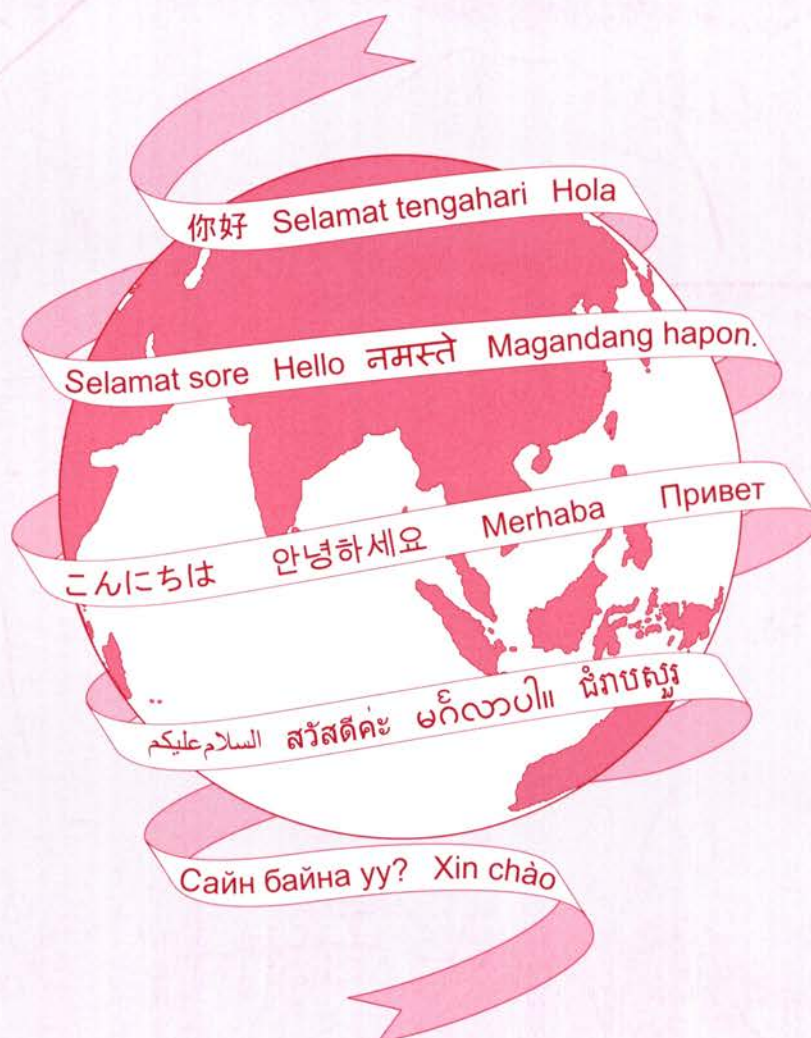


AAMT

Asia-Pacific Association for Machine Translation

Journal



June 2019

No.70

アジア太平洋機械翻訳協会

目 次

巻頭言：AAMT 初代会長の長尾真先生の文化勲章ご受章を祝して	辻井 潤一、井佐原 均、中岩 浩巳、隅田 英一郎	1
イベント紹介：第2回自動翻訳シンポジウム～自動翻訳と翻訳バンク～報告	内山 将夫	4
参加報告：2019年2月AAMTセミナー セッション2 参加報告	目次 由美子	10
シンポジウム報告：第5回特許情報シンポジウム 開催報告	須藤 克仁	12
シンポジウム報告：IWSLT 2018 参加報告	須藤 克仁	15
研究報告：ユーザーから見た NMT の使い勝手と活用の展望	浜口 宗武、宮城 三次、宮本 伸也、湯浅 豊裕、平林 千春	17
参加報告：LocWorld38 Seattle 参加報告	目次 由美子	26
参加報告：Women in Localization Japan 第16回イベント報告	目次 由美子	28
シンポジウム参加報告：平成30年度産業日本語研究会・ワークショップ参加報告	小谷 克則	31
研究会紹介：産業日本語研究会の取組について	清藤 弘晃	35
サービス・製品紹介：ニューラル機械翻訳を活かす翻訳支援ツール GreenT	新田 順也	36
会員の広場（個人会員）	藤崎 恵介	40
会員の広場（法人会員）	日本特許翻訳株式会社	42
事務局からのお知らせ：協会活動報告（2018年9月～2019年5月）	AAMT 事務局	44
編集後記	宇津呂 武仁	52

CONTENT

Foreword: Congratulations on the acceptance of the Order of Culture by Dr. Makoto Nagao, the first president of AAMT	<i>J. Tsujii, H. Isahara, H. Nakaiwa, E. Sumita</i>	1
Report: Report of 2nd Machine Translation Symposium -- Machine Translation and	<i>M. Utiyama</i>	4
Report: Report on Session 2 at AAMT Seminar in February 2019	<i>Y. Metsugi</i>	10
Report: The 5th Symposium on Patent Information Processing	<i>K. Sudoh</i>	12
Report: IWSLT 2018	<i>K. Sudoh</i>	15
Report: USABILITY AND FUTURE PROSPECTS OF NEURAL MACHINE TRANSLATION – From the viewpoints of users –	<i>M. Hamaguchi, S. Miyagi, S. Miyamoto, T. Yuasa, C. Hirabayashi</i>	17
Report: Report on LocWorld38 Seattle	<i>Y. Metsugi</i>	26
Report: Report on the 16th Women in Localization Japan event	<i>Y. Metsugi</i>	28
Report: Report on the 30th Meeting and Workshop of the Technical Japanese Association	<i>K. Kotani</i>	31
Report: The efforts of Technical Japanese Association	<i>H. Kiyoto</i>	35
Products :GreenT: CAT Tool for Leveraging Neural Machine Translation	<i>J. Nitta</i>	36
AAMT Members:	<i>K. Fujisaki</i>	40
AAMT Members:	<i>NPAT</i>	42
AAMT Activities: AAMT Activities (from September 2018 to May 2019)		44
Editor's Note:	<i>T. Utsuro</i>	52

産業技術総合研究所 辻井 潤一
豊橋技術科学大学 井佐原 均
名古屋大学 中岩 浩巳
情報通信研究機構 隅田 英一郎

1. はじめに

機械翻訳の研究開発をその黎明期から主導され、当協会、アジア太平洋機械翻訳協会（AAMT）を創設された^①長尾真初代会長が、2018年11月3日に文化勲章を受章されました。

ここでは、歴代会長が長尾先生のご功績を振り返りながらご受章をお祝いいたします。

2. お祝いの言葉（辻井潤一）

私は、1971年から1988年まで、長尾先生の指導のもとに言語処理研究に従事してきました。その間に実施された、先生をリーダーとし、私も参加させていただいた機械翻訳の国家プロジェクト（Muプロジェクト、1982–1986年）は、技術だけでなく、日本の機械翻訳を遂行する人材や社会的な基盤を作ったという意味でも、画期的なものでした。

この人がいなければ、ある分野の歴史が大きく変わっていたと思うような研究者は、科学技術の全分野を見渡しても、また、国際的に見ても、それほど多くおられるものではありません。先生は、機械翻訳という技術分野において、そのような人であったと思います。よく知られているように、1965年のALPACレポートの後、機械翻訳技術に対する懐疑的な見方が蔓延し、その技術開発が世界的に放棄された時期が続いていました。その状況を打ち破ったのが、前述のMuプロジェクトでした。このプロジェクトに刺激され、ヨーロッパのEurotraプロジェクトが開始されます。また、米国は、日本・ヨーロッパの動きに刺激されると同時に、この2つのプロジェクトとは違った発想で翻訳に取り組みをはじめ、統計的機械翻訳という流れを作り出します。この流れに先鞭をつけたものの一つに、長尾先生の「例に基づく翻訳」という考え方がありまし

た。これは大量の翻訳コーパスを基盤とする経験主義的な機械翻訳の流れとなり、現在のニューラル翻訳へとつながっています。

また、先生は、機械翻訳の実用化には、ユーザである翻訳家、開発運用を行う民間企業、研究資金を提供する政府機関の関与が不可欠であるという強い思いから、1987年に、日本主導で第1回のMTサミットを箱根で開催されます。このサミットを契機に日本機械翻訳協会¹が設立され、それに続く形でヨーロッパ機械翻訳協会（EAMT）、アメリカ機械翻訳協会（AMTA）が設立され、現在に至っています。

機械翻訳は、それを支える組織まで含めて、日本が国際的な流れを作った稀有な例の一つですが、それを一人でけん引されたのが先生でありました。先生の数多い学術への貢献が今回の文化勲章のご受章につながったものと思いますが、機械翻訳への貢献一つをとっても、十分にご受章の理由になると思っています。

先生、ご受章、おめでとうございます。

3. お祝いの言葉（井佐原均）

私は学部3年生だった1975年に初めて長尾先生の教授室にお伺いして以来、長年にわたって自然言語処理の研究について、ご指導いただけてきました。当時の長尾研究室には文系の先生方も出入りしておられ、さらには国際会議のお手伝いや、海外の研究者のお相手など、多くの経験をさせていただきました。私が電子技術総合研究所を経て、通信総合研究所（現、情報通信研究機構）に異動し、研究グループを立ち上げた時に文系の研究者を多く受け入れたこと、その頃から現在まで続く中国の多くの研究者との交流やタイでのラ

¹ 1991年に「日本機械翻訳協会」として設立、1992年に「アジア太平洋機械翻訳協会」と名称変更。

ボの設立などは、このような学生の時の経験が元になっております。

機械翻訳に関する長尾先生のご貢献につきましては、辻井先生が述べておられる通りですが、個人的には機械翻訳国際連盟 (IAMT) が立ち上がった 1991 年の第 3 回 MT サミット や、それに続く北米視察にご一緒させていただいたことが懐かしく思い出されます。

今回の受章は情報工学分野では初めての受章ですが、社会に使ってもらえることを重要視して研究に取り組んでこられた長尾先生の受章が、機械翻訳の社会実装が進む今であることに巡りあわせを感じます。

長尾先生の素晴らしいところは学術のみではなく、その基盤となっている考え方にあると思います。私自身が十分に理解できているかは分かりませんが、何か日本文化といったものが根底にあるように感じております。現在、私が所属する大学においても、長尾先生の書かれたものに感動した、長尾先生のお話を聞くと元気が出る、と語る教員たちが居りますが、専門分野に限らず感銘を与えるのは、このような基盤があつてのことと思っております。

これからも変わらず、我々に指針を与えてくださいますよう、お願い致します。長尾先生、文化勲章の受章、おめでとうございます。

4. お祝いの言葉 (中岩浩巳)

私は、大学院修了後 NTT に入社し、機械翻訳の研究をスタートしました。まず機械翻訳の基礎を学ぶために読んだ機械翻訳入門書が長尾先生の書かれた「機械翻訳はどこまで可能か」⁽²⁾でした。当時は、ルールベース翻訳方式が全盛で、NTT もルールベースの翻訳システムを研究開発しておりましたが、そのシステムで構築されてきた機械翻訳用の大規模辞書データを、長尾先生に評価していただき、是非公開すべきであるとの強いご助言があつたのがきっかけで、岩波書店より「日本語語彙大系」を出版することが出来ました。企業が開発した機械翻訳システムの辞書情報を書籍出版するということは、当時画期的なことでした。その後の私にとってこの出版の実現は、機械翻訳に関わって

いくキャリアにとって大変重要なものとなりました。

長尾先生は、Mu プロジェクト、MT サミット、機械翻訳協会と、日本を中心とした機械翻訳の国際的活動の流れを作ってこられました。長尾先生のご尽力があつたからこそ、日本発の機械翻訳の国際連携の流れが生み出されたものと思います。また、私が会長を就任していた 2012 年から 2018 年の間では、機械翻訳の主要ユーザの 1 つである翻訳業界との連携を強化してまいりましたが、その際にも、長尾先生からは温かいサポートをいただきました。そのご助力もあり、2017 年に日本での 24 年ぶりの開催となりました MT サミット XVI では、機械翻訳業界だけではなく翻訳業界の方々からもご協力いただき、成功裏に会を終えることが出来ました。

長尾先生の文化勲章受章は、AAMT 会員のみならず、機械翻訳や自然言語処理の研究開発、また、その活用などにかかわる多くの方々にとって励みになるとともに大変喜ばしいこととなりました。これからも、お元気で我々をご指導いただければ幸いです。長尾先生、ご受章おめでとうございます。

5. お祝いの言葉 (隅田英一郎)

私は Mu プロジェクトのルールベース翻訳方式の日英論文翻訳システム開発に携わらせていただきました。Mu プロジェクトの成果であるプログラムと辞書と技術と経験は、研究者が各社に持ち帰り、自社製品の開発に役立て、多数の商品が生まれました。

長尾先生はルールベースの Mu プロジェクトを指揮しておられましたが、ルールベースと全く異なる「例に基づく翻訳」というアイデアを少し早い時期の 1981 年に発表されました。これは現在の経験主義的な（統計的機械翻訳やニューラル翻訳等の）手法の魁になります。1989 年から、私は、長尾先生がテーマ設定に深く関わられた国際電気通信基礎技術研究所で、音声翻訳の国家プロジェクトに従事することになりました。そこで、私は「例に基づく翻訳」にのめりこみ、ルールベースでは翻訳出来ないことで有名な『日本語の「A の B」という形の名詞句を英語に翻訳する問題』を「例

に基づく翻訳」で解けることを発表し長尾先生にお認めいただき、その後の研究の糧となりました。「例に基づく翻訳」は世界に広がり、統計的機械翻訳の手法と融合して発展しました。

しかしながら、他の言語対に比べると翻訳が難しい日本語と英語の対においては、やがて精度改善速度が鈍り、天井にぶつかりました。深層学習によるニューラル翻訳が登場し、蓄積されてきた対訳データを用いて、一気に先の天井を打ち破る著しい精度改善を示しました。ニューラル翻訳は燎原の火となり瞬く間に全世界に広がりました。儼然なかった日英の機械翻訳も十分に高精度化し、AAMTにおいて、研究者・開発者・利用者が『活発に緊密に』連携する時代を迎え、AAMTの重要性が一気に高まりました。一文単位の翻訳は非常に高精度化できましたので、さらに、文を越えた情報の利活用の研究として、主語の補完等の所謂文脈処理技術や文外の情報を扱うマルチモーダル技術を取り込んだ異次元の機械翻訳も実現の兆しを見せています。

長尾先生の文化勲章受章は、機械翻訳の研究者・開発者・利用者のフォーラムである AAMT にとって、大変喜ばしいことでした。長尾先生、文化勲章のご受章、おめでとうございます。これからも、長尾先生には末永くご指導いただき、機械翻訳の進化をお楽しみいただきたいと思います。

6. おわりに

以上のように長尾先生のご業績は広く深いです。

歴代会長一同は、初期から機械翻訳の研究をリードされ、ヴィジョンをもって AAMT を創設された長尾先生のお仕事に感謝し、受章を心よりお祝い申し上げます。

参考文献

- (1) 「情報を読む力、学問する心」長尾 真、318 ページ、ミネルヴァ書房 (2010)
- (2) 「機械翻訳はどこまで可能か」長尾 真、157 ページ、岩波書店 (1986)

