2021年度AAMT/Japio特許翻訳研究会
報 告 書

機械翻訳及び機械翻訳評価に関する研究
並びに
第9回特許・科学文献翻訳ワークショップ
（PSLT 2021）開催報告

2022 年 3 月

一般財団法人　日本特許情報機構

目　　次

AAMT/Japio 特許翻訳研究会委員名簿

（敬称略・順不同）

委 員 長	辻井　潤一	国立研究開発法人 産業技術総合研究所 人工知能研究センター 研究センター長/ 東京大学大学院 名誉教授
副委員長	宇津呂武仁	筑波大学システム情報系教授
	須藤　克仁	奈良先端科学技術大学院大学 准教授
委　　員	今村　賢治	国立研究開発法人情報通信研究機構 先進的音声翻訳研究開発推進センター 主任研究員
	越前谷　博	北海学園大学大学院 教授
	江原　暉将	元・山梨英和大学 教授
	黒橋　禎夫	京都大学大学院 教授
	後藤　功雄	NHK 放送技術研究所スマートプロダクション研究部 主任研究員
	綱川　隆司	静岡大学学術院 情報学領域 講師
	中澤　敏明	東京大学大学院情報理工学系研究科 NEDO プロジェクト 実データで学ぶ人工知能講座 AI フロンティアコース 特任講師
	二宮　　崇	愛媛大学大学院 教授
オブザーバー	岡　　俊行	有限会社アジア産業 研究開発部長
	高　　京徹	株式会社高電社 経営企画部 部長
	園尾　　聡	東芝デジタルソリューションズ株式会社
	王　　向莉	株式会社ディープランゲージ

オブザーバー（(一財) 日本特許情報機構）：

大塩　只明	特許情報研究所 調査研究部 研究企画課
長部　喜幸	特許情報研究所 調査研究部 部長
木下　　聡	特許情報研究所 調査研究部
久々宇篤志	特許情報研究所 調査研究部 研究企画課 課長
小林　　明	専務理事/特許情報研究所 所長
土屋　雅史	情報活用支援部 情報活用支援課 係長
船戸さやか	特許情報研究所 調査研究部 研究企画課 主任
星山　直人	情報システム部 システム企画課 係長
三橋　朋晴	特許情報研究所 調査研究部 研究管理課 課長

事 務 局　　株式会社インターグループ

2021 年度 AAMT/Japio 特許翻訳研究会・活動履歴

2021 年 4 月 23 日
第 1 回 AAMT/Japio 特許翻訳研究会
（於オンライン開催）

2021 年 6 月 25 日
第 2 回 AAMT/Japio 特許翻訳研究会
（於オンライン開催）

2021 年 8 月 16 日
第 9 回特許・科学文献翻訳ワークショップ（PSLT 2021）
The 9th Workshop on Patent and Scientific Literature Translation (PSLT 2021)

2021 年 8 月 20 日
第 3 回 AAMT/Japio 特許翻訳研究会
（於オンライン開催）

2021 年 9 月 24 日
第 4 回 AAMT/Japio 特許翻訳研究会
（於オンライン開催）

2021 年 12 月 10 日
第 5 回 AAMT/Japio 特許翻訳研究会
（於オンライン開催）

2022 年 2 月 18 日
第 6 回 AAMT/Japio 特許翻訳研究会
（於オンライン開催）

1.　はじめに

AAMT－JAPIO　特許翻訳委員会委員長
産業技術総合研究所人工知能研究センター　センター長
辻井　潤一

ニューラル翻訳が大きな技術革新となり、機械翻訳が様々な場面で実用に供されるようになった。私も、海外の研究者と国際会議開催の準備をするときなど、まず、日本語で考えを整理し、それを機械翻訳で英語に翻訳してから、次に英語の文脈で考えることで英語のテキストを作る、といった使い方で機械翻訳のお世話になっている。

ただ、道具としての利便性が向上した半面、現在の機械翻訳技術に限界があることも確かである。自分の書いた日本語が頭の中にあり、翻訳された英語を無意識に日本語に直して読むために、英語としては意味が分からないテキストになっていることに気が付かないこともある。また、逆に、現在のニューラル翻訳の技術的な特質から、英語のテキストとしては自然なのだが、原文の意味からずれてしまっていることもある。原言語と相手言語でのコミュニケーションの在り方の差による問題、翻訳における意味の等価性とは何かといった翻訳論の問題にかかわる問題もある。今後は、人間の翻訳家と機械翻訳との共同作業の在り方、あるいは、一般のユーザによる機械翻訳の使い方という、翻訳における人間と機械の共同の在り方についての議論が必要になろう。

特許翻訳における道具としての機械翻訳も、このような広い観点からとらえ、次の技術的な課題を考えていく必要がある。

特許翻訳を科学・技術におけるコミュニケーションの観点からとらえると、専門用語の問題がある。ターミノロジーの専門家が指摘するように、科学技術分野の円滑なコミュニ―ケーションにとっては、専門的概念の言語表現への変換がその科学技術コミュニティの中で安定していることが必要である。同一概念が多様な言語表現を持ったり、逆に、異なった概念が同じ言語表現に縮退したりということが無秩序に起こると、科学・技術分野でのコミュニケーションは混乱する。

このことは、科学技術分野での翻訳においても同様である。原言語でのある専門用語を相手言語のどの専門用語に対応させるかは、分野専門家がコミュニティとして熟慮の上に決定している。翻訳においてこの対応が乱されると、その分野のコミュニケーションに混乱を生じる。科学技術分野での翻訳では、個別テキストがテキストとして理解できる翻訳が出力されるだけでは不十分で、専門用語が安定的に適切に翻訳されていることが不可欠である。

本委員会では、特許翻訳における専門用語の重要性を認識し、用語辞書の整備を行ってきた。その成果をどのように機械翻訳に系統的に反映させるかは、今後の大きな技術課題となる。

同様に、特許に頻出する長くて複雑な文の取り扱いも、重要である。特許においては、情報内容の複雑な論理構造を表現しなければならない要請があり、翻訳においても、原文が持つ論理構造を翻訳文が正確に反映していることが必須となる。長く複雑な複文構造をその背後の論理構造を保ちながら、翻訳文に移すことはそれほど簡単ではない。相手言語で理解できるテキストになっていることと、それが原言語テキストの論理構造を正確に反映しているかは別の問題であり、これをチェックする人間翻訳家のコストは非常に高いものとなる。複雑な論理構造を明確に表現する原言語での制限言語的なものの設計を含めて、長く複雑な文の取り扱いも、今後の大きな技術課題となる。

ニューラル翻訳は、大規模な翻訳対のデータを使うことで、原言語文の構造や意味を明示的にとらえることなく相手言語へと変換することで、大きな成功を収めた。ただ、このことから、原言語にはない情報が相手言語の文に付け加わったり、逆に、原言語にある情報（区や節）が相手言語のテキストですっぽり抜け落ちたりすることがある。情報の過不足、すなわち、原言語と相手言語での情報の等価性が大きく損なわれていることを検知する技術も、機械翻訳と人間の共同を円滑にするためには不可欠な技術となる。

機械翻訳技術が急速に発展したことから、翻訳に関する技術課題は解決したかのように思われている。しかし、上で述べたように、この技術がより広い分野で本格的に使われるためには、多くの挑戦的な課題が残されている。機械翻訳は、このような技術課題を整理し、新たな研究を組み立てる時期に来ている。今後の本委員会での活発な議論と技術の進展に期待している。

2. 年間報告書

2.1 機械翻訳のためのバイリンガルサブワード分割に関する研究

奈良先端科学技術大学院大学　出口　祥之
同志社大学　田村　晃裕
愛媛大学　二宮　崇
情報通信研究機構　内山　将夫
情報通信研究機構　隅田　英一郎

2.1.1 はじめに

現在、ニューラルネットワークを用いた機械翻訳（ニューラル機械翻訳）が機械翻訳の主流となっている。注意機構に基づく LSTM（Long Short-Term Memory）(Luong et al. 2015) は初期のころから広く使用されてきたニューラル機械翻訳モデルである。このモデルは、原言語文（翻訳元言語の文）内の単語と目的言語文（翻訳先言語の文）内の単語間の関係を捉える言語間注意機構を用いることで高い精度を実現した。また、近年、トランスフォーマー（Transformer）モデル (Vaswani et al. 2017) が LSTM や畳み込みニューラルネットワーク（Convolutional Neural Network; CNN）を用いた手法と比べて高い精度を達成し、注目されている。トランスフォーマーモデルは、従来の言語間注意機構に加えて、同じ文中の単語間の関係を捉える自己注意機構を導入している。

これらのニューラル機械翻訳モデルは高い精度と流暢性を同時に実現するが、予め指定した語彙に基づいて学習および翻訳を行うため、翻訳時の入力文に低頻度語や未知語が現れると翻訳精度が低下する問題が知られている。このような語彙の問題に対処するため、バイトペア符号化（Byte Pair Encoding、以下 BPE）(Sennrich et al. 2016) やユニグラム言語モデル (Kudo 2018a) などによるサブワード分割が現在広く用いられている。サブワードは、一般に単語よりも短く、文字よりも長い単位であり、例えば、「pretraining」は「pre」と「train」と「ing」の 3 つのサブワードに分割されることが考えられる。サブワード分割は、形態素解析のように言語学的知見や規則に基づいた分割を行うのではなく、語彙量を指定した上で、コーパスから自動的にサブワード語彙を抽出し、分割することを特徴とする。語彙に「pretraining」という単語が登録されていなくても、「pre」「train」「ing」がサブワードとして語彙に登録されていれば、「pretraining」を完全な未知語として扱うのではなく、「pre」「train」「ing」をニューラル機械翻訳の入力/出力トークンとすることができ、未知語の問題を大きく緩和することができる。

