

AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2018 **No.68**

アジア太平洋機械翻訳協会

目 次

巻頭言：AAMT 会長としての 6 年間の活動を振り返って.....	中岩 浩巳..... 1
エッセー：科学技術情報の機械翻訳.....	江原 暉将..... 3
AAMT 交流会報告：アジア太平洋機械翻訳協会交流会.....	葉 文君..... 6
シンポジウム参加報告：国際会議 AMTA2018 参加報告.....	中岩 浩巳..... 22
セミナー参加報告：2017 年度 JTF 第 6 回翻訳セミナー報告.....	目次 由美子..... 24
レポート：日本語オノマトペ共起表現レキシコン.....	首藤 公昭..... 27
研究報告：翻訳ソフトとマクロによる多言語文書同時作成システムの試作.....	保立 浩一..... 33
研究報告：手軽な英日機械翻訳.....	吉澤 忠義..... 39
会員の広場（法人会員）.....	インフォトランス株式会社..... 43
会員の広場（個人会員）.....	大塚 英..... 45
事務局からのお知らせ：協会活動報告（2018 年 3 月～2018 年 4 月）.....	AAMT 事務局..... 47
編集後記.....	宇津呂 武仁..... 50

CONTENT

Foreword:The past six years' activities and the future of AAMT.....	<i>H.Nakaiwa</i> 1
Essay : Machine Translation for Scientific Documents.....	<i>T.Ehara</i> 3
Symposium Report: AAMT Event Report.....	<i>W.J. Ye</i> 6
Symposium Report : AMTA 2018.....	<i>H.Nakaiwa</i> 22
Research Meeting Report : Report on Sixth JTF Translation Seminar 2017.....	<i>Y.Metsugi</i> 24
Report : A Lexicon of Japanese Onomatopoeic Collocations·JMWel_onomatopoeic.....	<i>K.Shudo</i> 27
Research Report:Prototyping of a multi-language documentat drafting system with translation software and a macroprogram.....	<i>K.Hotate</i> 33
Research Report:Easy English-Japanese machine translation.....	<i>T.Yoshizawa</i> 39
AAMT Members:.....	<i>InfoTrans Co., Ltd.</i> 43
AAMT Members:.....	<i>H.Otsuka</i> 45
AAMT Activities:AAMT Activities (from March 2018 to April 2018)..... 47
Editor's Note:.....	<i>T.Utsuro</i> 50

AAMT 会長としての 6 年間の活動を振り返って

中岩 浩巳

名古屋大学大学院情報学研究科 / アジア太平洋機械翻訳協会 (AAMT) 会長

1. 6 年間の主な活動

私は 2012 年 AAMT 総会で同協会会長に就任し 3 期 6 年続けてきたが、2018 年 6 月の AAMT 総会をもって、会長の任期を満了することとなった。本稿では、今まで 6 年間の AAMT 活動を振り返るとともに、機械翻訳に期待することを述べたい。

6 年前の機械翻訳を取り巻く状況を振り返ると、以前主流であったルールベース翻訳技術から統計翻訳技術に移行が始まり、英日翻訳において性能向上に貢献した語順並び替え技術が生まれるなど、日本語の関連する翻訳技術の本格的活用が始まった時期であるといえる。当時は統計翻訳の活用を前提とした、様々な技術が提案されていた。その性能向上は限定的ではあったが、特許を代表とする大量の対訳コーパスが存在する分野では統計翻訳がルールベース翻訳の翻訳性能を凌駕する性能になった。この状況変化を踏まえ、機械翻訳技術を利用する側の代表ともいえる翻訳業界では、当時から機械翻訳技術への注目が高まっていた。この流れは 2016 年 11 月の Google 翻訳へのニューラル機械翻訳 (NMT) 技術の導入に伴う翻訳性能の大幅向上により、決定的なものとなった。

このような状況変化を踏まえ、AAMT としては、会員や会員候補である翻訳業界を中心とした積極的な情報発信を進めた。具体的には、AAMT 総会に併設する機械翻訳フェアを開始し、最新の機械翻訳の動向や関連する招待講演、また、会員企業及び個人会員による展示する場を提供した。さらに、会員交流会を開始し、会員を中心とした最新動向を情報提供する場も新たに設けた。また、翻訳業界に翻訳技術を紹介・啓蒙する

ことを目的として、翻訳業界最大の団体である日本翻訳連盟 JTF との連携を強化し、JTF 主催の翻訳祭での AAMT 企画講演や、展示、また JTF 総会での基調講演や JTF セミナーでのパネルなど、翻訳業界を意識した機械翻訳に関する情報発信を積極的に行った。その結果、例えば、2016 年及び 2017 年の JTF 翻訳祭の AAMT 企画講演では、会場に入りきれないくらいの集客があった。

また、2017 年 9 月には国際機械翻訳連盟 IAMT が主催で AAMT がホストした第 16 回国際会議機械翻訳サミット (MT Summit XVI) が 24 年ぶりに日本で開催された。MT サミットは 2 年に一回、AAMT、欧州機械翻訳協会 (EAMT)、米州機械翻訳協会 (AMTA) が順番にホストしその地域で開催する会議であるため、AAMT は 6 年に一回ホストを担当することになる。今回の会議では、1 日目及び 5 日目には、ワークショップ 3 件 (うち 1 件は AAMT-Japio 特許翻訳研究会主催、1 件は JTF との共催) とチュートリアル 3 件が、2 日目からは 4 日目は本会議が (研究発表 48 件、技術展示 10 件、企業展示 10 件、招待講演 4 件、招待パネルディスカッション 1 件)、があり、機械翻訳の最先端技術からその活用方法まで幅広くカバーしたプログラムで編成された。本会議には、総計 300 名 (国内 235 名、外国 20 か国 65 名) が参加するとともに、①7 社のスポンサーと、経済産業省、総務省を含む、5 団体の後援をいただいた。第 15 回大会 (米マイアミ)、第 14 回大会 (仏ニース) とも 200 名程度の参加者数であったことを考えると、今回の MT サミットは、機械翻訳への国内の注目度を反映したものと考えられる。

2. 今後の展望と AAMT への期待

も AAMT へのご支援をお願いしたい。

前述のとおり、NMT 技術は、様々な新たな研究成果が生みだされ続けており、翻訳性能は短時間で急速な性能向上を続けている。また、NMT 技術は実サービスに導入され始めており、NMT 特有の訳出傾向を反映した活用方法に関する検討も始まっている。今後数年間はこの傾向が続くものと予想される。NMT は、統計翻訳時代のようにテキストだけではなく、音や画像、レイアウト情報等、他のメディアの情報もうまく統合的に取り扱える技術であるので、従来の機械翻訳では活用できていなかった情報を活用した翻訳性能の向上（例えば、イントネーション等の音情報と画像情報を融合活用した通訳）に期待したい。また、利用の場面では、今まで指摘されている NMT の様々な課題（訳抜け、重複訳出、訳語の不統一等）を克服する試みが始まっている。統計翻訳の時代にも技術の課題を克服する活用方法が考えられ、システム化が進んだが、NMT でも同様に、NMT の訳出傾向にあった活用方法や翻訳支援環境が定着していくのではないかと。

上記のように、機械翻訳技術は今後さらに広く活用されて行くとともに、活用方法も変化していくことが予想されるので、AAMT としては、最新の技術や活用方法をタイムリーに様々な形態で情報提供し、会員の皆様がメリットを感じるように、活動を強化していくべきである。また、長年の懸案事項である、日本だけではなく「アジア・太平洋」地域の協会としての活動も積極亭に進めてほしい。例えば、AAMT 主催のシンポジウムのアジア太平洋地域での開催や、既存の機械翻訳関連国際会議との共催など、具体的な実現に向けて検討を進めてほしい。

2020 年の東京オリンピック・パラリンピックを契機としたインバウンド需要増加の流れは、2020 年以降も継続することが予想される。その意味で、AAMT の存在意義は、これからも今まで以上に大きくなるのではないかと。会員の皆様の期待に応えられるよう、AAMT は事業の見直しと強化を進める計画であるので、今後

科学技術情報の機械翻訳

江原暉将

江原自然言語処理研究室

1. 機械翻訳の目的

機械翻訳の目的は人間が行う翻訳を支援することにある。人間が行う翻訳の目的は異なる言語話者間のコミュニケーションの支援にある。つまり言語の機能にコミュニケーションがあり、言語が異なることによってコミュニケーションが阻害される。それを取り除くのが翻訳や機械翻訳の目的である。

コミュニケーションを行うだけが言語の機能であるならば、人間社会は単一の言語を用いる方向に向かうはずである。単一の言語を用いた方がコミュニケーションには便利だからである。しかし世界には多様な言語がある。その理由を、筆者は、言語にディスコミュニケーションの機能があるためと考える。情報交換したい相手には理解でき、したくない相手には理解できないような言語を用いたい、という願望である。ディスコミュニケーションの機能は、隠語や業界用語の利用などに見られる。一部の若者言葉の利用もそうかもしれない。情報交換したくない相手とは、つきつめると「敵」である。すると翻訳とは敵とのコミュニケーションを支援することになる。敵との決定的な対立である戦争を避けるために対話(コミュニケーション)を行う。その支援のために翻訳や機械翻訳があると考えられる。

長尾真元会長は文献[1]の中で

「機械翻訳は研究として面白く、人間頭脳の働きをモデル理論的立場から解明することができるし、実用システムとしてビジネスにつながって行くだろうといったことを越えて、世界中の人達が自由にコミュニケーションすることによって世界に平和をもたらす最強の道具であるという使命感をもって取り組んで来た。

これから 15~20 年もすればかなり良いシステム

が出来ることは間違いないので、これを世界中が使うことによって少しでも平和な世界が実現されることを望んでいる。」

と述べている。田中穂積元会長は文献[2]の中で

「情報ネットワークに機械翻訳システムを重ね合わせると、世界の国々の人々との相互理解を深めることが可能になる。それにより機械翻訳システムが間接的に世界平和の貢献に役立つことになる。こうした夢を早く実現したいものである。」

と述べている。筆者自身、文献[3]の中で

「私がなぜ機械翻訳の研究をしているのかをつきつめて考えると、自分の誕生に遡る。私は 1944 年(昭和 19 年)11 月に現在の台湾・台北市で生まれた。私の生まれた日の新聞には「陸海の神鷲相次ぐ必殺の體當り一相呼應してレイテ灣の敵艦船を猛攻」とある。私は戦争の恐怖におびえながら、最幼年期を過ごしたことになる。以来、いかにして争いのない世界を作り出していくかを求めて、今に至っている。機械翻訳の研究もその一環である。」

と述べている。

2. 科学技術情報の機械翻訳

科学技術情報の機械翻訳の目的は、やはり言語の違いを越えて情報を交換することにある。科学技術情報の典型例の一つとして特許情報がある。日本の特許法の第一条には「この法律は、発明の保護及び利用を図ることにより、発明を奨励し、もつて産業の発達に寄与することを目的とする。」と書かれている。「保護」と「利用」は、ある意味矛盾している。第 1 章に示したディスコミュニケーションは保護に、コミュニケー

ションは利用に対応する。科学技術情報のもう一つの典型例の「論文」の場合も、著者のオリジナリティの保護と他の研究者による当該論文内容の利用という二つの側面がある。

機械翻訳は、言語の違いを超えて科学技術情報の「利用」を促進するためにある。一方で機械翻訳など用いず、科学技術分野での事実上の世界共通言語である英語で情報交換すれば良い、という意見もある。しかしこれは科学技術情報の利用のみに焦点をおいたものであり、狭い考え方である。言語、文化、思想は相互に密接に関係しており、科学技術分野でも英語を母語としない人が自分の母語で科学技術情報を発信するのは新規性あるいはオリジナリティの保護の観点から重要であろう。「保護」というより「確保」という言い方が良いかもしれないが、母語での発信は、文化や思想に直接関係する人文科学、社会科学だけでなく自然科学の分野でも大切であろう。さまざまな母語で発信された科学技術情報の利用促進に機械翻訳が果たす役割がある。

周知のように技術はもろ刃の剣であり、使い方を誤ると悪い結果を招く。機械翻訳技術も例外ではない。我々は機械翻訳技術が良い方向に使われているか常に自問して進まなければならない。ただ、何が良くて、何が悪いかは難しい問題である。歴史に学ぶことも、ある程度は可能だが万全ではない。特に現代は歴史上はじめて経験する事柄が多く、結局のところ自ら良く考えて善悪を判断するよりしかたがないと思う。

3. 機械翻訳の進歩

機械翻訳は、規則方式(RBMT)、用例方式(EBMT)、統計方式(SMT)、ニューラル方式(NMT)と進歩してきた。Workshop of Asian Translation (WAT) [4]の中日特許翻訳タスクに参加した筆者のシステムでも表1に示すように自動評価値(BLEU と RIBES)および人手評価値 (pairwise evaluation (HUMAN) と JPO adequacy [5])ともに値が向上してきている。

表1 機械翻訳方式の進歩と評価値の変化¹

	BLEU	RIBES	HUMAN	JPO adequ.
RBMT	15.77	0.7217	----	----
SMT	40.95	0.8280	39.00	----
NMT	46.52	0.8596	69.75	4.31

近年のニューラル翻訳の結果を見ると流暢性が高く良い翻訳に見える。しかし原文と詳しく比較して見ると誤っている場合も少なくない。「なぜこんな単純な誤りをするのか」というようなケースもある。では、機械翻訳の今後の進歩のためには何が必要なのだろうか。

古くから指摘され[6]、最近も指摘されている[7]ように、意味や文脈(コンテキスト)の利用ということになるだろう。機械翻訳は単語情報、構文情報、格情報、と利用する情報の範囲を徐々に広げてきて今日に至っている。意味情報についても、例えば文献[8]で利用されている。文献[8]では「意味類型」と「類推思考」という概念を用いて「言語の意味的等価変換方式」による機械翻訳を提案している。近年のニューラル翻訳では、単語や文を実数値ベクトルで表わしており、これは分布としての意味情報[9]を部分的に利用しているといえる。

文脈情報(コンテキスト情報)についてはどうだろうか。コンテキストは「言語的コンテキスト」と「非言語的コンテキスト」に分けられる[10]。現在機械翻訳で利用している範囲は言語的コンテキストの一部である。ニューラル翻訳にしても、基本は1文単位の翻訳であり、言語的コンテキストは大量の訓練データから推定されたモデルの中に埋め込まれている。今後は、1文単位の処理を越えて、特許文献や論文を文書単位(ドキュメント単位)で扱うなど言語的コンテキストを直接利用することが必要であろう。さらに翻訳対象文書を越えて、当該文書中の参考文献や関連文書にまで処理対象を広げることが考えられる。これは人間の翻訳者であれば、日常に行っていることであろう。

現状では、機械翻訳が非言語的コンテキストを利用するという事はほとんど行われていない。しかし、

¹<http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html>

例えば文献[11]の次の文

「帯状土塊を破砕するとともに、培土作業を停止すると抱き込まれていた土塊が自動的に落下する構造の培土器を提供する」

において「帯状土塊を破砕するとともに」が並列する相手は「培土作業を停止する」のか「土塊が自動的に落下する」のか、あるいは他の動詞句なのかを判定するには、「農作業の中での培土作業」という非言語的コンテキストを熟知していることが必須のように思う。

非言語的コンテキストを利用するには、画像情報や音響情報など、言語情報を越えてさまざまな情報を言語処理の中に取り込む必要がある。

さらに、もう一つ重要になると考える情報源が脳科学の知見であろう。工学の基礎に物理学があるように情報工学の基礎には脳科学があると考えられる。言語処理の目から見ると、まだまだ初歩的な脳科学の知見ではあるが、今後、それらの知見に沿って言語処理の処理系を設計する必要性が出てくるだろう。「人工知能」の研究も重要であるが「自然知能」の研究も重要であると考えられる。

4. まとめ

翻訳や機械翻訳の目的、特に科学技術情報の機械翻訳の目的について考察し、今後の機械翻訳の発展方向を展望した。

まとめると、機械翻訳を含む言語処理技術は、関連する諸分野との連携を一層深めてゆく、また深めなければならないと考える。そのために AAMT はじめ学協会の果たす役割も大きいものがある。

参考文献

- [1] 長尾真: AAMT 会員のひろば, AAMT ジャーナル, No.41, Dec., 2007.
- [2] 田中穂積: 機械翻訳の夢, AAMT ジャーナル, No.13, page 1, Jan., 1996.

[3] 江原暉将: AAMT 会員のひろば(第 3 回), AAMT ジャーナル, No.43, page46, Nov., 2008.

[4] Toshiaki Nakazawa, Shohei Higashiyama, Chenchen Ding, Hideya Mino, Isao Goto, HidetoKazawa, Yusuke Oda, Graham Neubig and SadaoKurohashi: Overview of the 4th Workshop on Asian Translation, Proceedings of the 4th Workshop on Asian Translation (WAT2017), pp.1-54, Nov., 2017.