第2回自動翻訳シンポジウム～自動翻訳と翻訳バンク～報告

内山将夫

情報通信研究機構

1. はじめに

当シンポジウムは、2019年3月6日にコングレスクエア日本橋（東京）で開催されました。会場は満杯で、非常に盛況でした。

主催者挨拶では、総務省総務大臣政務官の國重氏より、政府として多言語対応を進めるために、自動翻訳の精度向上が必要であるため、総務省とNICTが翻訳バンクを運営しているとの説明がありました。翻訳バンクは、世の中一般から対訳データをご提供いただくことにより、その対訳データを活用して、自動翻訳精度を向上することを目的としています。

2. 基調講演：「データがAIの燃料」となる時代の研究開発戦略／国立情報学研究所（NII）所長・東大生産技術研究所教授 喜連川 優 氏

喜連川氏の講演は次のようなものと感じました。（ほかのまとめについても同様ですが、これはあくまで報告者の主観によるまとめです。）

現在のITを俯瞰すると、データの時代と言えます。英語では「Data Fuels AI」といいますが、これは、「データはAIの燃料です」と言えます。自動翻訳においては、対訳データが自動翻訳の燃料と言えます。

IoTでデータを収集して、AIでデータ解析をします。この二つがそろって、Big Data と言えます。これは、自動翻訳においては、コーパス増強がデータ収集に相当するといえます。このデータ収集はIoTですと限らず、その収集のデザインが重要であり、翻訳バンクはそのデザインのひとつと言えます。

データの領域は米国NFSでも重要視されていて、



会場内の模様



展示会場の模様

米国では、Big Dataは登場から5年以上を経ています。一方、日本のSociety 5.0はデータにドライブされたSuper Smart Societyといえます。なお、日本では、Big Dataの前に、情報爆発プロジェクトが実施されているので、Big Dataは、日本にとっては、なんら新しいものではなく、登場前からのなじみのある概念です。

現状では、AIがうまくいく領域もあればうまくいかない領域もあります。基などでは上手くいきますが、自動運転などではまだまだといえます。

たとえば、画像からキャプションを生成するAIでは、かなり上手にキャプションを生成している例が知られています。けれども、ちょっと変わった入力に対して、たとえば、看板を冷蔵庫と間違えたりします。

これは、学習時には想定外の入力に対して弱いといえます。たとえ、その想定外の入力が、人間にとっては、常識であってもです。

N I I は、医療ビッグデータクラウド基盤構築として、医学関係の学会（複数）と連携し、医療画像データを学会が匿名化して収集し、それをN I I のクラウドにおいて、A I 画像解析研究者が医療画像の研究を進めるプロジェクトを進めています。

このクラウド基盤構築においては、A I は、ほんの一部です。その構築には、ネットワーク、クラウド、セキュリティ、データベース、キュレーション、システムソフトウェア等が必要です。また、A I に取り入れるデータにおいては、90%がデータの前処理に使われる労力です。そのため、A I を活用するためには、クラウド作成やデータ作成などでのA I 以外の部分が非常に重要です。A I は、システム全体のわずかな一部です。

A I 画像分析は、医師の判断の補助として大変役立ちます。

A I 画像分析においては、アルゴリズムによる性能の差は少なく、データを増やすことにより、性能が顕著に増加します。この類推から、自動翻訳においても、対訳データを増強することにより、性能が向上すると思います。これは、翻訳バンクと自動翻訳に対するエールです。

3. 翻訳バンクの概要説明／情報通信研究機構 隅田 英一郎 氏

自動翻訳が役立つ事例として、VoiceTraやみんなの自動翻訳@TexTraを紹介しました。また、自動翻訳の活用により、個人差はありますが、プロの翻訳者の効率が改善することを、ポストエディットにおける翻訳スピードの向上で示しました。

次に、自動翻訳が典型的なA Iであることを、自動翻訳の歴史により説明しました。自動翻訳は1946年より研究されていましたが、Deep Learningによる進化によりNMTが一気に使えるようになり

りました。

翻訳精度は次式できまるという紹介がありました。

「対訳データの量 x 質 x アルゴリズム」

対訳データの量が多ければ多いほど性能が向上しますが、質も重要なので、自動翻訳にも得手・不得手が出てきます。そのため、大量の良質なデータを集めることが重要です。これまでは、Webからそれほど良質でないデータを大量に集めていましたが、翻訳バンクでは、政府・会社等から大量の良質なデータを翻訳バンクに蓄積します。

日本では年間2000-3000億円が翻訳に使われており、それにより5億文程度が翻訳されていると推定されます。それを翻訳バンクで再利用することにより、高精度自動翻訳エンジンを作成できます。

N I C T では、業界ごとに最高精度の自動翻訳を構築し続けています。これまでに特許・I T で高精度エンジンを作っており、今後、製薬・金融等に進めていきます。

翻訳バンクという名のエコシステムにより、ポジティブスパイラルとして、対訳寄付と自動翻訳高精度化が進みます。今のペットボトルなどのリサイクルと同じ感覚で、対訳のリサイクルが常識となればよいです。対訳の形式としては、XMLが理想です。

最高精度の自動翻訳を作るベストな方法は、競争ではなく協調です。散在している対訳を一か所に集めることにより、良い自動翻訳エンジンを作りたいと思います。そのために、翻訳調達の仕様書に次の1行を加えていただきたいと思います。

「原文と翻訳文の対を翻訳バンクに提供」

4. 講演1. 医薬品開発向けA I 翻訳の評価と導入／アストラゼネカ株式会社 田中 倫夫氏

アストラゼネカ（AZ）では、翻訳品質に関する要求性能が非常に高く、以前は、自動翻訳は使えないという判断でしたが、最近の精度向上により注目しだして、現時点では、実際に活用しています。

医薬品開発のトレンドは世界同時開発ですが、日本では日本語の治験実施計画書（CSP）やサマリー文書（CTD）が必須となります。そのため、高品質の翻訳を短納期で実施する必要があります。

AZでは、年間に、A4換算で8000ページくらいの翻訳需要があります。この需要に対して、製薬業界の翻訳業務におけるビジネスニーズは、品質・コスト・スピードです。

AZでは、翻訳バンクの枠組みにより、100万文対を翻訳バンクに提供し、それによりみんなの自動翻訳における汎用NMTをアダプテーションしました。

アダプテーションというのは、「iPhone case」が、一般的には「iPhoneのケース」であるのに、訴訟の分野では、「iPhoneに関する訴訟」となるように、分野毎に訳語を最適化する仕組みです。100万文は、汎用NMTをアダプテーションするには、十分であったようです。

AZとNICTの協業の枠組みは、AZが対訳を提供し、NICTがアダプテーションエンジンを作成（Adapt）し、AZがAdaptの品質評価をするというものでした。

評価としては、訳漏れ、過剰訳、誤訳、文法、一般用語、専門用語、流暢さの観点から評価しました。評価結果により、5段階で3以上が使えるという判断のところで、5と4の評価が多く、翻訳結果が使えることが分かりました。ただし、長文が多い文書や特殊領域（品質関連の文書で対訳データが少ない領域）においては、翻訳性能が他の領域と比べて悪かったです。また、英日方向の性能の方が、日英方向の性能よりも高かったです。

BLEUスコアについていえば、アダプテーションの前後で、日英は18.68→41.94、英日は21.94→50.52のように向上しました。これらは、一般的なBLEUスコアの解釈によれば、かなり高いBLEUスコアといえます。けれども、人手による修正は必要です。

結論として、すべての評価項目について許容できる

品質を認めましたが、翻訳者の品質に到達するには、ポストエディット（PE）が必要です。また、翻訳に要する時間は、PEを実施しても、NMTを使うほうが短いです。

アダプテーションエンジンの最新バージョンは、上記の評価のときよりも更に性能向上しています。アダプテーションにより、専門用語の翻訳が向上しています。加えて、患者同意説明文書では、語尾が「ですます」調になるなど、読んでいて気持ちの良い翻訳が出てきました。

CSPの翻訳におけるNMT使用事例では、NMT導入前では、約150ページの文書について、外部翻訳者による翻訳が3週間＋内部での最終化に1週間、合計で4週間かかりました。一方、NMT導入後は、2週間で最終化まで終了しました。

まとめとして、ポストエディットが必要ですが、品質・コスト・スピードの観点から、AI翻訳の有用性が確認されました。

5. 講演2：著作権法改正がAI開発に与える 衝撃／STORIA法律事務所 柿沼 太一 氏

AI開発のフェーズは、生成フェーズと利用フェーズの2フェーズに分けて考えられます。生成フェーズでは、データを集めて学習モデルを作ります。このとき、データの権利者と、学習モデル作成者の権利が問題となります。利用フェーズにおいては、学習済みモデルが生成した出力の権利等が問題となります。その他にも様々な権利等が問題となります。

AIの生成に関する法律問題としては、生データについては、「データの種類」x「取得方法」で整理するとわかりやすいです。

データに関する規制としては、「法律」「契約」の2通りについて、「実施してよいか」「実施して悪いか」の2通りの組み合わせがあるので、全部で4種類の場合があります。

たとえば、漫画のデータについては、「古本屋から購入したものを自分でデジタル化」した場合と、「デジタ

ルデータを他者から契約により取得」した場合には、データの種類は同じであっても、契約のあるなしにより、データに関する規制が異なります。具体的には、前者は著作権法のみが規制しますが、後者では、法律と契約の双方により規制を受けます。

具体例として、MakeGirs.moeでは、AIが萌え画像を生成します。このときの学習データはGetchuというサイトからクロールにより収集したと思われます。これにより様々な権利に関する論点が生じますが、ここでは、「仮に生データを権利者の許諾なくダウンロードしてモデルを生成していた場合、生データの著作権侵害にならないのか」を検討します。

学習データの準備は、生データのコピー等、いずれの作業も原則として著作権者の承諾なく行うことはできません。けれども、AI開発においては、次の例外規定により、著作権者の承諾なく可能です。これを、権利制限規定といいます。

具体的には、著作権法30条の4により、情報解析に用いる場合には、著作物を自由に利用しても構いません。著作権法30条の4には次の特徴があります。

* AI生成にとって非常に重要

* 非営利目的に限定されていないのが最大の特徴

* 「いずれの方法によるかを問わず、利用することができる。」とされているので、情報解析における広範な作業が可能になった。

ただし、著作権法は「サーバーの所在地」に適用されると考えられていますので、国外にあるサーバーを利用すると日本著作権法30条の4が適用されません。そのため、次のことが言えます。

『機械学習するなら(物理的な意味で)日本において。』

日本著作権法30条の4で以下のことが可能になりました。

- ① 自分がWeb上の自然言語データから対訳データを作って、学習済みモデルを作成
- ② 自分がWeb上の自然言語データから対訳データを作って、それを販売

③ 自分が上記のように作成した対訳データと学習済みモデルを販売

④ 複数の参加者が学習用データセットを作成し共有→大きいインパクトとして、AIデータを作成して販売するビジネスが可能になった。

けれども、第三者が著作権を有している生データから適法に学習用データセットや学習済みモデルを生成するには次の点に注意が必要です。

- ① 契約が有効に締結されている場合は、当該契約(ライセンス)違反となる可能性が残る。
- ② 「情報解析を行う者の用に供するために作成されたデータベースの著作物」については無断で利用できない。
- ③ 30条の4但し書きの「当該著作物の種類及び用途並びに当該利用の態様に照らし著作権者の利益を不当に害することとなる場合」は不可

このように、著作権法の改正により、AI開発が相当自由に実施できることとなりました。

6. パネルディスカッション「AI時代の研究開発、ビジネス及び法律問題」

隅田氏がモデレーターで、上記講演者+NICT理事長の徳田英幸氏がパネリストでした。ディスカッションも非常に盛り上がりました。以下では、報告者の印象に残った発言のみを挙げます。

柿沼氏：AIのアルゴリズムはコモディティ化してきており、データの収集が重要になってきている。

医療においては、データ収集の仕組みが完成しているわけではなく、医療機関とAIベンダーが組となって動いているが、個人情報保護法においてはAIを考慮しておらず、その点がAI開発の問題意識として持っている。

隅田氏：音声も個人情報ということから、音声翻訳において、音声データの取り扱いがとても難しくなっている。

徳田氏：テクノロジーイノベーションのスピードと法律・社会制度のイノベーションのスピードがちぐ

はぐになっている。たとえば、自動運転についてテクノロジーの進歩があったとしても、社会制度のイノベーションがなければ、実際に社会で使うことができない。そのような、社会制度のイノベーションに関する定期的なレビューがあると良いと考える。

徳田氏：テクノロジートランスファーがスムーズに行くような新しい試みをEncourageしてムーブメントを起こしていくとよいのではないか。

隅田氏：データに関して、オープンなデータとクローズなデータのバランスは？

田中氏：機械翻訳では、過去のオープンデータのみで高精度になると理解している。AIは内部にデータを持たないので、クローズデータも含めて学習して、性能を高めることができるだろう。

田中氏：製薬会社では一般的にAIのためにデータを共有することはできない。

徳田氏：NICTにもデータは多数ある。NICTでは、政府予算から構築されたデータについては、API等の何等かの手段でオープンに使えるようにしている。データの種類として、オープン・クローズ等だけでなく、時間軸も考えて、たとえば、数か月後には公開できるとかになると、研究だけでなく、産業利用も活発になる。

隅田氏：機械翻訳においてもオープンなデータが活用でき、それにより毎年機械翻訳精度が向上しており、研究者だけでなく、ビジネスにとってもメリットがあると考えられる。

田中氏：AIによりみんなが同じになったら、他社との優位性はどのように出すのか？時間差によるデータの公開は、無条件のデータ共有と比べて、現実味のあるソリューションといえる。

柿沼氏：私取り扱い案件では1対1での開発が多い。データの権利とモデルの権利は無関係である。問題点は、データの開発コストをどのように回収するかにある。その点から、たとえば時間的な制限を受けたモデルの独占権などが一般的になる可能

性がある。

喜連川氏：データの利用に関して、バランスの取れた議論が必要。

喜連川氏：データの共有に関しては、難しい議論が残っている。コンテキストに応じて丁寧に整理する必要がある。

田中氏：日本発のAIは世界で使えるか？

柿沼氏：モデルが日本で生成されれば、国外にもっていっても大丈夫。データセットについては、国外に持ち出せるかが問題となると思われる。

隅田氏：日本のAIのデータ活性化においてデータ共有スキームの方向性は？

徳田氏：カナダのベンチャーの自動運転では、データ合成によりAIの学習をしている。自動翻訳でもデータ合成ができればよいのではないか

喜連川氏：基でデータ合成ができて、サッカーでそれができるまで時間がかかるだろう。日本は、いろいろな工夫をしなければAIが賢くならない領域を頑張るとよいのではないだろうか。

喜連川氏：ワトソンの勉強法とMTの関係性は？ワトソンは高度な意味処理を利用した勉強をしているが、そのような勉強方法をMTに取り入れることができるか。

隅田氏：特許自動翻訳は3.5億文で学習しているが、人間の翻訳者はもっと少ないデータ量でもっと高精度の翻訳ができるようになってきている。これから考えて、両者の原理は違う。そのため、今後、人間の原理を研究することにより、自動翻訳の性能は飛躍的に向上すると思う。

喜連川氏：特許は特殊な文体であり、それを本当に学習可能か。

隅田氏：完全に学習可能かはともかくとして、中日自動翻訳において、特許実務に役立つようなものは構築可能であり、実現している。

7. 展示

次の10件の展示がありました。これら展示の概要に

については、当シンポジウムサイト

<http://h-bank.nict.go.jp/seminars/index.html>

をご参照ください。

- 74言語対応 夢の通訳機 P O C K E T A L K
(ポケットーク) (ソースネクスト株式会社)
- 事業者様・医療機関様向け多言語通訳タブレット
(コニカミノルタ株式会社)
- 機械・音声翻訳で多言語対応をサポート (凸版印刷株式会社)
- 鉄道のサービスを向上する多言語音声翻訳ソリューション「駅コンシェル」(株式会社日立超L S I システムズ)
- 特許庁 機械翻訳システムに採用された高精度文書翻訳ソリューション (東芝デジタルソリューションズ株式会社)
- 企業向け機械翻訳サービス M i r a i T r a n s l a t o r ™ (株式会社みらい翻訳)
- みんなの自動翻訳@KI (商用版) (株式会社川村インターナショナル)
- 「T - t a c t」および十印の「みんなの自動翻訳 十みん」(株式会社十印)
- 特許と科学技術文献・技術文書に特化した N e u r a l T r a n s l a t o r サービス (日本特許翻訳株式会社)
- 翻訳者・専門家のためのニューラル機械翻訳支援ツール「G r e e n T」(みんなのワードマクロ)

2019 年 2 月 AAMT セミナー セッション 2 参加報告

目次 由美子

XTM International Ltd.