BPE によるサブワード分割は事前トークナイズを必要とするのに対し、ユニグラム言語モデルは生文からサブワード列に直接分割するため、日本語や中国語といった分かち書きされない言語においても形態素解析器を必要としない。BPE やユニグラム言語モデルはどちらもデータ圧縮に基づいたアルゴリズムであり、語彙量の上限を制約としたトークン数の最小化を行っている。語彙量を減らす方法としては文字単位に分割するという方法も考えられるが、文字単位の分割を用いると文全体のトークン数が増える（系列長が長くなる）ため、系列長に依存した計算量が増加する。サブワード分割によって、語彙量の上限を制約として満たす中でトークン数を最小化する

ことで、トレードオフの関係にある語彙量とトークン数（系列長）の問題に対処しているといえる。

しかしながら、これらの分割法は対訳関係を考慮せず、各言語ごとにサブワード分割を学習するため、機械翻訳タスクに適したサブワード分割になるとは限らない。例として、日英翻訳において「設計法（design method）」と「計測装置（measurement instrument）」という複合語が訓練データに多数出現する場合を考える。従来のサブワード分割法はデータ圧縮技術に基づきトークン数の最小化を行うため、これらの複合語が 1 つのサブワード単位に結合される。したがって、これらの訓練データは「計測法」という語の翻訳の学習に寄与しない。

本稿は、国際会議「The 28th International Conference on Computational Linguistics (COLING'2020)」および国内ジャーナル「自然言語処理」において我々が提案したバイリンガルサブワード分割法（Deguchi et al. 2020; 出口ら 2021）について報告する。提案法であるバイリンガルサブワード分割法は、ユニグラム言語モデル（Kudo 2018a）に基づき、対訳コーパスを用いて行うサブワード分割手法である。具体的には、ユニグラム言語モデルによって得られる原言語文と目的言語文それぞれの分割候補から、お互いのトークン数の差が小さくなるサブワード列を選択する方法となっている。提案法はユニグラム言語モデルに基づくため、分かち書きされない言語にも適用可能である。

提案法では、ユニグラム言語モデルから得られる原言語文と目的言語文の最尤解を比較し、トークン数が多い言語側のトークン数に近づけるように、より細かい単位のサブワード分割を複数分割候補から選択する。提案法を用いることで、原言語文と目的言語文のトークン数の差が小さくなり、言語間でトークンが 1 対 1 に対応付けされやすくなる。そのため、従来のサブワード分割法よりもニューラル機械翻訳に適した分割が得られることが期待される。提案法では日本語文と英語文のサブワードトークン数を近づけるため、課題例として挙げた「設計法」と「計測装置」という複合語は「設計（design）」と「法（method）」、「計測（measurement）」と「装置（instrument）」、それぞれ 2 トークンに分解される。これにより、ニューラル機械翻訳において、「設計」と「法」、「計測」と「装置」というそれぞれのサブワードの訓練データが「計測法」という語の翻訳にも活用できるようになると考えられる。

本手法は原言語文と目的言語文の分割数を比較しながらそれぞれの文を分割するため、原言語文単体では分割ができない。ニューラル機械翻訳の訓練時には原言語文と目的言語文の分割数を比較するために対訳コーパスを用いることができるが、翻訳時には原言語文に対応する目的言語文が存在しないため、原言語文を分割することができない。そこで提案法では、対訳コーパスを用いてサブワード分割した訓練データの原言語文から LSTM ベースのサブワード分割器を予め学習し、翻訳時において訓練時の分割に近い候補を選択することで、訓練時と翻訳時の分割のギャップを小さくして翻訳性能の低下を防ぐ。具体的には、翻訳時に、学習した LSTM ベースのサブワード分割器により原言語文のサブワード分割候補をリランキングし、スコアが最大となる分割を選択する。

WAT Asian Scientific Paper Excerpt Corpus（以下、ASPEC）(Nakazawa et al. 2016) 英日・日英・英中・中英翻訳タスクと WMT14 英独・独英翻訳タスクにおいて、従来法と提案法を用い

た翻訳性能を比較したところ、トランスフォーマーニューラル機械翻訳モデルの性能が最大 0.81 BLEU ポイント改善した。

2.1.2 ユニグラム言語モデルに基づいたサブワード分割

本節では提案法の基礎となるユニグラム言語モデルに基づいたサブワード分割法（Kudo 2018a）について説明する。ユニグラム言語モデルでは各サブワードが独立に生起すると仮定し、サブワード列 $s = (s_1, s_2, \cdots, s_N)$ の生起確率 $P(s)$ を次式により表す。

$$P(s) = \prod_{i=1}^{N} P(s_i)$$

$$\sum_{v \in V} P(v) = 1$$

ただし、V は語彙集合（サブワード辞書）である。各サブワードの生起確率 $P(s_i)$ は EM アルゴリズムによって周辺尤度 L を最大化することにより推定される。

$$L = \sum_{x \in D} \log P(x) = \sum_{x \in D} \log \sum_{s \in S(x)} P(s)$$

ただし、D は対訳コーパスであり、x は D 中の原言語文または目的言語文であり、$S(x)$ は x のサブワード分割候補集合である。

x を入力文としたとき、x に対する生起確率が最大となるサブワード列（最尤解）は次式によって得られる。

$$s^* = \underset{s \in S(x)}{\operatorname{argmax}} P(s)$$

また、k-best 分割候補も入力文 x に対するユニグラム言語モデルによって計算される確率 $P(s)$ に基づいて得ることができる。ただし、サブワード列の生起確率は各サブワードの尤度の積の形で表されるため、系列長の短い（トークン数の少ない）サブワード列が高い確率を持つ傾向がある。

このユニグラム言語モデルによるサブワード分割は生文から直接学習できるため、日本語や中国語といった分かち書きされない言語においても単語分割器や形態素解析器を必要とせずに分割できるという特長がある。

2.1.3 バイリンガルサブワード分割

本節では、対訳文からサブワード列を得る提案手法を示す。我々の提案法では対訳文対でサブワードトークン数の差が最小になるような分割を行う。具体的には、原言語文と目的言語文それぞれのユニグラム言語モデルの最尤解のうち、トークン数の少ない（系列長の短い）側の文を、トークン数が多い側のトークン数に近づけるよう、より細かく分割された分割候補からサブワード列を選択する。ただし、ニューラル機械翻訳の訓練時には対訳コーパスを利用できるが、翻訳時（評価データ）には対訳文が存在しない。そこで、ニューラル機械翻訳の訓練時と翻訳時で異なる方法によりサブワード列を得る。図 1、2 にニューラル機械翻訳訓練時のサブワード分割と翻

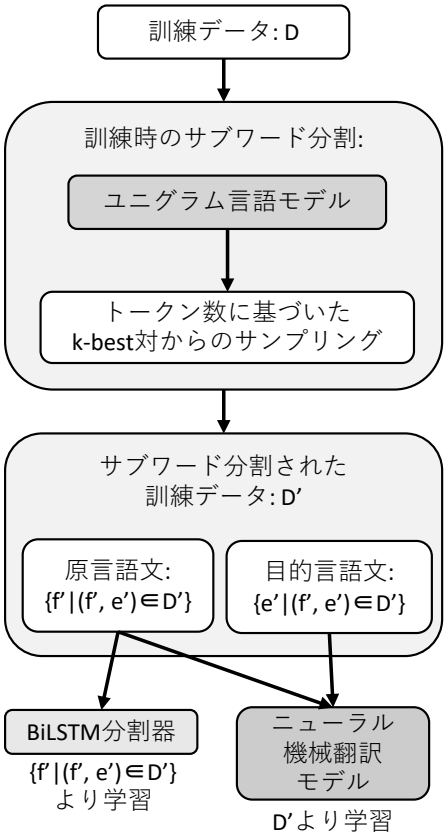

図 1:　訓練時のサブワード分割

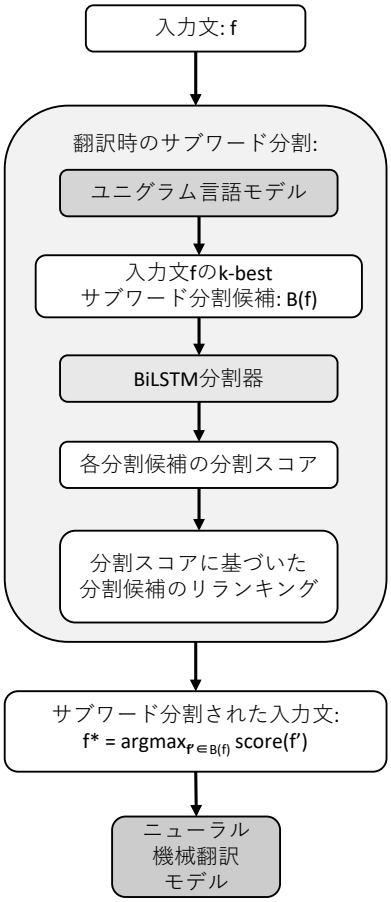

図 2:　翻訳時のサブワード分割

訳時のサブワード分割をそれぞれ示す。ニューラル機械翻訳モデルの訓練時は、図 1 の通り、対訳データに基づくサブワード分割結果を用いてニューラル機械翻訳モデルを学習する。一方で翻訳時には、図 2 の通り、対訳データのサブワード分割結果内の原言語文だけから予め学習しておいた LSTM ベースの単語分割器を用いて、翻訳対象の原言語文のサブワード分割候補をリランキングする。