[5] 特許庁: 特許文献機械翻訳の品質評価手順について, June, 2014.

http://www.jpo.go.jp/shiryoutoushin/chousa/tokkyohonyaku_hyouka.htm

[6] 田中穂積: 機械翻訳システムの今後, AAMT ジャーナル, No.16, pp.2-3, Sept., 1996.

[7] 長尾真×賀沢秀人: ニューラル翻訳を超える未来へ, JTF JOURNAL, No.294, pp.8-19, March-April, 2018.

[8] 池原悟: 究極の翻訳方式の実現に向けて—類推思考の原理に基づく翻訳方式—, AAMT ジャーナル, No.32, pp.1-7, March, 2002.

[9] 池上嘉彦: 意味論, pp.55-64, 大修館書店, May, 1975.

[10] 亀井孝, 河野六郎, 千野栄一 (編著): 言語学大辞典【術語編】, pp.595-596, 三省堂, Jan., 1996.

[11] 安久津昌義: カルチベータ用培土器, 公開特許公報, 特開 2003-000001, Jan., 2003.

アジア太平洋機械翻訳協会

交流会

葉 文君

株式会社クロスランゲージ

この度、アジア太平洋機械翻訳協会 (AAMT : Asia-Pacific Association for Machine Translation) によって、機械翻訳の最新動向の把握や開発者とユーザの交流を目的とした「AAMT 交流会」が開催された。

【日時】 2018 年 2 月 15 日 (木) 13:30-19:30

13:30—17:00 第一部 講演会

17:30—19:30 第二部 懇親会

【会場】 IT ソフトウェア健保会館・会議室 (大久保)

【プログラム概要】

1. 開会挨拶
中岩浩巳 (AAMT 会長/名古屋大学)
2. EAMT2017, MT Summit XVI 報告
中岩浩巳 (AAMT 会長/名古屋大学)
3. AAMT 課題調査委員会の活動
長瀬友樹 (課題調査委員会/富士通研究所)
4. 自動翻訳技術の最新動向
内山将夫 (情報通信研究機構)
5. パネル「翻訳業界での自動翻訳の活用」
モデレータ :

隅田 英一郎 (情報通信研究機構)

登壇者 :

- ・中山 雄貴 (ヒューマンサイエンス) :
エンジニアの立場から
- ・三笠 綱郎 (十印) :
翻訳者・品質管理者の立場から

・森口 功造 (川村インターナショナル) :

経営者の立場から

6. 質疑応答

1. 開会のご挨拶

中岩浩巳

AAMT 会長、名古屋大学

アジア太平洋機械翻訳協会 AAMT の交流会は 2017 年から始まった。最初の目的は、AAMT の会員の方々への情報提供や意見交換する場を提供することと、AAMT が調査・検討した内容、及び、機械翻訳に関連する動向を紹介することだった。今年に関しては、1 年位前に、ニューラル機械翻訳が Google 翻訳に導入され、広く使われるようになったことから、AAMT 会員外の方々の機械翻訳に対する期待が極めて高まっているという状況があり、AAMT 会員以外の方々にも声掛けをして、今回 AAMT 交流会を開催することになった。

今回のプログラムの構成は、AAMT が企画、もしくは、関係する組織が運営した国際会議における研究動向、特に利用者側にとって重要な翻訳品質等に関する発表の紹介や、AAMT の課題調査委員会の活動報告、さらには、昨今の自動翻訳の技術の状況もしくは技術の概要といった内容になっている。

パネルディスカッションでは「翻訳業界での自動翻訳の活用」をテーマに、NICT の隅田氏にモデレータをして頂き、ヒューマンサイエンスの中山氏、十印の三笠氏、川村インターナショナルの森口氏にそれぞれご講演いただくことになった。

最後に、「この交流会は AAMT の活動を理解して頂く機会としてこの場を提供させて頂いている。これからもぜひ AAMT の活動へのご支援・ご協力をお願いすると同時に、非会員の皆様にご入会していただけると幸いです」とのメッセージで挨拶が締めくくられた。

2. EAMT2017, MT Summit XVI 報告

中岩浩巳

AAMT 会長、名古屋大学

中岩会長より、2017 年 5 月にプラハで行われた国際会議 The 20th Annual Conference of the European Association for Machine Translation (EAMT 2017) 及び、2017 年 9 月に日本で開催された The 16th Machine Translation Summit (MT Summit XVI) に関する報告があった。

・ EAMT 2017 について

- 日時:

2017 年 5 月 29 日～31 日

- 場所:

チェコ・プラハ

- 特徴:

ヨーロッパ機械翻訳協会主催で、年に 1 回開催。

今回は、機械翻訳の研究開発及び機械翻訳の利用に関する最先端の事例が紹介された。

会議のキーノートスピーチは João Graça, CTO and co-founder of Unbabel (Lisboa, Portugal) による「How to combine AI with the crowd to scale professional-quality translation」だった。

・ MT Summit XVI について:

- 日時:

2017 年 9 月 18 日～22 日

- 場所:

名古屋大学東山キャンパス豊田講堂

- 特徴:

欧州・米国・アジアの順に 2 年に 1 回開催。アジア開催は 6 年に 1 回で、日本開催は 24 年ぶり。

機械翻訳の技術的発表だけでなく、利用法・性能等の利用者側の発表もあった。

会議の目玉は、主要 3 社 (Baidu, Google, Microsoft) による講演であり、この講演についての紹介があった。

3. AAMT 課題調査委員会の活動

長瀬友樹

AAMT 課題調査委員会、富士通研究所

長瀬氏から、まず、下記の図を使用して AAMT 課題調査委員会の活動報告があった。

機械翻訳課題調査委員会の活動

月に一回東京で定例会を開催(2017年度は10回開催)



続いて、ニューラル機械翻訳の訳文評価に関する紹介があった。

ニューラル機械翻訳 (NMT)、ルールベース機械翻訳 (RBMT)、統計機械翻訳 (SMT) を定量評価・定性評価で比較を行った。その結果は下記とおりである。

- ・ NMT は、正確さ・流暢さともに、SMT, RBMT よりも優れている。

- ・ NMT の短所として、訳漏れ、過剰訳出、誤訳が起きやすい。

- ・ NMT は、予測不能な挙動をする (たとえば、原文にコンマの有無で、訳語が大きく変わる)。

- ・ NMT は、SMT, RBMT に比較すると、文が長くなっても安定した性能・品質を保つ。

- ・ どの分野でも、NMT の訳質は、SMT, RBMT の訳質よりも良い。特に特許分野は、突出して良い。

- ・ NMT は口語の原文でも良い訳を出す。

- ・ 同一ベンダの NMT と SMT を比較すると、NMT のほうが正確さ、流暢さともに圧倒的に良い。

最後に、以前他の講演でその場にいる実務翻訳者に NMT と RBMT のどちらかを使いたいかという口頭調査では、意外に半々だったことを紹介。実務での利用にはまだまだ課題があるという興味深い話で講演を終えた。

4. 自動翻訳技術の最新動向

内山将夫

情報通信研究機構

内山氏より、NMT を中心とした自動翻訳技術の最新動向に関する講演があった。また、NMT の仕組みについても説明した。その要点は下記とおりである。

- ・ NMT は、「コーパスに基づく自動翻訳技術 (コーパスベース MT) の最新パラダイム」で

ある。

- ・ NMT は、Google の発表が最もインパクトあり。

- ・ NMT の新しい点は、注意機構付きの「系列から系列への学習」であること。また GPU による高速計算。

- ・ NMT の良い点は、流暢な訳文生成、個別分野へのアダプテーションが容易。悪い点は、単語の脱落がある、GPU が必要。

- ・ 汎用対訳コーパスが汎用 NMT には必要。対訳コーパスは散らばっているが、それを集めるのがチャレンジ。NMT は、コーパスサイズが 100 万文を超えると、性能が急に上がる。

- ・ コーパスを増やす方法として、総務省と NICT が翻訳バンクを始めた。2023 年に 20 億文を目指している。

最後に、知財専門の弁護士から、「原文が公知の場合、原文と翻訳文を翻訳バンクに提供し、その対訳データを利用して NMT の精度向上をしても、特に問題がない」という意見書を頂いた。これを参考に、会場の皆様にご協力頂きたいということをお話した。

5. パネルディスカッション「翻訳業界での自動翻訳の活用」

モデレータ：

隅田英一郎

情報通信研究機構

パネルディスカッションでは、モデレータ隅田氏の司会により、機械翻訳のニーズ側の方々から話題提供をいただいた。登壇者はそれぞれ特徴の違う翻訳会社のエンジニアの方、翻訳者であり、かつ翻訳の品質管理をされている方、翻訳会社の経営者の方という視点が全く違う三

人だった。それぞれの講演の後で質疑応答し、そして3つの講演が終了後、全体の質疑応答をするという構成で進められた。

5.1. エンジニアの立場から

中山雄貴

ヒューマンサイエンス

産業翻訳の中で、実際にMTエンジンを使う際に問題となる課題を詳しく説明して頂いた。

翻訳の仕事をするうえで必要なこととして例に挙げたのは、タグ、用語／ユーザインタフェース、スタイル。これに対応しないと、どんなに機械翻訳の精度が上がったとしても、それをそのまま産業翻訳で使うことは難しいという。

産業翻訳ではHTMLやXMLというフォーマットで翻訳することが非常に多く、それらのフォーマットの中には、「タグ」がある。原文の中にあるタグを、訳文側に正しく持ってくる必要があるわけだが、多くの機械翻訳エンジンでは、タグを入れるとタグの前後で意味が切れてしまったり、タグそのものが消えてしまったり、もしくはタグの中身（例：<product>）を「<製品>」と訳してしまったりする。このような問題は、翻訳者やポストエディタと呼ばれる人が手直しすることで解消することができるが、これでは、機械翻訳エンジンを入れることで却って手間が増えてしまう。タグに対応することが、実際の業務で機械翻訳エンジンを使うにあたって、非常に重要なポイントになる。

このようなタグの対応は、Googleで試したところ、うまくいかなかったが、ヒューマンサイエンス社では対応可能な形で実装しているという。

含まれている文字	翻訳業者	例	Google	Google + HS
なし	なし	<ul style="list-style-type: none"> 製品名は「<product_name />」です。 → The product name is "<product_name />". 詳しくは、<url />を参照。 → For more information, see <url />. 	可	可
あり	なし	<ul style="list-style-type: none"> 製品名は「<product_name />」です。 → The product name is "<product_name />". 詳しくは、<url>example.com</url>を参照。 → For more information, see <url>example.com</url>. 	不可	可
あり	あり	<ul style="list-style-type: none"> 製品名は、「ABC」です。 → The product name is "ABC". 詳しくは、<url>こちらのサイト</url>を参照。 → For more information, see <url>this website</url>. 	不可	可

続いて、中山氏は用語／ユーザインタフェースの訳について紹介した。

産業翻訳では、製品名、サービス名、機能名などを統一する必要があるため、用語集が非常に重要だという。ソフトウェアマニュアルを翻訳する際には、ユーザインタフェースの訳は決まっている。もし機械翻訳で勝手にそれを翻訳されると、実際のメニュー名と違う訳になってしまう可能性がある。したがって、ソフトウェアマニュアルを機械翻訳にかけた場合、UI用語集に則って、ユーザインタフェースと思われるところはUI用語集の訳を引っ張ってくる必要がある。

講演では用語集とUI用語集の例を下記図のように挙げた。

種類	例	Google	Google + HS
用語集	<ul style="list-style-type: none"> AAMT 用語集あり → AAMT 用語集なし → アジア太平洋機械翻訳協会 	不可	可
UI用語集	<ul style="list-style-type: none"> 「完了」ボタンをクリックします。 UI用語集あり → Press the "Finish" button. UI用語集なし → Press the "Done" button. 	不可	可

・用語集

「AAMT」という原文に対して、日本語では「アジア太平洋機械翻訳協会」と訳したい場合に、用語集なしだと、原文のまま。用語集ありだと、「アジア太平洋機械翻訳協会」と置き換えることが可能になる。Google翻訳は、当然用語集機能がないので、このような置き換えはできない。ヒューマンサイエンス社では、お客様の要望に応じてこういった機能を追加して「Google+カ

スタマイズ」という形で提供しているという。

・UI 用語集

「完了ボタン」という原文の「完了」に対して、「Done」という訳にしたいのだが、機械翻訳エンジンではUI 用語集が登録されていないので、勝手に「Finish」と訳されてしまうという問題がある。UI 用語集があれば「Done」と訳される。こちらもヒューマンサイエンス社で独自にカスタマイズして、お客様に提供していると紹介。

最後に、スタイルについて紹介があった。

英語から日本語に翻訳する際、細かいスタイル規則が存在する。さらにお客様によっても指定が異なる。自動翻訳がこういったスタイルに則っていないと、スタイルを一つ一つポストエディタの人が修正することになってしまうので、手間が発生する。一般的な機械翻訳サービスでは全角文字と半角文字の間のスペースの有無や、英数字の全角半角は統一されていないことが多いが、ヒューマンサイエンス社は独自の実装で統一するようにしているという。こうすることで、ポストエディタの負荷を下げている。

5.1.1. 質疑応答

モデレータ：

「NMT は素晴らしい」という話が巷では話題になっているが、中山様の発表内容によると、それだけでは不十分で、タグやUI 用語集をちゃんとやらなければならないと却って邪魔になるという話が非常に印象深かった。会場からもそれに対してご質疑・ご討論をお願いしたい。

Q1 (Mik)：弊社でもこういう形で機械翻訳の出力を編集することは頻繁にやっているが、たとえば、タグの処理で、ある程度自

動化できる部分と、NMT エンジンでもできない部分があるという話だったが、一番右の列で、ヒューマンサイエンスさんで「可」となっているのは、人間が直すという部分と、機械で自動処理する部分とがあると思う。このあたりは、それぞれのタグについて、人間が対応して直しているという意味ではなくて、その精度とか、その後にポストエディットで人間のチェックを入れているのかについて質問したい。

A1 (Na)：ここで「可」とあるのは、機械的にプログラムで対応しているという意味で「可」にしている。ただし、ニューラル機械翻訳の特徴として、どうしてもブラックボックスのところがあり、動作が読めない場合がある。そのタグが、なぜか訳文で消えてしまうとか、なぜかタグが追加されている場合もある。「タグ」と言っているが、要は訳語と同じような問題が、タグに対しても現状出てしまう場合がある。したがって、ポストエディタ、翻訳者の方に、タグについて確認してもらい処理して頂く必要がある。

ただ、以前のタグに対応していない機械翻訳を使っていた翻訳者にとっては、多少なりとも自動化ができていたので、将来的には完全な自動化を目指してはいるものの、現状ではいくらか負荷を下げられているのではないかと考えている。

Q2 (Ho)：Google 利用のヒューマンサイエンスさんのシステムは、CAT ツールと連携してシステムを組み上げられているのか、オリジナルでそういうものを持たれているのか。

A2 (Na)：仕組みを説明すると、Google API を呼び出すコードを Amazon の AWS 上に置

いている。そこでいろいろな処理をやっている。Google API を呼び出す前や後で、いろいろな処理をしている。この AWS 上のプログラムに対して、SDL Trados Studio のプラグインであったり、Memsorce から呼び出すようなプラグインであったり、その CAT ツールの部分はいくつか対応をしている。メインの機能は、あくまで AWS 上に実装しているので、プラグインは必要になればすぐに実装が可能だ。

Q3(Mi) : SDL 社や Memsorce の機能呼び出すという話だったが、SDL 社の AdaptiveMT を使っているのか。

A3(Na) : SDL 社の AdaptiveMT は、また別の機械翻訳エンジンサービスであり、Google とは別のものになっている。現在、弊社の NMT ソリューションでは AdaptiveMT を使っていない。

5.2. 翻訳者・品質管理者の立場から

三笠綱郎

株式会社十印

三笠氏は、翻訳サービスの品質管理全般、機械翻訳導入・活用を担当されている。翻訳の品質という観点から、生産性等についてどう考えていくかという話をされた。

まず、十印社では生産性に関する調査を行った。具体的には、ベテランの人間翻訳者が翻訳したときと同じ品質を確保するときに、それをポストエディットするときにとどれくらいの生産性向上があったかの検証だという。

最初に説明されたのは、従来のポストエディットでは、ある程度の品質に関する妥協をして、

「この程度の品質でよいから速く、生産性を上げよう」という形で、いろんなリクワイアメントを細かく定めるのだが、そういうことを一切しないで、「実際に経験豊富な翻訳者が普通に翻訳するときと全く同じレベルのものを作ってください」という条件で実施したということだった。さらに、「なるべく早く仕上げてください」とか、「MT の出力をできる限り利用して、ポストエディットしてください」というようなガイドラインを通常は作るのだが、そういうことも一切しなかったという。

検証の結果、人間翻訳者の生産性は、1 時間に 360 語のスピードで翻訳できた。Google のニューラル機械翻訳結果をポストエディットした場合、1 時間に 587 ワードのスピードで仕上げることができた。