AAMT セミナー セッション 2 参加報告

開催日：2019 年 2 月 12 日（火）

開催場所：主婦会館プラザエフ

セッション 2：『MT×ポストエディット -最前線-』

講演 1：『インバウンドシーンで実用化された機械翻訳
+ポストエディットサービス事例：ジャパリンガル』

野阪 知新氏（凸版印刷株式会社 情報コミュニケーション事業本部 ソーシャルイノベーションセンター 情報インフラ本部 地方創生・観光事業推進部 係長）

府内 翼氏（東日本電信電話株式会社 ビジネス開発本部 第三部門 IoT サービス推進担当）

講演 2：『ジャパリンガル API を利用した CMS での多
言語化サイト運用への適用事例』

早瀬 将一氏（シックス・アパート株式会社 製品企画
マネジャー）

報告者：目次 由美子（XTM International Ltd.）

【講演 1】

2014 年以降、凸版印刷では翻訳システムの社会実装を
中心に、以下のような活動を実施してきた。

店舗ごとの需要に応じて個別展開されていた翻訳案
件を擁する百貨店に対して翻訳の一元管理を提案し、
用語統一によるブランド管理や過去訳流用によるコス
ト削減を実現したり、外国語でのやり取りが求められ
る地方自治体の窓口業務を支援するための音声翻訳シ
ステムの研究開発を NICT からの委託の下、複数の自
治体にて展開している。また、音声翻訳システムを日
本郵便の約 2 万局に導入し、外国人の訪問目的を確認
するなどのコミュニケーションで活用したり、ベトナム
で現地の技能実習生を対象とした講習センターや外
国人労働者を受け入れる日本企業の工場などでも活用

している。

このような自動翻訳の社会実装の取り組みが評価され
2018 年に第 13 回長尾賞を受賞している。



東京でオリンピック・パラリンピックが開催される
2020 年に向けて訪日外国人の増加は顕著であるも、訪
日外国人の多くは言葉の壁に悩まされているという実
情を否めない。外国人客を受け入れたい、受け入れる
準備を進めたいという日本企業の多くは「翻訳」とい
う業務の経験に乏しく、発注時の必要事項も把握でき
ていない。そこに凸版印刷は「ワンストップで課題を
解決する仕組み」の必要性を見出した。凸版印刷の翻
訳者ネットワークと独自の翻訳プラットフォームを
NTT 東日本の機械翻訳エンジン技術と結合させ、イン
バウンド翻訳サービス「ジャパリンガル」としてサー
ビス提供を展開している。

ジャパリンガルでは Web 上にて 24 時間 365 日、翻
訳の発注と納品の受け取りが可能であり、NTT 東日本
の翻訳エンジンを活用して低価格・短納期・高品質を
実現している。2 千文字の翻訳は翌日に納品可能との
こと。機械翻訳を適用した後の事後編集作業にもオプ
ションがあり、ネイティブチェックを含めた高品質の
翻訳も提供している。

ジャパリンガルの機械翻訳として採用している NTT 東日本の「ひかりクラウド cototoba」は、「訪日外国人観光客の“おもてなし”の実現」を目指した AI 機械翻訳サービスであり、特に文化・観光の分野における強さを特長としている。具体的には、文化・芸術、観光・交通などの 6 ジャンル、28 万語彙を含む日本の固有名詞を大量に学習している。また、固有名詞の抽出及び高品質なコーパスを作成する際にも NTT グループの AI 技術 (corevo) が活用されている。また、API 提供する事で様々なパートナー企業様とのシステム連携も可能であるサービスである。



【講演 2】

シックス・アパートでは WebCMS で運用される多言語サイトにてジャパリンガル API を利用している。同社が提供している Web サイト構築のための CMS プラットフォーム「MovableType.net」(MT.net) では、HTML などの技術的な部分とコンテンツ管理が分離されているため、更新管理が実施しやすい仕組みが実現されている。日本語用ページや英語用ページなど、ボタンを

クリックするだけで表示言語が切り替わる多言語の Web サイトは珍しくない。MT.net で管理されているコンテンツを、ジャパリンガル連携機能を活用して翻訳する一連の操作を早瀬氏は実演してくれた。



従来の翻訳のフローでは、条件に見合う翻訳会社を探す 1 つめの手順から、打ち合わせやサンプル提供を経て、翻訳済みテキストを CMS に反映させて公開する 8 つめの手順にて完成に至っている。

MT.net の管理画面では、コンテンツに対して [ジャパリンガルへ依頼] するための機能が搭載されており、翻訳の見積もりや依頼ができるだけでなく、納品された翻訳の受け取りや確認、オリジナルの記事への反映もできる。翻訳済みテキストを差し戻すような必要が発生しない限り、一連の翻訳は MT.net 上で完結させることができる。



第5回特許情報シンポジウム 開催報告

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. 開催概要

AAMT/Japio 特許翻訳研究会の活動の一環として、翻訳を中心とする特許情報処理に関する情報交換と議論の場を提供するための「特許情報シンポジウム」を、平成30年12月7日(金)午後にビジョンセンター浜松町において開催した。本シンポジウムは2010年に第1回を開催、以後2年ごとの開催で今回が第5回となる。シンポジウムの企画および運営は実行委員長を宇津呂武仁先生(筑波大・AAMT/Japio 特許翻訳研究会 副委員長)、副委員長を須藤とし、研究会委員で組織する実行・プログラム委員会を中心に行った。一般講演の受付や選定については綱川隆司先生(静岡大・AAMT/Japio 特許翻訳研究会 委員)にご尽力いただいた。

シンポジウムは招待講演3件、特別講演1件と一般講演4件の構成とし、研究会からの報告は行わなかった。参加者は講演者を含めておよそ100名となり、大変盛況であった。

2. 講演の概要

2.1 招待講演(1): 松谷 洋平 様

(特許庁 総務部総務課 特許情報室 室長補佐)

「特許庁における機械翻訳活用の現状と今後の課題」

内容: 2010年頃を境にしての中国における特許出願件数の突出した伸び、また日本から海外への出願件数の増加を受け、機械翻訳の重要性は増す一方である。WIPOやEPOではニューラル機械

翻訳による翻訳サービスを提供しているが、日本特許庁でも審査における海外特許情報の参照や審査結果の海外発信のために機械翻訳を積極的に活用しており、また機械翻訳の更なる充実へ向けて辞書やコーパスの整備、技術調査を継続的にを行っている。今後はニューラル機械翻訳に基づく機械翻訳プラットフォームの提供、また審査書類の機械翻訳に関する事業を実施する予定である。

2.2 招待講演(2): 井佐原 均 様

(豊橋技術科学大学 情報メディア基盤センター長/情報・知能工学系 教授)

「ニューラル翻訳時代の機械翻訳システムの社会実装」

内容: ニューラルネットワークによる機械翻訳の急速な精度向上によって機械翻訳のニーズが高まっている。豊橋技科大では機械翻訳の社会実装のため、機械翻訳エンジンのベンダー、応用システム開発のベンダー、そしてユーザとなる企業や自治体との連携による共同開発・実証実験プロジェクトを進めている。自治体における情報提供では分野特化型の高品質な翻訳が求められるため、ドメイン内の対訳データの学習に加え辞書登録機能が実装され、システムの運用・評価・効果検証を進めている。

2.3 招待講演(3): 田村 晃裕 様

(愛媛大学 大学院理工学研究科 助教)

「ニューラル機械翻訳の研究動向」

内容: ニューラルネットワークによる機械翻訳の技術について初期の方法から最新の研究までを

概観する。2014 年の回帰型ニューラルネットワーク(RNN)による系列変換モデルの提案に始まり、訳文中の対応関係をモデル化する注視(attention)機構の導入、8 層にも及ぶ多層化によって高精度化を達成した Google NMT、現在標準的な手法と認識されている Transformer までのわずか 3 年ほどで翻訳精度は大きく向上した。さらにサブワードの利用、逆翻訳による単言語データの活用、さらには対訳文データを要しない教師なし機械翻訳、画像の情報を活用するマルチモーダル翻訳等、様々な新しい技術とその応用が急速に広がりつつある。

2.4 特別講演：日本知的財産翻訳協会 (NIPTA) 特許機械翻訳研究会 (奥山 尚一 様[全体紹介]、上野 哲也 様[化学分科会]、湯浅 豊裕 様[機械工学分科会]、菊地 公一 様[電気・電子工学分科会]、宮城 三次 様[知財法務実務分科会]、新田 順也 様[ツール開発])

「ユーザーから見た NMT の使い勝手と活用の展望」

内容：ニューラル機械翻訳によって特許等の翻訳においても機械翻訳が高い精度でできるようになりつつあり、NIPTA では実務において現状のニューラル機械翻訳の技術水準の評価を行いその活用方法を検討するための研究会を立ち上げ、複数の分科会に分かれて活動している。各分科会において NMT を翻訳実務に利用可能であるかどうかの検討を行った。NMT では SMT と異なる訳文の自然性が大きく向上したが、知財翻訳においては NMT で頻出する訳抜けや湧き出しの問題に加え、専門用語の訳出や用語の統一等の面で多くの課題が残されているのが現状である。しかしながら様々な工夫によって NMT の能力を活かすことができ、知財翻訳実務において有益であると考えられる。

2.5 一般講演

一般講演は発表募集を行い、8 件の発表申込に対して以下の 4 件を採択の上、質疑込みで各 15 分の発表とした。いずれも特許の翻訳や検索に関する実用視点からの発表であり、産業における特許情報処理の重要性を感じさせるものであった。

- ・「特許明細書の翻訳で注意すべきこと」
吉川 潔 様 (翻訳業)
- ・「請求項の記載を構造化する提案」
小池 誠 様 (弁理士)
- ・「用語集による原文の一括置換と正規表現による一括並べ替えを用いた特許明細書の高品質翻訳」
杉山 範雄 様 (株式会社杉山特許翻訳)
- ・「知的財産リスクマネジメントにおける効果的な特許検索」
宮崎 祐 様、河野 京子 様
(ヤフー株式会社)

3. 所感

前回 2016 年の開催時と比較し、ニューラル機械翻訳の技術が広く浸透しつつあること、そして翻訳業界での利用が急速に進みつつあることを改めて強く感じさせる内容であった。技術の進展に合わせ、各国特許庁による機械翻訳利用が活発化し、企業等の知財実務・知財翻訳・産業翻訳にも変化をもたらしつつある。

今回は知財行政・研究開発に関する招待講演に加え、特別講演としてユーザの立場から見たニューラル機械翻訳の利点と欠点について詳細なご報告があり、機械翻訳の研究に携わる者として非常に興味深かった。実用のニーズがさらに高まることは間違いなく、研究会としてもこうしたニーズの高まりに対して学術的研究の視点からどういった価値を提供できるか、また技術と実用との橋渡しをすることができるかが重要であろう。

最後に、シンポジウム開催に際しご支援を賜った Japio、AAMT、JTF に感謝を申し上げたい。

IWSLT 2018 参加報告

須藤 克仁

奈良先端科学技術大学院大学

1. ワークショップ概要

IWSLT (International Workshop on Spoken Language Translation)は音声翻訳を主たるトピックとする国際ワークショップで、2004 年の第1 回から毎年開催され今回で 15 回を数える（余談だが筆者は今回で 8 回目の参加であった）。2018 年は自然言語処理の重要国際会議である EMNLP (Conference on Empirical Methods in Natural Language Processing) がブリュッセル（ベルギー）で開催されるのに先立ち、2018 年 10 月 29 日・30 日の両日に同国のブルージュで開催された。参加者は 70 名程度であった。

IWSLT は研究論文の発表やパネルディスカッションを通じた研究議論の場というだけでなく、当初から Evaluation Campaign として音声翻訳・話し言葉翻訳の共通タスクを通じて、機械翻訳や音声認識の研究の加速に貢献してきた。最初は旅行会話コーパス BTEC を利用した共通タスクから始まり、当時発展期を迎えていた統計的機械翻訳と、ルールベース機械翻訳、用例ベース機械翻訳、あるいはそれらのハイブリッド構成システム等、非常に多様な機械翻訳技術が用いられていた。その後音声入力に対する翻訳や音声認識のタスクも追加され、統計的手法の急速な発展に合わせ性能向上を果たしていった。2010 年からは TED Talks の講演音声・字幕データが共通タスクの中心となり、さらに最近は大学の講義音声も追加され、旅行会話のような比較的短い発話を対象としたものから、講演のような長くかつ専門的な発話を扱うものに変化してきている。さらにニ

ューラル機械翻訳の技術が 2015 年に本格的に利用され始め上位を占めるようになり、多言語データを持つ TED Talks の特徴を活かした多言語機械翻訳タスクが 2017 年に実施される等、技術のトレンドに合わせて重要な課題に取り組み続けている。

2. ワークショップ内容

ワークショップでは招待講演と、Evaluation Campaign の報告（主催者による全体説明と上位チームによる口頭発表および全参加チームのポスター発表）、研究論文の発表、パネルディスカッションが行われた。

2.1 招待講演

ワークショップ冒頭に Imperial College London の Lucia Specia 教授の招待講演が行われた。内容はマルチモーダル機械翻訳であり、現在の主流である、画像を利用して文を翻訳するというものから進めて映像や音声の情報を活用する研究についての紹介があった。画像・映像情報は言語の解釈における曖昧性解消に貢献し得るため、適切な利用により機械翻訳を頑健にする効果が期待できること、また音声認識とそれを介した機械翻訳を同時に学習（マルチタスク学習）させることによる性能向上についての議論がなされた。

2.2 Evaluation Campaign

2018 年の Evaluation Campaign では TED

talks のデータを利用したバスク語から英語への翻訳タスクである Low Resource Translation と、英語の講義音声をドイツ語に翻訳する Speech Translation の 2 タスクが実施され、それぞれ 7 チーム、10 チームが参加した。

