人アジア太平洋機械翻訳協会(AAMT)、パネリストは、中山雄貴氏／株式会社ヒューマンサイエンス、成田崇宏氏／株式会社ホンヤク社、恒川正志氏／個人翻訳者、山田優氏／関西大学。

冒頭で、モデレーターの安達氏より、以下の説明がありました。

AAMTでは今年がPE元年と位置付けている。議論の前提として、次のことに留意してほしい。まず、MTが出力するのは、翻訳文「らしき」単語列であり、人間がチェックしないと翻訳文とならない。また、翻訳とPEを対象としたISO規格がそれぞれあるが、翻訳の規格でHTであった一次翻訳をMTに置き換えたも

のがPEの規格となっている。PEの規格でも、レビューの工程は必ず人が携わる必要がある。HTではTMが使われていたが、マッチ率が低くてもTMの代わりにMTが使えるようになった、と捉えるのが良い。

それぞれのパネリストの話題について、前述の講演資料より次にまとめます。

◎ポストエディットに対するニーズとトレンド ～翻訳会社の視点から～：中山雄貴氏

産業翻訳におけるMT/PEのトレンドとして、2000年代後半からSMTを活用してIT分野で導入が始まり、2016年にNMTが実用化されるとともに、IT以外の分野に広がった。2020年現在は、医療などの分野でもMT/PEが導入されている。MT/PEは、短期間・低価格というニーズから導入が広がっている。

MT/PEでは品質要件の合意と共有が一番重要である。翻訳を始める前に、依頼側と翻訳側で品質を合意する。品質基準として大きく、正確さ・流暢さ・用語・スタイルの基準があり、それぞれ詳細を定め、翻訳する文書に応じて要件を決めている。ポストエディターは翻訳者が担当することが多い。MTを巨大なTMと捉えれば、CATツール未経験の場合以外は、翻訳から新たなスキルは必要ないが、作業内容と割合が変わる。HTは訳文の入力に一番時間がかかっており、PEでは訳文の確認に時間がかかる。MT訳の問題点としては、翻訳が文単位のため、文脈に応じた表現や用語で訳せない。また、係り受けが間違っていることや、用語・スタイルが統一できないことも問題となる。翻訳者から「MT/PEを受けたくない」という声があるが、その理由は、作業内容が異なること、単価が低いこと、用語やスタイルの修正が手間であることが多かった。

翻訳部門や翻訳会社が翻訳者のためにできることとして、品質要件を明確化し翻訳者と共有すること、最善のMTエンジンを使い正確さや流暢さを改善すること、MTエンジンの用語機能やスタイル機能を確認し活用することが挙げられる。用語やスタイルの改善に、PEで2割の時間がかかっていたため、ヒューマンサ

イエンスでは、Memsource/Trados 向けにスタイルの後編集や用語を置き換えるプラグインを開発した。ある案件では、用語やスタイルの修正が 109 文中 95 文 661 カ所で必要だったものを、プラグイン利用により自動で修正できた。

◎ポストエディット受託側から見た問題点：成田崇宏氏、恒川正志氏

PE 案件受託側から見た問題点は「翻訳会社の仕切りの問題」である。翻訳会社が PE 案件の経験不足から案件を仕切れていない。翻訳会社は、顧客の過大な期待、技術的な問題、案件における QCD³のバランスを仕切るべきである。

翻訳者にとっての PE のメリットは、訳文のブラッシュアップ作業に専念でき、省力化・時間短縮につながる点がある。PE のデメリットは、MT 出力の質次第でコスト削減のしわ寄せがポストエディターに来ることや、顧客要求が低コスト・短納期の場合に、翻訳者が品質を妥協せざるを得ない点である。翻訳者やポストエディターが翻訳会社に求めるものとして、案件における MT/PE 工程の適否判断、実工数に基づいた合理的な条件、仕上がりに関する目標地点の合意がある。

翻訳会社はその期待に応えられないと、PE 業務の魅力がなくなり、PE 案件をやりたい人・やれる人がいなくなってしまう。ISO では PE について「翻訳品質の適切な見極めができる人のみが行える特別な仕事」と定めており、不適切な運用を行うと、翻訳業界全体の品質を崩しかねない。翻訳会社は、PE 案件に対して持続可能にするという意識を持ち、顧客・翻訳会社・翻訳者／ポストエディターの三方良し、をめざすべきである。

そのために、翻訳会社は顧客の過大な期待に対して、メリット・デメリットや MT/PE の適否を提示し、技術的な問題に対応できるようエンジンやプラットフォームを研究し、案件の QCD バランスと顧客ニーズをマッチングする必要がある。また業界全体として、

基準やベースを定め、ISO の浸透、各案件の SLA 締結、SOW⁴明確化を進めるべきと考える。

中長期的には、PE のメリットとして、ポストエディターがブラッシュアップに専念でき、省力化・時間短縮の効果が高めることで、PE が魅力的になると考える。また、PE の効率化により、報酬が割に合うようになり、文をより良くすることにやりがいを得られ魅力的になると考えられる。

◎ポストエディットのある翻訳業界、世界：山田優氏

持続可能な翻訳業界を考える上で、今年が PE 元年ということで、翻訳者とポストエディターは仕事が違うという議論があり、支払い面や社会的環境など、まだ業界として受け入れる環境が整っていない。何を直していくのかについて、いくつかの学術研究踏まえて、今後の PE を考えるヒントにできれば、と考えている。

現状把握として、Commonsense Advisory によると、アクティブなプロの翻訳者は約 20 万人で、1 日に 4 億ワード翻訳しかない。MT 含めると 1 日に 1,400 億ワード翻訳されているという報告もあり、翻訳需要は 1%以下しか満たされていない。

Full PE と Light PE の定義があいまいに捉えられていることも問題となるため、定義を確認する。Full PE は HT と見分けがつかない訳文に仕上げるために、正確性と流暢性・スタイルについて注意を払う項目が定義されており問題はない。一方、Light PE は正確性のみで良い、という定義だが、これが本当にデリバリーできるのか、という問題があると考えている。PE 結果で MT をドメインアダプテーションし、Full PE と Light PE を組み合わせる使い方も考えていく必要がある。

翻訳者が文化的な資本としてリスペクトされているのに対し、ポストエディターがコスト削減を担う立場として、社会的地位が低くみられる傾向があるという報告^[40]もある。これに対して、HT と PE に求められるスキルが違うのか、という点を研究し、両者に求め

³ Quality (品質)、Cost (コスト)、Delivery (納期)

⁴ statement of work / 作業範囲

られるスキルは同じという結果が出た。これに関する研究を3つほど紹介する。

SMTとNMTの結果を学生がPEした結果を比較した研究^[41]からは、NMTの方が高品質でPE結果も良かったが、PE作業の負荷は軽減できなかった、という結果になった。エラーの種類としては、NMTでは意味的なエラーと用語エラーが多い。あまり実力のない翻訳者は用語エラーを直せるが意味的なエラーを直せないため、正しく直せずに終わってしまう。