NICT の「みんなの自動翻訳」の統計翻訳結果をポストエディットした場合、1 時間に 466 ワードのスピードで仕上げることができた。

いずれも、人間翻訳者よりも速いスピードで、同等レベルの翻訳文を仕上げることができた。

なお、翻訳スピードは人によってかなり異なるが、いずれも同じ傾向を示しているとのことだ。

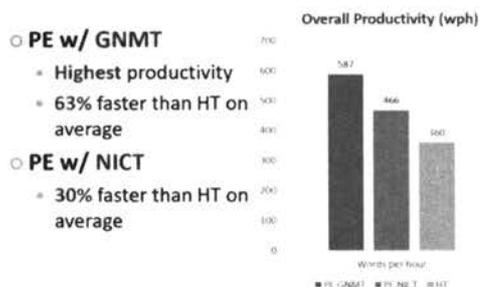
十印社の調査では、Google のニューラル機械翻訳結果を使ったポストエディットが最も速い生産性を達成した。

また、先ほどの講演ではセグメント長が品質に影響を与えるとあったが、十印社の調査ではそのような相関は見られなかった。文が長くても、逆に文が長いと若干速くできるくらいの感じだったという。

結論として、Google 翻訳では、生産性向上率として人間翻訳よりも 63%速いスピードで

同じ品質レベルのものを仕上げるができる。NICT のポストエディットでは 30%の向上という結果が得られた。

GNMT&NICTエンジンでのFull PE生産性調査結果



出典: Tsuruo Misao and Nobuko Kasahara, "99 Neural MT Is a Breakthrough in Terms of Post-Editing Productivity in English-to-Japanese Technical Translation", MT Summit XV

続いて、マーケティング翻訳をご紹介いただいた。

三笠氏は、人間翻訳者の間では、広告文のようなすごく難しいものは、いくら MT が進化しても人間でなければできないし、よほど人間のほうが速くできるというような観方がだいぶ多いという。そこで、下記の調査を行い、人間翻訳者の考え方が正しいかどうかを検証した。

調査方法は、Microsoft, Google, Apple の新製品紹介の英語ページを、下記の 4 種類と比較し、どちらの精度がよいかを決める。

機械翻訳 1 : Google NMT

機械翻訳 2 : NICT ニューラル版

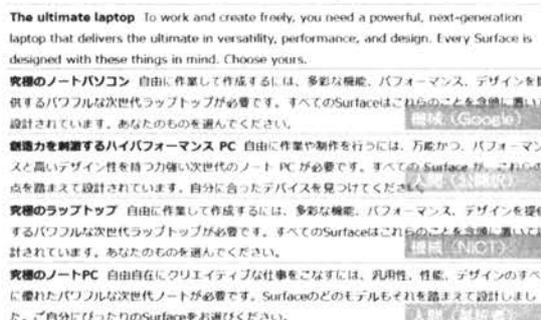
人間翻訳 1 : ホームページに公開された日本語

人間翻訳 2 : 某訳者が訳したもの

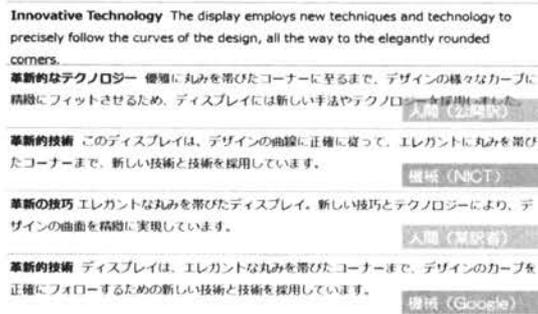
最初のテーマは、Microsoft の Surface という三つの新商品のノート PC の紹介ページで、「The ultimate laptop」は翻訳が難しいと指摘。これを 2 種類の機械翻訳結果と、2 種類の人間の訳を示した。どちらの機械翻訳もかなり良く訳せていた。ただ、人間訳の特徴として、原文は「The ultimate laptop」だが、公開訳

は「創造力を刺激するハイパフォーマンス PC」になっていて、原文とまったく違うものになっている。これは翻訳ではない。翻訳業界では「トランスクリエーション」と呼ぶが、つまり翻訳とコピーライティングを組み合わせたものだ。このようなトランスクリエーションは機械には絶対にできないことだと強調された。

某訳者の訳では、第 2 文目の「Surface のどのモデルもそれを踏まえて設計しました」は、原文は受身になっているが、メッセージ性が高まるのではないかと、あえて設計者が書いたような能動態にした。こういうのは、人間でないとできないかもしれないという。最後は、「Choose yours.」としか書いていないが、機械ではどれももうまく訳せていない。「あなたの Surface を選んでください」という意味だが、ここは人間の訳のほうが勝っている。



二つ目のテーマは、iPhone X の広告。厳密には機械翻訳はどちらも誤訳になっている。意味的によく読むと、係り受けなどが間違えている。ただ、訳語の選択という点において、機械翻訳はかなり適切な訳語が選ばれている。人間翻訳した公開訳は、誤訳ではないが流暢さという点において欠けている。某訳者の訳は「トランスクリエーション」的な要素を入れて、ちょっとした工夫でより読みやすく・理解しやすい日本語になっている。というような報告があった。



三つ目のテーマは、Google Home。紹介文の「Say it to play it」は決して難しい文章ではなく、前後の文脈を把握すれば簡単に理解できるはずだが、機械翻訳はその点において弱い。そのため、全くうまく訳せていなかった。一方人間翻訳はすべての文において正解ではないが、前後の文から伝えたいことを分かりやすく伝えることができたと話され、講演を終えた。



5.2.1. 質疑応答

Q1 (Ka) : やはり、マーケティングコンテンツは、人の感情を動かさなければならないようなところがあるので、コピーライティング要素と切り分けるのは、なかなか難しいと思う。逆に、マニュアルや契約書のように感情を入れる要素の無いもののほうが、機械翻訳は進んでいきやすいと思う。三笠さんがご覧になって、今後、このような二極化や、今後 1, 2 年で機械翻訳がどのように進むかというような予測はあるか？

A1 (Mik) : 仰るとおりだ。こういうレベルの感情表現は機械にはまだ難しい面があると思う。ただ、それは私のような人間翻訳者にとっても難しいというのも事実。私どもがマーケティング系コンテンツの翻訳をする際に、とるプロセスの一つとして、まずは直訳調に普通に訳す。その後に、コピーライティングのような能力を持つライターが、アピールする文にリライトする。そういった2工程、あるいは3工程がある。たとえば、その最初の、あるいは最初の二つの工程を MTPE に置き換えることで、現時点で十分な生産性向上効果、あるいはコスト削減効果が得られる可能性は高い。最後の仕上げの部分は、やはり人間がやる必要がある。

あと、たとえば、マーケティング系コンテンツの特徴の一つとして、先ほどの中山さんの発表にあったようなタグ処理や、TM の大々的な活用といった側面の制約が小さい案件が多いので、そういった意味では MTPE の導入が、逆にマーケティング系コンテンツのほうがしやすいという側面もかなりある。

Q2-1 (A) : 生産性の調査のところで、ある意味、通常のポストエディットの制約とかを与えないで翻訳してくださいという内容だが、その場合、原文と機械翻訳の訳文は、ペアでお見せして、それとは別のところに訳を作ってください、という感じにしているのか、あるいは機械が出した訳文を上書きする形で修正するという形で訳文を生成してもらったのか、どちらか？

A2-1 (Mik) : ポストエディットのほうについては、機械が出した出力を上書きで編集する。人手で翻訳する場合は、機械の訳は出

てこない。空のままになっていて、シンプルなユーザインタフェースで、原文と訳文を入れるボックスがあって、PE のほうは出力がでてきてそれを編集する。

Q2-2(A) : 原文と訳文が表示されて、それを参考にしながら、新たにあなたの訳文を作ってくださいというような感じか？

A2-2(Mik) : そうではなく、人間の翻訳では原文だけが出てくる。この英語を、何も参照しないで日本語に訳しなさいと伝えている。

Q2-3(A) : 訳文は提示していたのでは？

A2-3(Mik) : それはポストエディットだ。2種類のポストエディットと、機械翻訳に関係のない普通の翻訳を、同じコンテンツに関して生産性を計測した。原文と訳文の二つのボックスしかなくて、ポストエディットのケースでは、訳文のボックスにまずは機械翻訳の出力が出てくる。そしてそれを上書きし、適宜修正する。

Q2-4(A) : 場合によっては、出ている対訳を参考的に見て一からリライトする場合もあるのかなと思った。

A2-4(Mik) : そういうふうに指示したので、それもあると思う。

Q2-5(A) : 要するに、その二つの要因があって、上書きする場合もあるし、上書きではなくて、その対訳を参考にしながら、一から訳文を作っていくこともある。

A2-5(Mik) : その場合も、「ポストエディットをした」という区分で計測している。「これは使えない」と思って、パッと消して上書きしたのも、「ポストエディットをした」

という分類をして集計した。



会場からの質問(撮影：阿部編集委員：以下同様)

5.3. 経営者の立場から

森口 功造

川村インターナショナル

森口氏は今後、あるいは今、翻訳業界が、機械翻訳を利用したソリューションをどう見ていくべきか、あとは、この先どういうところに注力していくべきかについて話をなされた。

2015年頃はSMTの全盛期で、ある程度の品質までは行っているが、それを実用化するためには、かなりいろいろな工夫をしなければいけない状態だったと過去を振り返った。今なお、機械翻訳は人手の翻訳に比べれば質が落ちるし、やらなければならないこともまだまだある。では、ポストエディットというサービスは、ここにいる三社はどこもやっているが、それを付加価値として出すことで済んでいいのか。それともその先を業界として何か考えるべきなのか。そこで、森口氏はSWOTの考え方をを用いて分析を行った。

MT 活用 - 戦略策定

	MTPESP 目標での MT・PE の強み・弱み	
	Strength	Weaknesses
MTPESP 目標その外に強み	事業チャンスを作る	弱みを克服して事業チャンスを作る
Opportunities		
Threats	脅威を回避して事業チャンスを作る	撤退/回避すべきだが・・・イノベーションはここから？

「MTPESP」は、「マシントランスレーションとポストエディットのサービスプロバイダ」の略で、ポストエディットというソリューションを含めた機械翻訳の強み・弱みということと、サービスプロバイダの目線でみたときの今の市場の Opportunities と Threats のことを意味する。通常は、「Strength と Opportunities を攻めましょう」となる。それ以外に、Weakness を克服できるものなら、そこは事業チャンスになるし、脅威があるところはそれを解決・回避することで、ビジネスチャンスとなる。「MTPESP」において、それぞれは何なのかを整理していけば、攻めるべきポイントが分りやすくなると話した。

・強み(Strength)

今の機械翻訳の強みは「スピードが速い」、「コストが低減できる」、「質が向上した」の3点。特に、訳質の向上は非常に大きな強みだ。SMT は癖があるが、「それがそういうものだ」ということを理解しながらなら PE もできた。一方で、今の NMT はすごく簡単にできてしまう。ボタン一つで誰でもできるようなレベルだ。

つまり、以前はそこに対してのビジネスチャンスがあったが、今は Google NMT の登場によって、スタートラインは皆一緒になってしまったとのことだ。

・機会(Opportunities)

外部環境として、翻訳をしてほしいテキスト

はずっと増え続けている。そこに追いついていないことは、この業界にとってはチャンスだ。

機械翻訳はその解決策として考えられたが、そこに対しての技術的ハードルや、対訳データをどう集めるかなどが原因で、参入がなかなか難しかった。Google や Microsoft を使うということを見ると、その障壁はすごく下がったという。また NMT のハードウェアとかのリソース的な部分でお金が非常にかかるというのはあるが、技術的なところは非常にシンプルだったので、ある程度勉強された方にはできる。

あとは、「Google は使えそう、機械翻訳いけるじゃないか」と皆が思い始めたマインドの変化は、我々 MT サービスプロバイダには追い風だ。更に、ポストエディットプレイヤーとしては、競合する相手が少ないことも機だ。

その中重要なことは、「わかりやすさ」と「顧客体験」。

- わかりやすさ

「ポストエディットと人手による翻訳の違いがよく分らない」ということのないようにしたい。そのサービス提供方法を確立したい。後で少し説明がある、キーワードとしては、やはり、「シンプルに、セキュアに」というところが重要だと思うと森口氏は語った。

- 顧客体験

どうということが機械翻訳を使って、顧客・ユーザーとして素晴らしい経験につながるのだろうかと考えると、自分たちで翻訳が終えられ翻訳会社が要らなくなる。ただ、それはまだまだ先であり、それを目指していくことが翻訳業界の中で必要だと思っているとのことだった。そこに仕事がたくさん眠っている。要は、セルフサービスで機械翻訳はすごく簡単に使える。そうすると、「機械翻訳はいいじゃないか」とい

うことが分ってくるし、よくよく聞くと「セキュアです」となると、「じゃあ、使おう」というふうになっていく。それが体験として、「このレベルであれば自分でちょっと直せば使えるな」と、しかし「この文章を機械翻訳にかけたら使えないなあ。じゃあ、翻訳会社に出そう」となる流れ（仕掛け）を作らねばならない。ここでは、TM や用語集という言葉は要らない。それを実現するのが、将来的な目標だという。

・弱み(Weakness) と回避

よく「機械翻訳は安全なの？」と言われる。Google も Microsoft もすごく明確に「データを収集しません」ということを書いてあるにも関わらず、ユーザ側はまだ慎重だ。セキュリティに関しては、情報漏洩だけでなく、送ったデータがそのまま収集されることを懸念するだけでなく、IT 部門のセキュリティに対する懸念が多い。そういったところも含めて、今は情報漏洩だけになっているが、セキュリティというキーワードに関してうまく説明して、「そんなことはないですよ」ということが理解頂ければ、お客様も「じゃあ、ちょっと使ってみよう」となる。そのときにシンプルに提供できれば、機械翻訳の結果をすぐに見ることができる。そして、見て頂くと、このままでは使えないことが分る。そして、自分たちで直すか、翻訳会社に直してもらおうか、というところの仕掛けを作ることが重要だ。

そこを回避する方法として、オンプレミスだ。また、タグの処理などを含めて、まだ NMT で解決できていない品質の課題を回避・補完するため人間が直さなければならない。やはりポストエディットは大事だ。さらに、希少言語対応や、医療、特許を含めて分野的専門性等々はビジネスチャンスだと森口氏がいう。

・脅威(Threats) と回避

翻訳業界では、翻訳するテキスト量は増えているが、予算が増えたことは一度もなく、減り続けている。それを制約と受け止めるが、別の方法で解決する方法を考えるか。あとは、上記述べた参入障壁の低下。翻訳会社はどこでもポストエディットができるようになっているので、そこは脅威だ。

さらに、Google や Microsoft のような情報資産ホルダーは、とてつもない優位性を持つので、そこは当然戦えない。それをどう考えて付き合っていくかを考えなければならない。

また、この業界に特有な労働集約型産業のところ、生産性の向上が非常に重要になってくる。これをうまく説明しないと、翻訳者にはただ単に価格を安くされたと思われることもある。そのケアが必要だ。そこがケアできれば、次のポストエディタ不足も解決できると思うと森口氏は語った。

ここで、その場にいる翻訳会社の方、研究者、ツールベンダの方に要望を述べられた。

・専門性・残留品質課題を解決できる技術革新が欲しい。

例えば、タグの処理。産業翻訳の中ではタグの処理が完璧にできると、それだけでそのツールを使おうというモチベーションにつながることもある。

・サポートする文書タイプの増加。

Word, Excel, PowerPoint くらい普及している文書であればよいのだが、産業翻訳では、InDesign, XML, FrameMaker などで作られているものが非常に多い。それをサポートできるようにしてほしい。

・情報資産の収集の効率化。

情報資産ホルダーと組んで、対訳データを提供して精度を上げていく努力はしなければならない。

- ・情報資産の量によらない革新が欲しい。
- ・PEの自動化。

これは翻訳業界が何かしらの協力をしなければならない部分である。

次に、森口氏は自動翻訳とPE活用のキャズムの壁に言及した。今までは、「キャズム16%の壁」というのがあったが、川村インターナショナル社が既にその壁を超えたという。それが普及率とすぐに結び付くというのはかなり乱暴な理屈だが、要は、ポストエディット、機械翻訳を採用していこうと市場が考えている。そこはメインストリーム市場に入っていこうとしているところで、乗り遅れると勿体ないということで、そこが一番のビジネスチャンスだ。我々は乗り遅れたくないと思っている。それには、セキュリティも非常に重要な点になると話した。

自動翻訳+PE活用：キャズム（16%の壁）を超えた？
すでにメインストリームと考えるべきか



最後に、森口氏らの夢として、Word, Excel, PowerPoint だけではなく、すべてのドキュメントを、お客様が簡単に試せる環境を作りたい。それを作って、機械翻訳が適用されて、『これだったら自分たちでいい、これだったら翻訳会社に出さなければ無理だ』というようなことができるようなプラットフォームを作りたいと将来への思いを語り、講演を終えた。