Low Resource Translation タスクでは、Samsung とエディンバラ大学のチームがバスク語の対訳資源の不足を別分野や別言語のデータ、また機械翻訳結果をニューラル機械翻訳の学習に活用することで好成績を達成した。Speech Translation タスクでは音声認識と機械翻訳を組み合わせるパイプライン型システム（搜狗 (Sogou) と精華大学のチームが最高精度）が、英語音声を英語のテキストを介さず直接ドイツ語テキストに翻訳する end-to-end 型システム（中国科技大と iFLYTEC のチームが最高精度）の性能を上回った。

2.3 研究論文

研究論文発表は口頭発表 9 件とポスター発表 3 件が行われた。ほぼすべての論文がニューラルネットワークの技術に基づく研究によるものであり、対訳データの獲得や選別、ニューラル機械翻訳のアルゴリズム、多言語翻訳の手法、分野適応、翻訳結果の評価・分析等幅広い課題について扱っていた。

筆者は奈良先端大の研究グループで行った音声翻訳研究のポスター発表を行ったが、ポスター会場は共通タスク参加システムの報告含め 16 件の発表があり、非常に熱気のある議論が進められていた。ニューラル機械翻訳の技術が広く使われるようになったことで技術的な課題についても共有され、同様の問題意識を持つ研究者と深い議論をすることができたと感じている。

また会期 2 日目の朝のセッションでは Best Student Paper Award 候補の 2 論文の発表があり、筆者の研究グループの学生が受賞することと

なった。次選となった 1 件は大阪大学と NTT の研究グループによるもので、いずれも修士課程の学生を中心とした研究成果であった。日本の機械翻訳研究は国際的には苦戦を強いられているが、このように成果が認められることで今後の研究の活性化に繋がっていくことを願っている。

2.4 パネルディスカッション

パネルディスカッションでは大学の教員や企業の研究者を含めた 5 名のパネリストにより、機械翻訳・音声翻訳の今後について議論された。

パネリストに共通した認識として、音声翻訳における重要な課題は自然発話や対話の処理であり、フィラーや言い淀み等も含めた音声の処理、対話における文脈の処理の必要性が議論された。現在共通タスクとなっている TED talks は通常十分に練習を重ねた単独話者の講演であって、これらの問題が考慮されていない状況である。

もう一つの課題は招待講演でも触れられたマルチモーダル情報の活用である。人間はマルチモーダル環境下で思考し発声することを鑑みれば、特に対話を翻訳するような状況では環境の情報としてのマルチモーダル情報は不可欠であるという意見が出された。

他にも離散的な情報をニューラルネットワークで連続量として扱うことの難しさや、「数」の翻訳が実用上非常に厄介だ等の議論があった。

3. 所感

音声翻訳は実用的にも身近な課題であり、Microsoft の Skype や Powerpoint に音声翻訳機能が搭載される等一般に手の届くものになってきている。研究レベルではより「賢い」翻訳の実現のためマルチモーダル情報や文脈情報の活用に向けた試みが続いており、今後も注目するとともに自らもさらに積極的に関わっていきたい。

ユーザーから見た NMT の使い勝手と活用の展望

浜口宗武 宮城三次 宮本伸也 湯浅豊裕 平林千春

特定非営利活動法人 日本知的財産翻訳協会 (NIPTA)

特許機械翻訳研究会

日本知的財産翻訳協会について

1. 活動趣旨

我が国の国際競争力を維持増大するために、企業・研究機関などが持つ知的財産の海外展開を進めることの重要度がますます高まっている。一方、これを支える知財語学スペシャリストは一朝一夕には育たず慢性的に足りていない。日本知的財産翻訳協会 (Nippon Intellectual Property Translation Association:: NIPTA) は、日本における知的財産関連の翻訳力の拡充や翻訳者の評価育成を目的に 2004 年に設立された特定非営利活動法人 (NPO) で、現在会員数は特許事務所、翻訳会社、個人特許翻訳者など約 150 法人個人。

2. 主な事業

(1) 「知的財産翻訳検定試験」の実施

特許庁の後援を得、2004 年 12 月の第 1 回試験を皮切りに過去 27 回の検定試験を実施。1 級・2 級・3 級がある。プロレベルを認定する 1 級は「知財法務実務」、「機械工学」、「電気・電子工学」、「化学」、「バイオテクノロジー」の 5 分野に分かれる。試験は年 2 回、春 (日本語から英語) と秋 (英語・中国語・ドイツ語から日本語) にインターネット経由で行われ、最近では海外からの受験も目立つ。出題内容、参考解答、講評などが都度協会ウェブサイトに掲載されるほか、可否を問わず各受験者に採点者 (複数) のコメントが送られる。1 級合格者には ISO17100 適合翻訳者として (一財) 日本要員認証協会が運営する翻訳者評価登録センターへの登録する途が開ける。なお、この試

験は、2019 年 5 月実施の第 28 回試験から、NIPTA と (一社) 日本翻訳連盟 (JTF) が合同で運営する。

(2) 啓蒙活動

月刊機関誌「知的財産翻訳ジャーナル」は、知的財産の世界の情報を「語学」という切り口で解説提供するもので、内容は翻訳学習から外国における重要知財判決や法改正情報に至るまで広範囲にわたる。会員外の購入も可能である。そのほかセミナーなども非定期に開催している。

(3) 特許機械翻訳研究会の活動

2016 年秋の Google NMT のリリースがきっかけとなり、特許翻訳における NMT 活用の気運が俄かに高まる中、NIPTA では、2017 年 10 月に「NIPTA 特許機械翻訳研究会」を立ち上げた。この訳研究会はオープンな組織で、特許事務所や翻訳会社などの NIPTA 会員のみならず、JTF、日本翻訳者協会 (JAT) などの翻訳関係組織やソフト開発者、NMT 有識者など 20 人が参加している。具体的には、「知的財産翻訳検定試験」の過去問題を題材にして現在利用可能な複数の NMT エンジンを用いた試訳を行い、開発者や学識者などのゲストを招いた隔月のミーティングで NMT ユーザーの視点から問題点の抽出や開発者側への提言のとりまとめを行ってきた。今回「第 5 回特許情報シンポジウム」で成果発表の機会をいただいたことは大変有難く関係者各位に謝意を表する次第である。

(文責: NIPTA 常務理事兼事務局長 浜口宗武)

NMT 特許翻訳評価分析 1 知財法務実務

機械翻訳の高速処理を利用した翻訳作業効率化のささやかな試み

宮城三次

一色国際特許業務法人パートナー

Google が NMT を採用した機械翻訳 (MT) サービスを公開した後、社会には一種の MT フィーバーとも言うべき状況が現出した。NMT のメカニズムとその限界が知られるにつれてその熱狂は沈静化しているようであるが、なんらかの形で MT が翻訳に取り入れられ、影響を与え続けていくという全体的なトレンドは今後も持続すると考えられる。

このような翻訳を巡る趨勢に鑑み、日本知的財産翻訳協会 (NIPTA) の機械翻訳研究会では、機械翻訳 (MT) の有効利用を図るべく活動してきた。その中で、筆者らは、NIPTA 検定試験の「知財法務実務」分野に対応する、裁判例等の日本語テキストを素材として翻訳結果の分析などを進めてきたが、このようなテキストは個々の文章が長文であり、主述の関係が入り組んでいるなど、複雑な構造を持っていて、現在一般に利用可能な MT で処理しただけでは実用的な結果が得られないことが明らかとなった。そこで、MT 処理後のテキスト翻訳をチェックする労力を極力省けるようにして効率的な訳文作成を目指し、MT 処理前のテキスト編集であるプレエディット (PE) の効果を確認した。

まず、PE を行う際の指針として、「産業日本語・特許ライティングマニュアル」(産業日本語研究会) が提唱する、「短文にする」、「省略しない」、「読点を工夫する」等の 7 つのポイントを採用することとした。試行対象の文例を以下に示す。

「原告は、各基礎出願明細書の図 2 から、螺旋溝 27 の傾斜角度が「約 20 度」であることが読み取れ、当

該「約 20 度」を中央値として「±10 度」とすることで、「10 度から 30 度」の範囲が設定できることは、図面および明細書の記載から当業者にとって自明な事項である旨主張する。」

この例文を例えば Google 翻訳を用いて英語に翻訳させると、「原告は、…である旨主張する。」という読解の基本となる主述対応を読み取ることができない等の問題が生じて、効率化に有用と考えられる翻訳結果を得ることはできなかった。そこで、筆者らは、先の PE 指針に沿って、次のような処理を実行した。

・入れ子構成を下記 A～C のように解消する

A…原告は、(B、C) (である) 旨主張する。

B… (C) (である) ことは、図面および明細書の記載から当業者にとって自明な事項である。

C…各基礎出願明細書の図 2 から、螺旋溝 27 の傾斜角度が「約 20 度」であることが読み取れ、当該「約 20 度」を中央値として「±10 度」とすることで、「10 度から 30 度」の範囲が設定できる。

・C 部を分割するとともに、「中央値として、「±10 度」とする」という表現を言い換える。

このような PE により、一例として以下の MT 入力用テキストを得た。これを Google 翻訳に入力したところ、基本的な主述関係の読み誤りを始めとして誤りが低減され、翻訳の効率化に資すると考えられる結果を得た。

「原告は、下記を主張する。

以下のことは、図面および明細書の記載から当業者にとって自明な事項である。

各基礎出願明細書の図 2 から、螺旋溝 27 の傾斜角度が「約 20 度」であることが読み取れる。したがって、当該「約 20 度」を中央値と設定して、その中央値から 10 度増減することで、「10 度から 30 度」の範囲が設定できる。」

上記のような PE を適用しながら MT の速度を活用して結果を参照しながら繰り返し原文の形式を整えていくことで、効率的な翻訳処理を実現可能との見通しを得た。

なお、周知のように Google の利用規約により、投入したテキストには Google に対して広範な利用許諾を与えたことになるので、処理対象のステータス（内容が公知であるか否か）には十分に留意する必要がある。

（初出：JTF Journal #298、2018 年 11/12 月号）

NMT 特許翻訳評価分析 2 電気・電子工学

特許翻訳における NMT の活用—その課題と提言—

宮本伸也

日本ビジネス翻訳株式会社

1. NMT に対する実用性評価

電気分科会では、最初に 3 社の機械翻訳エンジンを
使い NIPTA の過去問を翻訳して、それらの結果を評
価してみた。誌面の都合上、詳報は割愛するが下記に
各翻訳エンジンの傾向をまとめてみた。

N 社: 同じ翻訳文が繰り返し出現、脱訳等がみられた。

R 社: 「です・ます」調の翻訳文の混在、並列関係が
不適切な翻訳文がみられた。

G 社: 不適切な用語選択、文章として成立しない翻
訳文がみられた。

次に、実用性の評価としては、各翻訳エンジンとも、
その出力はより自然で流暢な英文であり、読みやすさ
ひとつ取れば、SMT との比較で言えば隔世の感を禁じ
得なかった。他方、実用性という点では、各エンジン
とも大同小異で品質面における安定性の問題が散見さ
れた。NMT と雖もそれらの出力を最終成果物として
直接利用することは、現段階での完成度を見る限り時
期尚早との心証をもった。もっとも、後述するポスト
エディット (PE) を導入することにより、これらを人
間翻訳と同等レベルの翻訳文に編集することは可能で
ある。以上の観点から、NMT の翻訳エンジンは、電
気・機械分野で限って、尚且つ PE を密接不可分な後
工程と位置付けさえすれば、概ね実用レベルに達して
いるとも言えよう。因みに、本研究会の活動を通して
得られた PE の作業項目についての試案を下記に紹介
する。

(共通項目)

用語の統一、表現の統一、訳漏れのチェックと加筆、
湧き出しのチェックと修正、全角文字の検出

(英日翻訳)

文体の統一: 「です・ます」調→「である」調

(日英翻訳)

冠詞: 初出/既出の整理、単数/複数の整理

2. MT プロバイダーへの提言

上述した通り、現段階において NMT 出力をそのま
ま「翻訳文」として使用するにあたっては、まだ解決
すべき課題も多く、また PE を行う者への作業負荷も
決して軽いものとは言えない。そこで、差し出がまし
いとは思いつつ、ユーザの立場から MT の開発に携わ
っている方々への提言として以下の問題に対する改善
策を提言したい。

(1) 湧き出しへの問題

(2) 脱訳・訳漏れの問題

(3) 用語登録オプション

まず、(1) と (2) に対する改善策としては、該当しそ
うな箇所に何らかの注記・警告らしき表示があれば、PE
作業中に注意を払って便利と思われる。また、(3) につい
ては、既に各翻訳エンジンとも用語登録機能を備えており、
それらはユーザにとって用語統一を図る上で大変有難い
機能であることは認めつつも、登録用語によっては翻訳の
結果が大きく左右されてしまうことを指摘したい。

(例文) 送風機は、ファンの回転数を増減させるインバータを備える。

(A) 用語登録なし

The blower includes an inverter for increasing or decreasing the number of revolutions of the fan.

(B) 用語登録あり (増減: variation)

The blower includes inverters for variation the rotational speed of the fan.

このように用語登録をしたのは良いが、NMT が品詞の解釈を誤ってしまうと、(B) のように反って NMT の品質低下を招く場合もあることを申し上げておきたい。そこで、せっかくの機能を有効に活かす上でも、登録する用語に対する品詞指定ができればかかる問題を解決できるのではないかと思った。

3. ユーザ側で対応可能な改善策

最後に、我々ユーザ側で対応可能な改善策としては、NMT に不向きな日本語を事前に編集することが挙げられる。例えば、複数通りの係り受けの解釈が可能となるような「ネスト構文」を避けるだけでも、翻訳精度は向上し PE の作業負荷もかなり低減される。他にも、人が速読して読み難いと感じるような表現箇所を、解釈し易い日本語文に書き換えることも有効である。その際、特許ライティングマニュアル（日本特許情報機構編）を活用されることを推奨したい。

(初出: JTF Journal #298、2018 年 11/12 月号)

NMT 特許翻訳評価分析 3 機械工学

NMT を使ってみました。

湯浅豊裕

WIS 知財コンシェル株式会社

1. はじめに

知的財産翻訳検定 1 級 (機械工学) の過去問を NMT にかけてみました。1 級に合格できる品質ではありませんが、得意なところ、苦手なところを理解することで、ツールとして使うことはできます。

以下、得意なところと苦手なところをまとめました。例文はすべて知的財産翻訳検定の過去問です。NMT 訳は、NMT (一般に利用可能な NMT エンジンを使用しました) の出力をそのまま掲載しています。

2. 得意なところ

概して、自然なことば (流暢) に翻訳されています。

(例文)

Having two stems 11 enables easier access if the component 1 is located in a hard-to-reach place or is placed close to a wall.

(NMT 訳)

2 つのステム 11 を有することにより、コンポーネント 1 が届きにくい場所に配置されている場合、または壁の近くに配置されている場合にアクセスが容易になる。

(コメント)

無生物主語などの構文もきちんと訳されます。

3. 苦手なところ (形式的なところ)

訳語が揺れることがあります。

(例文)

蒸着工程が繰り返される過程でチャンバ 20 と噴射ノズル 10 は、繰り返し熱によって膨張と収縮が繰り返される繰り返し荷重に露出し、(中略)、噴射ノズル 10 とチャンバの外壁 21 との間には微細な隙間が形成される可能性がある。

(NMT 訳)

In the process of repeating the vapor deposition process, the chamber 20 and the spray nozzle 10 are exposed to repeated loads in which expansion and contraction are repeated by repeated heat, and (中略) there is a possibility that a minute gap is formed between the injection nozzle 10 and the outer wall 21 of the chamber.