提案法はニューラル機械翻訳モデルや訓練法を修正する必要がなく、従来のサブワード分割法を置き換えるだけで適用可能である。

2.1.3.1　訓練データのサブワード分割

対訳コーパスの訓練データに対するサブワード分割では、ユニグラム言語モデルによる分割候補からトークン数が近い候補対を選択することで、対訳文対の分割を得る。

具体的には、訓練データ D 中の各原言語文 f と目的言語文 e の対に対し、次の手続きを行うことによって、サブワード分割を得る。

1. ユニグラム言語モデルにより、原言語文 f と目的言語文 e それぞれの k-best の分割候補 $B(f)$、$B(e)$ を得る。
2. $B(f)$ の最大確率のサブワード列を f^*、$B(e)$ の最大確率のサブワード列を e^* とする。

3. f^* と e^* のトークン数（サブワード数）を比較する。
   (ア) f^* と e^* のトークン数が同じ場合は、(f^*, e^*) を出力する。
   (イ) f^* の方が長く（トークン数が多く）、e^* の方が短い（トークン数が少ない）場合は、$B(e)$ の中で f^* とトークン数が最も近い候補の中で最大確率のサブワード列 e^{**} を探し出し、(f^*, e^{**}) を出力する。
   (ウ) f^* の方が短く、e^* の方が長い場合は、(イ)における f^* と e^* を入れ替えて同じ処理を行う。$B(f)$ の中で e^* とトークン数が最も近い候補の中で最大確率のサブワード列 f^{**} を探し出し、(f^{**}, e^*) を出力する。

トークン数が多いということはより細かいサブワードに分割されている、ということである。3.のステップにおいて、原言語文 f と目的言語文 e に対する最大確率のサブワード列 f^* と e^* のうち、より長い方のサブワード列を出力としてまず確定する。つまり、より細かく分割されているサブワード列を出力として確定する。短い方の最大確率のサブワード列は採用せず、分割候補の中からより細かく分割された（より長い）サブワード列を探索し、長い方のサブワード列と同じ長さのサブワード列探し出し、その中で最大確率のサブワード列を出力する。

上記の提案法によって訓練データ D の各対訳文をサブワード分割した訓練データを D' とする。ニューラル機械翻訳モデルは訓練データ D' から学習される。

2.1.3.2 翻訳時のサブワード分割

翻訳時は入力文 f に対する対訳文 e が存在しないため、サブワード分割の入力に対訳文を用いることができない。そのため、予め 2.1.3.1 節で作成した訓練データ D' の原言語文だけを集めたデータから、文字ベースの双方向 LSTM（Bidirectional LSTM、以下 BiLSTM）を用いたサブワード分割器（以下、BiLSTM 分割器）を学習しておく。翻訳時の分割は、ユニグラム言語モデルの k-best 分割候補を BiLSTM 分割器に入力し、各分割候補に対してスコア付けを行いリランキングすることにより得られる。

BiLSTM 分割器は、n個の文字からなる入力文字列 $c = (c_1, c_2, \ldots, c_n)$ に対して、サブワードの開始文字か否かを表す境界タグを割り当て、サブワードの境界点を 2 値分類として識別する。BiLSTM 分割器は以下のような構造のニューラルネットワークである。

$$Z = Embedding(c)$$
$$H = BiLSTM(Z)$$
$$B = softmax(HW)$$

ただし、$Embedding$ は文字埋め込み層、Z は文字列 c の d 次元埋め込み表現、$BiLSTM$ は BiLSTM 層、H は BiLSTM の中間表現、$softmax$ は softmax 関数、B は BiLSTM 分割器の出力、$W \in R^{d \times \{0,1\}}$ は中間表現から境界タグ次元に写像するパラメータ行列である。ベクトル $B_i = (b_{i,0}, b_{i,1})$ は文字 c_i がサブワードの開始点か（$b_{i,0}$）、開始点でないか（$b_{i,1}$）の確率分布を表現している。BiLSTM 分割器は、2.1.3.1 節の方法でサブワード分割された訓練データ D' 中の原言語文 f' $((f', e') \in D')$ について、以下の目的関数 $L_{segment}$ を最大化することにより学習される。

$$L_{segment} = \sum_{i=1}^{n} \log B_{i,r(i)}$$

$$r(i) = \begin{cases} 0 & if\ c_i \text{ はサブワードの開始点} \\ 1 & otherwise \end{cases}$$

翻訳時は次のようにして入力文 f のサブワード分割を行う。はじめに、ユニグラム言語モデルを用いて入力文 f の k-best サブワード分割候補 $B(f)$ を得る。次に、各分割候補 $f' \in B(f)$ のスコア $score(f')$ を、予め学習しておいた BiLSTM 分割器によって以下のように算出する。

$$score(f') = \sum_{i=1}^{n} \log B_{i,r(i)}$$

最後に、最大のスコアを持つサブワード列を選択し、出力とする。

$$f^* = \underset{f' \in B(f)}{\operatorname{argmax}}\, score(f')$$

以上により得られたサブワード列 f^* をニューラル機械翻訳モデルに入力し、翻訳を行う。

2.1.4 実験

2.1.4.1 実験設定

提案法と従来法（ユニグラム言語モデル（Kudo 2018a））の翻訳性能を比較した。また、従来法として、ユニグラム言語モデルによって得られる複数のサブワード分割候補について周辺尤度を最大化する「サブワード正則化（Kudo and Richardson 2018b)」とも性能を比較した。複数サブワード分割候補を得るためのユニグラム言語モデルには Sentencepiece[1]を用いた。全実験において、ニューラル機械翻訳システムとして Transformer base モデル（Vaswani et al. 2017）を用いた。

翻訳性能は WAT ASPEC 日英・英日（以下 ASPEC 日-英）翻訳タスク [2]（Nakazawa et al. 2016）を用いて評価した。ユニグラム言語モデルの学習は、原言語側と目的言語側でそれぞれ独立に行い、サブワードの語彙量は、原言語側と目的言語側でそれぞれ 16,000 になるように設定した。ミニバッチの大きさは約 10,000 トークンになるよう設定した。ニューラル機械翻訳モデルの訓練には訓練データの上位 150 万文対を使用し、データの前処理は WAT ベースラインシステム [3]に従った。開発データと評価データのデータ数はそれぞれ 1,790、1,812 文対であった。

全ニューラル機械翻訳モデルにおいて、パラメータ最適化には Adam（Kingma and Ba 2015）を用い、$\beta_1 = 0.9$、$\beta_2 = 0.98$ とした。モデルのパラメータ更新は 10 万回行った。学習率は 4,000 回更新時で $5e-4$ となるように線形に増加させ、以降は更新回数の逆平方根に比例して減衰させた（Vaswani et al. 2017)。ドロップアウトの確率は 0.1 に設定した。ニューラル機械翻訳モデルの損失関数にはラベル平滑化交差エントロピー（Szegedy et al. 2016）を用い、平滑化 ε は 0.1 に設定した。パラメータの更新 1,000 回毎にモデルを保存し、性能評価時には、訓練終了時点から前 5 つ分のモデルパラメータを平均化したモデルを用いた。翻訳文の生成にはビーム探索を用い、

[1] https://github.com/google/sentencepiece
[2] http://lotus.kuee.kyoto-u.ac.jp/ASPEC/
[3] http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2019/baseline/dataPreparationJE.html

表 1: ASPEC 日-英における翻訳性能の比較（BLEU(%)）

	日英	英日
ユニグラム言語モデル	28.58	43.19
サブワード正則化	28.86	43.10
BiSW（提案法）	$^{+,++}$**29.39**	$^{+,++}$**43.29**

ビーム幅は 4、文長正則化パラメータは 0.6 (Wu et al. 2016) とした。

提案法のハイパーパラメータに関して、ユニグラム言語モデルから得るサブワード分割候補数 k は開発データで調整し、5 に設定した。BiLSTM 分割器の埋め込み次元数は $d = 256$ とし、BiLSTM 層は 2 層スタックした。文字埋め込み層、BiLSTM 層、出力層のパラメータは全て $[-0.1, 0.1]$ の一様分布で初期化した。BiLSTM 分割器のパラメータ最適化には Adam を用い、$\beta_1 = 0.9$、$\beta_2 = 0.98$ とした。モデルのパラメータ更新は 10 エポック分行った。学習率は$5e-4$、ドロップアウトの確率は 0.1、ミニバッチの大きさは約 256 文にそれぞれ設定した。

サブワード正則化を用いたモデルでは、提案法と条件を揃えるため、ユニグラム言語モデルの最大スコアのサブワード列を翻訳する 1-best デコードを使用した。

2.1.4.2 実験結果

表 1 に実験結果を示す。表中の 「ユニグラム言語モデル」、「サブワード正則化」、「BiSW」はそれぞれ、ユニグラム言語モデル、サブワード正則化、提案法を用いたニューラル機械翻訳モデルを示している。翻訳性能は BLEU (Papineni et al. 2002) で評価し、評価方法は WAT Automatic Evaluation Procedures[4]に従った。また、ブートストラップ再サンプリングによる有意差検定 (Koehn 2004) を実施し、有意水準は 5%とした（$p \leqq 0.05$）。表 1 中の「+」は「BiSW」が「ユニグラム言語モデル」に対して、「++」は「BiSW」が「サブワード正則化」に対して、有意に高いことを示す。

表 1 から分かるとおり、提案法「BiSW」は日英、英日翻訳の両言語方向において 「ユニグラム言語モデル」および 「サブワード正則化」より性能が改善されている。「BiSW」を用いることで「ユニグラム言語モデル」に対し、日英、英日翻訳においてそれぞれ 0.81、0.10 BLEU ポイント、「サブワード正則化」に対し、それぞれ 0.53、0.19 BLEU ポイントの性能改善が確認された。また、両言語方向において、提案法はベースラインの「ユニグラム言語モデル」、及び「サブワード正則化」より有意に性能が高く、提案法の有効性が確認できる。