別の研究として、英語から中国語（繁体字）とスペイン語への翻訳を対象に、翻訳時の負荷をアイトラッカーで記録しつつ、HTおよびMT結果に、Multidimensional Quality Metrics (MQM)でタグ付けを行い、作業者のタグ一致率を分析した研究^[42]がある。正確性エラーは付与者間で一致するが、流暢性エラーは付与者間で一致しない結果となり、流暢性は翻訳者の好みによることが多いことが分かった。また、MTの正確性エラーが起きた箇所は、HTとPE双方で作業負荷が高かったことから、HTとPEのスキルの共通性が垣間見える。

別の実験^[43]でも、翻訳力・言語力に加え、情報検索能力が重要であることが分かっている。学生とプロの翻訳者の比較から、正確性エラーに関して学生が辞書しか参照せず間違えていたことが分かった。また、様々な検索エンジンや翻訳対象の文献を参照することで、誤訳を避けられることが分かった。HTもPEも同様に高度なスキルが求められることから、PEの社会的な価値や対価を考えていく必要がある。

議論のための情報提供として、MTでどの程度効率が上がるかを、翻訳者にTMとMTという情報を与えずに評価した実験^[44]で、80%マッチ以下のTMよりMTの評価が良かった、という結果となった。このことからPEは80%マッチのTMより効率的に作業できる可能性がある。

◎ディスカッション

安達氏：

中山さん成田さんに質問が入っています。

MT/PEの適否を決定する際のMT品質の見極めはどのような基準で行っていますか。機械的な判断や人手による判断ですか。効率的な運用はありますか。

中山氏：

両方使っている。機械的な判断として、月に一度、ITと医療系で機械的な評価をしている。不定期ですが人で評価も行っています。人手評価の方が確実ですが、BLEU値などの機械的評価も人手評価に相关している感触があるので、併用しています。

成田氏：

両方やっている。機械的評価ではBLEUを使っており、人手評価は人によって評価に差が付かないように工夫している。

安達氏：

山田先生に質問が来ています。

総合的にHTがPEになることによって、作業時間がどの程度削減できるとお考えでしょうか。

山田氏：

数字を言ってしまうと独り歩きしてしまうので怖いですが、英語とスペイン語の例では80%マッチより効率が良い場合がある。MT/PEは効率が上がると信じているかどうかで、差が出るという研究もある。個人的には2割くらいと考えている。

中山氏：

翻訳会社の翻訳フローとして、PMの作業や翻訳後のレビューがある。翻訳全体のコストが2割削減されるわけではないことは留意いただきたい。

安達氏：

それぞれの立場で、こうすべきというご意見がありましたが、それを実行に移すために、誰に何を言うのが最短でしょうか。

中山氏：

MT エンジンを開発している方々には、より良いエンジンを開発いただきたい。また、一般向けの MT が多く、産業翻訳向けに機能が足りない部分があるため、産業翻訳にも目を向けていただきたい。翻訳者に向けては、一から翻訳したい、という欲求さえ抑えられれば、PE も割と楽しめると考えているので、意識を変えていただきたい。

成田氏：

翻訳会社が MT/PE を使ったサービスを提供するのであれば、顧客に対して MT/PE の良いところ悪いところ含めて伝えていくのが大事だと考えている。PE をやりたがらない翻訳者の方も一定数いるが、翻訳会社が環境整備できる部分もまだあると考えている。翻訳会社が PE の価値を認めた上で、どうサービス提供するか、顧客に説明するか、考えていく必要がある。それにより、ポストエディターが増え、クリエイティブなことが起き、好循環が生み出せばよいと考えている。

山田氏：

現状は、翻訳会社が顧客に説明する部分が一番重要な部分と考える。アカデミア、産学の共同研究的な立場からは、2 つ言いたいことがある。Light PE で流暢性不要とあるが、人間が流暢性を気にせず作業できない、というのが研究で分かっており、流暢性を無視して正確性を高めるのは負荷が高くなる懸念がある。正確性に関してどうしても間違っはいけないことがある場合に、リスク管理や MT 品質の研究を進めるべきと考えている。

恒川氏：

翻訳者としては、顧客に PE の価値を認めていただくには、品質を認めていただくことが大切だと考えている。それにより、業界に貢献できればと考えている。

安達氏：

今後、翻訳業界で顧客含めた取り組みを考えていければよいと考えている。

5. おわりに

初めてのオンライン開催となったにもかかわらず、大きなトラブルもなく、活発な議論が行われました。また、ウェビナーの録画を 1 週間オンデマンドで視聴可能とすることや、参加者同士のオンライン交流企画など、新たな試みも行われました。本会は第 2 回 AAMT 年次大会でもありました。今後の年次大会にもご期待ください。

参考文献

- [1] Ghazvininejad, M., Levy, O., Liu, Y., and Zettlemoyer, L. (2019). "Mask-Predict: Parallel Decoding of Conditional Masked Language Models". Paper presented at the Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China.
- [2] Shu, R., Lee, J., Nakayama, H., and Cho, K. (2020). "Latent-Variable Non-Autoregressive Neural Machine Translation with Deterministic Inference Using a Delta Posterior". Proceedings of the AAAI Conference on Artificial Intelligence, 34(05), 8846-8853.
- [3] Lee, J., Shu, R., and Cho, K. (2020). "Iterative Refinement in the Continuous Space for Non-Autoregressive Neural Machine Translation". Paper presented at the Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online.
- [4] 龍, 梓., 木村, 龍., 飯田, 頌., 宇津呂, 武., 三橋, 朋., and 山本, 幹. (2019). "フレーズ・トークン込み NMT モデル及び SMT による大語彙フレーズ翻訳によるハイブリッド翻訳方式". 電子情報通信学会論文誌 D, J102-D, 104-117.
- [5] Luong, T., Sutskever, I., Le, Q., Vinyals, O.,

and Zaremba, W. (2015). "Addressing the Rare Word Problem in Neural Machine Translation". Paper presented at the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China.

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). "Attention is all you need". *Advances in Neural Information Processing Systems*, 5998--6008.

[7] Song, K., Tan, X., Qin, T., Lu, J., and Liu, T.-Y. (2019). "MASS: Masked Sequence to Sequence Pre-training for Language Generation". In C. Kamalika & S. Ruslan (Eds.), *Proceedings of the 36th International Conference on Machine Learning* (Vol. 97, pp. 5926--5936). *Proceedings of Machine Learning Research: PMLR*.

[8] Sennrich, R., Haddow, B., and Birch, A. (2016). "Neural Machine Translation of Rare Words with Subword Units". Paper presented at the Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany.

[9] Kudo, T., and Richardson, J. (2018). "SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing". Paper presented at the Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Brussels, Belgium.