(コメント)

1 つの文の中でも、訳語が統一されないことがあります。

脈略のない語が出てくる場合があります。

(例文)

As will be appreciated from the above, turning

one of the stems 11 will cause the ball 7 to rotate between a closed position, where the bore 9 is perpendicular to the main bore 4, and an open position, where the bore communicates with the main bore.

(NMT 訳)

上記から理解されるように、ステム 11 の 1 つを回転させることにより、ボア 9 が主ボア 4 に対して垂直である閉位置と、ボアが主ボア 4 に連通する開位置との間でボール 7 が回転する。ボア。

(コメント)

取って付けたように「ボア。」という単語が出現します。長いフレーズが出てくることもあります。

訳抜けすることがあります。

(例文)

The invention will now be described, by way of example, with reference to the accompanying drawings, in which: Figure 2a is a side-sectional view of a valve component in accordance with a first embodiment of the invention; Figure 2b is a side perspective view of the valve component shown in Figure 2a.

(NMT 訳)

本発明を、添付の図面を参照して例として以下に説明する。本発明の第 1 の実施形態による弁構成要素の側断面図である。図 2b は、図 2a に示される弁構成要素の側面斜視図である。

(コメント)

「Figure 2a」が抜けています。文単位で訳抜けすることもあります。

4. 苦手なところ (実体的なところ)

字面を追った解釈をすることがあります。

(例文)

ゴルフクラブ (以下クラブ) には飛んで曲がらない
(曲がりにくい) ことが最も要求されている。

(NMT 訳)

It is most demanded that a golf club (hereinafter referred to as a club) does not bend by bending (hardly bending).

(コメント)

ゴルフクラブ自体が曲がらないと解釈しています。技術 (内容) を理解していないと翻訳できない箇所は苦手なようです。

5. まとめ

エラーにもパターンがあります。これらのパターンを理解することが、効率的なプレエディットやポストエディットにもつながります。

上述の NMT 訳は、説明したエラーの箇所以外はそのまま使える部分もたくさんあります。NMT を活用したさまざまな翻訳の方法論が、これから提案されてくると思われます。

機械分科会メンバー

奥山尚一 (久遠特許事務所)

浜口宗武 (株式会社知財コーポレーション)

新田順也 (エヌ・アイ・ティー株式会社)

鈴木彰浩 (トランスユーロ株式会社)

湯浅豊裕 (WIS 知財コンシェル株式会社)

(初出: JTF Journal #298、2018 年 11/12 月号)

NMT 特許翻訳評価分析 4 化学

化学特許とニューラル翻訳のケミストリー

平林千春

平林特許・翻訳事務所

1. はじめに

化学分野の NMT 評価は、梶木正紀氏 (MK 翻訳事務所)、上野哲也氏 (MK 翻訳事務所)、柴田知恵氏 (知財コーポレーション)、岡本光弘氏 (トランスユーロ株式会社) 及び小職の 5 名で行った。NIPTA 知的財産翻訳検定の過去の試験問題を評価用の原文として使用した。

2. 予想以上に上手い！？

次の英文クレームを 3 つの NMT エンジンに入力した。

1. A multi-polymer hydrogel article comprising a first polymeric, water-swella
ble material and a second polymeric material, wherein a first region of the article substantially comprises the first polymeric, water-swella
ble material, a second region adjacent the first region comprises a mixture of the first polymeric, water-swella
ble material and the second polymeric material, and a third region adjacent the second region substantially comprises the second polymeric material, and wherein the second polymeric material exhibits an increasing concentration gradient moving from the first region, through the second region, to the third region.

エンジンによっては全く請求項の体をなしていない訳文を出力するものもあったが、請求項の翻訳に特化したエンジンからは、予想以上に上手い訳文が出力された。

1. 第 1 のポリマー、水膨潤性材料および第 2 のポリマー材料を含む多ポリマーヒドロゲル物品であって、前記物品の第 1 の領域が実質的に第 1 のポリマー、水膨潤性材料を含み、前記第 1 の領域に隣接する第 2 の領域が前記第 1 のポリマー、水膨潤性材料および前記第 2 のポリマー材料の混合物を含み、前記第 2 の領域に隣接する第 3 の領域が実質的に前記第 2 のポリマー材料を含み、前記第 2 のポリマー材料が、前記第 1 の領域から前記第 2 の領域を通して前記第 3 の領域に移動する増加する濃度勾配を示す、多ポリマーヒドロゲル物品。

「移動する増加する濃度勾配」や「第 1 のポリマー、水膨潤性材料」など、おかしい点が幾つかあるものの、構成要素の分節などは概ね正しいようで、非常に驚いた。

3. 略語が苦手？

化学案件に多くみられる conc.、sat.などのピリオド付きの略語の訳出が苦手であることが分かった。

例えば、「The organic phase is washed with

water (2×3 ml) and sat. NaCl (1×3 ml) and dried using magnesium sulfate, and the solvent is removed under reduced pressure.」という原文に対して、「有機相を水(2×3ml)および飽和 NaHCO₃ で洗浄する。NaCl(1×3ml)および硫酸マグネシウムを用いて乾燥させ、溶媒を減圧下で除去する。」という訳文が出力された。原文にない NaHCO₃ が突然現れたのである。重大な受け係りミスも犯している。

また、「The mixture is allowed to warm to room temperature overnight, 5.2 ml of water, 5.2 ml of methyl tert-butyl ether and 3 ml of conc. HCl are subsequently added.」という原文に対して、「混合物を室温まで一晩、5.2mlの水、5.2mlのメチル tert-ブチルエーテルおよび3mlの濃度まで加温する。続いてHClを添加する。」などの訳文が出力された。一旦 conc. で文を区切り、改めて「HCl を添加する。」と訳出してしまうのである。

使用した3つのエンジン全てにこの傾向がみられた。

4. 訳語の統一が苦手？

例えば、同じ「熱収縮率」という原語に対して、「thermal shrinkage」と「heat shrinkage ratio」というバラバラの訳語が出力された。どのエンジンも訳語の統一が不得意のようであった。

5. 化合物名も苦手？

「Polyvinyl chloride」を「塩化ポリビニル」と訳してしまったり、「5-Bromo-1,3-difluoro-2-[(R)-3-methylpent-1-enyl]benzene」を正しく訳出できず「5-ブロモ-1,3-ジフルオロ-2-[(R)-3methylpent1enyl]ベンゼン」と訳してしまったりする場合もあった。

6. まとめ

NMT を実務で使用するのであれば、その弱点をよく理解した上で使うべきであると考え。

(初出：JTF Journal #298、2018年11/12月号)

LocWorld38 Seattle 参加報告

目次 由美子

XTM International Ltd.

LocWorld38 Seattle

Preconference: 2018 年 10 月 17 日(水)

Main Conference : 2018 年 10 月 18 日 (木) 19 日(金)

開催場所 : 米国ワシントン州シアトル市、Bell Harbor
International Conference Center

The Localization Institute の創設者である Ulrich Henes 氏と Multilingual 誌の Donna Parrish 氏がローカリゼーション業界のためのイベントを企画しようと 2003 年に初めて開催されたその地シアトルにて、本年米国での LocWorld (LW) が執り行われた。近年では 1 年に 3 回、アメリカ、ヨーロッパとアジアの各地で開催されている。お気に入りの開催地とも考えられるシアトルへ、2012 年の第 20 回以来、実に 6 年ぶりに帰って来たといった雰囲気さえ漂っていた。世界有数のトップ企業が本社を擁する都市での久しぶりの開催には

「Digital Transformation」というテーマが掲げられた。

LW は日本でも開催されているが、まず規模が違った。それは会場の大きさにも表れていた。展示会場では翻訳会社や翻訳ツールベンダーを含め、40 以上の企業によるブースが構えられた。



参加者で賑わう展示会場

Main Conference は 6 つのトラックに分かれ、朝 9 時から夕方まで、合計で 40 を超えるセッションが開催された。TAUS や CSA といった翻訳業界団体、日本でも知名度が高いグローバル企業である Microsoft 社、Cisco 社、Expedia 社、シアトルに本社を擁する Starbucks 社、Tableau 社を始め、翻訳会社や翻訳ツールベンダーからももちろん、120 名を超える豪華な登壇者があった。

プログラム上、セッションには以下のカテゴリーが示されていた。

- Advanced Localization Management
- Community
- Content Management
- Core Competencies
- Global Business
- Inside Track
- Keynote
- TAUS
- Technical
- Unconference

Keynote セッションが行われる時間帯は、ほかのトラックではセッションがプログラムされない。National Geographic 社による Keynote セッションは「Change is the Only Constant」と題されており、LW38 のテーマである Digital Transformation (デジタルの変遷) に直結しており、モバイルや SNS などにも言及があり、非常に興味深い内容であった。

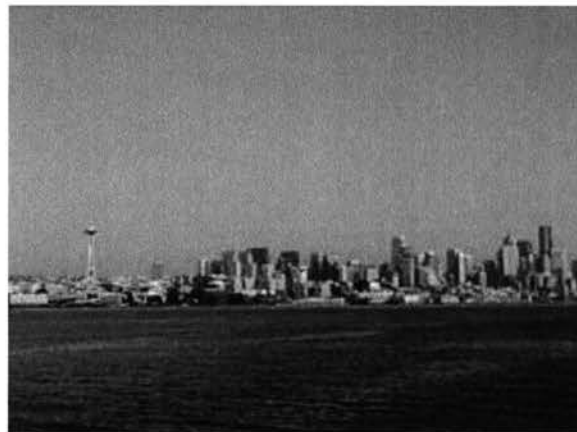
日本でも集客率に影響を与えるキーワードである「AI」や「機械翻訳」、「スマートスピーカー」に関するセッション、翻訳業界ならではのとも言えるプロジェクトの取り組み方やベンダー管理、GDPR (General Data Protection Regulation、EU 一般データ保護規則) に関するセッションなども見受けられた。

セッションや展示ばかりでなく、ネットワーキングの機会が豊富に設けられているのも LW の特徴と言える。Opening Reception は翌日から開催されるイベントへ向けてのキックオフミーティングとも考えられ、Main Conference 当日は夕食会だけでなく、After-Dinner Drinks まで企画されていた。ランチや展示会場の休憩スペースも、参加者どうしの交流を深める機会として大きく機能していた。



シアトル市内のレストランで開催された Conference Dinner

今回の LW では、日本のローカライズ担当者向けに特別なツアーが企画されたことを特筆したい。LocWorld オフィシャル パートナーである Learn for Yourself 事務局 (lw38-tour@l4y-edu.com) によって企画された「LocWorld38 Seattle ツアー&企業視察とワークショップ」は、17 日(水)に開催され、Tableau 社、Starbucks 社、Big Fish 社を訪問し、Facebook 社、Nintendo of America 社、Eventbrite 社の担当者のごく少数の参加者を対象としたセッションが開催され、きわめて充実した内容のツアーであった。この詳細は Multilingual 誌 (<https://multilingual.com/>) 2019 年 1 月 2 月号に掲載される予定であり、そちらも参照されたい。



海から見たシアトルの美しい街並み

Women in Localization Japan 第 16 回イベント報告

目次 由美子

XTM International Ltd.

Women in Localization Japan 第 16 回イベント報告

日時：2018 年 11 月 30 日（金）19:00 ～ 21:30

開催場所：ネットアップ株式会社 会議室

セッション I

テーマ：Continuous Localization Delivery (CLD)

登壇者：Dominick Kelly, XTM International Ltd.

セッション II

テーマ：伝わる日本語のスタイルとルール ～翻訳しやすい日本語・やさしい日本語の観点から～

登壇者：高橋 慈子 Shigeko Takahashi, 株式会社ハーティネス

Women in Localization Japan 事務局メンバー：

上田 有佳子、千葉 容子、澤村 雅以、加藤めぐみ、森 みゆき

ローカリゼーションに携わる女性のための非営利団体として 2008 年に米国加州で設立された「Women in Localization」の第 16 回イベントが、2018 年 11 月 30 日に忘年会も兼ねて開催された。日本支部は 2015 年に発足し、3 ヶ月ごとに勉強会兼ネットワーキングのイベントを開催している。発足当初は約 10 名であった会員は今や 100 名を越え、クライアント企業のみでなく、翻訳会社や個人翻訳者、翻訳ツールベンダーを含む。ローカリゼーション関連の新しい技術や共通の問題をトピックとしたセッションを行っている。本年は、米国での同団体発足から 10 周年を迎え、各支部でも記念すべきイベントが開催されてきた。来年に向けて、日本支部でも新たな運営や、イベントレポートを含む Web サイトの拡充 (<http://womeninlocalization.com/>) が予定されていることが発表された。そして、今回の 2

つのセッションでは、翻訳管理システムの特長機能や、日本語のライティングに関連する取り組みが紹介された。



Session I

“Continuous Localization Delivery (CLD)”

Mr. Dominick Kelly from XTM International Ltd.

Women in Localization Japan chapter welcomed speaker online from Dublin, Ireland. Dominick Kelly from XTM International Ltd., the world's leading developer of innovative Translation Management System, was happy to talk about a feature called “Continuous Localization Delivery” (CLD).

At first, Dominick reminded us that CLD is not required for everyone. However in today's world we find content is now broken up into smaller chunks with increased frequency. An example would be software and game content. This can be generated on daily or weekly basis; involving the development team cycles, creative team updates, and Localization Quality Assurance team bug fixes, for example. Thus the contents are continuously updated, and translation requests are frequently made. However, this is not the case for instance to legal documents and the process of creating this content is not agile.

Then Dominick gave us more details to this concept of CLD with the specific names of related tools such as Git for Source Control System, Sitecore for Digital Content Management, JIRA for Bug tracking System, and XTM for Translation Suppliers' System. There are the solutions that aide agile the localization tasks with frequent updates; it may involve with more than 30K files with 40 target languages. Those tools would help us not to miss any item, not to miss any file, or not to miss any deadline for the localization requirement.

Also, Dominick emphasized that the quality must not be missed – regardless of the tight deadline, the big quantity, and multiple rounds of agile cycles. Then he explained us about the importance of meta data which assures the translation quality. With XTM, an image can be displayed for each segment. This helps translators to provide accurate translation. In addition, the string IDs are as well displayed for each segment, and they are taken into account for leveraging translation memory.