2.1.4.3 提案法によるサブワード分割の例

本節では従来法「ユニグラム言語モデル」と提案法で得られるサブワードの違いを実例で確認する。表 2 に、APSEC 日英の訓練データに対して従来法と提案法をそれぞれ適用した実際の例を示す。表 2 より、従来法では複数の意味から成るサブワードが 1 トークンに結合されているの

[4] http://lotus.kuee.kyoto-u.ac.jp/WAT/evaluation/index.html#automatic_evaluation_systems.html

表 2: ASPEC 日英翻訳の訓練データにおけるサブワードの例

ユニグラム言語モデル	BiSW	訳文中の対応箇所
helper	help er	ヘル パー
basically	basic ally	基本的に は
focused	focus ed	に着目し た
popularization	popular ization	普及
第三 者	第 三 者	the third person
骨密度	骨 密度	bone density
設計法	設計 法	design method

表 3: ASPEC 日英翻訳の評価データにおけるサブワードの例

ユニグラム言語モデル	BiSW	訳文中の対応箇所
密度分布	密度 分布	density distribution
分散型	分散 型	dispersion type
透 水性	透 水 性	permeability

に対し、提案法ではそれらが分解されていることが分かる。また、表中に訳文中の対応箇所を示す。表より、「設計法」が訳文中の「design method」と対応付いて「設計」と「法」に分割されていることが確認できる。これは、従来法が生起確率のみに基づいて分割されるのに対し、提案法では対訳相手の分割情報を参照しているためであるといえる。これにより、原言語文と目的言語文間でサブワードが対応付けられやすくなり、ニューラル機械翻訳モデルの学習を支援できるようになると考えられる。ただし、「popularization」の訳文中の対応箇所は「普及」という 1 トークンであるのに対し、提案法による分割では「popular」と「ization」に分割されている。これは、単語単位ではなく、文単位でトークン数を近づけるため、「普及」以外のトークンの分割を参照し、訳文中の対応箇所とトークン数の対応がない分割を行ったと考えられる。

表 3 に APSEC 日英翻訳の評価データ（評価データの日本語文）に対して従来法と提案法をそれぞれ適用した実際の例を示す。表 3 より、評価データにおいても、対訳文（参照訳）を参照することなく、BiLSTM 分割器により、言語間で 1 対 1 のサブワードの対応付けを取りやすい単位に分解されていることが分かる。ただし、提案法において「透 水 性」と分割された例は、訳文中の対応箇所が 「permeability」であるのに対し、従来法の「透 水性」よりもトークン数の差が大きくなっている。これは、文単位でトークン数を近づけた訓練データから BiLSTM 分割器を学習させているため、もしくは、BiLSTM 分割器が誤って予測しているためであると考えられる。

2.1.4.4 分かち書きされた言語対及び分かち書きされない言語対に対する提案法の有効性

本節では、英独翻訳のような分かち書きされている言語対、及び、日中翻訳のような分かち書きされない言語対の翻訳に対する提案法の有効性を検証する。

表 4: WMT14 英-独及び ASPEC 日-中翻訳における提案法の有効性（BLEU(%)）

	WMT14 英-独		ASPEC 日-中	
	英独	独英	日中	中日
ユニグラム言語モデル	26.45	30.62	35.21	47.59
BiSW（提案法）	**26.77**	**30.64**	**35.33**	**47.71**

翻訳性能の評価には、それぞれ WMT14 英独・独英（以下 WMT14 英-独）翻訳タスク [5]と ASPEC 日中・中日（以下 ASPEC 日-中）翻訳タスクを用いた。

本実験におけるユニグラム言語モデルの学習は、原言語側と目的言語側で辞書を共有して行った。サブワードの語彙量は、WMT14 英-独で 37,000、ASPEC 日-中で 16,000 に設定し、ニューラル機械翻訳モデル内の原言語側と目的言語側の埋め込み層を共有した。ミニバッチの大きさは WMT14 英-独で約 25,000 トークン、ASPEC 日-中で約 6,000 トークンになるよう設定した。WMT14 英-独の訓練データにおいて、各文をサブワード分割した後、250 トークンを超える文と原言語/目的言語文のトークン数の比が 1.5 を超えるものを除去した。ハイパーパラメータ k は開発データで調整し、2 に設定した。

表 4 に WMT14 英独・独英翻訳、ASPEC 日中・中日翻訳の実験結果を示す。表 4 より、両言語方向において、提案法「BiSW」を用いることで従来法「ユニグラム言語モデル」と比べて翻訳性能が改善されることが確認された。具体的には、英独、独英、日中、中日翻訳において、「BiSW」は「ユニグラム言語モデル」と比べてそれぞれ 0.32、0.02、0.11、0.12 BLEU ポイント性能が改善された。実験結果より、分かち書きされた言語対及び分かち書きされない言語対に対しても提案法の有効性が確認された。

2.1.5 関連研究

BPE (Sennrich et al. 2016) とユニグラム言語モデル (Kudo 2018a) はサブワード分割法として広く用いられている。BPE は辞書式圧縮に基づいたサブワード分割アルゴリズムであり、指定した語彙量を上限として、出現回数順に隣接するサブワードを再帰的に結合する。BPE は簡単なアルゴリズムで実装が容易なため多くの NMT システムで採用されているが、決定的アルゴリズムであるため複数の分割候補を得ることができない。

ユニグラム言語モデルは尤度に基づいたサブワード分割アルゴリズムである。各サブワードの生起確率は EM アルゴリズムによって推定される。ユニグラム言語モデルは BPE と比べてアルゴリズムが複雑であるが、尤度に基づいた複数のサブワード分割候補を得られ、かつ、事前トークナイズを必要とせず生文から直接学習できるという特長がある。本研究の実験ではユニグラム言語モデルのリファレンス実装である SentencePiece (Kudo and Richardson 2018b) を用いた。

サブワード正則化 (Kudo 2018a) は複数のサブワード分割候補を用いたニューラル機械翻訳の訓練法であり、サンプリングされた分割候補の周辺尤度を最大化する。サブワード正則化をニュ

[5] https://www.statmt.org/wmt14/translation-task.html

ーラル機械翻訳に組み込むには、訓練時にパラメータを更新するごとに動的にサブワード分割をサンプリングする必要があり、ニューラル機械翻訳の訓練処理を修正する必要がある。

BPE-Dropout (Provilkov et al. 2020) はサブワード正則化を用いられるように BPE を拡張した手法である。BPE-Dropout では、隣接サブワードの結合を確率的に棄却することで複数のサブワード分割候補が得られる。ただし、$P(s|x)$ のような尤度に基づいたk-best 候補を得ることはできない。

単語やサブワードへの分割を行わずに文字単位で翻訳を行うニューラル機械翻訳モデルも提案されている。Cherry ら (2018) は単語単位やサブワード単位のニューラル機械翻訳よりも文字単位のニューラル機械翻訳の翻訳性能が高くなると報告している。ただし、Cherry らは文字単位のニューラル機械翻訳の問題点として計算量の多さとモデリングの難しさがあることも述べている。我々の手法はニューラル機械翻訳モデルの入出力の粒度について文字単位のニューラル機械翻訳の長所と短所（翻訳性能とモデリング、計算量）のバランスをとったものと考えられる。

Ataman ら (2017)、Ataman と Federico (2018b)、Huck ら (2017) は言語学に基づくサブワード分割を提案している。Ataman ら (2017)、Ataman と Federico (2018b) は教師なし形態学習に基づく「Linguistically Motivated Vocabulary Reduction (LMVR)」を用いることで BPE より翻訳性能が向上することを示した。Huck ら (2017) はサブワード分割においてステミングや複合語分割などによる言語学的な知識を用いた分割を組み合わせることで、翻訳性能が改善することを示した。また、Ataman と Federico (2018a) は単語をn-gram 文字で分解することで形態学的にリッチな言語を含む翻訳が改善することを示している。

2.1.6 おわりに

本稿は､我々が提案するバイリンガルサブワード分割法 (Deguchi et al. 2020; 出口ら 2021) について紹介した。提案法は、対訳文からサブワード列を得る、ニューラル機械翻訳のためのサブワード分割法である。WAT ASPEC 英日・日英・英中・中英翻訳タスクと WMT14 英独・独英翻訳タスクの実験を行い、提案法を用いることでトランスフォーマーニューラル機械翻訳モデルの性能が最大 0.81 BLEU ポイント改善されることが示された。これらの実験結果より、対訳文とのサブワードトークン数の差を小さくすることで翻訳性能が改善されたと考えている。今後は他の言語対でも提案法の有効性を確認していきたい。

謝辞

本稿は、国際会議「 The 28th International Conference on Computational Linguistics (COLING'2020)」に採択された論文 (Deguchi et al. 2020) および国内ジャーナル「自然言語処理」に採択された論文 (出口ら 2021) に基づいて、それらの論文を再構成し、解説したものである。

これらの研究成果は、国立研究開発法人情報通信研究機構の委託研究（課題 197「多言語音声翻訳高度化のためのディープラーニング技術の研究開発」）により得られたものである。ここに謝意を表する。

参考文献

D. Ataman, M. Negri, M. Turchi and M. Federico. (2017). Linguistically Motivated Vocabulary Reduction for Neural Machine Translation from Turkish to English. arXiv preprint arXiv:1707.09879.

D. Ataman and M. Federico. (2018a). Compositional Representation of Morphologically-Rich Input for Neural Machine Translation. In Proc. of ACL 2018 (Volume 2: Short Papers), pp. 305–311.

D. Ataman and M. Federico. (2018b). An Evaluation of Two Vocabulary Reduction Methods for Neural Machine Translation. In Proc. of AMTA 2018 (Volume 1: Research Papers), pp. 97–110.