[10] Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., and Kaiser, L. (2019). "Universal Transformers". *7th International Conference on Learning Representations, ICLR 2019*. New Orleans, LA, USA.

[11] Dabre, R., and Fujita, A. (2019). "Recurrent Stacking of Layers for Compact Neural Machine Translation Models". *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 6292-6299.

[12] Devlin, J. (2017). "Sharp Models on Dull Hardware: Fast and Accurate Neural Machine Translation Decoding on the CPU". Paper presented at the Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark.

[13] Aishwarya, B., Vamsi, S., Deepthi, K., Vivek, M., Sun, C., Kushal, D., and Vikram, S. (2019). "Efficient 8-Bit Quantization of Transformer NeuralMachine Language Translation Model". *ICML Joint Workshop on On-Device Machine Learning & Compact Deep Neural Network Representations*.

[14] Koehn, P., and Knowles, R. (2017). "Six Challenges for Neural Machine Translation". Paper presented at the Proceedings of the First Workshop on Neural Machine Translation, Vancouver.

[15] Sennrich, R., and Zhang, B. (2019). "Revisiting Low-Resource Neural Machine Translation: A Case Study". Paper presented at the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy.

[16] Duh, K., McNamee, P., Post, M., and Thompson, B. (2020). "Benchmarking Neural and Statistical Machine Translation on Low-Resource African Languages". Paper presented at the Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France.

[17] Costa-jussa, M. R., Casas, N., and Melero, M. (2018). "English-Catalan Neural Machine Translation in the Biomedical Domain through the

cascade approach". Paper presented at the workshop "MultilingualBIO: Multilingual Biomedical Text Processing" of the 11th Edition of the Language Resources and Evaluation Conference.

[18] Nguyen, T. Q., and Chiang, D. (2017, nov). "Transfer Learning across Low-Resource, Related Languages for Neural Machine Translation". Paper presented at the Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Taipei, Taiwan.

[19] Murthy, R., Kunchukuttan, A., and Bhattacharyya, P. (2019). "Addressing word-order Divergence in Multilingual Neural Machine Translation for extremely Low Resource Languages". Paper presented at the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota.

[20] Dabre, R., Fujita, A., and Chu, C. (2019). "Exploiting Multilingualism through Multistage Fine-Tuning for Low-Resource Neural Machine Translation". Paper presented at the Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China.

[21] Firat, O., Cho, K., and Bengio, Y. (2016). "Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism". Paper presented at the Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, California.

[22] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viegas, F., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J. (2017). "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation". *Transactions of the Association for Computational Linguistics*, 5, 339-351.

[23] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002, jul). "Bleu: a Method for Automatic Evaluation of Machine Translation". Paper presented at the Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, USA.

[24] Doddington, G. (2002). "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics". *Proc. ARPA Workshop on Human Language Technology*, 2002.

[25] Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). "A study of translation edit rate with targeted human annotation". Paper presented at the Association for Machine Translation in the Americas.

[26] Echizen-Ya, H., and Araki, K. (2007). "Automatic evaluation of machine translation based on recursive acquisition of an intuitive common parts continuum". *Proc. 10th Machine Translation Summit (MT Summit XD)*, 2007.

[27] Isozaki, H., Hirao, T., Duh, K., Sudoh, K., and Tsukada, H. (2010). "Automatic Evaluation of Translation Quality for Distant Language Pairs". Paper presented at the Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA.

[28] Banerjee, S., and Lavie, A. (2005). "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments". Paper presented at the Proceedings of

the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, Ann Arbor, Michigan.

[29] Lo, C.-k., and Wu, D. (2011). "MEANT: An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility based on semantic roles". Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA.

[30] Popovic, M. (2015). "chrF: character n-gram F-score for automatic MT evaluation". Paper presented at the Proceedings of the Tenth Workshop on Statistical Machine Translation, Lisbon, Portugal.

[31] Stanchev, P., Wang, W., and Ney, H. (2019). "EED: Extended Edit Distance Measure for Machine Translation". Paper presented at the Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1), Florence, Italy.

[32] Lo, C.-k. (2019). "YiSi - a Unified Semantic MT Quality Evaluation and Estimation Metric for Languages with Different Levels of Available Resources". Paper presented at the Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1), Florence, Italy.

[33] Echizen'ya, H., Araki, K., and Hovy, E. (2019). "Word Embedding-Based Automatic MT Evaluation Metric using Word Position Information". Paper presented at the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota.

[34] Mathur, N., Baldwin, T., and Cohn, T. (2019).

"Putting Evaluation in Context: Contextual Embeddings Improve Machine Translation Evaluation". Paper presented at the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy.

[35] Tianyi, Z., Varsha, K., Felix, W., Kilian, Q. W., and Yoav, A. (2020). "BERTScore: Evaluating Text Generation with BERT". Paper presented at the International Conference on Learning Representations.

[36] Zhao, W., Peyrard, M., Liu, F., Gao, Y., Meyer, C. M., and Eger, S. (2019). "MoverScore: Text Generation Evaluating with Contextualized Embeddings and Earth Mover Distance". Paper presented at the Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China.

[37] Stanojevic, M., and Sima'an, K. (2014). "BEER: BETter Evaluation as Ranking". Paper presented at the Proceedings of the Ninth Workshop on Statistical Machine Translation, Baltimore, Maryland, USA.

[38] Ma, Q., Graham, Y., Wang, S., and Liu, Q. (2017). "Blend: a Novel Combined MT Metric Based on Direct Assessment -- CASICT-DCU submission to WMT17 Metrics Task". Paper presented at the Proceedings of the Second Conference on Machine Translation, Copenhagen, Denmark.

[39] Mathur, N., Baldwin, T., and Cohn, T. (2020). "Tangled up in BLEU: Reevaluating the Evaluation of Automatic Machine Translation Evaluation Metrics". Paper presented at the Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online.

[40] Sakamoto, A. (2019). "Why do many

translators resist post-editing? A sociological analysis using Bourdieu's concepts". *The Journal of Specialised Translation*, 31, 201-216.

[41] Yamada, M. (2019). "The impact of Google Neural Machine Translation on Post-editing by student translators". *The journal of specialised translation*, 31 (Jostrans 31)(11), 87-106.

[42] Carl, M., and Baez, M. C. T. (2019). "Machine Translation Errors and the Translation Process: A Study across Different Languages". *The Journal of Specialised Translation*(31), 107-132.

[43] Onishi, N., and Yamada, M. (2020). "Why translator competence in information searching matters: An empirical investigation into differences in searching behavior between professionals and novice translators". *Invitation to Interpreting and Translation Studies*, 22, 1-23.

[44] Sanchez-Gijon, P., Moorkens, J., and Way, A. (2019). "Post-editing neural machine translation versus translation memory segments". *Machine Translation*, 33(1-2), 31-59.