セッション II

『伝わる日本語のスタイルとルール ～翻訳しやすい日本語・やさしい日本語の観点から～』

株式会社ハーティネス 高橋 慈子氏



プレゼンテーションの冒頭で高橋氏は、分かりにくい日本語の文書は生産性を低下させ、機会が損失されるといったリスクを指摘した。そして、分かりやすい文書の価値として、文書の目的が達成できるばかりではなく、信頼性を高める、組織を守る記録となる、業務に関する知識が共有されることなどを列挙した。

現在の社会で求められる文章を中心とするアウトプットについて、「スピード」というキーワードが含まれていた。これは文章を早く書くということではなく、読み手が短時間で理解できる文章であることを意味する。そして、多くの企業で「英文にしやすいシンプルな和文」が求められているようだ。特に多言語翻訳を伴うマニュアルを日本語で執筆する担当者には、このような意識が芽生え、一文が長く、理解しにくい日本語の文章は回避される傾向が見受けられるようだ。さらには DITA (Darwin Information Typing Architecture、技術情報を制作・管理・配信・再利用するための XML に基づいたアーキテクチャ) のようなトピックライティングを実行することで、簡潔に完結している情報を効率的に発信できるという。

文書の品質を向上させるには、「スタイルガイド」を作成してルール化することだとも言及された。スタイルガイドの項目例として、一文は原則として 50 文字以内にすることが挙げられていた。製品の仕様や運用に

ついて執筆する際、特定の担当者に依存するのではなく分業させることで効率が向上するだけではなく、製品に対する知識やライティングのノウハウが共有される。また、トピックの品質を標準化することによって、全体的な品質の向上を図ることができるそうだ。さらに、スタイルガイドの目的は作成することではなく、運用することであるとも高橋氏は力説された。実運用に向けては、ライティングスキルを高めるためのライティングトレーニング実施が必須になるという。

さらに、高橋氏は「やさしい日本語での説明文研究」に関する活動についても紹介した。日本語を母語としていない人たちが日本での生活で目にする「説明文書」を、分かりやすい表現にするための、あるべき表現のルールやスタイルを検討してまとめているとのこと。今後、外国人や高齢者、発達障害などを含めたコミュニケーションが取りにくい人などの多様な労働者が増えていくであろう日本の社会では、その需要が大きくなることも予想される。

登壇者プロフィール

Dominick Kelly

XTM International Ltd.

Client Solutions Manager - EMEA



Dominick is an experienced consultant with over 16 years in the translation and localization industry, providing language solutions across a range of major industry sectors. He is a qualified software engineer that has worked for Enterprises, translation suppliers and language software developers. A true problem solver and innovator within the translation technology space

高橋 慈子

株式会社ハーティネス

代表取締役



東京農工科大学卒。技術系出版社を経て、テクニカルライターとしてフリーランスで取扱説明書の作成に関わる。1988年テクニカルコミュニケーションの専門会社として、株式会社ハーティネス設立。

企業の取扱説明書やマニュアル制作のコンサルティングや人材育成に関わる。テクニカルライティング研修のほか、ビジネス文書研修の企画や実施、各種セミナーで文章の書き方指導を展開。大妻女子大学、立教大学、慶應義塾大学 非常勤講師。

情報処理学会ドキュメントコミュニケーション研究会幹事。

著書に『技術者のためのテクニカルライティング入門講座』（翔泳社）、『人より評価される文章術』（共著・宣伝会議）、『SEの文書技術』（日経BP）など。

<http://www.heartiness.co.jp/>

平成 30 年度産業日本語研究会・ワークショップ参加報告

小谷克則

関西外国語大学

1. はじめに

産業日本語研究会は明瞭な日本語文の作成と高品質な翻訳文の低コスト作成を目標とし活動する研究会です。本稿は本研究会平成 30 年度ワークショップを報告いたします。

本ワークショップは東京大学本郷キャンパスに於いて平成 30 年 12 月 26 日（水）14 時から 17 時まで開催されました。筆者のオブザーバー参加を認めていただいた関係者の皆様にこの場をお借りして深く感謝申し上げます。

ワークショップの流れは次の通りでした。開会の挨拶を小林明氏（日本特許情報機構専務理事）が執り行い、岡智之先生（東京学芸大学留学センター教授）、野中尋史先生（長岡技術科学大学情報・経営システム工学専攻准教授）が講演を行い、本研究会の各分科会（ライティング分科会、特許文書分科会、文書作成支援分科会）の各主査が活動報告を行い、小林氏の挨拶により閉会しました。本稿では講演内容と分科会での様子を報告します。

2. 講演 1：岡智之先生

岡先生の講演タイトルは『「場」でわかる日本語～場の言語学への招待～』でした。日本語の理解には文字列の理解だけでは到底事足りず「場」の理解が重要、との見地に至る背景として、日本語研究に加え、留学生への日本語指導も挙げられました。

講演では、まず、「場」の概念の解説をされ、この概念を端的に表す表現として「KY(場の空気が読めない)」を挙げられました。「場」は「場所の哲学(西田幾多郎)」

として哲学において研究されており、岡先生は「場所の言語学」を提唱されてきました。

次に「場」に基づく分析として、日本語・英語の比較、日本語格助詞の意味分析、「やりもらい」表現などの分析を挙げられました。

その一例として、日本語・英語の主語分析を取り上げます。一般的な分析において英語主語は顕在的であるのに対し、日本語主語は潜在的、つまり削除された要素と分析されます。しかし、場所の言語学によると英語は主体の論理に従うため主語が生じるのに対し、日本語は主体ではなく場所の論理に従うためそもそも主語が生じない、潜在的でもないと分析されました。

講演後の質疑応答は以下の通りでした。

3. 講演 2：野中尋史先生

野中先生の講演タイトルは『特許文書におけるコンテンツ／ネットワークマイニング』でした。野中先生は NEDO フェロー制度のもと豊橋技科大に於いて学内の知的財産や産学連携に従事した際、出願戦略・研究開発戦略の策定等に有用なパテントマップ（特許出願や権利化動向を可視化したもの）に着目し、特許情報解析の研究に取り組まれました。本講演で紹介された研究は、野中先生を始め、酒井浩之先生（元豊橋技科大、現成蹊大）や小林暁雄先生（理化学研究所）との共同研究によるものとのこと。

講演ではまず、パテントマップを解説されました。多種多様なパテントマップがあるなか、技術とその効果を示す「技術－効果型」パテントマップの有用性に着目されているとのことでした。このパテントマップにより、ある技術とその効果に対する出願状況を把握

でき、出願戦略に有用な指針となります。

パテントマップの作成には、技術の効果を表す表現をキーワードとして抽出する必要があります。例えば、特許文書中の記載には、「ガスの吹込み技術」の効果として「それにより亜鉛やマグネシウムなどの不純物を除去できる」とある場合、この記載から「不純物の除去」をキーワードとして抽出します。

「コンテンツマイニング」とは上記のキーワードを抽出する技術を指します。野中先生らの提案手法は文書中の手がかり表現を網羅的に抽出します。具体的には、まず、「できる」などの効果を表す初期手がかり表現と係り受け関係にある共通頻出表現集合（「小型化すること」など）を同定し、次に、「できる」以外の表現を手がかり表現集合（「が可能となる」など）として同定します。そして、手がかり表現獲得と共通頻出表現獲得を循環的にを行います。抽出表現の選別に統計的基準をエントロピー法により設けることで、一定の成果が得られたとのことでした。今後の課題として、ブートストラップ法に基づく抽出法の改良、意味的統合などを挙げられました。

「ネットワークマイニング」とは特許文書の価値を得点化する技術を指します。この技術は、特許情報評価に留まらず、その特許を有する企業の株価予測や財

務指標予測にも貢献すると期待されています。したがって、知財・投資戦略支援技術ともいえます。提案手法は特許データベースから審査官引用ネットワークを作成し、ネットワーク埋め込みを用いて技術分野を検出します。そして、LSTM (Long Short-term Memory) を用い、技術分野の成長性を時系列に沿って評価します。評価指標に HITS (Hyperlink-Induced Topic Search) /PageRank を用いることで、評価時の価値といった静的な評価において正解率 80%に到達したとのことです。今後は、評価対象の特許技術の成長・衰退の度合いも併せて動的な評価を目指し、改良を図るとのことです。また、様々な価値評価指標との統合も検討されているとのことでした。

最後に、日本特許情報機構 JAPIO を始めとする関連団体への要望を挙げられました。その要望とは特許解析技術の評価用データを蓄積・公開するといった環境整備でした。特許解析研究のより一層の飛躍には、評価用データの環境の整備が急務とのことでした。

4. 分科会活動報告

4.1. ライティング分科会

東京外国語大学教授／ライティング分科会主査佐野

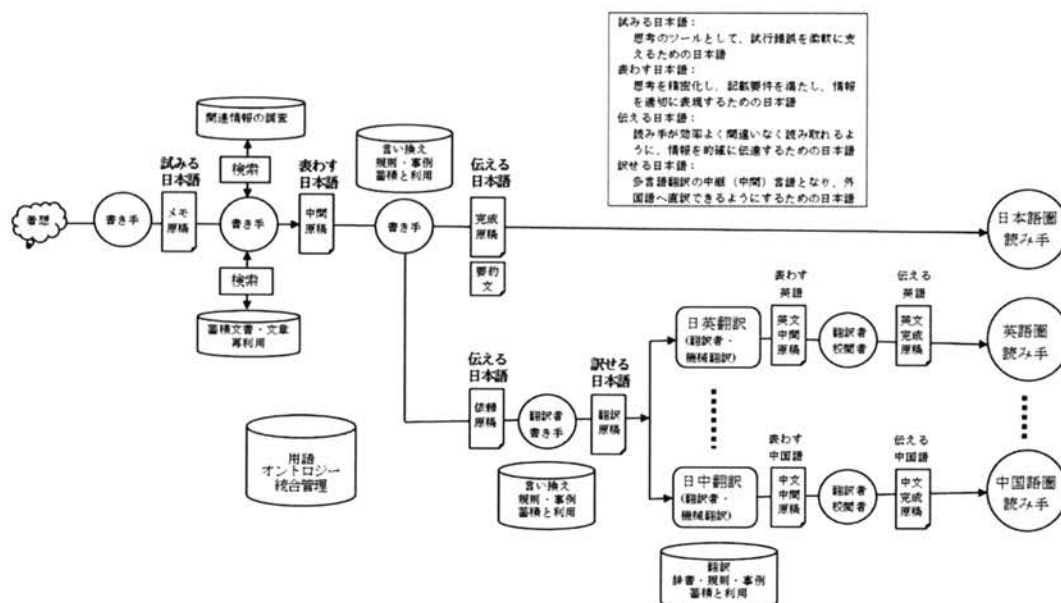


図1 文書作成モデル（日本語マニュアルの会 2016）

洋先生より本分科会での今年度の活動中間報告がありました。

まず、本分科会の概略が紹介されました。活動の目標「日本語の書き方マニュアル」のもと、文書作成モデル（図1）の作成を進め、今年度は「伝える日本語」のステージの検討を行ったとのことでした。

報告事項の第一点目は、日本語と英語における説得型文書のアプローチを比較結果でした。この比較を通じて、日本語文書は書き手が読み手の共感を高め、納得を得ようとする共感型の説得を行うのに対し、英語文書は書き手が読み手に意志を伝達し納得を得ようとする意志型の説得を行っていることが判明したとのことでした。

報告事項の第二点目は、日本語の書き方規則の検討結果でした。この書き方規則は共感型文書向け規則、説得型文書向け規則、そして共通規則に細分化されます。各規則の一例として「主張文」の違いを挙げられました。説得型主張文が根拠の存在性を主張するのに対し、共感型主張文は話題との関連性を主張するとのことでした。

今後の課題として、書き方規則の詳細化を図り、パンフレット形式にまとめることを挙げられました。

4.2. 特許文書分科会

IRD 国際特許事務所所長・弁理士／特許文書分科会主査谷川英和先生より本分科会での今年度の活動中間報告がありました。

まず、本分科会の概略が紹介されました。特許活動の一層の躍進を目標に、特許文書の標準化に取り組まれているとのことでした。特許文書の標準化が必要な理由として、特許実務には様々な立場の人が携わるため、立場毎に評価結果が異なるといった問題を挙げられました。

報告事項の第一点目は、特許文書の特性の分析結果でした。特許文書評価として人手・自動評価への展開を想定し、特許文書を多階層（大分類、中分類、小分

類）の特性により分類することが適切とのことでした。昨年度は特許文書の特性を示した表の初版を作成し、今年度は改良を施し公開したとのことでした。加えて、提案特性に基づき定性・定量的評価などを行ったとのことでした。

今後の課題として、評価特性の普及に向け、定性的／定量的評価事例の作成、産業日本語研究会ホームページを通じて意見募集、弁理士会、知財協等でのセミナー開催、日本知財学会などでの公開などを挙げられました。

4.3. 文書作成支援分科会

東京大学教授／文書作成支援分科会主査橋田浩一先生より本分科会での今年度の活動中間報告がありました。

まず、本分科会の研究背景として、人間の文章読解力が紹介されました。これまでの様々な研究が示すのは読解力の低さであり、これは子供に留まらず大人も該当するとのことでした。その事例として、人間の文章読解力が AI に劣る、米国の大学在学中に批判的思考力の向上が見られない（米国の大学でも無理なら日本の大学でも無理とも）といったことが紹介されました。本分科会は、文章理解の支援を通じ、読解力の向上、さらには作文力の向上も課題として活動を進めているとのことでした。

活動例として、文書理解の支援策が提示されました。文書理解を妨げる要因としてリニア（線形）な文配列に着目し、最適な文配列の構築を課題として位置付けられました。リニア配列では文間のつながり、つまり談話関係が不明瞭となるため、談話関係を明示する手段としてグラフ（図式）による文配列を検証されてきたとのことでした。これまでは既存の文書（特許請求項など）をグラフ化してきたため、今後は新規文書をグラフ構造により作成し、検証することで、文書作成の支援にもつなげていくとのことでした。

現在、グラフ配列のためのセマンティックエディタ

ツールを開発中とのこと。このツールは入力した単文をノード化し、ノード間の連結において談話関係（命題関係であり、照応関係は除く）を提示する機能を有するとのこと。

5. おわりに

本ワークショップへの参加を通し、産業日本語研究会の弛まない研究活動、そして、その研究射程の広さを実感しました。

岡先生の「場」に関する講演では、コミュニケーション支援においては、文字列で表される言語記号に加え、「場」を始めとする事態概念の記号化と解析が必要であることを確認しました。

また、野中先生のパテントマップに関する講演では、自然言語処理技術が学術的、社会的のみならず、知財・投資戦略といった産業的な分野でも貢献していることを確認しました。

そして、分科会の報告を通し、産業日本語の研究には、文書、言語を超え、人間の認知も取り込む必要性があることを実感しました。

産業日本語研究会のホームページ (<https://www.tech-jpn.jp/resources/>) から産業日本語研究会・シンポジウム資料・予稿集、そして、研究会の活動報告を確認できます。

参考文献

Hammarström, H. 2015. The basic word order typology: An exhaustive study. Paper presented at Diversity Linguistics: Retrospect and Prospect Conference, Leipzig, 3 May.