C. Cherry, G. Foster, A. Bapna, O. Firat and W. Macherey. (2018). Revisiting Character-Based Neural Machine Translation with Capacity and Compression. In Proc. of EMNLP 2018, pp. 4295–4305.

H. Deguchi, M. Utiyama, A. Tamura, T. Ninomiya, E. Sumita. (2020). Bilingual Subword Segmentation for Neural Machine Translation. In Proc. of COLING 2020, pp. 4287–4297.

M. Huck, S. Riess and A. Fraser. (2017). Target-side Word Segmentation Strategies for Neural Machine Translation. In Proc. of WMT 2017, pp. 56–67.

D. P. Kingma and J. Ba. (2015). Adam: A Method for Stochastic Optimization. In Proc. of ICLR 2015. arXiv:1412.6980.

P. Koehn. (2004). Statistical Significance Tests for Machine Translation Evaluation. In Proc. of EMNLP 2004, pp. 388–395.

T. Kudo. (2018a). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. In Proc. of ACL 2018, pp. 66–75.

T. Kudo and J. Richardson. (2018b). SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In Proc. of EMNLP 2018: System Demonstrations, pp. 66–71.

T. Luong, H. Pham and C. D. Manning. (2015) Effective Approaches to Attention-based Neural Machine Translation. In Proc. of EMNLP 2015, pp. 1412–1421.

T. Nakazawa, M. Yaguchi, K. Uchimoto, M. Utiyama, E. Sumita, S. Kurohashi and H. Isahara. (2016). ASPEC: Asian Scientific Paper Excerpt Corpus. In Proc. of LREC 2016, pp. 2204–2208.

K. Papineni, S. Roukos, T. Ward and W.-J. Zhu. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In Proc. of ACL 2002, pp. 311–318.

I. Provilkov, D. Emelianenko and E. Voita. (2020). BPE-Dropout: Simple and Effective Subword Regularization. In Proc. ACL 2020, pp. 1882–1892.

R. Sennrich, B. Haddow and A. Birch. (2016). Neural Machine Translation of Rare Words with

Subword Units. In Proc. of ACL 2016, pp. 1715–1725.

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna. (2016). Rethinking the Inception Architecture for Computer Vision. In Proc. of CVPR 2016, pp. 2818–2826.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin. (2017). Attention is All you Need. Advances in Neural Information Processing Systems 30, pp. 5998–6008.

Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Ł. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes and J. Dean. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv preprint arXiv:1609.08144.

出口祥之, 内山将夫, 田村晃裕, 二宮崇, 隅田英一郎. (2021). ニューラル機械翻訳のためのバイリンガルなサブワード分割. 自然言語処理, Vol. 28, No. 2, pp. 632-650.

2.2 単語分散表現モデルに基づく自動評価法への文ベクトルの適用

北海学園大学　越前谷　博

2.2.1　はじめに

近年、ニューラルネットワークに基づくディープラーニング技術の発展により機械翻訳の精度が大きく向上している。ニューラル翻訳[1][2]では単語や文の意味を捉えた翻訳が可能となったことでより高度な翻訳が可能となった。それに伴い、自動評価法もまたBLEU[3]やNIST[4]のような従来の参照訳と訳文間の表層的な一致に基づく手法からニューラルネットワークの技術を取り入れた、より高度な手法へと進化している。

ニューラルネットワーク技術を導入した自動評価法においては、どのようなモデルからベクトルを取得するかによって分類される。事前学習済みのモデルを用いる場合にはプレトレーニングモデルに基づく自動評価法となる。また、翻訳評価に関するデータを学習して構築されたモデルを用いる場合はトレーニングモデルに基づく自動評価法となる。プレトレーニングモデルに基づく自動評価法としては、単語分散表現モデルであるword2vec[5]やfastText[6]、そして、汎用モデルであるBERT[7]を用いる自動評価法などがある。これらの自動評価法の特徴としてはモデルの学習を必要としないため迅速な評価が可能である。しかし、必ずしも翻訳評価を対象としたモデルではないため、モデルにバイアスがかかる可能性がある。トレーニングモデルに基づく自動評価法としては、ニューラル翻訳のアーキテクチャを利用してモデルを学習する手法[8][9]や回帰モデルを学習する手法[10]などがある。これらの自動評価法の特徴としては、翻訳評価に特化したモデルを学習することで評価精度は向上するが、学習のために多くの時間を要することが問題となる。

そこで本報告ではプレトレーニングモデルに基づく自動評価法に翻訳評価データを用いて学習したニューラルネットワークモデルより得られる文ベクトルを適用した新たな自動評価法を提案する。提案手法では、翻訳評価データを用いて学習したseq-to-seqモデル[11]のencoderが出力する文ベクトルより求めたスコアを事前学習済みの単語分散表現モデルを用いて得られる評価スコアに対する重みとして付与することで最終的な評価スコアを求める。ここで用いられるモデル構築のためのデータは参照訳と原文のペアのみであるため学習に多くの時間を要しない。性能評価実験の結果、WMT20のMetrics task[12]においてセグメントレベルの“out-of-English（英語から他言語の訳文）”を除き、提案手法の評価精度は他の自動評価法に比べて上位に位置することが明らかとなった。

2.2.2 提案手法

2.2.2.1 単語分散表現モデルに基づく自動評価法

プレトレーニングモデルによる自動評価法には、これまでに提案している単語分散表現モデルに基づく自動評価法WE_WPI[13]を用いる。本手法では、事前学習済みの単語分散表現モデルより得られる単語ベクトルを参照訳と訳文間の単語アライメントに用いる。さらに、単語ベクトルをEarth Mover’s Distance（EMD）[14]の特徴量に用いると共に、単語アライメントの結果をEMD

の特徴量間の距離計算に取り入れている。また、EMD の重みには文レベルの *tf・idf* 値を用いている。単語アライメントの結果より得られる参照訳と訳文間の語順の違いを距離計算に反映させることで評価精度の向上を図っている。

2.2.2.2 attention 付き seq-to-seq モデルに基づく文ベクトルの利用

トレーニングモデルによる評価スコアを得るために参照訳と原文のペアを学習データとして、attention 付きの seq-to-seq によりモデルを構築する。そして、そのモデルの encoder に対して、参照訳と訳文を入力することでそれぞれの文ベクトルを得る。このようにして得られた 2 つの文ベクトルのコサイン類似度をスコアとして用いる。モデルの構築は参照訳と原文のペアのみを学習データとしているため小規模であり、多くの学習時間を要しない。

2.2.2.3　最終的な評価スコアの算出

本報告では単語分散表現モデルに基づく自動評価法により得られる評価スコアを pre-training スコア（以降、*p-score* と記す）、また、attention 付き seq-to-seq モデルに基づく文ベクトルより得られるスコアを training スコア（以降、*t-score* と記す）とする。提案手法では、*p-score* に対する重みとして *t-score* を付与することで最終的な評価スコア（以降、*final-score* と記す）を求める。*final-score* の計算式を以下の式（1）に記す。

$$final - score = \sqrt{t - score} \times p - score \quad (1)$$

2.2.3 性能評価実験

2.2.3.1 実験手順

性能評価実験は WMT20 の Metrics task データを用いて提案手法による評価スコアおよび人手評価との相関を求めた。提案手法における attention 付き seq-to-seq モデルの構築に用いる学習データは Metrics task データの参照訳と原文のペアを 10 回複製したものを用いた。学習時のエポック数は 80 とした。また、単語分散表現モデルに基づく自動評価法においては事前学習済みの単語分散表現モデルとして fastText を用いた。

さらに本実験では提案手法を含めたメタ評価実験も行うことで WMT20 の Metrics task に参加している様々な自動評価法に対して提案手法がどのような位置づけにあるかを調査した。

2.2.3.2 実験結果

表 1 から表 4 に各自動評価法における相関係数に基づくメタ評価の結果を示す。表 1 はシステムレベルの“to-English（他言語から英語への訳文に対する評価）”、表 2 はシステムレベルの“out-of-English（英語から他言語への訳文に対する評価）”のメタ評価結果である。表 3 はセグメントレベルの“to-English”、表 4 はセグメントレベルの“out-of-English”のメタ評価結果である。表 1 から表 4 の自動評価法 IMPACT[15]は表層レベルで決定したチャンクに基づき評価スコアを得る。