5.3.1. 質疑応答

Q1-1 (Mi) : 御社は翻訳会社で、医療関係の翻訳も扱っていると思うが、今後は医療関係

にも機械翻訳を使っていく予定か？

A1-1 (Mo) : すでに使っている。医療に関しては、どちらかというと、海外のサプライヤのほうがすごく積極的だ。なので、このエンジンを使ってくれということで、我々はポストエディットだけを提供するということは、すでに始まっている。

Q1-2 (Mi) : それは SYSTRAN か？

A1-2 (Mo) : それは回答不可としたい。

Q1-3 (Mi) : 以前、研究者の方に、「医療分野の翻訳に、機械翻訳はどうか？」とたずねたところ、「医療分野に関しては、製薬会社が、機密が厳しくてデータを出したがないので、機械翻訳のエンジンに必要なデータ量が集らないので、精度が出ない」と言っていた。あとは機密の問題もあると思うが、その2点についてはどう思うか？

A1-3 (Mo) : そのケースに関しては、そのお客様が提供されているので、それは当たらないと思う。要は、医療製薬会社のお客様の中でも、特にオランダなどの欧州の外資では、対訳データを活用するという考え方がかなり進んでいるので、それを自ら提供して、Adaptation した MT でやり始めている。質としては、それを使って実際にポストエディットしているわけではないが、そんなに悪くないと聞いている。製作の現場の中では、「かなり良いですよ」ということは言っていた。

Q1-4 (Mi) : それは既に商用レベルということか？

A1-4 (Mo) : 実際にビジネスになっているので、商用レベルだと思う。

Q2-1 (A) : 先ほどのスライドにもあったが、予算は減りつつある。もちろん、一から Human translation をするよりは、ポストエディットをするほうが、お客様のほうも

安くて大丈夫ですよという感じできていると思うが、その場合の翻訳品質は、たとえば機械翻訳を使おうが使うまいが、一から人間が翻訳したものと変わりませんよという感じでしているのか、まあそれよりもちょっと良ければいいですよ、たとえば松竹梅があればこのレベルで品質確保して頂いたら大丈夫ですよ、みたいなのはあるのか？

A2-1(Mo):それはある。たぶんここにおられる会社の方もいろんなポストエディットのレベルを作っていると思うが、弊社は、最近それは避けている。たとえば日英翻訳の場合は、書式を直さないとか。そういうところで意外とポストエディットを言ったりするが、実際はお客様にも非常に分りづらいので、弊社としてはプロセスですよ。翻訳会社の付加価値としては品質管理の部分がいちばん大きいとっていて、品質管理のために別のプロセスを入れるわけだ。ISO17100も、そうだし。だから、それを入れた場合と、今のポストエディットというものを人手の翻訳と同じにするという場合の人手が何を指しているのか。基本的に、人間が一から翻訳したものが人手翻訳だ。しかし翻訳会社が提供するものは、それをバイリンガルチェックしたものだ。「どちらが人手翻訳でしょう？」といったときに、どちらも人手翻訳だ。だから、ちゃんと間違っていないかどうかを確認するプロセスが入っていれば、当然そこにコストがかかる。

Q2-2(A):要するに、お客様が受け入れる品質基準は、何をしようが、変わっていないということですね？

A2-2(Mo):仰るとおりだ。そしてプロセスが

違うので、リスクがあることはもちろん説明するし、ウチの場合は基本的にサンプルを出す。「このレベルで出しますよ」というサンプルを出して、合意頂いたケースしか受注しないことに決めている。

Q2-3(A):そこのところのお客様とちゃんと合意が取れていないと、「こんな筈じゃなかった」とか、「これじゃあ困りますよ」と言われるようなことが起きると思う。すでに16%を超えておられるようなので、そういうのがかなりボディブローのように効いてくるのかなと思った。

A2-3(Mo):不思議なことに、今のところ無い。幸せなことに、今のところクレームになっていないが、今後出てくるかもしれない。そこに関しては、我々としては、品質管理を継続的に改善する部分になるので、体制としてどうするかとか、自動化ツールを活用するか、いろいろ考えないといけないと思う。

Q2-4-1(A):いっしょに作業する翻訳者の方々が、単価は下がるが、質は変わらないということで、そこがかなり問題になりそうな気がする。「量をこなすから単価が安くてもよいのでは」という話もあるが、ただ実際に品質のところ、結局これだけのレベルをクリアしてもらわないと困りますという話になると、さっきちょっと言ったが、「であれば、リライトしたほうがいいよね」という話になる。しかし、リライトするのなら、「この値段ならちょっとね」みたいな。「この金額なら、このレベルだよ」みたいなのを、翻訳者のほうは、たぶん考えていると思う。

A2-4(Mo):仰るとおりだ。

Q2-5(A):お客様と翻訳者との間でギャップが出ると、大変だと思う。

A2-5(Mo):そこは大変だと思う。

Q2-6(A):そこはきちっとお互いに合意を取りながらやらないと、ビジネスとして大変だ。

A2-6(Mo):それは本当に仰るとおりだ。ちょっと前に書いたが、ビジネスモデルが非常に重要で、生産性が上がって、ポストエディットということをや加価値としていかなければならない。「生産性が倍になったので半分にしてくれ」というお客様はお断りしている。

Q2-7(A):そのようなお客様は必ずいる。

A2-7-1(Mo):「それは無理なのです」と言って断る。ポストエディットはそれだけ価値の高いソリューションであるし、それを理解してもらうための我々側の役割も大きい。歯に物が引っ掛かった感じの言い方だが、我々は海外のお客様が結構多いので、海外主導でポストエディットしてくれと言われると、単価を向こうが決めてくる。そのケースがいちばん辛い。予算もあるので、こんどは翻訳者に対して、「これでやってもらえるか？」という話になる。実際にそういうお客様がいて、1年以上お付き合いがあるのだが、つい最近断った。なぜかというと、機械翻訳の質が悪すぎる。やはり翻訳者からも、クレームにつながるだろうなということは分っていたし、「生産性が倍だから半分だ」というのもどうかなというので、いま弊社が外向けに提案するのは、やはりそこに付加価値を付けている。そこが、「いけるな」という感覚があるの

で断るというケースがあった。

A2-7-2(Mi):ひとことだが、要求品質レベルと翻訳サービスをどういうプロセスで確保し、どういう対価を頂いてお客様に提供するかという話と、MTPEの話はもちろん密接に結びついているが、本来切り分けて考えるべきと思う。たとえば、さきほど森口さんが仰っていたが、人間翻訳限定しても、分野によって、お客様によって、いろんな品質要求レベルがある。実際のところ、そういうお客様のニーズに応じて、あるときは本当にチェック無しで安く提供する。そういったフレキシビリティを持たせて、私どもは提供しているわけだ。それがいちばん露わに現れているのがMTPEのところ、それが露見するので、それを組み合わせた議論が行われているが、本来は別の話で、PEという切り口だけでそれを語ると、ちょっとミスリーディングだと思う。ただ、ISOの規格とかが、いま本当に高い品質レベルの翻訳サービスを提供するというような高い基準のものしかないの、そこらへんを逆にPEの規格の制が進めば、そこらへんをPEという切り口ではなくて、品質レベルをもっと柔軟に課していくという方向で考えるべきだと私は思う。

6. 全体質疑応答

モデレータ:パネラーのどなたに対する質問でもよいので、全体的質問時間とします。

Q1-1(An):川村インターナショナルさんが「キャズムの16%の壁を越えた」ということだったが、ヒューマンサイエンスさんと十印さんは何%くらい浸透しているのか?

A1-1(Mo):それは、私が勝手に言っただけなので、フェアでないかもしれない。

Q1-2(Mi):何の割合なのか？

A1-2-1(Mo):我々の場合は、英日・日英の翻訳が、比率でいうと80%が人手翻訳で、それ以外がポストエディットになってきた。私は経営判断で言ったが、言えなければ、「言えない」と言ったほうがよい。

A1-2-2(Na):答えられる範囲で答えると、ヒューマンサイエンスの私の部署は、IT系の英日翻訳をしているが、ITマニュアルに限定すると、それこそ8割は超えているような状況だ。ITのマーケティングに関していえば、ゼロだ。

A1-2-3(Mi):数字は控えさせて頂くが、私どもはかなりPEを導入している。両社よりは低いくらいだ。お客様からの要望に合わせて直すのをいちばんの売りにしている。お客様が高い品質を求める場合は、PEという選択肢を取らないという場合が多かったりする。ただ、非常に多くのプロジェクトで、ポストエディットも検討するという一方で、検討する割合といえは16%を超えていると思う。

Q2-1(Nak):高品質の翻訳ということかもしれないが、意味が通る翻訳ということと、品格のある翻訳というのがあると思う。たとえば、相手会社の重役が読むというときは、やはり品格のある、あるいはどなたかが仰っていたような「感性を揺さぶるような」翻訳が求められることがあると思う。現時点で、翻訳会社の皆様方は、そういう要求があったときは、人がやるのか？ それとも、機械翻訳になんとかそれを実現させようと思っているのか？ そのあたり、現在と今後について、どのようにお考えか。

A2-1(Na):ヒューマンサイエンスでは、先ほど申したようにマーケティング系、その中に

は、手紙やメールのような人に対してその個人の方に訴えかける必要があるような場合は、機械翻訳を使っていない。これは、一から翻訳者が翻訳するという形を取っている。

Q2-2(Nak):機械翻訳には、それを期待しないのか？

A2-2-1(Na):実際に試してはいて、その品質を見ると、「意外に良いな」というものはある。ただ、現状だと、マーケティング翻訳ということで、コスト削減のプレッシャーが少ないので、そこに取って機械翻訳を入れる必要はないと思う。

A2-2-2(Mi):プラクティカルにはそのとおり。私どもでも、品格が高い訳を望むお客様や、メッセージ性の強い先ほどマーケティング翻訳と呼んだ分野では、実際のところ、MTやPEの導入率は低い。ただ、それを活用してコスト削減や効率向上に使える部分はかなりあるとは思っている。あと、実際の出力でも、私自身も翻訳者なので、ニューラルのMTにかけると、自分が思い付かなかったようなすごく良い訳を出してくることが、あまり多くはないが、たまにある。なので、そういう意味ではすでに機械に頼っている。

A2-2-3(Mo):私どもは、ヒューマンサイエンスさんと同じで、機械翻訳を活用してはいない。ただ、従来のやり方でも、人が翻訳したものをわざわざ別のライターさんが直すという作業になるので、それを考えたときに、本当に翻訳って要るのかなと近頃感じている。要は、求めているものがそもそも翻訳ではないわけだ。だから、そこはひとつ考えてもよいところかなと思う。機械翻訳のほうが、「なるほど」という流暢な翻訳を急に出したりする。自分ではやらないな

というような、綺麗な訳語をだすことがある。そういう意味では、今までマーケティングの翻訳は、ちょっと誤解されているのは、読みやすさとか正確性を重視することによって堅苦しい文章になっているのが、読みやすいものになっていると喜ばれることもある。実はもっと喜ばれるのは、それを無視して、お客様にアピールできるもの。これって、本当に翻訳なのかなと思うところがある。



会場の質疑応答も含めてパネル討論

2) 本報告は、当日の議事録（事務局管轄）も参照させていただいた。



講演会後の交流会の挨拶に耳を傾ける参加者の皆様

7. 記載上の補足説明

1) 質疑応答部分の略称について

質疑応答部分は略称で記載させていただいた。

各略称は以下のようにになっている（表示順）。
姓のみの表示とさせていただく。

Mik: 三笠氏（十印）

Na: 中山氏（ヒューマンサイエンス社）

Ho: 本多氏（アイ・ディー・エー株式会社）

Mi: 三浦氏（個人翻訳者）

Ka: 加藤氏（Memsorce）

A: 安達氏（サンフレア）

Mo: 森口氏（川村インターナショナル）

An: 安藤氏（アイ・ディー・エー株式会社）

Nak: 長尾氏（AAMT 初代会長）

国際会議 AMTA2018 参加報告

中岩浩巳

名古屋大学情報学研究科／アジア太平洋機械翻訳協会 (AAMT) 会長

1. はじめに

本稿では、2018年3月にアメリカ・ボストンにて開催された米州機械翻訳協会 (AMTA) 主催の第13回目の国際会議 AMTA2018 に、中岩が参加したので、会議の概要、及び、AAMT 課題調査委員会が進めている機械翻訳の自動評価に関する発表について述べる。

2. 発表と会議

AMTA2018 は AMTA 主催で隔年開催の会議であり、機械翻訳の研究開発 (research track)、機械翻訳の利用 (user track) に加え、政府による機械翻訳プロジェクト (government track) に関する発表も含まれるのが本会議の特徴である。今回は、本会議で、research track が口頭発表 15 件及びオープンソース NMT の展示が 6 件、user track が口頭発表 17 件、government track が口頭発表 16 件及びパネルディスカッション 1 件が行われた。また企業や研究機関などによる展示会であるテクノロジーショーケースには、15 件の展示と 10 件の口頭発表 (AAMT は展示と口頭発表両方) が行われた。また、全体セッションとして、基調講演 5 件、パネルディスカッション 1 件が行われた。参加者数は本会議 175 名、チュートリアル 89 名、ワークショップ 40 名であり、欧州を中心に米国以外からの参加者も意外と多かった (4 割弱) が印象的であった。以降、基調講演を中心に主だった発表について概説する。

Linden 大学の Arianna Bisazza による基調講演 “Unveiling the Linguistic Weaknesses of Neural MT” では、SMT の技術 (語順並び替えなど) を紹介しながら NMT に言語構造などの言語的特徴を扱うこと

の有効性を示し、現在主流の RNN ではなく Convolutional NN や Full Attention Network が主語と動詞の一致に効果があることが紹介された。

Google の Macduff Hughes による基調講演 “Machine Translation Beyond the Sentence” では、NMT の性能は英仏、英西ではほぼ人間と同じレベルに達成しているがこれは 1 つのマイルストーンであり最終目標ではないこと、gender が無い言語からある言語への翻訳において区別が困難であること、NMT はマイナー言語において特にインパクトが大きいこと、今後は 1 文を超えた処理が必要であることなどが紹介された。

IARPA の Carl Rubino による基調講演 “MATERIAL, designing an MT Program for IARPA” では、多言語への拡張が容易な言語非依存の翻訳システム MATERIAL についての紹介があった。IARPA の様々な目的で活用 (機械翻訳に加え、言語横断情報検索、言語横断要約、音声翻訳等) されており、特に言語資源の少ない言語への拡張を考慮したシステムになっているとのことであった。

パネルディスカッション “Deploying Open Source Neural Machine Translation (NMT) toolkits in the Enterprise” ではオープンソース NMT を開発している Systran, Google, Microsoft, Amazon の各担当者から各社のシステムの紹介が行われたのち、会場を交えた質疑応答及び議論が行われた。オープンソースにするメリットは外からのフィードバックが得られ最新の技術が反映できること、エンタープライズでの実装のベストプラクティス、今後の展開については SMT における Moses 同様に、オープンソース NMT が技術を深化させるのではないかなどの議論がなされた。

Microsoft の Glen Poor による基調講演“Use more Machine Translation and Keep Your Customers Happy”では、MicrosoftOffice の多言語化に Microsoft Translator を活用していること、その過程で機械翻訳の活用に関する工夫 (MT 出力を活用できる品質の閾値設定や訳のリサイクル等) などが紹介された。

Harvard 大学の Alexander (Sasha) Rush による基調講演“Towards Easier Machine Translation”では、NMT の歴史を紹介したのちに、NMT のメリット (学習の容易さやインプリメントの容易さなど) を示し、計算量の多さを解消する工夫 (NMT on the Edge; モバイルデバイスでも動く) や、LSTM の状態を可視化するツール等を紹介した。

テクノロジーショーケースでは、AAMT 課題調査委員会で検討を続けてきた AAMT テストセットに基づき機械翻訳出力を文法的特徴別に自動評価する技術と、それを可視化して提示する英語版ウェブインターフェースを紹介した。この展示では、MT の自動評価の研究を行っている多くの方々がブースに来られ、積極的な議論をすることが出来た。特に、我々と同様のアプローチをとっている独 DFKI/CSA の方からは、是非今後情報交換をさせてほしいとの要望をいただいた。NMT 主流となった今、標準的な自動評価法である BLEU は人間評価との相関が低いことが様々な場で報告されているタイミングでもあり、我々の研究も含め、評価法に関する研究に、以前より注目を集めている印象を受けた。

上記以外にも、本会議の一般講演や、ワークショップ“Translation Quality Estimation and Automatic Post-Editing”にも参加した。これら AMTA2018 の発表内容に関しては AMTA2018 のウェブサイトプロシーディングスやスライドのほとんどが掲載されているので、それをご覧いただきたい[1]。

3. おわりに

AMTA は、EAMT や MTSummit と同様に、機械翻

訳の研究開発者と利用者、また、政府関係者が集い、それぞれの側面からの最新の成果を共有・意見交換できる貴重な場である。進歩の激しい NMT 技術の活用面での最先端動向を把握するためには、欧州機械翻訳協会が開催する EAMT と併せて、継続的に参加する必要があることを痛感した。また AAMT はアジア太平洋地域の機械翻訳の団体であるが、AMTA や EAMT のように、独自の国際会議を持たない。AAMT が真のアジア太平洋地域の協会となるためには、AAMT という国際会議もしくはシンポジウムの開催を具体化し、アジア太平洋地区の活動を広くアピールする場を設けるべきことを再確認させられた。

参考文献

[1]<https://amtaweb.org/amta-2018-proceedings-for-the-conference-keynotes-workshops-and-tutorials/>



ボストンの街並



会場のホテル (Aloft Boston Seaport)

2017年度 JTF 第6回翻訳セミナー報告

目次 由美子

XTM International Ltd.