日本語マニュアルの会. 2016. 日本人のための日本語マニュアル：言葉の仕組みを学び、外国語との対照を通じて日本語スキルを磨く.
<http://ngc2068.tufs.ac.jp/nihongo/htdocs/>

産業日本語研究会の取組について

清藤 弘晃

一般財団法人日本特許情報機構 特許情報研究所調査研究部長

1. 産業日本語とは

「産業日本語」とは、「産業・技術情報を人に理解しやすく、かつ、コンピュータ（機械）にも処理しやすく表現するための日本語」として、一般財団法人日本特許情報機構（以下、Japio）が定義した造語です。

産業・技術情報を表現する際に利用される日本語を、「人に理解しやすく、かつ、コンピュータ（機械）にも処理しやすい」日本語とすることで、関係者間で情報を客観的かつ正確に伝達し、また、機械翻訳等の言語処理技術の活用をスムーズにして、情報伝達における関係者間の壁を越えと共に、これらの技術情報をより広く活用していくことを目指し、その研究が進められてきました。

2. 産業日本語研究会における取組の紹介

Japio では平成 19 年度より、特許情報の専門機関として、自然言語処理、特許翻訳、知的財産、情報工学等の専門家の協力を得ながら、特許文書に注目した「特許版・産業日本語」に関する検討を進めていましたが、より広い知見を有する各界の専門家の皆様による、さらなる研究・普及が必要と考え、平成 21 年度、高度言語情報融合フォーラム（ALAGIN）との相互協力のもと、「産業日本語研究会」を発足させました。

当該研究会では、特許分野に限らない産業日本語全般に関する活動を行っており、「ライティング分科会」、「文書作成支援分科会」、「特許文書分科会」の 3 つの分科会において、それぞれ異なる切り口で産業日本語に関する研究を進めています。また、毎年度、産業日本語研究会・シンポジ

ウムを開催し、研究会の活動報告、有識者による講演や最新トピックスの紹介等を行っています。

年に 4～5 回程度開催される各分科会での研究内容は、年度ごとに報告書にまとめられ、シンポジウムやワークショップ資料等と共に産業日本語研究会のホームページに掲載されています。

3. 今後の課題

産業日本語に対する機械処理は、「ビッグデータの普及」、「人工知能技術（AI）の発展」等により、今までと大きく様変わりしています。

機械翻訳アルゴリズムは、従来のルールベース翻訳や統計翻訳と全く異なる、ディープラーニング技術を活用した「ニューラル機械翻訳」の時代となっており、日本語情報を対象とする機械処理自体も、特許情報をはじめとしたビッグデータの分析・活用に向けて、テキストマイニングの発展や AI 適用等、様々な拡がりを見せています。このような新技術に対しても柔軟に対応可能な日本語の研究を進めることが、今後の重要課題です。

また、本取組は、その研究内容もさることながら、産業日本語の思想をその利用者や作成者の皆様に広め、次々に発生する文書自体を変えていくことが、重要なポイントとなります。そのため今後は「普及」という観点でも、よりいっそう積極的な取組を進めてまいりたいと考えています。

産業日本語研究会ホームページでは、特許ライティングマニュアルの無料 DL や研究内容に関する意見募集も行っています。皆様もまずは一度、上記サイト (<https://www.tech-jpn.jp/>) にアクセスし、産業日本語の世界に触れてみてください。

ニューラル機械翻訳を活かす翻訳支援ツール GreenT

新田順也

エヌ・アイ・ティー株式会社

1. GreenT のコンセプト

弊社はこれまで Word アドインとしての翻訳支援ツールや翻訳チェックツール（製品名：色 de チェック）などを翻訳者目線で開発をしてきた。このたび、ニューラル機械翻訳（NMT）の支援ツールである GreenT（グリーンティー）を開発したのでその特徴を紹介する。GreenT は翻訳者や専門家をターゲットユーザーとした翻訳支援ツールである。複数種類の機械翻訳エンジンの出力結果を比較しながら翻訳者目線で訳文作りに集中できる環境をユーザーに提供する。ウェブブラウザでもメールソフトでも利用可能なため、各種 CAT ツール（翻訳支援ツール）と併用できる。GreenT は NMT を用いて訳文を出力するが、用語集を適用できる。さらに、数字や用語の QA チェック機能も搭載しており、オールインワンの翻訳支援ツールである。



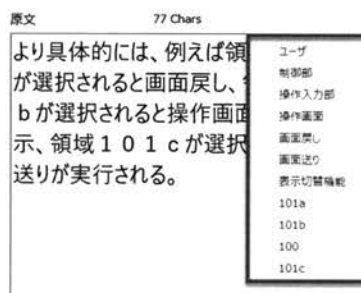
GreenT の最大の特徴は、NMT の出力をユーザーが積極的に制御できる点である。工夫次第で修正が最小限で済む訳文を出力できる。

2. 原文を編集する仕組み

最近の NMT は性能が向上し、原文によってはある程度正確な訳文を出力できる。しかし、原文の書き方

によっては出力される訳文が破綻することもあり、正確な訳文を出力するためには原文の編集が不可欠な場合がある。翻訳者である筆者は、対象となる原文を自由にかつ迅速に編集できるようにすることを目指し GreenT を開発してきた。日英翻訳における誤訳の原因は NMT の能力に起因する場合もあるが、原文の情報不足に起因する場合もよくある。和文では自明な主語や目的語を省略して表記することがある。そのため、翻訳に必要な情報が書かれておらず、NMT の出力がおかしくなるのだ。また、日本語における修飾語の記載順序の自由度が高いために修飾語の係り受けが分かりづらいことがある。通常翻訳者はこの係り受けを文章の内容や文法により特定して訳している。

この文章の修正要否の判断については、2019 年 4 月現在では AI 等の既存の技術では対応ができないと思われる。人間の判断の正確さとスピードを上回る技術がないわけだから、無理をしてコンピューターに任せない方がよい。翻訳者が頭の中で行う「必要な主語や目的語を補う作業」を GreenT の原文欄で実際にテキストを書き換えて実行する。このことで、出力文の品質が向上する。このような語句の補足支援として、対象原文とその前後の文中の名詞句を一覧にする機能がある。ユーザーはリストから選択するだけで、主語や目的語を入力できる。



また一般に機械翻訳しやすい日本語にするために、原文を短くすれば出力される訳文の品質が上がるということが知られている。そのため、必要に応じて訳文を分割もできる。分割機能として、句読点（コンマ・ピリオド）で改行を挿入する機能や、括弧内の文字列を文章の末尾に移動する機能もある。

このように自由に文章を分割・編集した場合でも、GreenT が翻訳メモリに蓄積する文章はあくまでも編集前の原文と最終的な訳文であるため、翻訳メモリ用のデータ管理面でも支障はない。

3. 原文の表現の変更例

和文の表現を別の訳しやすい表現に変える例をいくつか紹介する。このような表現は分野毎に異なり書き手によっても異なる。よって、万人に共通する変換というものは限られている。そのため、ユーザーが案件毎に出力される訳文を見比べながら登録していくことが望ましい。筆者が専門とする特許翻訳では以下のような変換をする。

原文表記	修正案
用いることとしてもよい	用いてもよい
することとしてもよい	してもよい
文献 1 の	文献 1 に記載された
に記載の	に記載された
設けた構成を有する	を含む
を含む構成を有している	を含む
実験を行う	実験をする
実験を実行する	実験をする
場合については	場合は

上記のような表現が原文で使われていると GreenT の原文編集欄（ブリエディット欄）に候補が表示される仕組みになっている。注意喚起としてコメントを表示してもよいし、そのまま使える修正案を提示してもよい。以下の場合、[適用] ボタンをクリックすると、選択箇所が「含む」に修正される。

原文 46 Chars
これは、前述した第 1 の実施の形態に係る画像表示装置を設けた構成を有するデジタルカメラである。

ブリエディット

ダブルクリックで次の検索

を設けた構成を有する

を含む

[Punctuation]

[Insert Paragraph]

有する

含む

これは

【これ】は何？

ユーザー設定

リセット

適用

4. 翻訳ノウハウで訳文の文体を変更

上記のような表現の修正ではなく、構文の修正案を提案する機能もある。翻訳者が NMT の出力文を嫌う理由の一つに、自身が使う文章構文とかけ離れた訳文が出力されるということがある。この場合、訳文全体を編集しないと思いつきの訳文にならないため、編集のストレスや負荷が大きくなってしまう。

この例として、「無生物主語の取り扱い」を紹介する。日本語の技術系の文章は受動態になりがちである。ところが、英文では対象物が生物ではない場合でも、対象物を動作の主体として能動態で表記することがある。いわゆる、無生物主語を用いた文である。このような特徴に基づき、日英翻訳では和文の操作対象を無生物主語として英文を書くことが翻訳技術として知られている。この知識を用いて原文の和文を無生物主語で書き直し、出力される英文を英語らしい表現に変更する例を紹介する。

(原文)

この論文では、消費増税の是非について議論されている。

(出力)

In this paper, the pros and cons of the consumption tax increase are discussed.

上記のような文章の場合、Google 翻訳では俗に言う「原文が透けて見える英文」が出力されることがある。つまり、日本語と語順が同じで、主語が長く動詞が文末にある英文である。

GreenT では、上記のような翻訳者の知識を原文の修正に反映できる。ブリエディット欄に修正案が表示されるので、ユーザーは候補から文脈にあわせて選択するだけで修正ができる。

J2E E21

新へ(J)

改へ(B)

Translate

QA

Style

TM

MemoSource

Setting

原文 26 Chars

この論文では、消費増税の是非について議論されている。

ブリエディット

ダブルクリックで次の検索

[Punctuation]

[Insert Paragraph]

について議論され

を議論し

について

が

では、

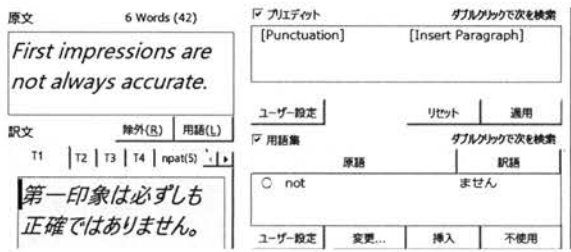
は、

ユーザー設定

リセット

適用

また、NMT では否定語の訳もれがときどき発生するため、原文も訳文ともに否定語がある場合には文字が斜体になり視覚情報で注意喚起がなされる。



8. 訳文の自動修正例

翻訳者の判断を必要とする箇所については、上記のように候補が表示されたり注意喚起がされたりするが、出力された訳文が明らかに間違えている場合は自動的に修正できる。この自動修正についてもユーザーが定義できる。以下は英日翻訳での修正例である。

訳文表記	修正案
ニューラル	ニューラル
ニューラルネットワーク	ニューラル・ネットワーク

9. 翻訳者と NMT とのつきあい方

Google 翻訳のような分野に特化していない汎用 NMT では、出力された訳文ではそのまま使えない場合が多い。しかし、用語集を適用し数字が正しく出力できるようにチューニングした翻訳エンジンであれば、出力文をそのまま使える場合が増えてくる。たとえば特許翻訳における図面の説明部分であれば、日英翻訳・英日翻訳ともにかなり正確な訳文を出力できる。英日翻訳で構造体を含む要素を説明するシンプルな文章 (A includes B, C and D.) であれば、用語集を適用できる GreenT のようなツールを使えば特許明細書で使えるレベルの翻訳ができてしまう。つまり、分野や言語方向によっては、その文章の一部で NMT 出力そのままでも利用可能な場合があるということだ。

今後は NMT が使える・使えないの大雑把な議論ではなく、翻訳支援ツールと組み合わせてどの分野のどのような文章で NMT を活用できるという議論になっていくと思われる。翻訳に求められる品質はクライアントや分野や言語方向により異なるので、NMT が使えるのかどうかは翻訳者自身が判断する必要があると思う。また、既存の翻訳手法・翻訳支援ツールとの親和性も翻訳全体の効率化と関連してくるので検討が必要になるだろう。そして翻訳エンジンによっても性能が異なるため、ツールと翻訳エンジンの組み合わせの結果を評価することにもなると思われる。

10. GreenT の今後の改良の方針

GreenT は Windows 上のあらゆるソフトウェアで呼び出して使えるようになっている。つまり、翻訳者が普段使っている翻訳支援ツール・CAT ツール (Memsource や Trados、MemoQ など) で GreenT を使える。各種ツールにあった細かな使い方には対応していないので、今後チューニングをしていく。また、GreenT で利用可能な翻訳エンジンの種類を増やし、分野毎に適切な翻訳エンジンを使えるようにしていきたい。

これ以外にも、プリエディットやポストエディットでの修正方法の研究を重ね、使いやすい候補を増やしていく。さらに、訳文の表記の統一方法は分野毎にルールがあるため、今後の改良で特許以外の分野にも広げていきたい。

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 のスタートから第二十五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。

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

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

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

個人会員（敬称略・50 音順）

会員名

藤崎 恵介

GI Japan 株式会社 (www.gi-japan.com/)

営業担当 / 翻訳コンサルタント

自己紹介

これまで印刷会社、DTP 制作会社での中国赴任経験を経て、2007 年に翻訳業界に転身。営業のポジションでは様々な業界の方とお話しする機会を得られ、さらに実際の業務でお役に立てることが大きな喜びです。

昨今の機械翻訳の潮流に関し、お客様の疑問に正しく自身の言葉でお答えするため、またお客様により良い選択をお示しできるように、貴団体で学ぶ機会を得られればと考えております。

何卒よろしくお願いします。

MT/翻訳とのかかわり

受注産業であり、フルオーダーメイドであり、効率化の波を被ってもなお労働集約型の典型的のような翻訳業界。他の産業と比較して一番近いのはどの産業だろうかよく考えますが、的確なものが思いつきません。私自身は、社内ではお客様の代弁者として、お客様には翻訳側の代弁者としての役割を担うよう心がけています。

MT につきましては、あくまで自身の新し物好きとしての興味の対象としての関わりが最初です。ビジネスに関するものとして自身で積極的に意識し始めたのはつい最近の 2017 年初旬からです。

MT および翻訳業界に期待すること

あくまで営業のポジションとしての立場では、お客様が喜んでくれるもの、満足してくれるものであって欲しいと願っています。また一方で、正直に申しますと MT が一般的に利用されるようになった場合の業界の行く末も大変気になります。

研究者さんや翻訳者さん、制作サイドの方だけでなく、私のような営業としての立場の者が正面から向き合い、理解を深め、そして業界外への周知・波及の一助になればと思います。

AAMT への要望

様々な立場の方々交流を通じ、研究開発の方向性のヒント、テクノロジーの活用方法のアイデア、新たなビジネスモデルの確立、新たな需要の発掘など、人や業界や、より広く社会の発展に貢献できる団体であればと思います。またそのために私も何かお役にたつことができればと考えています。

やや話が大き過ぎるかもしれませんが、そのような先々を見据えた活動を切に願います。

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 のスタートから第二十五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

日本特許翻訳株式会社

自己紹介

富士電機株式会社で感光ドラムの研究開発を行った後、富士ゼロックス株式会社で、有機感光体の開発・生産に携わり、2000 年から特許情報検索システム DocuPat の事業立ち上げ・推進をおこない、その後日本発明資料にて語順変換統計的機械翻訳エンジンを用いた英日・中日翻訳システムを開発、特許庁中国実案抄録翻訳事業に採用され、現在 100 万件以上 J-PlatPat で利用されています。

2015 年 2 月に日本特許翻訳株式会社を設立、会社設立して間もなく、ご縁があって、日本翻訳連盟の機関紙 JTF ジャーナルに 2 年間「機械翻訳の近未来」というタイトルで連載させていただきました。この記事がきっかけとなり、昨年 11 月には、「特許実務者・研究開発者のための AI 機械翻訳の最前線」を出版させていただきました。

弊社では、化学情報協会の JAICI AutoTrans サービス、株式会社プロパティの AI-TRANSEER/Infotranserをはじめ、翻訳会社、民間企業に NICT NMT エンジンと独自前処理、後処理技術で並列翻訳をおこなう Neural Translator システムを提供しています。また、特許庁から昨年に続き「英文抄録データ作成事業」を受注、今年 3 月に納品させていただきました。