第7回アジア翻訳ワークショップ (WAT2020) 開催報告

中澤 敏明

東京大学

1. はじめに

本稿では WAT2020[1]の開催報告を行う。アジア翻訳ワークショップ (Workshop on Asian Translation, WAT) はアジア言語を中心とした評価型機械翻訳ワークショップであり、2014 年に第1回 (WAT2014) を開催して以降、毎年開催している。2016 年の第3回 (WAT2016) 以降は自然言語処理の国際会議との併設ワークショップとして開催しており、2020 年の第7回 (WAT2020) は ACL-IJCNLP2020 の併設ワークショップとして、2020 年 12 月 4 日に開催された。今回はコロナ禍のためオンラインでの開催となった。招待講演は行わず、機械翻訳評価に参加した各チームによる翻訳システムの説明のほか、機械翻訳やアジア言語処理に関する一般研究論文発表が行われた。

最近のオンライン国際会議では時差を考慮して各発表 2 回ずつの QA セッションが割り当てられていることが多いが、ACL-IJCNLP2020 では全ての発表者は事前に発表内容を録画したビデオの提出に加えて、オンラインでのライブ発表及び QA を 1 回のみ行うことが要請されていた。そのため会議のプログラムは発表者の居住地のタイムゾーン (所属から推測した) を考慮して組む必要があった。ビデオはほぼ全件事前提出されたが、当日の発表時間に発表者がその場にはいないケースが半分程度あり、事前提出ビデオを流すことで対応せざるを得なかった。なお会議の参加者は入れ替わり立ち替わり、常時 20 名程度が入っている感じであった。

WAT で行われた翻訳タスクの変遷を図 1 に示す。最初の WAT では科学技術論文の日英・日中翻訳タスクのみであったが、年々言語対やドメインが増えてきたことがわかる。WAT2020 では新たに以下の翻訳タスクが追加された (図 1 の赤い矢印で示されたもの)。

クのみであったが、年々言語対やドメインが増えてきたことがわかる。WAT2020 では新たに以下の翻訳タスクが追加された (図 1 の赤い矢印で示されたもの)。

- ・7つのインド諸語 ⇄ 英語翻訳
- ・オリヤー語 ⇄ 英語翻訳
- ・ヒンディー/タイ/マレー/インドネシア語 ⇄ 英語 IT&Wikinews 翻訳
- ・日本語 ⇄ 英語 マルチモーダル翻訳
- ・日本語 ⇄ 英語 ビジネス対話翻訳

ドメイン	言語対	2014	2015	2016	2017	2018	2019	2020
科学技術論文	Ja ⇄ En/Zh							
特許	Ja ⇄ Zh							
	Ja ⇄ Ko							
	Ja ⇄ En							
	Id ⇄ En							
新聞	Ja ⇄ En							
	Ja ⇄ Ru							
	Hi ⇄ En							
	Hi ⇄ Ja							
混合	7 Indic ⇄ En							
	Ta ⇄ En							
	My ⇄ En							
	Km ⇄ En							
	Odia ⇄ En							
レシピ	Ja ⇄ En							
通時開示文書	Ja → En							
IT&Wikinews	Hi/Th/Ms/Id ⇄ En							
マルチモーダル	En → Hi							
ビジネス対話	Ja ⇄ En							

図 1: 翻訳タスクの変遷

また WAT2020 では新たな試みとして、科学技術論文翻訳及びビジネス対話翻訳において、文書単位での翻訳タスクも設定した。これは 1 文を超えた文脈情報も考慮することで、全体としてより適切な翻訳を出力することを目的としたものである。科学技術論文の文書単位データとして ParaNatCom (<https://www2.nict.go.jp/astrec-att/member/mutiyama/paranacom/>) を利用したが、開発データ及びテストデータのみの提供であったためか、残念ながら今回参加者はいなかった。ビジネス対話の文書単位データとしては BSD コー

パス (<https://github.com/tsuruoka-lab/BSD>) を利用し、4 チームの参加があった。

図 2 に各年の翻訳タスクの参加チーム数を示す。昨年は参加チーム数が非常に多かったが、今回は例年通りの水準にとどまった。海外で最も参加が多かった国はインドの 5 チームで、これはインド諸語を対象としたタスクが増えたためと思われる。他には中国、シンガポール、スイス、アイルランドからの参加があった。

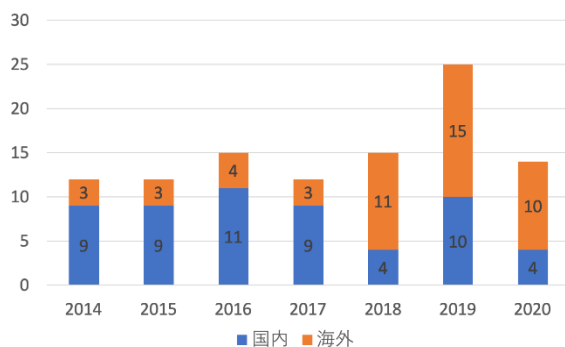


図 2：翻訳タスクの参加チーム数

また一般研究論文の募集も行い、12 件の投稿のうち 7 件を採択（うち 1 件は後日、著者により取り下げ）した。採択した論文のタイトルを以下に示す。

- An Effective Optimization Method for Neural Machine Translation: The Case of English-Persian Bilingually Low-Resource Scenario
- Transformer-based Double-token Bidirectional Autoregressive Decoding in Neural Machine Translation
- Translation of New Named Entities from English to Chinese
- A parallel evaluation data set of software documentation with document structure annotation
- Inference-only sub-character decomposition improves translation of unseen logographic characters
- An Error-based Investigation of Statistical and Neural Machine Translation Performance on Hindi-to-Tamil and English-to-Tamil

2. 翻訳評価

翻訳評価は自動評価と人手評価を行った。これまで人手評価は一対比較評価と正確性評価の 2 段階での評価を行なってきたが、今回は翻訳結果の提出が多いタスクがなかったため、日英マルチモーダル翻訳を除いて正確性評価のみを行なった。また英語からヒンディー語へのマルチモーダル翻訳タスクにおいては、WMT (Conference on Machine Translation) でも採用されている Direct Assessment (DA)[2] という評価指標を用いている。DA は評価者が各翻訳に対して 0 から 100 の間での翻訳の質を整数値で付与することで評価を行うものである。

2.1 自動評価

WAT では自動評価サーバーを用意し、参加チームが機械翻訳結果の自動評価結果をいつでも確認できるようにしている。自動評価システムはワークショップ終了後も利用可能であり、引き続き翻訳結果の提出・修正を受け付けている。

自動評価尺度としては BLEU[3]、RIBES[4] 及び AM-FM[5] を用いている。AM-FM は正確さと流暢さの両方を考慮したような評価手法である。日本語と中国語に関しては単語分割基準の違いによりスコアが変化するため、いくつかの単語分割ツールを使って翻訳結果を単語に分割し、それぞれの基準でのスコアを算出する（詳細は WAT のウェブサイト参照）。