2017年度 JTF 第6回翻訳セミナー報告

日時：2018年3月8日(木)14:00～16:40

開催場所：剛堂会館

テーマ：ニューラル翻訳（NMT）の実体とその周囲の実態に迫る ～NMT との良い関係の確立を～

登壇者：

内山 将夫 国立研究開発法人 情報通信研究機構
研究マネージャー

三笠 綱郎 株式会社 十印 品質管理・MT 戦略部マネージャー

隅田 英一郎 国立研究開発法人 情報通信研究機構
フェロー、JTF 理事

※ 本稿は、日本翻訳連盟が発行する『日本翻訳ジャーナル』（Web版）にも同時掲載いただいています。

約130名の参加者で溢れんばかりの会場にて、隅田氏のイントロダクションによってセミナーが開始された。機械翻訳は過去には使えないと言われたこともあり、崖を登っては転げ落ちるということを約60年間にわたって繰り返して来たそうだ。しかしながら、峠は越えたと認識しており、その確たるところを紹介したいと隅田氏は力強く語られた。そして、質疑応答を中心に本セミナーを進めたいと、参加者の積極的な発言を促した。

「NMTの利用方法」というスライドが映し出され、内山氏のセッションが始められた。自動翻訳サイト「みんなの自動翻訳」は2014年頃から製作し、公開しているそうだ。当初は統計的翻訳（Statistical Machine

Translation: SMT）であったが、昨年度からニューラル翻訳（Neural Machine Translation: NMT）も利用可能になった。登録にも利用にも費用は発生しない。このサイトで自動翻訳を利用するには、複数の方法がある。

1. テキストを入力して翻訳を実行する。
2. MS Word 文書などのファイルや、Web ページの URL を指定して翻訳を実行する。

読み込まれた文書はフォルダに格納され、左列にセグメント化された原文、右列に自動翻訳によって生成された訳文が対訳テーブルとして表示される。いわゆる「翻訳支援ツール」の機能を統合的に利用できる。翻訳メモリ（Translation Memory: TM）や用語集を活用でき、対訳の登録も可能。Word 文書などの場合は WYSIWYG モードとして別ウィンドウに表示しての編集も可能。

自動翻訳についてはカスタマイズも可能であり、翻訳前に原文を書き換える、翻訳後に訳文を書き換える、TM を利用した完全一致翻訳、訳語の指定などが設定できる。著名な翻訳支援ツールのみでなく WebAPI も利用可能であり、連携の幅広さも伺える。平均して1日100人以上のログインを確認しており、近い将来に翻訳速度の高速化が予定されている。

さらに、内山氏は「NMTの技術解説」を続け、NMTとはコーパスに基づく自動翻訳技術（コーパスベースMT）の最新パラダイムであると説明した。コーパスベースMTを構成する要素として、1. アルゴリズムとハードウェア、2. 対訳コーパス、3. 評価を挙げた。

1. アルゴリズムとハードウェアについては、以下の表が示された。

	1980年代	2016年まで	いま
アルゴリズム	EBMT*	SMT	NMT
ハードウェア	CPU	CPU	GPU

* EBMT = Example-based Machine Translation、用例ベース機械翻訳

NMTでは従来の約1000倍の計算量が必要であるため、計算処理能力はきわめて優れているが高額なGPUが必要とされるようだ。

2. 対訳コーパスについては、異なる言語で同一の意味を有する文章の組から成るデータベースであるとのこと。

3. 評価については、人間がどのように感じるかという観点からMTの良さが評価されてきたようだ。

コーパスベースMTはEBMTからSMTへ、さらにNMTへと新しいアルゴリズムが変遷を遂げるなか、翻訳精度は右肩上がりの向上を続けている。コーパスが大きい(例:100万文を超える)とき、SMTに比較してNMTはきわめて高精度であるとも指摘された。

さらに、NMTがSMTよりも優位である例として、SMTの語順変換を説明してくれた。SMTでは翻訳精度を上げるため、「語順変換」(1.文をフレーズに分割、2.フレーズの順番を変更、3.フレーズを翻訳)という手法が採択されてきた。結果として、言語対の構造が異なる場合では良好な結果を得ることは困難だったが、NMTでは自然な語順を生成することが可能とのこと。

NMTの良い点としては、流暢な訳文生成が可能であることと個別分野への適用が容易であること、悪い点としては、単語の脱落が発生することとGPUの需要が指摘された。結論として、NMTの改善によって翻訳の高精度化を達成することが最善と考えられている。

今後のコーパスベースMTの改善について、アルゴリズムやハードウェアは研究者・開発者が業務の一環として改善すべきであるとしながらも、汎用コーパスをより多く持つことによるベース精度の確保、分野を特化したコーパスによる汎用NMTに対する訓練(アダプテーション)、用語集の活用や前・後処理によるカスタマイズの重要性が指摘された。

三笠氏は、英日翻訳の品質評価に対する取り組みを紹介した。1つはWeb上で公開されているマーケティング系のコンテンツを対象に、もう1つは技術系コンテンツを対象にして、NMTと人間翻訳(Human Translation: HT)を比較している。NMTにはGoogle NMTとNICTニューラルを使用し、HTにはWeb公開テキストと実在する某翻訳者による翻訳を示した。

自らもIT系翻訳者として活躍してきた三笠氏は、機械翻訳には本当の翻訳はできないと言い続けるのではなく、なかなかやるじゃないか!というところを見て欲しいと言及した。

マーケティング系コンテンツのテキストには「The ultimate laptop」という一例があり、これは経験豊富な翻訳者にとっても良い訳を提供する難易度は高いように思われる。NMTの訳出は「究極のラップトップ」と「究極のノートパソコン」であり、HTでは「想像力を刺激するハイパフォーマンスPC」と「究極のノートPC」であった。三笠氏からは、HTでこのような訳を捻出するには長時間を要するがMTの訳出は秒単位で実行される、広告訳に対する経費として包括的に考慮するとMTを活用するという選択肢も見いだせるのではないかという指摘があった。

技術系コンテンツでは、約5000ワードのオンラインヘルプを翻訳対象としていた。10年以上の翻訳歴を有する4名の上級翻訳者がHTまたはポストエディット(PE、MTが生成した訳に対する人手による事後編集)をするに際の生産性を1時間あたりの作業ワード数として算出したとのこと。PEにはHTと同等レベルの品質を求め、急がせず、翻訳支援ツールは使用せず、TM、用語集、スタイルガイドも提供しなかったとのこと。MTはGNMTとNICTのNMTを利用した。

翻訳スピードは、いずれの翻訳者にもHTよりPE

が速いという結果であった。

三笠氏からは、翻訳業界ではこれまで PE が翻訳単価を切り下げる手段かのように捉えられてきた側面があるが、実際の生産性向上の効果を見極めてきっちりとした仕組みを作る必要があるのではないかと言及された。会場から原文を理解しない人に PE ができる将来が来るかとの質問が発されると、通常の翻訳プロセスにおける専門家レビュー (Subject Matter Expert: SME) を PE として組み込むことの可能性に話が及んだ。ただ、コスト面も含めて有効性を慎重に検討する必要があるだろうとのことだった。

当日のセミナー最後のトピックとして、隅田氏が再登壇し、「翻訳バンク」の概要を説明してくれた。これは、どんな文でも高精度に翻訳するシステムをみんなで作る方法とのこと。以下の形式の日・英、英・日の翻訳データの提供を広く募っている。

A. 翻訳メモリ、B. テキスト、C. Word (一段組図表なし)、D. Excel、E. その他 (応相談)

Web サイト「みんなの自動翻訳」には、対訳集の登録を受け付けるページが用意されている。また、情報提供者として NICT と二者間契約を締結することも可能。さらに、自動翻訳技術のライセンス料算定時に提供翻訳データを考慮することも可能とのこと。機密保持契約や著作権などに対する不安の声を聞くこともあるが、公知の翻訳テキストを NICT へ提供することは問題に該当しないという説明もあった。詳しくは、<http://h-bank.nict.go.jp/index.html> を参照されたい。

日本中からの寄付によって建立された奈良の大仏様のように、オールジャパンで世界一の自動翻訳を作ろう！と力説されていた。

日本語オノマトペ共起表現レキシコン

首藤 公昭

日本語処理研究工房ことばの森

<http://jefi.info>

はじめに

オノマトペ（擬音語、擬態語）の豊富さは日本語の特徴の一つであり、日本語を母国語とする人にとって有用・不可欠な語彙資源であるが、従来の機械翻訳等における日本語処理（NLP）では十分な対応がなされていない。近年、微妙かつ繊細な人の感覚を端的に表現するオノマトペの重要性が注目されており、特定領域文書に出現するオノマトペを調査・抽出する研究[1][2]をはじめ、オノマトペの意味に関する研究[3][4]、コーパスからオノマトペと他語の共起を調べる研究[5][6]、出現オノマトペを手掛かりに文書の種別判定をする研究[7]等々、種々の研究が報告されている。しかし、機械処理用のオノマトペ・コロケーション辞書を開発した報告は筆者の知る限り見られないようである。本稿では、筆者が開発した機械処理用の日本語オノマトペ共起表現レキシコン JMWEL_onomatopoeic v2.9（以降、本レキシコンと記す）の概要を紹介する。

本レキシコンの主な特徴は、

1. オノマトペと他語の共起表現中にギャップ（内部修飾句）が存在する可能性を記載している
2. オノマトペ（単体）、オノマトペ共起表現中の語彙に漢字・片仮名異表記を与えている
3. オノマトペの連体、連用、動詞化用法を体系化して記載している事である。

今世紀に入り、日常の言語にはコロケーション、決まり文句、慣用表現等の定型表現が予想外に多種、多量に使われていることが重視されるようになり、NLP 分野では複単語表現(Multiword Expression, MWE [8][16])、言語学では定型言語 (Formulaic Language[9])、単語連鎖 (Lexical Bundles [10])、構文文法 (Construction Grammar[11]) などの枠組みで

種々の研究が行われるに至っている。

当工房では、この種の日本語複単語表現の総合的なレキシコン JMWEL (Japanese MultiWord Expression Lexicon) の開発を進めてきた[12][13]。本レキシコンはその一部をなすものである。

収録表現

本レキシコンの見出しは、

- (1) オノマトペ（単体）約2,900種
- (2) オノマトペと他語が共起した日常よく現れる句（オノマトペ・コロケーション、以後、オノマトペ共起表現と記す）約31,000種であり、「悠々」、

「懇々」、「生き生き」のような漢語由来、漢字表記可能なもの、「オヤッ」のような感動詞に分類し得るものを含め、新聞、雑誌等の記事、小説、テレビ、ラジオの放送文から内省によって抽出したものを基本に、既存の不特定の辞典 [14][15] などを使って補強したものである。方言、古語や近年殆ど使われなくなったもの、特定年齢層でのみ使われるものは除外されている。

（注、(1)は複単語ではないが少数なので現在は便宜上、本レキシコンに含めている。）

記載情報

本レキシコンは、Microsoft Excel で作成した xls ファイルに纏められており、1行に割り当てた1個の見出しに対して、A~I 欄に以下の情報を記載している。

A 欄…オノマトペ

見出し表現中に使用されているオノマトペを片仮

名表記で与える。オノマトペからその共起表現を検索する際の指標として利用できる。

B 欄---見出し

オノマトペ (単体), オノマトペ共起表現ともに平仮名ベタ書きで見出しを与える。同音異義, 同音異機能オノマトペは原則として別見出しとする。例えば, 「ばらばら」は擬音と擬態で別見出し, 「こんこん」では, 擬音とは別に擬態の多義でも別見出しとする。

C 欄---分かち書き

オノマトペ共起表現に対し, その分かち書きを平仮名表記上にハイフン「-」で与える。分かち書き単位は, 単語, 接頭語, 接尾語, 接頭造語要素, 接尾造語要素とし, 活用語尾は形容動詞語尾「な」「に」「たる」「と」以外は切り離していない。造語要素は多くの場合, 音読みの一漢字であり, 例えば, 「緊張-感」の「感」は接尾造語要素である。複合語は基本的にアンダースコア「_」で切り離している。

D 欄---異表記

オノマトペ (単体), オノマトペ共起表現に対して, 漢字, カタカナなど, 異表記可能な部分には, C 欄の分かち書きの上で, 一種の正規表現によって選択肢を与える。例えば, 「ポッチャリ-と-した-(身)体_付き」という記載は「ポッチャリ-と-した-身体_付き」, 「ポッチャリ-と-した-体_付き」の2つの可能性, 「満(満/々)-たる-自信」は, 「満満-たる-自信」, 「満々-たる-自信」の2つの可能性を表す。ハイフンやアンダースコアで区切られた C, D 欄の記載から種々の表記を簡単に生成できる, 例えば, C 欄の「からだ_つき」と D 欄から得られた「身体_付き」, 「体_付き」から「からだつき」, 「身体つき」, 「身体付き」, 「からだ付き」, 「体付き」, 「体つき」の6通りの表記が得られる。

E 欄---構文的機能

採録しているオノマトペ (単体) は, 構文・形態上の機能により,

- 1 単純オノマトペ
- 2 連用オノマトペ (副詞的オノマトペ)
- 3 接頭オノマトペ
- 4 接尾オノマトペ
- 5 名詞性オノマトペ

に分類される。

1 は, 格助詞「と」を後接して連用修飾機能を持つものである。(注, 本レキシコンでは, 「ころっと」などの「と」を格助詞とみなしている。) 末尾促音型オノマトペの殆どは 1 に分類される。

2 は, そのままでも「と」を後接しても連用修飾機能を持つもの, 3, 4 は, 他語に接続して造語する機能を持つものである。表 1 に本レキシコンにおける 1~5 の分布と例を示す。E 欄には, 表 1 の記号が記載されている。1, 2 の機能は後述する H 欄の情報で詳細化される。いっぽう, 採録しているオノマトペ共起表現には,

- i 名詞句
- ii 動詞句
- iii 形容詞句
- iv 形容動詞 (語幹) 句
- v 連用修飾句
- vi 連体修飾句
- vii 名詞文形式

が有る。表 2 に本レキシコンにおける i ~ vii の分布と表現例を示す。オノマトペ共起表現の E 欄には表 2 の記号が記載されている。

F 欄---構文構造とギャップの可能性表示

オノマトペ共起表現に対して C 欄のハイフンによる分かち書きに基づき, 係り受け構造を修飾子, 被修飾子の対をカッコ[]で括って記載する。すなわち, 句 a の主辞が句 b の主辞を修飾して出来た句 aB の構造記述を a, B の構造記述 a, b を使って [ab] と

記載する。

要素単語の構造記述は、以下の通りとする。

- ・単純、連用、名詞性オノマトペ：O
- ・接頭オノマトペ：Op
- ・接尾オノマトペ：Os
- ・接頭語：P
- ・接尾語：S
- ・接頭造語要素：Q
- ・接尾造語要素：R
- ・名詞：N
- ・動詞：V（未然形V11,V12，連用形V22,V23，終止形V30，連体形V40，仮定形V50，命令形V60）
- ・形容詞：A（未然形A13，連用形A22，A23，終止形A30，連体形A40，仮定形A50，命令形A60）
- ・形容動詞（語幹）：K00
- ・副詞：D
- ・連体詞：T
- ・接続詞：C
- ・機能語及び機能性自立語：活用形も含め英小文字ローマ字綴り

文節内の語接続も便宜上、左2分岐句構造とみなして同様の記述を行っている。

例えば、オノマトペ共起表現「クンクン-と-犬-が-鳴く」の構造記述は「クンクン」=オノマトペ O，「と」=格助詞 to，「犬」=名詞 N，「が」=格助詞 ga，「鳴く」=動詞終止形 V30 であることから，[[Oto]*[[*Nga]*V30]] と記載する。図1にその意味する構文木と係り受け構造を示す。（注、この表現に対する2018/04/19におけるGoogl*の英訳結果は，“Cunkand dogs cry”であった。）