MTおよび翻訳業界に期待すること

機械翻訳の高精度化のためには、翻訳者が個人として NMT を使いこなすノウハウと同時に、AAMT が翻訳精度の向上のため、業界としての方向性をまとめて、国レベルの施策を提言していくことも重要です。

たとえば特許翻訳についてみると、特許翻訳の特にクレーム部の機械翻訳では、構文解析が重要です。この構文解析のために、公報の xml に主題部タグ、主題部の前提部タグ、構成要素タグ（階層化可能形式）、構成要素間の説明部タグさらに課題タグ、用途タグ、作用効果タグ、請求項引用タグなどの新たな構造タグが翻訳精度にどのようにかわるかを調査し、効果が大きければ国レベルで標準化することが考えられます。

このようなタグ仕様を含む公報 xml 仕様が日本特許庁だけでなく外国特許庁（5 庁）間で共通化されると、①機械翻訳の高精度化（主題タグや構造タグ単位で翻訳して後処理すればよい）、②意味的な検索が可能となると考えられ、特許公報が技術文書としても活用されることが期待されます。

AAMT への要望

特許の機械翻訳は NMT の応用として先駆的な役割を担っており、特許庁でも J-PlatPat や OPD を通じたサービスが今年 4 月以降に実用化される計画です。特許翻訳のさらなる高度化のために、出願明細書の XML にあらたに構造タグ（クレームの表記に、主題部、構成要素部ならびに明細書の表記に課題（目的）部、効果部、構成要素参照タグ）を付与することで、翻訳精度向上にどのように寄与するかを NICT などの研究機関と JAPIO や民間翻訳会社が協力して調査する分科会の設立を要望します。もちろん実施にあたっては、関係機関のご賛同がえられた場合という条件付きですが・・・

協会活動報告

(2018 年 9 月～2019 年 5 月)

機械翻訳課題調査委員会

2018 年 9 月 3 日 (2018 年度 第 5 回)

1. JTF 翻訳祭について

- ・ 展示コーナーの検討
- ・ 委員会成果の展示の検討
 - ・ 準備作業
 - ・ 当日対応者
- ・ 会員勧誘
- ・ アンケート

展示コーナーでのアンケートは実施しない

AAMT のセッション（黒橋先生のご講演）時の前後にアンケート配布を行うこととする。

講演の聴講者にアンケートを配布して回答をお願いする。

準備作業：アンケート作成

2018 年 9 月 20 日 (2018 年度 第 6 回)

1. JTF 翻訳祭の準備

- ・ 委員会成果の展示
 - ・ <準備作業>
 - ・ 自動評価サイトデモ
 - ・ サイトの動作確認、デモ手順の整備
 - ・ PC の準備 → AAMT の備品確認
 - ・ 実機動作確認
 - ・ 活動成果に関する資料を準備
 - ・ NMT の品質評価結果 (SMT 比較)
 - ・ 言語処理学会の論文
 - ・ UTX パンフレット
- ・ アンケート
 - ・ アンケート作成
 - ・ 印刷

- ・会場へ持ち込み
- ・当日配布・回収の方法
- ・持ち帰り、集計

2018 年 10 月 12 日 (2018 年度 第 7 回)

1. JTF 翻訳祭の準備
 - ・ガイドライン確認
 - ・展示
 - ・アンケート
 - ・当日担当、シフト
2. これまでの活動（成果）の残し方・活用方法
 - ・ UTX 仕様書
 - ・基本残す
 - ・仕様書の方針
 - ・パブリックドメイン
 - ・リエイティブコモンズ
 - ・ UTX の辞書
 - ・そのまま残す
 - ・基本文テストセット
 - ・HP から link を設定する
 - ・基本公開の方針（クリエイティブコモンズ？ 利用規約検討）
 - ・評価サイトリンクも作る
 - ・AAMT/JAPIO との関係整理

2018 年月 11 月 15 日 (2018 年度 第 8 回)

1. JTF 翻訳祭の振り返り
 - ・総括
 - ・アンケート結果
2. AAMT 会員交流イベントの準備
3. これまでの資産の残し方 (UTX)

2018 年月 12 月 13 日 (2018 年度 第 9 回)

1. JTF 翻訳祭のアンケート結果
 - ・グラフ化したものを共有
2. AAMT 会員交流イベントの準備について
 - ・イベント理事会作成の資料をレビュー
3. これまでの資産の残し方 (UTX) の確認

- ・クリエイティブコモンズ (CC0) で OK

2019 年 1 月 25 日 (2018 年度 第 10 回)

1. AAMT 会員交流イベントの準備について
2. 課題調査委員会のラップアップ作業の確認
 - ・ WG1、WG2、WG3

2019 年 2 月 28 日 (2018 年度 第 11 回)

1. WG1 の活動のまとめ (残作業)
 - ・テストセットの AAMT ホームページからの公開に向けた作業
 - ① ページ (HTML) の作成
 - ② HP への組み込み (→IWG に依頼)
2. WG2 の活動のまとめ (残作業)
 - ・ファイルの整理と共用フォルダの設定
 - ⇒Web 調査、時系列評価などは、AAMT の事業として仕切り直し
3. WG3 の活動のまとめ (残作業)
 - ・creativecommons の CC0 の説明文の追加 (ホームページ、コンテンツ内)

2019 年 3 月 21 日 (2018 年度 第 12 回)

1. 2018 年度報告案作成
 - ・総会向け報告案の作成と読み合わせ

インターネットWG

2018 年 9 月

1. ウェブサイトに AAMT 長尾賞及び AAMT 長尾賞学生奨励賞の英語版情報を掲示した。
2. ウェブサイトに JTF 翻訳祭の情報を掲示した。

2018 年 10 月

1. ウェブサイトの会長挨拶を更新した。
2. ウェブサイトの役員情報を更新した。

2018 年 11 月

1. ウェブサイトに長尾真先生文化勲章受章のお知らせを掲示した。
2. ウェブサイトの組織情報 (日本語版・英語版) を更新した。

2019 年 1 月

1. ウェブサイトにアジア太平洋機械翻訳協会 (AAMT) セミナーの案内を掲示した。
2. サーバー障害発生に伴うウェブサイトアクセスの不具合への対応及びウェブサイトにも本件のお詫びを掲示した。

3. ウェブサイトに AAMT 長尾賞と AAMT 長尾賞学生奨励賞の案内を掲示した。

2019 年 3 月

1. ウェブサイトの AAMT 長尾賞規約と AAMT 長尾賞学生奨励賞規約を更新した。

編集委員会

2018 年 11 月 2 日 (2018 年度 第 3 回)

1. No.70 (この時点では 2018 年 3 月発行予定)の企画を行った。
2. およびその後の編集作業を行った。

2019 年 3 月 1 日 (2018 年度 第 4 回)

1. No.70 は 2019 年 6 月総会号として発行することとなった。追加の記事の原稿の検討および依頼分担を行った。
2. その後、継続して No.70(2019 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10月12日(金) (2018年度 第4回)

議事次第：

1. 議事録の確認

2. 発表

山岸 駿秀 (首都大学東京 大学院システムデザイン研究科 自然言語処理研究室)

松村 雪桜 (首都大学東京 大学院システムデザイン研究科 自然言語処理研究室)

3. 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)
 - 次回研究会予定
 - 主な議題
- 新年度の活動計画書

2018年12月21日(金) (2018年度 第5回)

議事次第：

1. 議事録の確認

2. 発表

単語分散表現に基づく Earth Mover's Distance と文長の違いを考慮した対訳文の自動抽出手法

田上 諒 (首北海学園大学大学院工学研究科電子情報生命工学専攻)

3. 第5回特許情報シンポジウム 開催報告

4. MT サミットワークショップ 開催提案について

- 運営体制 (Workshop/Program (co-)chair, Publication chair, Program committee members)
- 全体構成
- 招待講演者選定

2019年2月8日(金) (2018年度 第6回)

議事次第：

1. 研究発表

文脈考慮型ニューラル機械翻訳における最適文脈文選択法

木村龍一郎 / 飯田頌平 / 崔鴻翌 / 洪博軒 / 宇津呂武仁 / 永田昌明

(筑波大学大学院システム情報工学研究科) (NTT コミュニケーション科学基礎研究所)

2. 報告書の執筆について

3. MT サミットワークショップ 開催提案について

2019年3月8日(金) (2018年度 第7回)

議事次第：

- ・多言語音声翻訳研究プロジェクトのご紹介 (東工大 岡崎先生)
- ・平成 30 年度報告書内容 (各執筆者から簡単な紹介)
- ・MT サミットワークショップについて (招待講演者他)
- ・平成 31 年度の活動について (研究会開催時期, 内容, 部会)
- ・Japioが抱える機械翻訳に関する課題 (Japio 木下聡)

2019年5月17日(金) (2019年度 第1回)

1. 議事次第：議事録の確認
2. 発表 愛媛大学大学院 二宮 崇先生
「係り受け構造に対する相対的位置を用いた Transformer モデル」
3. 今年度研究計画発表
MT サミットワークショップについて
4. 2018年度報告書の内容紹介
5. 次回の開催について
 - ・今後の研究会の日時予定 (場所)
 - ・次回研究会予定 (令和元年7月開催予定)
 - ・主な議題

AAMT/Japio 特許翻訳研究会 拡大評価部会

2018 年 10 月 12 日(金) (2018 年度 第 2 回)

議事概要：

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

2019年3月8日(金) (2018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. 次回開催日予定

第5 回 特許情報シンポジウム

2018年12月7日 ビジョンセンター浜松町

開会挨拶

宇津呂 武仁（シンポジウム実行委員長・AAMT/Japio 特許翻訳研究会 副委員長、筑波大学）

招待講演

招待講演1「特許庁における機械翻訳活用の現状と今後の課題」

松谷 洋平（特許庁 総務部総務課特許情報室 室長補佐）

招待講演2「ニューラル翻訳時代の機械翻訳システムの社会実装」

井佐原 均（豊橋技術科学大学 情報メディア基盤センター長／情報・知能工学系 教授）

招待講演3「ニューラル機械翻訳の研究動向」

田村 晃裕（愛媛大学 大学院理工学研究科 助教）

特別講演「ユーザーから見たNMT の使い勝手と活用の展望」

日本知的財産翻訳協会(NIPTA) 特許機械翻訳研究会

奥山 尚一（久遠特許事務所代表, NPO 法人 日本知的財産翻訳協会理事長）

上野 哲也（株式会社MK 翻訳事務所）

湯浅 豊裕（WIS 知財コンシェル株式会社 翻訳事業部）

菊地 公一（DEN IP ソリューションズ株式会社 代表取締役 弁理士）

宮城 三次（一色国際特許業務法人 パートナー弁理士）

新田 順也（ブログ「みんなのワードマクロ」管理人）

一般講演

一般講演1「特許明細書の翻訳で注意すべきこと」

吉川 潔（翻訳業）3

一般講演2「請求項の記載を構造化する提案」

小池 誠（弁理士）

一般講演3「用語集による原文の一括置換と正規表現による一括並べ換えを用いた特許明細書の高品質翻訳」

杉山 範雄（株式会社杉山特許翻訳 代表）

一般講演4「知的財産リスクマネジメントにおける効果的な特許検索」

宮崎 祐、河野 京子（ヤフー株式会社 知的財産部）

閉会挨拶

小林 明（日本特許情報機構）

18:30-20:30 懇親会 ビジョンセンター浜松町 4F I室

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 知能機能工学科
宇津呂 武仁

AAMT ジャーナル 70 号をお送りします。

今号の巻頭言は、歴代会長の辻井潤一先生、井佐原均先生、中岩浩巳先生、隅田英一郎先生より、「AAMT 初代会長の長尾真先生の文化勲章ご受章を祝して」と題しまして、長尾真先生文化勲章ご受賞のお祝いのお言葉を頂きました。

今号に先立ちまして、2月12日(火)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 セミナー」が開催されました。XTM International Ltd.の目次 由美子様からは、この会の参加報告を御寄稿頂きました。また、昨年12月7日(金)には、AAMT/Japio 特許翻訳研究会の主催で、第5回特許情報シンポジウムが開催されました。同委員会副委員長・奈良先端科学技術大学院大学 須藤 克仁先生には、同シンポジウム開催報告を御寄稿頂きました。また、イベント紹介として、情報通信研究機構 内山将夫様からは、本年3月に開催された「第2回 自動翻訳シンポジウム ～自動翻訳と翻訳バンク～」の紹介を御寄稿頂きました。

その他、イベント報告として、奈良先端科学技術大学院大学 須藤 克仁先生より、昨年10月に開催された IWSLT (International Workshop on Spoken Language Translation)国際ワークショップの参加報告を御寄稿頂きました。また、関西外国語大学 小谷克則先生より、昨年12月に開催された平成30年度産業日本語研究会・ワークショップ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、昨年10月に開催された LocWorld38 Seattle 参加報告、ならびに、昨年11月に開催された Women in Localization Japan 第16回イベント参加報告を御寄稿頂きました。

日本知的財産翻訳協会 (NIPTA) 特許機械翻訳研究会の浜口宗武様、宮城三次様、宮本伸也様、湯浅豊裕様、平林千春様からは、研究報告「ユーザーから見た NMT の使い勝手と活用の展望」の御寄稿を頂きました。また、研究会紹介として、日本特許情報機構 清藤弘晃様からは、「産業日本語研究会の取組について」の御寄稿を頂きました。

サービス・製品紹介として、エヌ・アイ・ティー株式会社新田順也様からは、「ニューラル機械翻訳を活かす翻訳支援ツール GreenT」の御寄稿を頂きました。

その他、「AAMT 会員のひろば」の企画におきましては、法人会員・個人会員各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、最新の機会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

Asia-Pacific Association for Machine Translation

入会申込書

以下の通り、アジア太平洋機械翻訳協会の会員申し込みを致します。

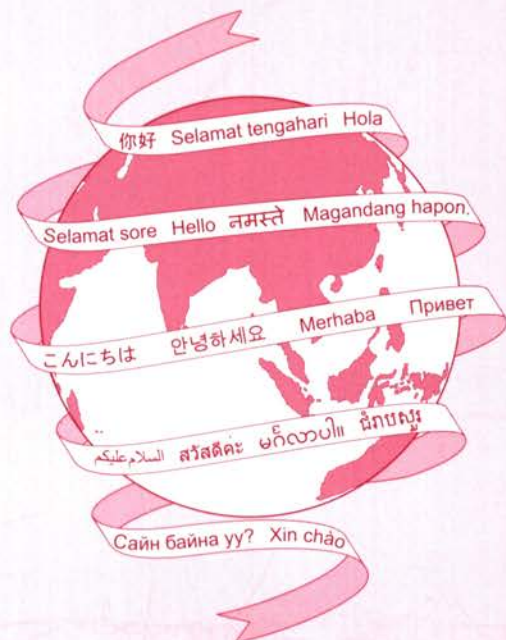
申込日 20 年 月 日

氏名（ローマ字）	
氏名（漢字）	
電話番号	
メールアドレス	
所属先	
所属先住所	〒
種別	ユーザ 研究開発者 その他
機械翻訳に関するお知らせメールの配信	希望する 希望しない

コメント

--

AAMT



AAMTジャーナル No.70

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：〒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

phone：0774-98-6332

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：石川 弘美 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 National Institute of Information and Communications Technology

3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, JAPAN

aamt-info@aamt.info

phone:+81(0)774-98-6332