2.2 人手評価（正確性評価）

人手評価は特許庁基準での専門家による正確性評価（JPO Adequacy Evaluation）により行なった。これは特許庁が公開している「特許文献機械翻訳の品質評価手順」の中の「内容の伝達レベルの評価」に従った評価であり、機械翻訳結果が原文の実質的な内容をどの程度正確に伝達しているかを、人手翻訳の内容に照らして、下記 5 段階の評価基準で主観的に評価するものである。

- 5: すべての重要情報が正確に伝達されている。
(100%)
- 4: ほとんどの重要情報は正確に伝達されている。
(80%~)
- 3: 半分以上の重要情報は正確に伝達されている。
(50%~)
- 2: いくつかの重要情報は正確に伝達されている。
(20%~)
- 1: 文意がわからない、もしくは正確に伝達されている重要情報はほとんどない。(~20%)

内容の伝達レベルの評価はテストデータからランダムに選択した 200 文を対象を絞って行った。また内容の伝達レベルの評価においては、2 名の評価者が独立に評価を行なっている。

次章からは翻訳タスクのうちのいくつか（日本語を含むもの）について、結果や知見などの概要を報告する。なお評価結果の詳細や各参加チームの翻訳システムの説明などは ACL Anthology (<https://www.aclweb.org/anthology/events/wat-2020/>) で確認することができる。また採択されたすべての研究論文も同ウェブサイトで開催されているので、参考にしていただきたい。

3. 評価結果と得られた知見

3.1 科学技術論文

今回は日⇄英翻訳には参加チームがなく、日⇄中翻訳に 1 チームのみの参加であった。また 1 章でも述べた通り、文書単位の翻訳タスクにも参加チームはなかった。図 3 および図 4 に評価結果を示す。自動評価結果を見ると毎年少しずつ向上しており、今回も過去最高値を達成しているのだが、人手評価結果を見ると残念ながら低下してしまっている。もちろん毎年評価者が異なるため直接比較することはできないかもしれないが、人手評価が低下している要因は別に存在する可能性がある。今回参加したチームは提供されている対訳コーパス以外の他の言語資源も利用したり、複

数の翻訳モデルを組み合わせたりするなどして翻訳精度の向上を図っている。これは推測だが、このように様々な言語資源やモデルを用いることの利点としては、翻訳モデルがより汎用的になり、生成される文の流暢性が向上する可能性があることが挙げられる。一方、欠点として、専門用語など定訳が存在するものであってもより一般的な語彙を用いて翻訳されてしまい、人手評価の観点で悪影響がある可能性があることが挙げられる。科学技術論文においてはこのような専門用語の割合が比較的多いため、全体的な流暢さの向上により自動評価が向上した一方で、専門用語の翻訳誤りにより人手評価が低下したのではないかと考えられる。

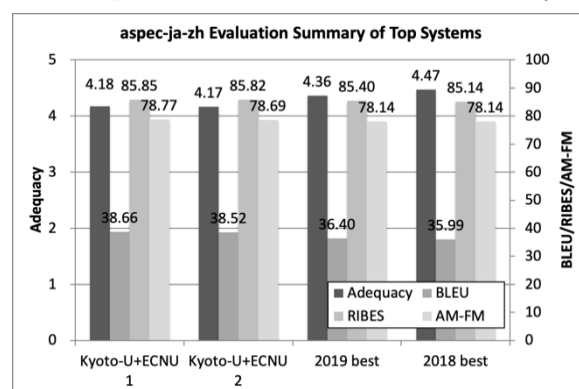


図 3：科学技術論文 日→中 評価結果

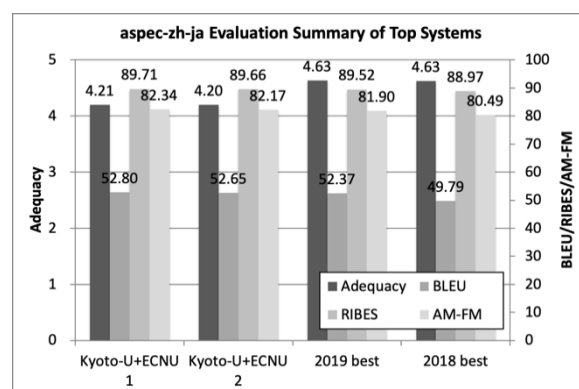


図 4：科学技術論文 中→日 評価結果

3.2 特許文

特許翻訳タスクには 2 チームの参加があった。全ての言語対の翻訳結果を提出したチームが 1 つ、もう一つは韓日翻訳サブタスクのみの参加だった。図 5 から図 10 に評価結果を示す。日中・日韓・韓日の 3 つの

サブタスクについては人手評価において昨年よりも良い結果となったが、それ以外のサブタスクでは昨年の方が良いという結果となった。今後の展開としては、訓練データに含まれない IPC 分類へのドメイン適応などが考えられる。