この例の如くF欄の構造記述内には適所にアスタリスク「*」が含まれており、その位置に直後の句の主辞に対する修飾句が入り得ることを意味する。従って、図1の構造記述 [[Oto]*[[*Nga]*V30]] は、「クンクン-と-朝-から-隣-の-犬-が-寂し-そう-に-鳴く」のような拡大表現の可能性を示している。図2にこの模様を示す。（注、この拡張表現に対する2018/04/19におけるGoogl*の英訳結果は，“Cunngun and morning

next dog crying lonesome”であった。）

JMWEL のこのようなギャップ付き構造記述は定形表現が部分的に持ち得る柔軟性を構文・意味解析機に反映させるためには非常に重要である。

G 欄…後方文脈条件

オノマトペ（単体）、オノマトペ共起表現に対し、文末側に呼応する語句がある場合にその情報を与える。例えば、「オチオチと」に対しては、文末側に否定句が要求されることを<negation>と記す。

H 欄…連体化、連用化、動詞化情報

オノマトペ（単体）に対し、E欄の構文機能情報を詳細化して与える。オノマトペを連体修飾、連用修飾に使用する場合と動詞化して使用する場合に必要となる後接語句を以下のように整理した。

- ・連体修飾：「な」、「の」、「たる」
- ・連用修飾：「に」、「と」、「ε」
- ・動詞化：「する」、「になる」、「とする」

εは空列を表し、オノマトペが後接語句なしで連用修飾できる事を表す。

例えば、「フラフラ」は、「フラフラの（姿）」で連体修飾、「ふらふらと（歩く）」、「ふらふら（歩く）」で連用修飾、「ふらふらする」、「フラフラになる」、「ふらふらとする」と動詞化すること、「ふっ」の場合は、「ふっと（気が付く）」で連用修飾する以外には考えにくいことを、それぞれ、後接する語句集合のトリプルによって{no}-{to, ε}-{suru, ninaru, tosuru}, Φ-{to}-Φ と記載する。Φは空集合を表す。トリプルのパターンは約100種である。

I 欄…擬音、擬態の別

オノマトペ（単体）に対し、擬音、擬態の別を「音」、「態」と記載する。

おわりに

「コロコロと音がする」、「コロコロ転がる」、「話がコロコロ変わる」、「目がコロコロと痛い」、「コロコロと笑う」などの共起表現によって「コロコロ」の多義が低減されるように、自然語文における語の意味は他語との共起によって相補的に定まる。従って、語の共起を如何に捉えるかが NLP における基本課題となるが、筆者は 1960 年代に始めた日英機械翻訳の研究を通して、語の共起を語類、意味属性レベルでルール化する事の現実的な難しさ、および、語の共起を表層レベルで辞書化しておくことの重要性を認識し、大規模フレーズレキシコン JMWEL の内省による開発を開始した。現在、JMWEL は見出し数が 10 万件(異なり)を超え、プロトタイプとして一定のレベルに近づいたと判断し、研究用、商用に公開している[13]。(注、2018 年 3 月、言語処理学会、言語資源協会(GSK)より 2018 言語資源賞を受賞[17].)

本稿で紹介したオノマトペ共起表現レキシコンは JMWEL のサブレキシコンであり、言語資源協会からも研究用として公開予定である。

本レキシコンが今後の機械翻訳、日本語処理、日本語教育・研究の発展に役立つことを期待したい。

参考文献

- [1] 高丸圭一,内田ゆず,乙武北斗,木村奏知.地方議会会議録におけるオノマトペの出現傾向に関する基礎的検討.言語処理学会第20回大会発表論文集,2014.
- [2] 井上音々,望月源. 日本語歌謡曲のオノマトペに関する調査. 言語処理学会第23回大会発表論文集, 2017.
- [3] 小宮嘉那子,佐々木稔,新納浩幸. 分散表現と文脈ベクトルによるオノマトペの分類の比較. 言語処理学会第22回大会発表論文集, 2016.
- [4] 福島弘識, 内田ゆず, 荒木健治. 2つの意味を持つオノマトペの意味判別における組成の検討. 言語処理学会第20回大会発表論文集, 2014.
- [5] 玉岡賀津雄, 木山幸子, 宮岡弥生. 新聞と小説のコーパスにおけるオノマトペと動詞の共起パターン. 言語研究139, 2011.
- [6] 乙武北斗, 内田ゆず, 高丸圭一, 木村奏知. 表層核に着目したオノマトペ共起語の抽出と分析. 言語処理学会第22回大会発表論文集, 2016.
- [7] 渡辺知恵美, 中村聡. オノマトペロリ: 味覚や食感を表すオノマトペによる料理レシピのランキング. 人工知能学会論文誌,30-1, 2015.
- [8] I. A. Sag, T. Baldwin, F. Bond, A. Copestake, D. Flickinger. A Pain in the Neck for NLP. Proc. of the 3rd CICLING, 2002.
- [9] R. Corrigan, E. A. Moravcsik, H. Ouali, K. Wheatley (eds.). Formulaic Language Vol.1, Distribution and historical change. John Benjamins Publishing Company, 2009.
- [10] D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan (eds.). Longman Grammar of Spoken and written English. Harlow: Pearson Education Limited, 1999.
- [11] C. J. Fillmore, J. Charles and P. Kay. A Construction Grammar Coursebook. Unpublished ms, University of California, Berkeley. 1995.
- [12] T. Tanabe, M. Takahashi, K. Shudo. A lexicon of multiword expressions for linguistically precise, wide-coverage natural language processing. Computer Speech and Language, 28-6, Elsevier, 2014.
- [13] <http://jefi.info>.
- [14] 小野正弘. 日本語オノマトペ辞典. 小学館, 2007.
- [15] 阿刀田稔子, 星野和子. 擬音語擬態語使い方辞典第2版. 創拓社出版, 2004.
- [16] "Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)". http://multiword.sourceforge.net/PHITE.php?sitesig=CONF&page=CONF_04_LAW-MWE-CxG_2018, COLING 2018.
- [17] <http://www.gsk.or.jp/event/nlp2018-言語資源賞/>

表1 採録したオノマトペの分布と例

Table 1 Morpho-syntactic categories, sizes and examples of onomatopoeias in the lexicon

種別	記号	見出し数	例
単純オノマトペ	O	1,576	ツルリ, ホワッ, ドッカーン, グネツ, ピューン, ゴロリ, ヒヤッ, ドブン
連用オノマトペ	AdvO	1,155	ドッカー, フラフラ, ミッチリ, フワフワ, チャリチャリ, ゴリゴリ
接頭オノマトペ	Op	140	ドタ, ジリ, グラ, ゴタ, ソヨ, ビリ
接尾オノマトペ	Os	30	タツプリ, タラタラ, モリモリ, ピカ
名詞性オノマトペ	NPO	25	ブツブツ, コリコリ, フリフリ, デコボコ
計		2,926	

表2 採録したオノマトペ共起表現の分布と例

Table 2 Syntactic categories, sizes and examples of onomatopoeic expressions in the lexicon

種別	記号	見出し数	表現例
名詞句	NP	3,649	サッパリと-し-た-性格, カリッと-し-た-口当り, ホクホク-し-た-食感, サラサラ-し-た-肌触り, ギリギリ-の-妥協
動詞句	VP	20,930	ドタッと-音-が-する, 鼻-先-に-人蔘-を-ブラ-(下/提)げる, 肌-が-バサバサ-に-乾く, 馬-が-ヒーン-と-嘶く, ニャーン-と-(ネコ/猫)-が-(鳴/啼)く, フツと-胸-に-浮かぶ, カツカツと-靴音-が-する, クラクラと-(眩暈/目眩)-が-する, キリキリ-痛む (深(深/々)/シンシン)-と-夜-が-(更/深)ける
形容詞句	AP	373	モチモチと-柔らかい, ポンポンと-威勢-が-良い, ガンガンと-痛い
形容動詞(語幹)句	AVP	183	愛嬌-タツプリ, フンワリと-柔らかか, ツンツンと-無愛想
連用修飾句	AdvP	3,708	キョトンと-して, ガラガラ-音-を-立て-て, 熟(熟/々)-思う-に, ワイノワイノと
連体修飾句	AdnP	2,180	シドロモドロ-の, グチョグチョ-し-た, ガチガチ-な
名詞文形式	NPS	13	英語-が-ペラペラ, 予定-が-ビッシリ, 収支-が-トントン
計		31,036	

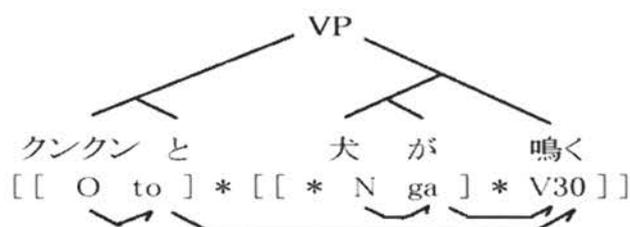


図1 オノマトペ共起表現「クンクンと-犬が-鳴く」の構造記述

Figure 1 Syntactic structure denoted by $[[Oto] * [[*Nga] * V30]]$ given to the verbal expression クンクンと-犬が-鳴く *kunkun-to-inu-ga-naku* "a dog whines". 注, 2018年4月19日における Google*オンライン英訳結果は "cunk and dogs cry"であった.

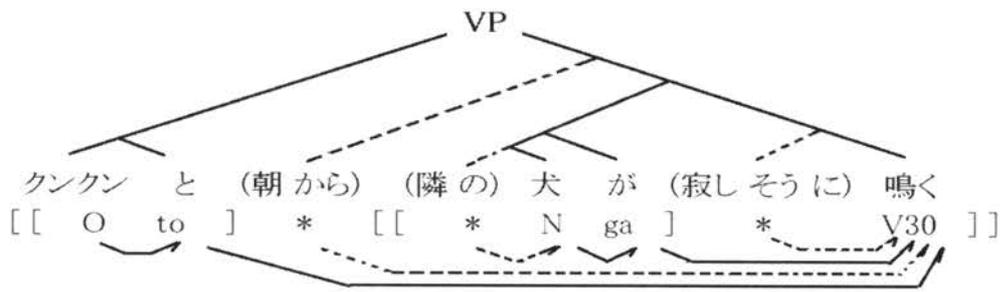


図2 「クンクンと朝から隣の犬が寂しそうに鳴く」の構造記述

Figure 2 Syntactic structure of expression クンクンと朝から隣の犬が寂しそうに鳴く
kunkun-to-asa-kara-tonari-no-inu-ga-sabisi-sou-ni-naku “a dog of the neighboring house whines sadly since the morning” derivable from the description $[[Oto] * [[*Nga] * V30]]$ given to the expression クンクンと犬が鳴く *kunkun-to-inu-ga-naku* “a dog whines”. 注, 2018年4月19日における Google*オンライン英訳結果は “Cunngun and morning next dog crying lonesome” であった.

翻訳ソフトとマクロによる多言語文書同時作成システムの試作

弁理士 保立 浩一

保立国際特許事務所

1. はじめに

企業活動のグローバル化に伴い、海外での特許取得活動の重要性が増している。各国は自国の言語で特許審査を行うから、当該国の言語で出願する必要があり、特許明細書の翻訳という作業が発生する。このため、多くの弁理士は、作成した特許明細書を英語等に翻訳する作業にも関与している。通常は翻訳業者に依頼して翻訳させるが、自ら翻訳する場合もある。

特許明細書の翻訳にはかなりのコストがかかっており、企業側には削減したいという強い要望がある。このため、機械翻訳（翻訳ソフト）を使用して特許明細書の翻訳を省力化する試みがかなり前から行われている[1]。筆者も、20年以上も前に、当時開発されていた機械翻訳ソフトを使用して特許明細書の英訳を試みたが、「本発明」を book invention と翻訳した時点でこれは使えないと思い、早々に断念した記憶がある。

実際、翻訳ソフトについてのユーザーの評価をネットなどで調べてみると、殆どが使えないという評価である[2][3]。短いありふれた文章の英日翻訳であれば翻訳ソフトはかろうじて使えるが、それ以外、特に特許明細書英訳のような専門技術的な文書の翻訳には全く使えない、というのが一般的な評価であろう。使えないというのは、翻訳ソフトで翻訳された英語を手直しして正しい翻訳とする手間と、翻訳ソフトを使わずに最初からマニュアルで翻訳した場合の手間が変わらないか、又は前者の方が大きな手間になってしまうということである。

それでも、短いありふれた文章であれば無料のウェブ翻訳サービスでもかなり高い精度の英訳ができるし、市販の翻訳ソフトでも、ユーザー辞書の登録や翻訳メ

モリを使いこなすことで、ある程度の翻訳品質が確保できる。

しかし、文章が長くなった際の、特に日英の翻訳では、構文解析に失敗するため、ほぼ100%の確率で誤訳となる。一例として、特許庁が提供する特許明細書のひな型の一文の翻訳結果を図1に示す。

「イメージ入力装置の中で、」は among image input devices と訳すべきといった構文以外の問題は別にして、無料ウェブ翻訳では、後半の so 以下の部分で構文解析に失敗している。A社製翻訳ソフトの場合、この部分は** is done...となっている。これは、「共」が「されている」と翻訳しており、「共」が辞書登録していないためである。「共する」の英訳として provide を登録しておくこと、it is provided to practical use as an input part of OCR for POS となり、完全な誤訳とは言えない内容に修正される。しかし、in the image input device が necessary information を修飾した状態となっており、この点のミスは致命的である。

(原文)

イメージ入力装置の中で、ハンドスキャナは、入力情報の記載された媒体の形状や媒体上の入力情報の位置を問わず、必要な情報のみを入力できる利点があるので、POS用のOCRの入力部として実用に共されている。



(無料ウェブ翻訳)

In the image input device, the hand scanner has an advantage of being able to input only the necessary information regardless of the shape of the medium in which the input information is described and the position of the input information on the medium, so the input of the OCR for POS It is shared with practical use as a department.

(A社製翻訳ソフト)

Because the hand scanner has the advantage that it can input only necessary information in the image input device regardless of the position of input information on shape and the medium, ** is done as an input part of OCR for POS by practical use.

図1 現状の機械翻訳の例

翻訳ソフトを使って特許明細書の英訳を行う場合、ユーザー辞書の登録を頻繁に行って翻訳品質を高めることが必須となるが、それでも構文解析のミス、修飾

関係の把握ミスがしばしば生じ、その都度、手作業で修正しなければならない。このため、最初からマニュアルで翻訳した方が早かった、という結果になるのである。

多くの翻訳ソフトでは、原文と正しい翻訳文とを対応づけて記憶する翻訳メモリの機能が備わっており、グレードの高い翻訳ソフトでは、一部をワイルドカード(変数)として登録する機能や、任意の一致率の設定を可能にして低い一致率の場合でもヒットするようにする機能も備わっている。このような機能は、同じような文章を翻訳する場合の多い技術マニュアルや取り扱い説明書等の翻訳にはある程度は有効かもしれないが、特許明細書のように以前とは異なる文章の翻訳が主である場合、翻訳メモリの有効性は限定的である。

2. 開発のヒントとコンセプト

このような状況ではあるものの、筆者は、特許明細書のような専門技術的な文書の英訳を翻訳ソフトの利用により大幅に省力化することを目差し、実用的なシステムの開発を意図した。この理由の一つは、近年における翻訳ソフトの機能向上には目覚ましいものがあり、「本発明」と book invention と翻訳していた時代とは隔世の感があるからである。特許明細書のような技術文書の場合でも、きちんとユーザー辞書を登録し、文章をできるだけ短くしていけば、驚くほど誤訳の少ない結果が得られるようになってきている。

もう一つの理由は、翻訳ソフトの拡張性の一つとして一般化しつつある Office 連携機能である。最近の翻訳ソフトは、Word や Excel といった Office 上のアプリのアドインとして利用できる機能が備わっており、Word 上で入力した文章を Word 上で翻訳したり、あるセルに入力した文章をそのまま Excel 上で翻訳したりすることが可能となっている。Office が使えるということは、マイクロソフト社が Office とともに提供しているマクロ言語である VBA(Visual Basic for Applications)が使えるということである。VBA が使

えれば、翻訳ソフトで足りない分を VBA で補うことができ、VBA 程度であれば、筆者のような専門のプログラマーではない者でも、何とか扱える。であれば、実用に耐え得る翻訳も行えるようになるのではないかと考えた次第である。

(1) 文章の各要素のセル化とセル毎の翻訳

短い文章なら精度の高い翻訳ができるのであれば、長い文章を幾つかの要素の分割し、要素毎に翻訳をし、翻訳後に各要素を結合すれば良い。イメージ的には、

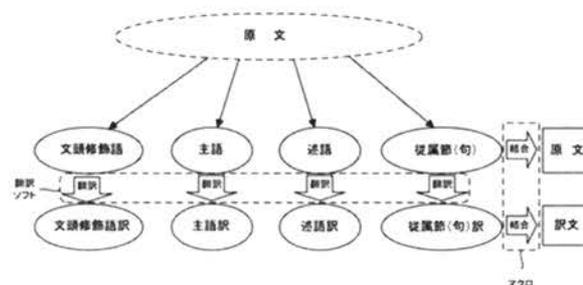


図2 要素毎の翻訳

図2のような感じになる。

要素毎の翻訳のメリットは、上述したような修飾関係のミスが基本的に生じないということである。要素内での修飾関係のミスは生じ得るが、要素をまたぐミス、つまり他の要素中の語を修飾する形の翻訳は絶対にされないので、上記のようなミスは生じ得ないということになる。つまり、「イメージ入力装置の中で、」を他の要素から切り離してそれだけで翻訳させ、文全体を修飾する文頭修飾句として後で結合すれば良いのである。