表 1　システムレベルにおける“to-English”の実験結果

	cs-en	de-en	ja-en	pl-en	ru-en	ta-en	zh-en	iu-en	km-en	ps-en	
n	12	12	10	14	11	14	16	11	7	6	Avg.
提案手法	0.837	0.998	0.969	0.576	0.937	0.928	0.969	0.788	0.992	0.917	0.891
BERT-base-L2	0.775	0.997	0.971	0.552	0.919	0.909	0.967	0.704	0.967	0.945	0.871
BERT-large-L2	0.784	0.990	0.974	0.520	0.925	0.901	0.962	0.744	0.959	0.950	0.871
BLEU	0.851	0.985	0.969	0.549	0.884	0.916	0.956	0.569	0.969	0.888	0.854
BLEURT	0.792	0.996	0.978	0.591	0.924	0.906	0.966	0.771	0.984	0.955	0.886
BLEURT-extended	0.771	0.985	0.961	0.551	0.900	0.897	0.945	0.789	0.985	0.942	0.873
CharacTER	0.844	0.998	0.970	0.522	0.927	0.965	0.964	0.763	0.977	0.841	0.877
chrF	0.872	0.997	0.968	0.528	0.890	0.951	0.976	0.729	0.978	0.898	0.879
chrF++	0.867	0.997	0.974	0.538	0.894	0.953	0.975	0.726	0.983	0.900	0.881
COMET	0.783	0.998	0.964	0.591	0.923	0.880	0.952	0.852	0.971	0.941	0.886
COMET-2R	0.777	0.998	0.964	0.584	0.924	0.881	0.949	0.872	0.970	0.949	0.887
COMET-HTER	0.738	0.995	0.912	0.446	0.867	0.726	0.809	0.770	0.901	0.862	0.803
COMET-MQM	0.728	0.991	0.906	0.424	0.858	0.767	0.784	0.841	0.914	0.880	0.809
COMET-QE	0.755	0.939	0.892	0.447	0.883	0.795	0.847	0.685	0.896	0.832	0.797
COMET-Rank	0.705	0.964	0.923	0.483	0.868	0.787	0.877	0.158	0.911	0.855	0.753
EED	0.884	0.997	0.974	0.538	0.926	0.958	0.956	0.821	0.990	0.930	0.897
esim	0.790	0.998	0.983	0.591	0.928	0.885	0.963	0.807	0.929	0.929	0.880
IMPACT	0.848	0.996	0.973	0.536	0.934	0.911	0.952	0.714	0.981	0.910	0.876
mBERT-L2	0.798	0.995	0.969	0.555	0.908	0.887	0.959	0.837	0.980	0.938	0.883
MEE	0.861	0.995	0.982	0.464	0.927	0.950	0.952	0.771	0.970	0.878	0.875
OpenKiwi-Bert	0.861	0.995	0.982	0.464	0.927	0.950	0.952	0.771	0.970	0.878	0.875
OpenKiwi-XLMR	0.760	0.995	0.931	0.442	0.859	0.792	0.905	0.271	0.880	0.865	0.770
parbleu	0.834	0.986	0.970	0.562	0.877	0.908	0.958	0.624	0.971	0.939	0.863
parchrf++	0.865	0.998	0.974	0.551	0.885	0.942	0.976	0.720	0.985	0.899	0.880
paresim-1	0.778	0.998	0.983	0.591	0.926	0.885	0.963	0.800	0.929	0.929	0.878
prism	0.818	0.998	0.974	0.502	0.908	0.898	0.957	0.833	0.950	0.966	0.880
sentBLEU	0.844	0.978	0.974	0.502	0.916	0.925	0.948	0.649	0.969	0.888	0.859
TER	0.845	0.993	0.974	0.586	0.904	0.805	0.956	0.733	0.973	0.935	0.870
WE_WPI	0.838	0.998	0.973	0.573	0.939	0.933	0.965	0.776	0.993	0.891	0.888
YiSi-0	0.876	0.998	0.972	0.453	0.938	0.968	0.956	0.831	0.986	0.932	0.891
YiSi-1	0.832	0.998	0.982	0.543	0.915	0.925	0.961	0.834	0.977	0.953	0.892
YiSi-2	0.764	0.988	0.971	0.437	0.825	0.849	0.964	0.676	0.790	0.942	0.821

表 2　システムレベルにおける“out-of-English”の実験結果

	en-cs	en-de	en-ja	en-pl	en-ru	en-ta	en-zh	en-iu_full	en-iu_news	
n	12	14	11	14	9	15	12	11	11	Avg.
提案手法	0.879	0.946	0.967	0.888	0.945	0.924	0.911	0.498	0.655	0.846
BLEU	0.825	0.928	0.945	0.943	0.980	0.880	0.928	0.163	0.074	0.741
BLEURT-extended	0.989	0.969	0.944	0.982	0.980	0.940	0.928	0.823	0.762	0.924
CharacTER	0.807	0.961	0.951	0.935	0.961	0.957	0.905	0.503	0.515	0.833
chrF	0.826	0.962	0.951	0.957	0.982	0.937	0.923	0.350	0.336	0.803
chrF++	0.833	0.958	0.952	0.956	0.983	0.929	0.878	0.328	0.315	0.792
COMET	0.978	0.972	0.974	0.981	0.925	0.944	0.007	0.860	0.858	0.833
COMET-2R	0.983	0.972	0.986	0.982	0.872	0.959	-0.066	0.848	0.867	0.823
COMET-HTER	0.976	0.951	0.989	0.974	0.803	0.925	-0.073	0.900	0.888	0.815
COMET-MQM	0.974	0.881	0.974	0.967	0.788	0.910	0.084	0.870	0.867	0.813
COMET-QE	0.989	0.903	0.953	0.969	0.807	0.887	0.375	0.905	0.928	0.857
COMET-Rank	0.959	0.877	0.931	0.957	0.676	0.876	0.540	0.283	0.392	0.721
EED	0.817	0.965	0.955	0.962	0.980	0.959	0.928	0.519	0.483	0.841
esim	0.908	0.979	0.993	0.969	0.967	0.937	0.972	0.814	0.760	0.922
IMPACT	0.861	0.932	0.932	0.939	0.961	0.954	0.913	0.405	0.430	0.814
OpenKiwi-Bert	0.920	0.852	0.363	0.903	0.834	0.846	0.551	0.573	0.808	0.739
OpenKiwi-XLMR	0.972	0.968	0.992	0.957	0.875	0.910	-0.010	0.513	0.680	0.762
parbleu	0.870	0.910	0.869	0.948	0.959	0.871	0.962	0.194	0.126	0.745
paresim-1	0.919	0.974	0.989	0.968	0.964	0.937	0.983	0.814	0.760	0.923
prism	0.949	0.958	0.932	0.958	0.724	0.863	0.221	0.957	0.945	0.834
sentBLEU	0.840	0.934	0.946	0.950	0.981	0.881	0.927	0.129	0.075	0.740
TER	0.814	0.941	0.297	0.893	0.064	0.870	-0.213	0.384	0.357	0.490
YiSi-0	0.797	0.953	0.967	0.953	0.971	0.929	0.362	0.525	0.505	0.774
YiSi-1	0.922	0.971	0.969	0.964	0.926	0.973	0.959	0.554	0.523	0.862
YiSi-2	0.714	0.899	0.854	0.470	0.584	0.922	-0.215	0.802	0.830	0.651

表 3　セグメントレベルにおける“to-English”の実験結果

	cs-en	de-en	iu-en	ja-en	km-en	pl-en	ps-en	ru-en	ta-en	zh-en	
n	14018	16584	8162	15193	3706	21121	3507	14024	12789	62586	Avg.
提案手法	0.105	0.481	0.220	0.242	0.230	0.097	0.141	0.136	0.225	0.152	0.203
BERT-base-L2	0.103	0.454	0.238	0.263	0.295	0.032	0.159	0.087	0.223	0.141	0.200
BERT-large-L2	0.102	0.456	0.251	0.262	0.314	0.044	0.151	0.094	0.245	0.133	0.205
BLEURT	0.126	0.456	0.258	0.265	0.327	0.057	0.207	0.093	0.230	0.137	0.216
BLEURT-extended	0.127	0.448	0.259	0.271	0.330	0.044	0.161	0.101	0.246	0.137	0.212
CharacTER	0.090	0.440	0.214	0.221	0.248	0.023	0.172	0.057	0.138	0.123	0.173
chrF	0.086	0.438	0.254	0.242	0.267	0.028	0.144	0.049	0.186	0.132	0.183
chrF++	0.090	0.435	0.246	0.245	0.275	0.034	0.145	0.054	0.186	0.130	0.184
COMET	0.129	0.485	0.281	0.274	0.298	0.099	0.158	0.156	0.241	0.171	0.229
COMET-2R	0.120	0.479	0.257	0.268	0.308	0.098	0.144	0.148	0.253	0.163	0.224
COMET-HTER	0.103	0.481	0.198	0.241	0.269	0.080	0.116	0.131	0.227	0.135	0.198
COMET-MQM	0.108	0.483	0.215	0.259	0.282	0.080	0.141	0.137	0.227	0.141	0.207
COMET-QE	0.091	0.410	0.031	0.153	0.148	0.039	0.092	0.084	0.163	0.088	0.130
COMET－Rank	0.099	0.470	0.188	0.235	0.228	0.073	0.107	0.118	0.199	0.142	0.186
EED	0.091	0.440	0.256	0.235	0.271	0.045	0.149	0.053	0.198	0.129	0.187
esim	0.110	0.454	0.241	0.239	0.300	0.058	0.147	0.084	0.208	0.138	0.198
IMPACT	0.070	0.427	0.188	0.194	0.243	-0.007	0.097	0.009	0.191	0.103	0.152
mBERT-L2	0.119	0.442	0.244	0.251	0.312	0.047	0.151	0.083	0.227	0.133	0.201
MEE	0.063	0.402	0.134	0.187	0.206	-0.084	0.078	-0.041	0.114	0.083	0.114
OpenKiwi-Bert	0.036	0.379	-0.005	0.110	0.168	-0.033	0.076	-0.033	0.118	0.029	0.085
OpenKiwi-XLMR	0.093	0.463	0.056	0.220	0.244	0.059	0.106	0.092	0.188	0.115	0.164
parbleu	0.058	0.415	0.167	0.198	0.203	-0.025	0.100	-0.011	0.159	0.095	0.136
parchrf++	0.096	0.436	0.232	0.247	0.267	0.027	0.147	0.044	0.184	0.132	0.181
paresim-1	0.105	0.464	0.249	0.242	0.292	0.066	0.149	0.089	0.213	0.139	0.201
prism	0.143	0.475	0.255	0.272	0.304	0.109	0.165	0.145	0.237	0.167	0.227
sentBLEU	0.068	0.413	0.182	0.188	0.226	-0.024	0.096	-0.005	0.162	0.093	0.140
TER	-0.04	0.355	0.021	0.044	0.125	-0.172	-0.036	-0.117	0.046	-0.01	0.022
WE_WPI	0.102	0.474	0.218	0.238	0.239	0.080	0.134	0.133	0.222	0.151	0.199
YiSi-0	0.072	0.441	0.261	0.241	0.268	0.035	0.140	0.065	0.183	0.127	0.183
YiSi-1	0.117	0.468	0.253	0.277	0.316	0.042	0.147	0.091	0.248	0.146	0.211
YiSi-2	0.068	0.413	0.039	0.204	0.214	0.048	0.073	0.070	0.199	0.116	0.144