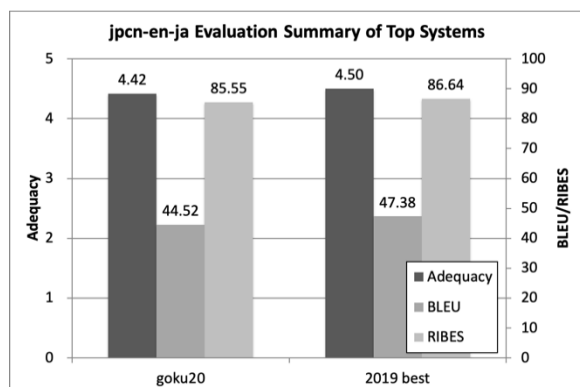


図 5：特許文 英→日 評価結果

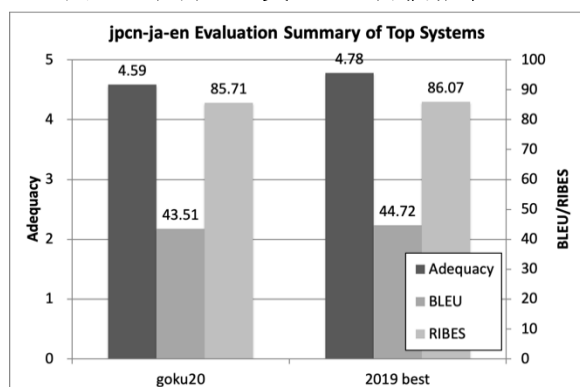


図 6：特許文 日→英 評価結果

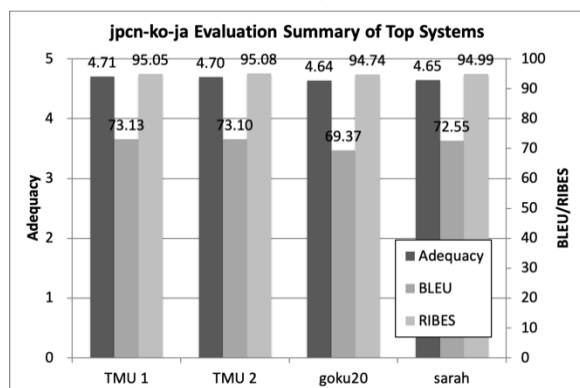


図 7：特許文 韓→日 評価結果

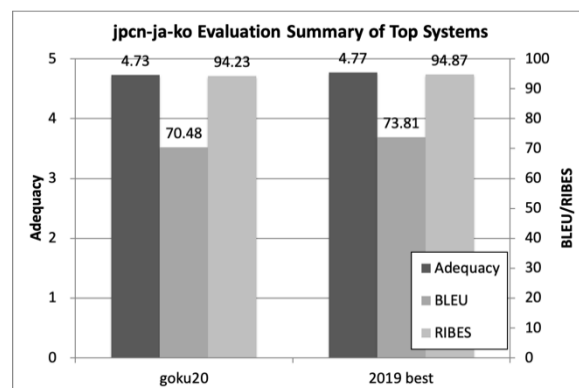


図 8：特許文 日→韓 評価結果

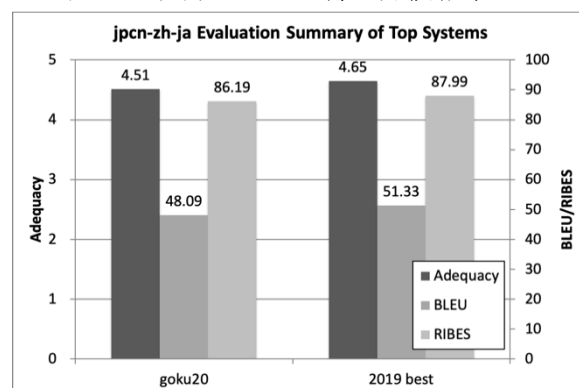


図 9：特許文 中→日 評価結果

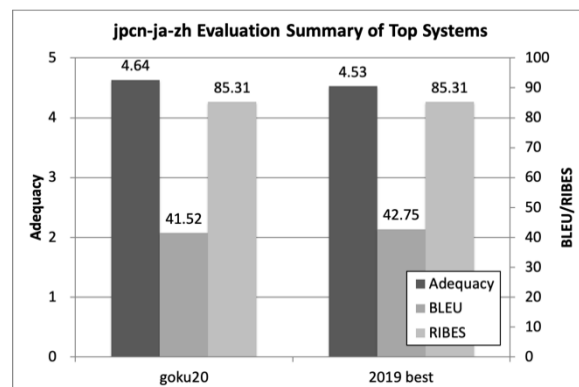


図 10：特許文 日→中 評価結果

3.3 ニュース記事

WAT2020 の時事ニュースの日英・英日翻訳タスクでは新たに文脈情報を用いたテストセットが追加された。参加者は翻訳対象文だけでなく、その文脈文の情報も利用することができる。ただし、訓練データはこれまで通り一文単位である。参加チームは残念ながら日英翻訳サブタスクに 1 チームだけであった。図 11 に評価結果を示す。人手評価の結果、文脈情報を用いるモデルの方が用いないモデルよりも若干精度が高い

ということがわかった。

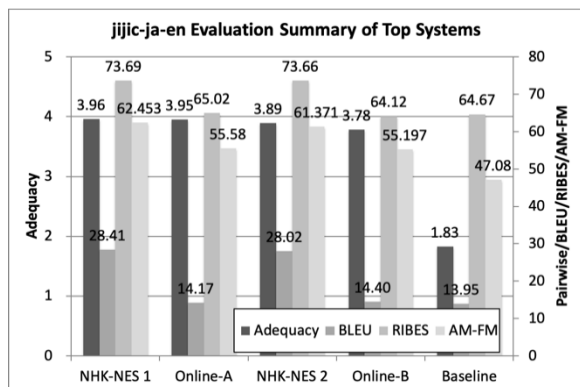


図 11：ニュース記事 日→英 評価結果

3.4 ビジネスシーン対話

本タスクには日英翻訳に 4 チーム、英日翻訳に 3 チームの参加があった。図 12 および図 13 に評価結果を示す。結果を見ると人手評価結果上位のシステムのスコアはとても近い値となっているが、システムの構成は大きく異なっている。文脈文として 3 文前までを考慮した mBART を採用したシステムや、大規模な外部リソースを利用し、BSD コーパスなどで転移学習したシステムなどがあった。

興味深い知見として、文脈を考慮したモデルの転移学習において、BSD データだけを用いた場合には翻訳精度の向上は見られなかったが、BSD コーパスに近いドメインの外部リソースも併せて転移学習に用いた場合には精度の向上が見られたとの報告がいくつかのチームからあった。つまり転移学習においてもある程度の量の対訳コーパスが必要となるということがこの結果からわかる。

また別のチームからは、サブワードの語彙サイズを小さく（報告では 6000 を用いていた）した方が翻訳精度が高くなるという報告もあった。これについてはその要因の分析まではされていないが、可能性としてはビジネスシーン対話において使われる語彙サイズがそれほど多くない可能性がある。

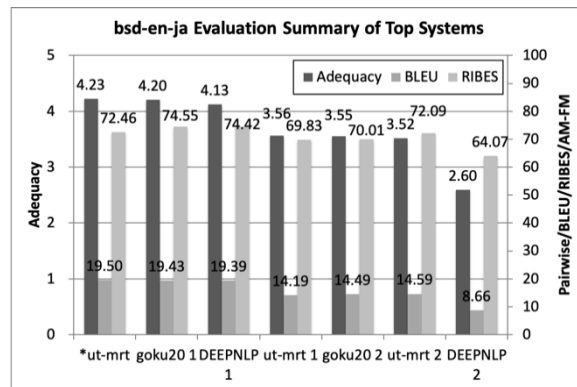


図 12：ビジネスシーン対話 英→日 評価結果

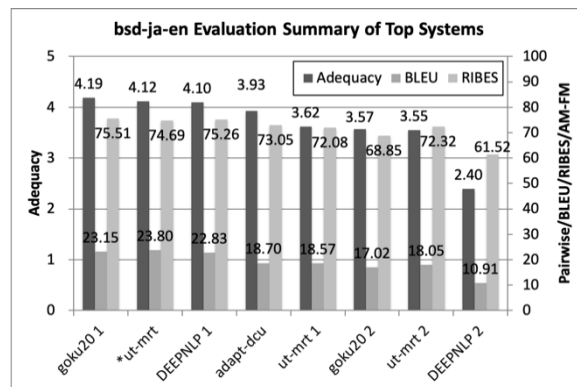


図 13：ビジネスシーン対話 日→英 評価結果

4. まとめ

本稿では WAT2020 の開催報告を行なった。WAT2020 は新型コロナウイルスの影響で、初めて完全オンライン開催となった。参加者の居住地のタイムゾーンを考慮しつつプログラムを組む必要があったり、発表者に事前に発表ビデオを提出してもらう必要があったりなど色々大変な開催であった。