(2) 各要素の機能のユーザーによる指定

要素毎に分けて翻訳し、翻訳後にマクロで結合するとしても、結合に際しては当該要素の文章中での位置づけの情報が必要になる。つまり、その要素が主語なのか、述語なのか、文頭修飾なのか、それと従属節なのか、といった情報(以下、文内機能情報という)である。これがないと、マクロは、翻訳後の各要素の結合ができない。

文内機能情報は、翻訳をする者がマクロに対して与える必要がある。このためには、主語ならこれ、述語ならこれ、といったように各機能に対して識別記号を

設定し、識別記号をとともマクロに渡す構成が考えられる。しかし、その分だけ翻訳者の労力が増えてしまい、省力化の意図とは真逆になってしまう。これを回避するための構成として、各要素を個別の入力欄（テキストボックス）に入力するようにし、入力欄の識別情報（VBA で言えば「オブジェクト名」）を文内機能情報として扱うのである。イメージ的には、図3のような構成となる。

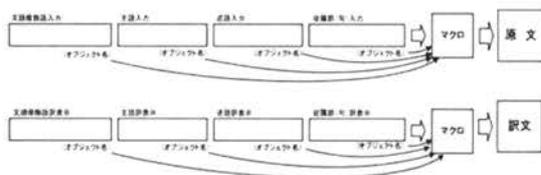


図3 マクロへの文内機能情報の渡し

ただ、この場合も、翻訳者は、各要素を各入力欄に入力していかなければならない。原語の文書の各文章について、各要素を各入力欄に copy & paste していくという作業が発生する。全てをマニュアルで翻訳する場合に比べれば、この部分は微々たる作業であるが、筆者は、ここでもう一段上のソリューションを意図した。

翻訳される原語（文章）も、元々は誰かが作っている。文章の作成者は、自らテキスト入力しているのであり、元々のテキストの入力者が各入力欄に入力するようにすれば、全体として作業量の増加はない。

つまり、日本語での文章の作成の際に最初から要素に分けて入力し、各要素を翻訳ソフトで英訳させた後、日本語、英語それぞれについてマクロで結合すれば良い。この場合、日本語の文章と英語の文章とが同時に出来上がる。つまり、日英文書同時作成である。

3. 構築したマクロ

上記のようなコンセプトの下、マクロを自作し、日英の多言語文書同時作成システムを試作した。

(1) 基本構成

このシステムは、基本的には Excel を利用しており、翻訳ソフトの Excel 連携機能と VBA で記述したマク

ロを追加している。図4に示すように、動作画面は原語（日本語）の入力フォームと訳語（英語）の表示フォームを含んでいる。図4の例は、基本文型としての S+V の文型で入力する例となっている。図4に示すように、原文の各要素の入力欄と訳文の各要素の表示欄が一对一で表示される画面レイアウトとなっている。



図4 基本構成

ユーザーは、自分が作成しようとしている文章の主語、述語、文頭修飾語、従属節（句）を頭に思い浮かべ、それぞれ入力欄に入力する。そして、翻訳ボタンをクリックすると、翻訳ソフトが起動し、各入力欄の日本語が翻訳ソフトにより翻訳され、その結果が訳語の表示フォームに表示される。

訳語の表示フォームは編集可能となっており、ユーザーは、翻訳ソフトの翻訳結果をマニュアル修正できる。適宜マニュアル修正をした後、結合ボタンをクリックすると、その時点で原語入力フォームで表示されている各要素がマクロにより結合されて原文が生成され、生成された原文が原文表示欄に表示されるとともに原語用のシートに記録される。同時に、訳語表示フォームに表示されている各要素がマクロにより結合されて訳文が生成され、生成された訳文が訳文表示欄に表示されるとともに訳語用のシートに記録される。

各文章の各要素の入力、要素毎の翻訳、各要素の結合による原文・訳文の生成、記録を逐次行い、全ての文章の生成、記録が終わったら、各シートの各行に記録されている文章を結合させることで、日本語と英語の双方で文書が出来上がる。つまり、日本語で文章を入力して文書を完成させると、同時に英語版の文書も

出来上がっているということである。

(2) 使用できる文型

基本型の他、複数述語や複数目的語といった他の各種文型でも入力、翻訳が可能となっている。文型の選択は、左側のメニューコラムに設けられた文型選択ボタンで行える。文型選択ボタンをクリックすると、選択可能な文型一覧が表示され、そこから一つを選ぶと、選んだ文型の原語入力フォーム、訳語表示フォームに切り替えられるようになっている。

(3) 要素単位の翻訳メモリの利用

今回試作したシステムで最も効果を発揮したのが、要素毎の翻訳メモリ登録である。翻訳メモリは、本来は一つの文全体として原文と訳文とを対にして登録するものであるが、このマクロでは、各要素の原語と訳語とを対応させて登録するのに使っている。例えば、図4の基本型の場合の訳語表示フォームで、文頭修飾語の下の翻訳メモリ登録ボタンをクリックすると、その時点で表示されている文頭修飾語の原語と訳語とが対となって翻訳メモリに登録される。前述した例で言えば、「イメージ入力装置の中で」の対訳として"among image input devices"を登録することになる。

同じような表現が繰り返されると想定される場合、その表現を最初に使用して訳語をマニュアル修正した際に修正後の訳を翻訳メモリ登録しておく、以後はマニュアル修正の必要はほぼなくなる。筆者がこのシステムを使用して実際に日英文書同時作成を試みたところ、要素単位の翻訳メモリ登録が省力化に非常に効果的であることが実感された。

(4) 翻訳の補助ツールとしての利用の考慮

日英同時作成の他、既に出来上がっている文書を英語にする、つまり翻訳業務を行う場合を想定した好適なソリューションも構築した。

このマクロを翻訳業務に使用する場合、原語入力フォームの各入力欄に各要素のテキストを入力する(copy & pasteする)という作業がどうしても伴う。これを少しで省略化するための構成を用意した。

翻訳業務に使う場合でも、ユーザーは文型を予め選

択する必要があるが、文型選択の際、文章中に簡単な識別子(区切り記号)を挿入することで原語入力フォームの各入力欄に自動的に貼付がされるようにシステムを構成した。図5に、区切り記号を挿入する画面の一例を示す。

ユーザーは、翻訳する文全体をテキストボックスにcopy & pasteする。そして、文頭修飾語を#bで囲み、主語を#sで囲み、述語を#vで囲み、従属節を#jで囲む。その上で、OKボタンをクリックすると、マクロにより、各要素が基本型の原語入力フォームの各入力



図5 文区切り入力

欄に自動的に貼り付けられる。

各要素のテキストをマニュアルで各入力欄に copy & paste するのがそう変わらない労力にも思えるが、マニュアルの場合、copy と paste でマウスの操作が2回必要である。上記各識別子の挿入の場合、1回で済む。

4. 評価

日本語での文書作成終了と同時に英語版も完成しているというのが理想型であるが、そのために負荷がどの程度増すかが問題となる。この点を確認するため、架空の特許出願依頼についてこのシステムを使って特許明細書の日英同時作成を試みたところ、全体としては20~30%程度の負荷アップであった。

この負荷アップをどうみるかは評価の分かれるところであるが、特許明細書の作成料の例えば30%で英語版の特許明細書が得られるとすれば、現状の英訳料

の相場感からすると、格安である。まだまだ改良の余地は多々あるが、筆者個人としては、このシステムのポテンシャルの大きさを実感した次第である。

5. 機械翻訳の次のテーマ（おわりに代えて）

当初は自らの仕事（翻訳）の省力化、効率化を目指した試みであったが、図らずも機械翻訳の分野にかなり首を突っ込んだ結果となってしまった。突っ込んだついでに、この分野の次のテーマについて少し考えてみたい。

周知のように、いま、この分野では、ニューラル機械翻訳（Neural Machine Translation, NMT）が席捲している。学会での発表テーマを見ても、NMT関連が殆どである。

しかし、個人的には、NMTには幾つかの深刻な限界があるように思われる。その一つは、クローズドなネットワークにはなじまないという限界である。NMTの特徴の一つとして、必要なのは対訳データのみであるというシンプルさが挙げられる。しかし、逆に言えば、対訳データしかないのであるから、精度の高い翻訳のためには膨大な量の対訳データが必要となる。十分な精度のNMTには数百万対の対訳データが必要と言われているし、グーグルは数億の対訳データを持っていると言われる[4]。ドメインを限定すれば比較的少ない対訳データでも流暢な翻訳が生成できるとの報告もあるが、それでも150万対が必要とされている[4]。

このような膨大な対訳データの集積は、クローズドな環境では不可能で、インターネットのようなオープンな環境が必須であると考えられる。つまり、ある会社が社内のネットワークに翻訳サーバを設け、そこにNMTを実装したとしても、対訳データが不足するため、NMTの特徴を活かした高品質な翻訳ができない可能性がある。かといって、社内のPCからグーグル翻訳のようなオープンなサーバに原語データを送って翻訳してもらうことも現実には難しい。社内のサーバ

で翻訳するのは、それが外に出したくない情報だからであり、そのような内部情報を翻訳するためにインターネット上のオープンなサーバに原語データを送ることはありえない。暗号化して送るとしても、サーバ上にデータが残るようでは、漏洩の問題がつきまとう。

NMTのもう一つの限界は、対訳データしか使用しないことによる別の限界である。対訳データしか参照しないということは、過去にあったこと（文）を参照して翻訳することしかできない、ということである。すなわち、新しい用語、新しい概念、新しい現象等を記述した文章を翻訳するときは、NMTは誤訳になり易いと推測される。例えば特許の世界では、パイオニア発明と呼び得るような発明を説明するとき、これまで使われたことのない用語を新たに作って概念を説明しなければならなかったり、これまでには知られていない新しい現象について説明しなければならなかったりすることがある。このような場合、NMTでは誤訳になり易いと推測される。

さらに、NMTが抱える問題としてしばしば指摘される点に、長文化した場合の誤訳の問題がある。実際、NMTを使用して多くの翻訳実験を試みたが、長文化すると必ずといっていいほど、構文解析に失敗したような誤訳となる。ピリオドもないのに途中で先頭文字が大文字になり、文を付け足したような訳になる。前述した図1の無料ウェブ翻訳の例もこの例である。

最初の二つの限界はそれを乗り越えるソリューションが幾つか想いつくが、最後の限界は、克服できないように思われる。というのは、この問題はNMTがいわゆるリカレントニューラルネットワークを行うことに起因しており、本質的と思われるからである。

尚、手前味噌になるが、本稿で提案されたシステムは、長い文章でも各要素に区分けして翻訳エンジンに渡すので、仮に翻訳エンジンがNMTサーバであったとしても、誤訳の発生は非常に少なくなる筈である。つまり、本稿で提案されたシステムは、NMTの欠点を補う目的でも使える。

NMTはさらに進歩を続け、いわゆる破壊的イノベ

ーションとして人間による翻訳をゼロにするという予測をする向きもあるだろう。しかし、ニューラルネットワークとはいっても、翻訳を行う人間の脳プロセスをそのままデジタルプロセスに置き換えている訳ではない。その意味で、NMTが翻訳者を不要すとか、NMTにより完全な機械翻訳が実現されるとかいったことにはならないと、個人的には予測する。

翻訳というのは、第一言語で書かれた文章を第二言語に変換するというだけでなく、その文章を書いた人が最初から第二言語で書いた場合を想定して文を構築する、という作業だと思う。リカレントニューラルネットワーク的なプロセスも局所的には存在するが、人間は、まず言いたいこと（書くこと）を決め、それを構成するブロックを一つ一つ考え、それを最適な位置に配置しながら文を生成する。だから、翻訳の場合も、その文章作成者の頭の中のブロック構造を学習して行うことが必要である。本稿でいえば、文章作成者がどの文型を意図し、原語の単語ないし単語群を文型のどのブロックに属させているのかを考える必要がある。

こう書くと、二世世代も前のRBM T(Rule Based Machine Translation)の時代に戻ったようであるが、構造解析をディープラーニングで正確に行うこと、これが機械翻訳の次のテーマであるという気もするのである。仮に、それが実現できれば、NMTの併用により完全無欠な機械翻訳の時代が到来するのかもしれない。

参考文献

- [1]熊野明. 特許明細書の日英機械翻訳における課題, Japio 2008 YEAR BOOK pp.88-92
- [2]<http://review.kakaku.com/review/newreview/Cate goryCD=0365/>
- [3] <https://www.translator.jp/soft/>
- [4] 永田昌明. ニューラル機械翻訳とは何か?, JTF Journal #288 2017 03/04 pp.10-11
- [5] 保立浩一. 翻訳ソフトとマクロによる多言語文書

手軽な英日機械翻訳

吉澤 忠義

1. はじめに

個人的に利用している手軽な US 式（上野・柴田）英日機械翻訳システム (US_EtranJ) について紹介します。この機械翻訳プログラムは、上野俊夫氏が 1986 年 4 月-6 月の「PC マガジン」誌上に連載された、パソコン英和機械翻訳プログラム (BASIC 言語: 500 ステップ位) を、元埼玉大学数学科教授柴田勝征氏が、3 年をかけて改良し、1990 年、著書「C 言語による英和翻訳システム」（ラッセル社）にまとめ発表しました。発表当時、中学の英語教科書の英文を訳すという規模で、非常にコンパクトなものでした。柴田氏は、「私はちょうど盆栽を育てるように、「US 式翻訳システム」をレベルアップするために、長い長い道程を着実に歩いて行くつもりです」と語っていました。私は、この翻訳プログラムに画面インタフェース等を付け Windows 版翻訳システムとして、現在、Yahoo ニュース (<https://www.yahoo.com/>) などの英文を翻訳させ、翻訳精度を高めることに趣味をもち、取り組んでいます。

2. 翻訳システムの仕組み

この機械翻訳の処理の流れは、下記に示すようになっています。

- ① 辞書引き…原文の語彙を辞書ファイルから検索し訳語候補を作成
- ② 適訳処理…訳語候補から品詞、訳語の適訳を選び出す
- ③ 文法処理…訳語並びを文法ルールに照らし合せ、訳語を結び付ける
- ④ 推敲処理…結び付けられた訳語の表現を見直す

それぞれの処理は、外付けの辞書、適訳ルール、文法ルール、推敲ルールのファイルから、翻訳知識を取り出し、英語原文の要素と変換情報要素とのマッチングを重ねながら変換して行きます。

このルールは、英文を単なる単語の羅列と見るのではなく、何らかの構造を持った集合と見なして、その構造を抽出し、これを「中間表現」に変換し、これを主要な補助手段として活用しながら翻訳文を生成するルールの総体となっています。この方法は、一種のトランスファー方式を規定しています。そして、翻訳知識 (ルール類) は、翻訳処理本体 (翻訳エンジン) から独立したものになっています。このことは、英文を翻訳しながら、上手く翻訳できないとき、翻訳エンジンを直すのではなく翻訳経過情報より、どのような翻訳知識 (ルール) が不足しているかを判断し、ルール類を補うことによって、翻訳精度を高めて行くことが出来る特徴を持っています。

翻訳経過情報は、語彙並びに対する適訳過程と翻訳 (結合) 過程でマッチングした適訳ルール、文法ルールテキストを順次ダンプされた情報と、そのルールによる処理結果情報の集まりです。

この翻訳知識 (ルール類) は、テキストタイプであり、エディタ等で編集、修正、追加を行っています。

現在の情報量は、次の通りです。

- ① 辞書…123,000 語彙
- ② 適訳…10,300 行
- ③ 文法…15,200 行
- ④ 推敲…600 行

この規模は、30 年で 10 倍に増やして来たのですが、他の機械翻訳システムと比べると、まだまだ不十分なものです。

3. 翻訳知識の拡充

翻訳知識の拡充は、翻訳経過情報を見て、英文をどのように訳したか、以下の点を調べます。

- ① 辞書登録情報が十分であるか
- ② 辞書引き後、語彙の適訳抽出が妥当かどうか
- ③ 訳語の結合順位が妥当であるかどうか
- ④ 補足的に推敲結果の訳語表現が適当か

これらの吟味は、翻訳処理で、フィットしたルールに落ち度がないか、抜けがないか、等吟味して辞書の補充、ルールの補正を繰り返し行います。

当然、ルールの変更、追加により、他の英文翻訳に及ぼす副作用も発生します。

4. 当面の翻訳知識拡充の目標

- ① 大規模な辞書構築…現在、Yahoo ニュース等の複数行の英文から、未知語が 20 から 30 語位見つかる、また、結合度の高い語彙の並び、熟語、固有名詞などを逐次登録していますが、これらの英文の翻訳に必要な語彙の数は計り知れないものと思われまます。
- ② 適訳ルールの拡充…英語と日本語の語彙の対応は 1 対 1 ではなく、普通は複数の選択肢があり、それらの中から正しい訳語をどの様を選び出すかが、大きなポイントになります。現在、辞書に登録されている語彙に品詞と簡単な属性を付け、この属性を手掛かりに前後の語彙の並びから訳語を選択するルールを与えていますが、十分に絞り込めているとは言えません。上手く確率論と結びつく解決方法が必要かもしれません。
- ③ 文法ルール…このトランスファー方式では、原文の構文木を下の方から作って行く「ボトムアップ」方式です。構文木をトップまで積み重ねてゆきたどり着いたものが翻訳結果となります。この構文木の構成をルールとして表しているものが文法ルールです。実際の英文では、多くの例外記

述があります。これらの記述に合わせたルールを補完して行く必要があります。

AI の時代といわれ、膨大な知識ベースを元に物事を分析し、人間に必要な指針を示すシステムの進化が急速に進んできています。しかし、機械翻訳の世界では、無限の宇宙に繰り広げられる言語表現をどうとらえ、認識し伝えていくか難しい問題を抱えています。

人間は、3 歳の子供が言葉を覚え、語彙を増やし、コミュニケーション能力を高めて行きます。

そのメカニズムを機械が実現できるか、機械翻訳が真に発達するために必要な技術がそこにあるような気がします。そして、自分で学習して賢くなるためには、異次元の方法論が必要となると思います。US 式には、これらの機能はないが、翻訳知識拡充を繰り返し、教えた範囲では賢くなるので、どこまで翻訳性能が高められるかを楽しんでいます。

5. 翻訳システムの性格

機械翻訳の結果は、そのシステムを手掛けた人の性格に依存します。どうゆう英文をどれだけ訳し、誰が翻訳知識を教え込んだかによります。

機械といえども俗人性が高く、間違いも多いのだと思います。

次の例は、Yahoo ニュースにあった記事の一部です。

- ① 英文 : Polar bears crowd on Russian island in sign of Arctic change

訳 : 北極グマは、北極の変化の兆候でロシアの島に集まります…「in sign of」前置詞として扱う

- ② 英文 : I read the book indicating the cold case with the notion that Kate might be the criminal.