表 4　セグメントレベルにおける“out-of-English”の実験結果

	en-cs	en-de	en-iu	en-ja	en-pl	en-ru	en-ta	en-zh	
n	21121	9339	13159	12830	17689	8330	9087	12652	Avg.
提案手法	0.473	0.327	0.208	0.504	0.267	0.184	0.371	0.382	0.340
BLEURT-extended	0.689	0.447	0.359	0.533	0.430	0.305	0.643	0.460	0.483
CharacTER	0.413	0.311	0.309	0.471	0.198	0.143	0.525	0.339	0.339
chrF	0.472	0.379	0.344	0.506	0.250	0.153	0.589	0.400	0.387
chrF++	0.478	0.367	0.338	0.506	0.255	0.156	0.579	0.388	0.383
COMET	0.668	0.468	0.322	0.624	0.462	0.344	0.671	0.432	0.499
COMET-2R	0.669	0.463	0.326	0.630	0.445	0.343	0.676	0.434	0.498
COMET-HTER	0.665	0.440	0.331	0.601	0.427	0.292	0.640	0.411	0.476
COMET-MQM	0.666	0.423	0.313	0.588	0.424	0.281	0.635	0.388	0.465
COMET-QE	0.614	0.347	-0.04	0.470	0.360	0.264	0.514	0.346	0.359
COMET-Rank	0.629	0.379	0.297	0.569	0.388	0.229	0.588	0.380	0.432
EED	0.458	0.363	0.361	0.515	0.248	0.155	0.587	0.393	0.385
esim	0.469	0.347	0.122	0.522	0.312	0.224	0.599	0.391	0.373
IMPACT	0.457	0.306	0.209	0.486	0.181	0.068	0.525	0.391	0.328
OpenKiwi-Bert	0.262	0.168	-0.115	-0.529	0.153	0.164	0.169	0.077	0.044
OpenKiwi-XLMR	0.607	0.369	0.060	0.553	0.347	0.279	0.604	0.377	0.400
parbleu	0.460	0.299	0.212	0.052	0.183	0.062	0.340	0.356	0.246
paresim-1	0.475	0.343	0.122	0.510	0.324	0.230	0.599	0.396	0.375
prism	0.619	0.447	0.452	0.579	0.414	0.283	0.448	0.397	0.455
sentBLEU	0.432	0.303	0.206	0.479	0.153	0.051	0.398	0.396	0.02
TER	0.317	0.182	-0.071	-0.591	0.003	-0.121	0.203	-0.36	-0.055
YiSi-0	0.432	0.349	0.362	0.484	0.233	0.151	0.547	0.319	0.360
YiSi-1	0.550	0.427	0.251	0.568	0.349	0.256	0.669	0.463	0.442
YiSi-2	0.187	0.296	0.146	0.383	0.115	0.146	0.545	0.152	0.246

2.2.3.3 考察

表 1 から表 4 の実験結果では WMT20 の Metrics task に参加した自動評価法において全言語間の相関係数が得られているもののみを記載している。例えば、WE_WP では“en-iu（英語からイヌクティトゥット語への翻訳）”においてイヌクティトゥット語の fastText による単語分散表現モデルが得られなかったため表 2 と表 4 には含めていない。また、それに伴い、提案手法においても表 2 と表 4 の“en-iu”の相関係数は fastText を使用できなかったため、式（1）の *p-score* の値を 1.0 とすることで *t-score* のみに基づき *final-score* を算出している。すなわち、attention 付き seq-to-seq モデルを用いた $\sqrt{t-score}$ のみにより得られた相関係数となっている。

表 1 よりシステムレベルの“to-English”において提案手法の“Avg.”（全言語間の相関係数の

平均）は全自動評価法 32 中で 2 番目に高い値を示した。システムレベルの“out-of-English”においては表 2 より提案手法の“Avg.”は全自動評価法 25 の中で 6 番目であった。また、表 3 よりセグメントレベルの“to-English”において提案手法の“Avg.”は全自動評価法 31 中で 9 番目に高い値を示した。セグメントレベルの“out-of-English”においては表 4 より提案手法の“Avg.”は全自動評価法 24 の中で 17 番目であった。したがって、表 4 のセグメントレベルの“out-of-English”を除き、WMT20 の Metrics task に参加した自動評価法に対して提案手法は上位に位置しているといえる。

表 4 のセグメントレベルにおける“out-of-English”の提案手法の相関係数が他の自動評価法に比べて低くなった要因としては“en-ta（英語からタミル語への翻訳）”の相関係数が非常に低かったことがあげられる。提案手法の“en-ta”の相関係数 0.371 は表層情報に基づく自動評価法 IMPACT の相関係数 0.525 よりも大幅に低くなっていることからも“Avg.”の低下に大きく影響したと考えられる。提案手法の“en-ta”に対する相関係数が低くなった原因は今後精査する必要がある。

2.2.4 まとめ

本報告では単語分散表現モデルに基づく自動評価法に対してニューラルネットワークに基づく文ベクトルより得られるスコアを重みとして用いる新たな自動評価法を提案し、その性能評価実験を行った。メタ評価の結果、提案手法はシステムレベルの“to-English”と“out-of-English”、そして、セグメントレベルの“to-English”において他の自動評価法に対し比較的高い評価精度を有していることを確認できた。

今後はより適切な文ベクトルの生成および単語分散表現モデルの選定により評価精度の向上を図る予定である。

参考文献

[1] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio (2014) “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp.1724-1734.

[2] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le (2014) “Sequence to Sequence Learning with Neural Networks,” Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS 2014), pp.3104-3112.

[3] Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu (2002) "BLEU: a Method for Automatic Evaluation of Machine Translation, " Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL02), pp. 311-318.

[4] George Doddington (2002) "Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics, " Proceedings of the Second International Conference on Human Language Technology Research (HLT02), pp.138-145.

[5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean (2013) "Efficient Estimation of Word Representations in Vector Space, " Proceedings of Workshop at International Conference on Learning Representations 2013.
[6] "Word vectors for 157 languages," https://fasttext.cc/docs/en/crawl-vectors.html
[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018) "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," https://arxiv.org/abs/1810.04805
[8] Nitika Mathur, Timothy Baldwin and Trevor Cohn (2019) “Putting Evaluation in Context: Contextual Embeddings improve Machine Translation Evaluation” Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL19), pp. 2799–2808.
[9] Brian Thompson and Matt Post (2020) “Automatic Machine Translation Evaluation in Many Languages via Zero-Shot Paraphrasing” Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP20), pp. 90–121.
[10] Hiroki Shimanaka, Tomoyuki Kajiwara and Mamoru Komachi (2018) “RUSE: Regressor Using Sentence Embeddings for Automatic Machine Translation Evaluation” Proceedings of the Third Conference on Machine Translation (WMT), Volume 2: Shared Task Papers, pp. 751–758.
[11] D. Bahdanau, K. Cho and Y. Bengio (205) “Neural Machine Translation by Jointly Learning to Align and Translate,” 3rd International Conference on Learning Representations (ICLR), 2015.
[12] Nitika Mathur, Johnny Tian-Zheng Wei, Markus Freitag, Qingsong Ma and Ondřej Bojar (2020) “Results of the WMT20 Metrics Shared Task,” Proceedings of the 5th Conference on Machine Translation (WMT), pp.688-725, 2020.
[13] Hiroshi Echizen'ya, Kenji Araki and Eduard Hovy (2019) "Word Embedding-Based Automatic MT Evaluation Metric using Word Position Information, " Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT19), pp.1874-1883.
[14] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas (2000) "The Earth Mover's Distance as a Metric for Image Retrieval, " International Journal of Computer Vision, 40(2), pp.99-121.
[15] Hiroshi Echizen-ya and Kenji Araki (2007) “Automatic Evaluation of Machine Translation based on Recursive Acquisition of an Intuitive Common Parts Continuum,” Proceedings of the Eleventh Machine Translation Summit, pp.151-158.

2.3 機械翻訳における文章中での訳語の統一

静岡大学　綱川　隆司
元・山梨英和大学　江原　暉将
東京大学　中澤　敏明

自然言語では、同一の物事を複数の異なる表現で表すことができる。一方で、論文などの説明的文章では、同一の意味を表すのに統一した表現を用いることで無用な誤解を防ぐことが一般的である。また、ある概念に対して決まった専門用語を用いるような場合には、用語集を用いて出力文中の語彙の統制が必要となる。

機械翻訳において訳語を統一するための方法としては、用語集・対訳辞書を用いる方法と、翻訳履歴や文脈を考慮する方法の二つが挙げられる。本稿では用語集を用いる方法について概要を述べる。訳出された文を後編集したり機械翻訳の過程で対訳辞書を適用したりすることにより訳語を統一する。

後編集は用いる機械翻訳システムに依らず適用ができるため一般的な戦略であるが、語のもつ曖昧性により誤った置換が生じるおそれや、そもそも訳出すべき語が抜けたり誤って訳された場合に対応できないことなどの課題に対処する必要が生じる。人手による後編集については、国際標準規格である ISO 18587:2017 によって規定されており、分野や翻訳依頼者が定める特定の用語集に応じて一貫性のある訳語を用いることも要件として含まれている。また、一貫性のある用語集を整備するための規格として、国際的に用いられている TBX や AAMT によるシンプルな用語集形式 UTX がある。これらにおいては、同じものを表す複数の用語候補の中で使用すべき用語や使用を避けるべき用語の情報を付加できるように設計されている。