WAT2021 は過去の開催時期よりも早く、2021 年 8 月上旬に開催される ACL-IJCNLP 2021 の併設ワークショップとしてタイのバンコクで開催予定である。

WAT2021 では以下の新たなタスクを実施する。

- * MultiIndicMT: Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Oriya, Punjabi, Tamil, Telugu ⇔ English Multilingual Task
- * Malayalam Visual Genome: English → Malayalam Multimodal Task
- * Ambiguous MS COCO: English ⇔ Japanese

* ArEnMulti30K: English ⇔ Arabic Multimodal Task

* Restricted Translation Task

なお翻訳評価にかかる費用等のためのスポンサーを募集しているので、興味のある方はご連絡いただければ幸いです。

参考文献

[1] Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Shohei Higashiyama, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar, Sadao Kurohashi. 2020. Overview of the 7th Workshop on Asian Translation. In Proceedings of the 7th Workshop on Asian Translation (WAT2020), pages 1-44.

[2] Yvette Graham, Timothy Baldwin, Meghan Dowling, Maria Eskevich, Lamia Tounsi, Teresa Lynn. 2016. Is all that glitters in MT quality estimation really gold standard? In Proceedings of the 26th International Conference on Computational Linguistics, Osaka, Japan.

[3] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of ACL, pages 311–318.

[4] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. 2010. Automatic evaluation of translation quality for distant language pairs. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 944–952.

[5] Rafael E. Banchs, Luis F. D’Haro, and Haizhou Li. 2015. Adequacy-fluency metrics: Evaluating mt in the continuous space model framework. IEEE/ACM Trans. Audio, Speech and Lang. Proc.,

第6回特許情報シンポジウム

須藤 克仁

奈良先端科学技術大学院大学 (AAMT/Japio 特許翻訳研究会 副委員長)

1. イベント概要

AAMT/Japio 特許翻訳研究会では、2010 年度より隔年で「特許情報シンポジウム」を開催しており、2020 年度は 2021 年 2 月 26 日（金）午後にオンラインで開催した。一般論文発表については論文募集・選考の期間の確保が難しかったため招待講演およびパネル討論による構成とした。オンライン開催でありながら最大 100 名程度の参加者を集めることができ、従来のオンライン開催と同程度の規模が維持できた。

2. イベント内容

今回のシンポジウムでは以下の 6 件の招待講演とパネル討論が行われた。内容を簡単に紹介する。

（1）「特許庁における機械翻訳の活用」 中西 聡 氏（特許庁）

近年の日本国内からの海外出願率の増加や全世界的な特許・実用新案の出願の急増を背景として、特許の審査情報を各国で活用するための機械翻訳技術はさらにその重要性を増している。日本国特許庁においても日本語と外国語双方向の機械翻訳を活用しており、現在はパブリッククラウドを用いた機械翻訳プラットフォームとして様々な翻訳エンジンを活用するに至っている。また機械翻訳の持続的改善に向けた各種調査活動も継続的に実施している。

（2）「令和 2 年度特許出願技術動向調査-機械翻訳-」 成瀬 博之 氏（特許庁）

特許庁では毎年いくつかの技術分野を対象とした出願動向調査を行い報告しているが、令和 2 年度の調査

対象の一つとして機械翻訳技術が選ばれた。本講演ではその途中経過が報告された。詳細は令和 3 年 4 月頃に報告書が公開される予定であり、本報告では詳細を述べないが、米国が市場・技術を主導する中で、中国籍の出願が急増しており注目に値する。また、日本で特徴的な技術領域や、総務省のグローバルコミュニケーション計画 2025 のような戦略プロジェクトが存在することを踏まえ、今後の機械翻訳関連技術の研究開発への提言が行われた。

（3）「機械翻訳コンペティション参加報告」 清野 舜氏（理化学研究所）

機械翻訳研究では長年に渡り様々なコンペティション（共通のデータを利用して性能を競う）が行われており、近年最も注目を集めている国際会議 WMT の 2020 年新聞記事翻訳タスクにおいて、東北大・理研・NTT の合同チームがトップの成績を収めた。技術としてはニューラル機械翻訳（NMT）に基づくものであるが、教科書的な単純な構成による翻訳システムではなく、様々な工夫を凝らした「泥臭い」技術の集合体とも言える大規模かつ複雑なものとなった。それ故に必要な計算資源も膨大となっているが、一方で機械翻訳に限らず様々な自然言語処理が同様のアプローチを用いて実現されるようになり、技術の垣根が下がった新たなチャンスが生まれてきているとも言える。

（4）「特許文書の品質評価の標準化について ～特許文書品質特性 モデル～」 谷川 英和 氏（IRD 国際特許事務所）

産業日本語研究会 特許文書分科会では特許文書を対象とした文書品質モデルの検討や評価を行っている。特許文書は発明を解説する技術文書という側面と知財

Event report: The 6th Symposium on Patent Information
Katsuhito Sudoh

Nara Institute of Science and Technology (Vice-chair of AAMT/Japio Special Interest Group on Patent Translation)

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

権を確保するための権利文書という側面があり、一口に品質といっても多角的に捉える必要がある。着目すべき品質特性は権利化に向けたフェーズ毎、また評価する立場や対象によって異なるため、的確な特許文書の作成・評価のためには品質特性を把握し活用できるようにならなければならない。特許文書分科会ではそのための助けとなる教育用テキストを作成しており、弁理士をはじめとした知財関連人材の研修目的等での展開を予定している。

（５）「COVID-19 に絡む特許情報の分析 ～ COVID-19(敵)を知り、己を守る～」 桐山 勉 氏（日本特許情報機構）

COVID-19 に関係する特許の分析を通じた提言。社会的影響が甚大であったため COVID-19 に関連する研究開発は世界各国で驚異的なスピードで進んでおり、関連する早期出願が急増する等熾烈なワクチン開発競争が行われていることが窺える。COVID-19 に関しては膨大な情報が存在するが、そうした中から重要な情報を特許情報で確認・検証することが重要である。その際に機械翻訳は有用なツールであり、世界各国のオリジナル特許の早期のアクセスを可能にする。