訳 : 私は、ケイトが犯人であるかもしれないと思われる迷宮入り事件を示している本を読みました。

…「with the notion that」を前置詞「>と思われる」として扱う

上記の例では、一般辞書にない語彙の並びを「前置

詞句」として扱っています。

6. 翻訳システムの評価

先日、ある翻訳会社の商品評価の担当者に「ポツダム宣言」の英文を翻訳したものを見ていただきました。

「全体の印象としては、日暮れてなお道遠しの感があります。」と評価されました。

翻訳会社の人々の間では、機械翻訳については、まだまだ使い物にならないと感じているようです。

また、顧客からも、どうゆう人たちが携わっているのが翻訳依頼の判断になっているようです。

7. 翻訳構文図の描画

US_EtranJ の翻訳方式は、句構造木を用いる独特のトランスファー方式です。この句構造木は、適訳処理で選択された語彙の集まりを、順次文法ルールにマッチングさせ新しい語彙の集まりに昇華させてゆき、最終的に、ひとつのものに結合されたときに翻訳が終了したと判断しています。この過程の状態を線で結んで表すと、一種の構文木として表せます。下図は、先の英文を翻訳して作成された構文木の図です。



また、別紙の記事は、平昌オリンピックで女子 500メートルスピード競技の Yahoo ニュースの一部です。オリンピック中継の TV 画面から受けた感動が表現されているでしょうか。言語表現は、人間が受けた印象や感動を言葉によって表現されたものであると思います。

また、添付されている構文図は、英文の網掛けされた部分を翻訳した時のものです。

機械翻訳ツールで、このような構文図を表示する機能を持っているのも、US_EtranJ の特徴です。

この図は、翻訳過程の誤り等を図から判断でき、その誤りを翻訳知識ベースに反映しやすいことです。

また、ある学生が、この図を見て、「翻訳ってこおやっするの！」と驚き表情を見せてくれたのも思わぬ副産物でした。

参考文献

ラッセルブック「パーソナルコンピュータによる機械翻訳プログラムの制作」上野俊夫 著 1986年 ラッセル社

ラッセルブック「C言語による英和翻訳システム」柴田勝征著 1989年 ラッセル社

AAMT 会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第二十三回を迎えることができました。今号では、個人会員一名、法人会員一社からのご寄稿をいただいております。

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

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

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

法人会員（敬称略・50 音順）

会員名

- インフォトランス株式会社 / InfoTrans Co., Ltd.
- <http://www.infotrans.jp>
- 【本社】〒162-0803 東京都新宿区赤城下町 30 番地 パールモア一交 3F
- 【ヨーロッパ本社】 InfoTrans B.V. Concourslaan 48 2132 DK Hoofddorp The Netherlands

自己紹介

当社は OA 機器や産業機械、医療機器などの製品に付属するマニュアルの多言語版を制作することを主な業務としております。クライアントより提供された英語または日本語のマニュアルを翻訳、DTP、データ編集等のローカライズを行うことをメインとし、その他ローカライズに関連する業務も請け負っております。オランダで 2001 年より業務を開始しましたが現在は日本国内にも本社をおき、オランダ側と協同しながら営業しています。社名は情報(Information)と翻訳(Translation)を合わせたもので、情報化社会において加速しつづけるグローバルイゼーションに対応したサービスを提供していきたいという目標を示しています。

創業してより世界各地の翻訳者とネットワークを構築し、高品質なサービスの提供を心がけてまいりました。また一つのプロジェクトをオランダサイドと日本サイドで分担することによる迅速かつ柔軟な対応をクライアントより評価いただいております。

MT/翻訳とのかかわり

MT および翻訳業界に期待すること

当社の設立時はクライアントの中で徐々に CAT ツールの存在が知られていくようになった段階でした。昔ながらのフローを好まれるクライアントもいる一方で、積極的に新しいフローを導入しようとするクライアントも現れるようになってきていました。我々もそうした流れに対応すべく設立当初より CAT ツールのさまざまな特質（クセ）を見極め、作業フローに組み込んでまいりました。結果として業務フローの大部分が CAT ツールの特質に依存するものになっていったのは皮肉なことです。それぞれのクライアントが持つ要求事項を踏まえ、種々の CAT ツールや周辺ツールの新規開発および組み合わせにもとづくフローの最適化を提案させていただいてきました。

近年 MT という存在が、そうした積極的に新しいフローを導入しようとするクライアントにとって非常に魅力的な「次のステージ」として受け取られている、ということを実感する機会に接することが多くなりました。ただ、CAT ツールが広まっていった時よりもさらに、過大ともとれる期待を MT に対してクライアントが抱いているようにも感じています。その結果として、最初はバラ色の夢を MT に抱いていたクライアントが、さまざまな検証を進めていく中で現実との乖離に直面するといったケースも現れています。当社もそうしたクライアントの相談に対して応えられるように種々の検証結果を積み重ねているという段階です。

とはいえ、過度な期待をはらみつつも MT の進歩スピードは非常に速いものであり、現在のクライアントがもつ期待を一部ではあれ満たしうる存在ともなりつつあります。また社内におきましても、既存翻訳のストックをより活用しうる可能性など MT には期待している部分があります。一方で従来のサービスにおいて品質を決定づけるプレイヤーが翻訳者であったとすれば、MT による彼らの業務内容の変化は CAT ツール導入時以上ともなりえます。したがってその品質をどのように維持・向上していくのか、随時変化していく情報をきちんと収集・分析していきたいと考えています。

AAMT への要望

MT に関してさまざまな形で情報が発信されていますが、単独での情報収集には限界があります。ですのでそうした種々の情報の集約箇所として AAMT が機能され続けることを期待しています。

AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第二十三回を迎えることができました。今号では、法人会員一社、個人会員一名の皆さまからのご寄稿をいただいております。

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

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

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

個人会員（敬称略・50 音順）

会員名

大塚英

MT ユーザー / フリーランス翻訳者（特許） / 翻訳スクール講師

自己紹介

オランダ、英国の大学で生物学研究に従事した後、フリーランス翻訳者になりました。主に医薬・化学分野の特許明細書を翻訳しています。翻訳校閲を行っていた時期を境に、その前は英文和訳、その後は和文英訳の仕事を受けています。翻訳スクールの講師になったのも校閲を行っていたときです。

スクールの受講生がその後どのようにキャリア形成していくのがよいか考えています。現在起こりつつある大きな技術革新・社会変化にどのように対応すべきかに特に興味があります。新たな技術を積極的に取り入れ、利用の仕方を模索・学習しようと思い AAMT の会員になりました。

MT/翻訳とのかかわり

翻訳者になる前の学習段階から現在まで一貫してルールベース機械翻訳（RBMT）ソフトウェアを用いています。

訳抜けなどのミス無くすために利用を始めました。翻訳を対訳形式で行い、RBMT の出力の編集を、削除ではなく置換で行うようにすることで訳抜けを防止しています。用語や訳出順序、言い回しなどを文書全体で統一するのも RBMT は役立っています。私自身が前にどう訳したかを覚えておくことはできませんので、RBMT がそれらを記憶し、確実に再現してくれるのには助かります。「機械」に期待する働きの第一は、このように自分（ひと）がおぼつかないところを確

実にする事です。

RBMT は、翻訳に要する時間の短縮にも役立っています。人手翻訳の律速段階はテキスト入力と調査ですが、RBMT の出力を用いることでテキスト入力の手間が減り、以前に自分が調査し決定した訳語が記憶されているので調査の手間も減ります。

一方で、編集にかなりの手間がかかるので時間短縮効果は限定的です。最終訳文に利用できる文字列が、RBMT 出力中で断片化されていると、その度合いに応じて編集の手間が増えます。ニューラル機械翻訳 (NMT) では、この断片化の度合いが減っていると考えられますので、この点において大きな時間短縮効果が得られるものと期待しています。

MT および翻訳業界に期待すること

翻訳者が利用することを考えると、NMT にはいくつかの問題があるようです。そのような問題が解決されることを期待しています。

訳抜けや訳ぶれの問題が深刻だと考えます。そのような誤りがどこにあるのか分かりにくいのも問題です。99%正しい訳文でも 1%の誤りがどこにあるのか分からなければそれを見つけるのに大きな手間が生じ、重大な誤りを見落としてしまう可能性もあります。誤りを見つけるための方法または誤りが生じないようにする方法を求めています。

訳語 (句、節、文) を指定できないのも困ります。特に、自分が決定した訳語 (表現) を以降に反映させられるようになると助かります。

様々な手段を統合することで、自分の作業 (思考) プロセスに沿った自動処理を行い、正しいと判断できる訳文を短時間で完成できるようにしたいと考えています。

MT に限らず、機械学習を取り入れた自動処理をひとの作業 (思考) プロセスに沿うように組み込んだシステムの構築方法を学びたいと思います。

翻訳会社が MT を利用する場合も、訳文を完成させるひとにとって作業がしやすいような、ひとを中心としたシステム (プロセス) となることを願っています。

AAMT への要望

AAMT ジャーナルや交流会で多くのことを学んでいます。例えば、NMT の欠点克服を目指した技術開発が進んでいること、前処理や後処理で用語等の適用・チェックを行う方法など、気になっていたことについて知ることができました。今後も技術開発の動向や MT の利用方法などを知ることのできる場を提供していただければと思っています。

協会活動報告

(2018年3月～2018年4月)

機械翻訳課題調査委員会

2018年3月9日(2017年度 第11回)

① 各WGの活動について(各WGに分かれて議論)

(WG1、WG2)

- ・言語処理学会発表について
発表スライドのレビュー
発表ポイントの確認
- ・AMTA発表について
英語サイトの動作確認
英語表記の確認&修正依頼内容まとめ
発表スライドレビュー
- ・18年度計画検討
成果と計画 報告文章案の確認
予算執行計画
- ・機械翻訳フェア
講演者
展示会

② 全体会議

- ・各WGの報告
- ・AAMT新体制
- ・その他

2018年4月13日(2018年度 第1回)

① 各WGの活動について(各WGに分かれて議論)

(WG1、WG2)

- ・言語処理学会発表の報告
- ・AMTA発表の報告
- ・評価サイト英語化について
- ・18年度計画検討
アンケートはプッシュ型の方法で、もっと多くの人にメール出す。

申込と同時にアンケート依頼を送る（リマインドメール、後でも）

アンケートの内容を決める（今年度共通の設問）

目標時期 5月下旬（翻訳フェアで配りたい）

・機械翻訳フェア

講演者

展示

(WG3)

・変換ツール

今後の機能追加案

ユーザー辞書が使えるサービスを調べてみる

MemSource の仕様調査と UTX との対応検討

UTX から MemSource に変換するのをまず作ってみたい

Google 翻訳者ツールキットの用語集形

インターネットWG

- ① AAMT ホームページの更新

編集委員会

- ① 2018年4月6日（2018年度 第1回）

No.68(2018年6月発行)の原稿依頼、編集作業を行った。

AAMT/Japio 特許翻訳研究会

2018年3月2日(金)(2017年度 第7回)

1. 議事録の確認
2. 「原言語側の欠落を考慮した Multi-Source NMT」発表
(奈良先端科学技術大学院大学 須藤克仁先生)
3. 来年度のシンポジウムについて
4. 次回の開催について
 - ・今後の研究会・拡大評価部会開催の日時予定(場所)
 - ・主な議題

AAMT/Japio 特許翻訳研究会 拡大評価部会

2018年3月2日(金)(2017年度 第3回)

1. 各グループの活動発表
 - ・自動評価グループ
 - ・人手評価グループ
 - ・テストセットグループ
2. その他
3. 次回開催予定

AAMT ジャーナル編集委員会委員長
筑波大学システム情報系知能機能工学域
宇津呂 武仁

AAMT ジャーナル 68 号をお送りします。

今号の巻頭言は、AAMT 会長の名古屋大学中岩浩巳先生より、御寄稿を頂きました。

研究報告として、江原暉将先生より「科学技術情報の機械翻訳」について、首藤公昭先生より「日本語オノマトペ共起表現レキシコン」について、弁理士の保立浩一様より「翻訳ソフトとマクロによる多言語文書同時作成システム」について、吉澤忠義様より「手軽な英日機械翻訳」について、それぞれ御寄稿頂きました。

一方、今号に先立ちまして、2月15日(木)に機械翻訳の最新動向の把握や開発者とユーザの交流を目的とした「AAMT 交流会」が開催されました。クロスランゲージの葉文君様からは、この会の参加報告を御寄稿頂きました。その他、イベント報告として、中岩浩巳会長より、本年3月に開催された米州機械翻訳協会 (AMTA) 主催の国際会議 AMTA2018 の参加報告を御寄稿頂きました。また、XTM International Ltd の目次由美子様からは、本年3月8日(木)に開催された2017年度 JTF 第6回翻訳セミナーの参加報告を御寄稿頂きました。

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

AAMT

Asia-Pacific Association for Machine Translation

AAMT 入会のご案内

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

機械翻訳にご関心のあるすべての方にご入会をお勧めします。

** AAMT 会員の特典 **

1. AAMT Journal の購読ができます。

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

2. 機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機械翻訳関連の情報をお届けします。

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

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

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

4. 関連機関の主催する国際会議に参加できます。

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

AMTA* : Association for Machine Translation in the Americas

EAMT** : European Association for Machine Translation

年会費は以下の通りです。

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

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

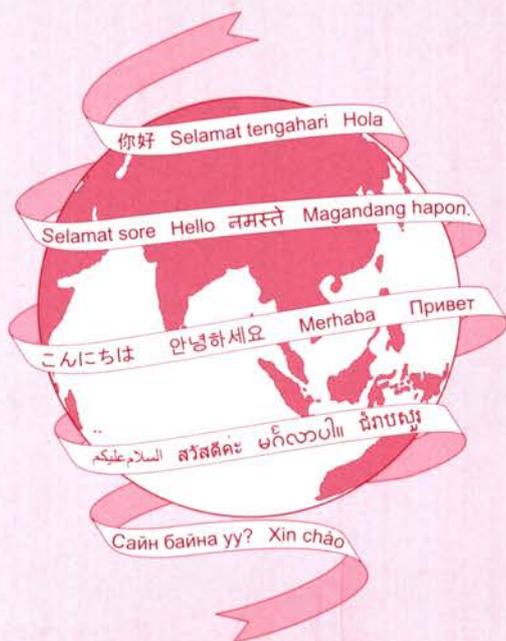
ご関心のある方は、事務局までお問い合わせください。

アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

電子メール：aamt-info@aamt.info

AAMT



AAMTジャーナル No.68

発行：アジア太平洋機械翻訳協会（AAMT）

ホームページ：<http://www.aamt.info>

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

phone：03-5951-3961 fax：03-5951-3966

編集委員会：宇津呂 武仁 小谷 克則 大倉 清司

阿部 さつき 釜谷 聡史 河野 弘毅

表紙(図部分)デザイン：阿部 さつき

事務局：神崎 享子 荻野 孝野

印刷所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)

c/o Japan System Application Co., Ltd.

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

aamt-info@aamt.info

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