一方で、並列コーパスにより翻訳器の学習を行うデータ駆動型の機械翻訳モデルでは対訳辞書を明示的に用いないため、ニューラル機械翻訳においては対訳辞書を考慮した翻訳モデルや、訳文生成時に語彙制約を課した探索を行う方法、翻訳器の学習に用いる並列コーパスを対訳辞書により事前に編集する方法が提案されている。ニューラル機械翻訳において外部辞書を考慮した訳文生成を行うモデルとして、Arthur et al. (2016) は出力層に語彙モデルを結合することで低頻度語の誤訳を低減する手法を提案している。

2.4 低リソース言語対対策

元・山梨英和大学　江原　暉将
情報通信研究機構　今村　賢治

2.4.1　低リソース言語対対策

低リソース言語対対策として、2020 年度は以下の 4 点を調査した。1) 間に第 3 の言語を挟んで高リソースとする、2) 単言語コーパスから疑似的な 2 言語コーパスを作成する、3) 高リソース言語対で事前訓練を行い低リソース言語対で微調整を行う、4) 多対多の翻訳システムの一部として低リソース言語対を含める。

2021 年度は、近年発展が著しい 5) 一般公開された事前訓練モデルを微調整する方法について調査した。この方法は大規模な単言語コーパスで学習した事前訓練モデルを元に、対訳コーパスで微調整するものであり、低リソースの弱点をカバーし、従来の翻訳品質を大幅に上回る精度を出せるようになっている。

2.5 サーベイ論文：特許機械翻訳関連の技術動向

北海学園大学　越前谷　博
奈良先端科学技術大学院大学　須藤　克仁
株式会社ディープランゲージ　王　　向莉

2.5.1 最新の機械翻訳評価の研究動向

現在の機械翻訳評価の研究において中心的な役割を果たしている国際会議WMTにおける機械翻訳評価共通タスク（Freitag et al. 2021a）において、従来の直接評価(direct assessment)に変わり、MQMと呼ばれる評価尺度が人手評価に利用されるようになった。MQMは翻訳の個々の誤りについて誤りの種別と深刻さの分類を行い、誤りの重み付けにより翻訳の評価スコアを付する方式であり、日本翻訳連盟(JTF)が策定した翻訳品質評価ガイドラインの基にもなっている。機械翻訳が人手翻訳と同様の厳しさをもって評価されるべきである、という時代が訪れつつあると言えよう(Freitag et al. 2021b)。

2.5.2 特許翻訳評価との関係

機械翻訳やその評価についての研究が加速する一方で、近年はかつてのNTCIR PatentMTやWAT JPOのように特許を対象とする機械翻訳を扱った研究は多くなく、現在の技術が特許翻訳の評価にどう寄与するかについて論じているものがない。本研究会として、特許庁における品質評価手順[1]や特許翻訳において重視される点についての現在の技術動向について調査・検討を行っていく予定である。

参考文献

Freitag, M., Rei, R., Mathur, N., Lo, C.-k., Stewart, C., Foster, G., Lavie, A., and Bojar, O. (2021a). “Results of the WMT21 Metrics Shared Task: Evaluating Metrics with Expert-based Human Evaluations on TED and News Domain.” In Proceedings of the Sixth Conference on Machine Translation, pp. 733–774, Online. Association for Computational Linguistics.

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., and Macherey, W. (2021a). “Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation.” Transactions of the Association for Computational Linguistics, 9, pp. 1460–1474.

[1] https://www.jpo.go.jp/system/laws/sesaku/kikaihonyaku/tokkyohonyaku_hyouka.html

2.6 デプロイ対策

国立研究開発法人　情報通信研究機構　今村　賢治
東芝デジタルソリューションズ株式会社　園尾　聡

2.6.1 例　デプロイ対策調査報告

昨年度に引き続き、ニューラル機械翻訳（NMT）を特許翻訳で実際に使用するときの問題点のうち、システムの翻訳速度、および使用メモリー量の削減に焦点を当てて調査を行った。本年度は、上記問題への対策技法を、以下のように整理した。

- モデル自体への対策（モデル縮小、サブワード、自己注視機構の置き換え、非自己回帰型デコーダー)、
- 訓練・テストでの対策（知識蒸留、ビーム幅、動的語彙選択、ミニバッチ構成法)、
- 実装での対策（半精度浮動小数点、8 ビット整数量子化、注視機構のキャッシュ、C++実装）

また、Workshop on Generation and Machine Translation [1, 2]において実施された翻訳コストの可視化法について調査を行った。

参考文献

[1] H. Hayashi, Y. Oda, A. Birch, I. Konstas, A. Finch, M. Luong, G. Neubig, and K. Sudoh. 2019. Findings of the third workshop on neural generation and translation. In Proceedings of the 3rd Workshop on Neural Generation and Translation, pages 1-14, Hong Kong, November.
[2] K. Heafield, H. Hayashi, Y. Oda, I. Konstas, A. Finch, G. Neubig, X. Li, and A. Birch. 2020. Findings of the fourth workshop on neural generation and translation. In Proceedings of the Fourth Workshop on Neural Generation and Translation, pages 1-9, Online, July.

3. ワークショップ開催報告

3 国際ワークショップ開催報告：PSLT 2021

静岡大学　綱川　隆司

3.1　開催概要

本研究会では 2005 年から特許翻訳に関するワークショップを主催しており、2015 年以降は特許・技術文書翻訳ワークショップ（Workshop on Patent and Scientific Literature Translation）として隔年で開催している。第 9 回を数える今回も前回から引き続き機械翻訳サミットとの併催として 8 月 16 日（月）に開催した。オーガナイザは Co-chair として綱川を筆頭に須藤・宇津呂の両名および後藤氏（NHK）を含む 4 名が務め、プログラム委員として研究会委員のほか 5 名の国内外の研究者にご協力いただいた。新型コロナウイルス拡大の影響により、本ワークショップにおいて初めての完全オンライン形式による開催となった。

例年通り一般講演の募集を行うとともに招待講演者の選定を行い、結果として知財関係公的期間および機械翻訳研究者のそれぞれから 1 名ずつに招待講演を依頼した。一般講演の募集は投稿期限の延長および投稿要件の変更を行ったものの投稿がなかったため、一般講演セッションは行わないこととした。また、プログラム委員により特許の分類に関する特別講演セッションを実施した。

ワークショップ当日は、当初開催予定であった米国オーランドおよびその他の地域からの参加を考慮して米国東部夏時間の午前 8 時（日本時間午後 9 時）から半日の開催としたため日本からは遅い時間帯での実施となったが、当日の参加者はオンラインで参加しやすいこともあり最大で 33 名程度であった。

3.2　ワークショップ報告

今回のワークショップは招待講演 2 件、特別講演 1 件の構成となった。

1 件目の招待講演は特許庁の相澤啓祐氏による、特許情報の翻訳に関する特許庁の取り組みに関する講演であった。特許庁ではパブリッククラウドに構築された機械翻訳プラットフォーム（MTP）を運用している。ニューラル機械翻訳をメインエンジンとして日本語と英語、中国語、韓国語との間の翻訳機能を搭載し、作成された翻訳文は特許情報プラットフォーム（J-PlatPat）やワン・ポータル・ドシエ（OPD）等を通じて一般利用者や国内外の審査官等に提供されていることが紹介された。

2 件目の招待講演は国立台湾大学（National Taiwan University）の Yvonne Tsai 氏による中国語および英語の特許関連テキストにおける特徴に関する講演であった。一般の翻訳者は誤りを避けるため、対象となる言語の規範よりも「安全策」をとることを優先するために、直訳調としたりよく現れる語彙等を用いたりする傾向がある可能性がある。本講演では英語と中国語の特許およびその英訳からなるコーパスを量的・質的に分析し、中国語の特許用語やその展開、バリエーションを調査し、また人手翻訳との比較を行っている。

特別講演は、イタリアで翻訳サービスを提供している企業である Aglatech14 の Elena Murgolo

氏により、言語サービス提供者の立場から特許分類の重要性についての議論をいただいた。特定の分野に特化した機械翻訳エンジンと一般の機械翻訳エンジンを用い、分野内・分野外のそれぞれのテキストの訳文を翻訳者が作成した場合の翻訳効率を比較する検証結果を示し、得られた結果についての分析を行った。本検証では分野特化のエンジンは分野内のテキストだけでなく分野外のテキストにも効果を示しており、その理由について議論が交わされた。

3.3 所感

ニューラル機械翻訳の発展により特許翻訳を含め実用的な検討がますます進んでいる中、従来から存在する分野適応、専門用語等の課題は研究対象としての興味が薄れていないことが本ワークショップを通しても垣間見ることができた。これは翻訳モデルの研究進展にもかかわらず、入手可能なリソースの問題、モデルの説明可能性の問題などから現状のモデルではなお特許をはじめ正確性が求められる機械翻訳の応用先では不十分な部分が残っているからであり、現状の課題を整理して今後の発展につなげることが期待される。

コロナ禍の影響により学術分野の会議でも国際・国内を問わずオンラインでの開催が多くを占めている状況が続いている。今後その影響が薄れても、オンラインまたはハイブリッド形式の会議が増えることが予想され、従来よりも細かいテーマを設定したワークショップでも参加者を集められる可能性がある一方、ワークショップ開催側としては問題の本質に迫るテーマ設定が一層求められるようになると思われる。

2021年度AAMT/Japio特許翻訳研究会報告書

発 行 日　2022 年 3 月

発　　行　一般財団法人　日本特許情報機構（Japio）
〒135−0016　東京都江東区東陽町 4 丁目1番 7 号
佐藤ダイヤビルディング
TEL：(03) 3615−5511　FAX：(03) 3615−5521

編　　集　一般社団法人　アジア太平洋機械翻訳協会（AAMT）

印　　刷　株式会社インターグループ