（６）「自動翻訳技術の概要：なにができるか／できるようになってきているか」 内山 将夫 氏（情報通信研究機構）

情報通信研究機構（NICT）では「みんなの自動翻訳 @TexTra®」という機械翻訳サイトを運営しており、現在は第２世代の NMT と言える Transformer が主軸を担っている。実用に資するためには対象に合わせたチューニング（アダプテーション）を重視し、また実務では同様の文を繰り返し翻訳する事例が多いことから用例ベース翻訳（EBMT）も併せて活用している。対訳データの収集にあたっては、「翻訳バンク」と呼ばれる対訳データの提供を機械翻訳技術のライセンス料負担軽減に繋げる仕組みによってデータの蓄積が進んでいる。こうした大規模データに加え分野アダプテーションや用語のカスタマイズ等を活用することで機械翻訳精度は大きく向上した。一方で依然として正確性

が保証されるとはなっておらず、利用には引き続き注意を要する。

（７）パネル討論

シンポジウムの最後に招待講演者をパネリストとしたパネル討論を行った。特許情報処理の課題や今後について議論すべく、次の３つのトピックを事前にパネリストに提示し、議論を行った。①特許のような専門的な業務において人と機械の得手不得手は何で、何を人が担い、何を機械が担うべきと考えられるか？②西欧言語・中韓以外の言語、特にアジア諸語をどう扱っていくか。仮に主要言語ほど精度が出なくても用途はあるか？また非主要言語におけるチャレンジとは何か？③特許という大規模な公開技術文書資源は広く諸用途に活用できないか？

トピックを機械翻訳に寄せたこともあり特許実務の先生方のご意見を十分くみ取れなかった点は反省点と言えるが、一口に特許と言っても様々な側面があり、一概に特許だからこうであるというほど単純に扱える問題ではないということも一つの結論である。ただ、近年の機械学習技術の進展は言語の違いや分野の違い、さらにはメディアの違いすらも超えて様々なタスクに「適応」が可能な汎用性を持つと謳うモデルを生むに至っており、特に資源の多い領域の知見を十分に集めることの重要性は変わらないと言えよう。

3. 所感

オンライン開催のシンポジウムであったが、様々な観点からの講演で充実したものであったと言って良いだろう。ただ、一つ開催側として心残りだった点は機械翻訳以外に特許実務と先端研究のギャップを埋めるような議論が今回提供できなかったことである。自然言語処理業界はこの３・４年で急速に変化を遂げており、国内外問わず技術開発競争が非常に激しくなっている。そうした観点からは特許実務における自然言語処理をはじめとした最新技術の適用可能性についてより踏み込んだ議論ができる場にしたい。

編集後記

早川 威士

AAMT 編集委員会

AAMT はこの 2021 年、創立 30 周年を迎えました。編集委員一同ここに祝意を表するとともに、会員の皆様をはじめ様々な形で AAMT の活動や機械翻訳技術の振興に貢献してくださった全ての方々にお礼を申し上げます。

巻頭言では、AAMT 隅田英一郎会長より「時は来たり」と題し、機械翻訳を通じて新しい翻訳の未来を創り出すという当会の決意をあらためて表明しています。機械翻訳の発展はこれまで様々な苦難を伴ってきましたが、深層学習の導入など技術的進歩によって大きく前途が開けました。しかし機械翻訳が真に普及する未来のための最後のひとつのピースは、それをいかに使いこなすかということであり、使う側である人間との協業（ポスト・エディット）が欠かせません。本年 AAMT は機械翻訳+ポスト・エディットにフォーカスを当てて活動を進める予定です。この取り組みが皆様の協力を得て、大きく展開していくことを期待しています。

本号では、解説記事を 3 報掲載しています。影浦峽様には、「人間の翻訳と機械の翻訳」シリーズの第 4 回として、(人文科学での) 翻訳論と機械翻訳研究で取り扱われている「翻訳」の違いについて教科書のコーパス分析をもとに考察いただいています。美野秀弥様には COLING2020 で発表された論文 (Mino et al., 2020) をベースに、外部情報として目的言語側の文脈情報を用いる効果について寄稿いただきました。宮田玲様、藤田篤様からは、EACL 2021 に採択された論文 (Miyata & Fujita, 2021) より、入力文を事前に編集するプリエディットについて、書き換えの傾向と出力への影響を解説いただいています。

続いて書評として、2 冊の書籍を紹介しています。中

澤敏明様には、『自動翻訳大全：終わらない英語の仕事が 5 分で片づく超英語術』（三オブックス）、山田優様には『機械翻訳：歴史・技術・産業』（森北出版）をレビューいただきました。いずれも機械翻訳を主題として取り上げた著作であり、自然言語処理や人工知能の中のひとつの分野ではなく独立したテーマとして取り上げられるようになってきていることは、機械翻訳に対する社会の関心の高まりを示していると言えるのではないのでしょうか。

イベント報告では、昨年 12 月に行われた AAMT2020, Online について、出内将夫様にレポートをご執筆いただきました。長尾賞の受賞者講演に加え、招待講演とパネルディスカッションが行われました。今回はコロナ禍の影響で初のオンライン開催となったにもかかわらず、運営委員会の尽力により大きなトラブルはなく、また議論も活発に行われ盛況のうちにイベントを終えることができました。ご参加くださった皆様には改めてお礼申し上げます。

また、中澤様には同じく昨年 12 月に開催されたアジア翻訳ワークショップ (WAT2020) についてご報告いただいています。新しく追加された言語対や翻訳ドメインのタスクを含め、興味深い評価結果が示されています。さらに須藤克仁様には、本年 2 月に AAMT/Japio 特許翻訳研究会が開催した特許情報シンポジウムより、特許分野に機械翻訳を導入する数々の取り組みについてレポートをお寄せいただきました。

この編集後記をしたためている 2021 年 4 月、新型コロナウイルス感染症は過去にない猛威を振るっています。最後になりましたが、1 日も早い終息と皆様のご無事ご健康を祈念し、編集後記の締めくくりとさせていただきます。

Editor's note

Takeshi Hayakawa

AAMT Editorial Board

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License. License details: <https://creativecommons.org/licenses/by-sa/4.0/>

AAMTジャーナル「機械翻訳」No. 74

- 【 発 行 日 】 2021年6月15日
【 発 行 】 アジア太平洋機械翻訳協会 (AAMT)
ホームページ: <https://aamt.info/>
【 住 所 】 〒619-0289
京都府相楽郡精華町光台3-5
国立研究開発法人情報通信研究機構 先進的翻訳技術研究室内
【 編 集 委 員 会 】 内山将夫 後藤功雄 中澤敏明 新田順也 園尾聡 森口功造 隅田英一郎
仲山裕子 小谷克則 山田優 石川弘美 早川威士
【表紙デザイン】 泉谷東十郎
【 題 字 】 長尾真
【 事 務 局 】 石川弘美
【 印 刷 所 】 株式会社 プリントパック

Asia-Pacific Association for Machine Translation (AAMT)
c/o National Institute of Information and Communications Technology
3-5 Hikaridai Seika-cho Soraku-gun Kyoto 619-0289, Japan



AAMT

Asia-Pacific Association for Machine Translation

アジア太平洋機械翻訳協会