ISSN 1883-1818
No.71
December 2019

AAMT Journal

Asia-Pacific Association for Machine Translation

機 械 翻 訳

AAMT Journal
Asia-Pacific Association for Machine Translation 機械翻訳
No.71

目　次

C O N T E N T S

巻頭言

標準語

長尾　真

古来日本列島と称される地域で話されてきた言葉は日本語と名付けられる言葉であるが、そこには種々の方言が含まれていた。鹿児島や琉球の言葉、奥羽地方の言葉、京や関東の言葉などは相互に通じにくいものだったに違いない。それでは支障をきたすということで、明治から昭和の初めにかけて「東京の山手の教養のある人たちの使う日本語」が標準語とされるようになっていったが（法律などで正式に決められたものではない）、学校で教えられ普及した。またラジオやテレビなどでこの言葉が使われた結果、この言葉がだれにでも通じ、またしゃべれるものとなっている。過去何千年も使われてきた各地の言葉が廃れて行き、たった百年ほどで標準日本語の時代になったのは見方によっては驚くべきことである。

このことを地球世界に当てはめてみよう。人類誕生以来民族によって使う言葉はいろいろであり、3000あるいは7000の言語が地球上にあるといわれている。しかしグローバル化が進み、人々の移動、交流が盛んになるにつれて、よく使われる言語は100以下、国連で使われる公用語は（国連が戦勝国によって作られたこともあって）たったの6つである。こうしてマイノリティ言語がどんどんと消滅して行っていることを心配する人たちも多い。

では今日のようにインターネットで世界中が結ばれた世界の出現によって世界の言語はどの様になって行くのだろうか。現在は英語が圧倒的に多くネットの場で使われているが、書き言葉としては絵文字での意味通信の場もあるようだし（？）、英語といってもかなり文法に沿わない表現が国際的には許容されてゆく時代になるかもしれない。

世界中の人が理解し発信できるような国際標準語が確立する可能性はあるのだろうか。またあるとすればそれは何時頃か。言語の歴史は何万年もあるのだから国際標準語が成立するとしても数百年後のことなのか、あるいはスピードの速い現代ではひょっとするとあと百年以内に実現する可能性も否定できないだろうか。

ポータブルの音声翻訳器が多くの言語で使えるようになりつつあるが、これが普及することになると、標準国際語は必要なくなるのか、あるいは全ての言語はこの言語を経由して他の言語に変換されてゆくようになるのか。そんなことを夢想しながら自然言語処理、機械翻訳、スマホなどの先行きを考えれば考えるほど面白い。日本語は日本人しか使わない言語であるから、日本語で発信される情報が標準国際語だけでなく各国語に直って世界中に浸透してゆくよういろんな手段を考える必要があるだろう。

それにしても標準語とは何なのだろう。明確な定義を与えるのが難しいのは当然としても、少なくともその存在と意義、それに従う努力の大切さを国民に知らせる努力があってもよいのではないだろうか。フランスのアカデミーはその努力をしていると聞くが、日本の場合は言葉の使い方を流れに任せて放置していて、いたるところに理解の難しい文章が氾濫していて、日本語が乱れていると言われている。例えば「レガシー」という言葉を何%の日本人が読んでスッと頭に入れられるだろうか。これを「遺産」という立派な日本語で表現すればよいのに、わざわざ意味不明確な表現にするのは日本人の性格によるものだろうか。ほかにも、例えば「三日前にお店に行った」というべきところを、「三日ほど前にお店に行った」といったりすることが多い。若者の一部にしか通じない方言も多い。規範というものの重要性を考えるべきなのである。

Standard Japanese
Makoto Nagao

第 6 回 AAMT 長尾賞学生奨励賞受賞記念

構文構造を導入したニューラル機械翻訳

江里口　瑛子

Microsoft, Corporation

1. はじめに

自然言語処理タスクの１つである機械翻訳タスクは，ある言語で記述された文を，文意を保持しながら，別の言語で記述された文へ変換するタスクとして定義される．機械翻訳タスクのように異種のデータを変換するモデルに，**エンコーダ・デコーダモデル**がある．エンコーダ・デコーダモデルでは，入力データを固定長ベクトルへ写像し（エンコード），その固定長ベクトルから異種の出力データを生成する（デコード）．機械翻訳タスクは，翻訳元言語の文（i.e. 単語の系列データ）を，別の言語で書かれた文（i.e.単語の系列データ）へ変換するタスクであるとして，エンコーダ・デコーダモデルが適用されている [1].

ニューラルネットワークを利用した機械翻訳（ニューラル機械翻訳）モデルは，長文（i.e. 単語の系列データサイズが大きい）の出力が課題となっていた．これに対して，デコード時に，エンコーダ側へのアクセスを許すアテンション機構 [3]を導入することで，ニューラル機械翻訳モデルの性能は飛躍的に改善された．アテンション機構を伴ったニューラル機械翻訳モデルは，翻訳タスクワークショップの１つである WMT'14 の英独翻訳において最高精度を報告している [4]．また，ニューラル機械翻訳モデルはオンライン機械翻訳システムとしても近年採用されてきており，様々な言語組において高い翻訳性能を示していることが報告されている[6,7].

従来のエンコーダ・デコーダモデルでは入出力データとして系列データを前提としているものが多かったため，言語データの有する構文構造情報などを考慮するようなニューラル機械翻訳モデルが存在しておらず，十分に研究されていなかった．他方，ニューラル機械翻訳が登場する以前の統計的機械翻訳モデルでは，英日などの遠縁の言語対に対しては，事前並び替えや構文構造を考慮することで翻訳精度が改善することが知られている [5].

本研究では，既存のアテンション機構を伴うニューラル機械翻訳モデル [4]を拡張し，エンコーダ側，デコーダ側にそれぞれ構文構造情報を導入した新たなモデルの提案を行った．これら研究成果は博士論文 ”Introducing Syntactic Structure into Neural Machine Translation”にまとめられている．

エンコーダ側への導入　既存のニューラル機械翻訳モデルのエンコーダ側に文の句構造情報を取り入れ，更に, その句構造へのアテンション機構を取り入れた，新たなニューラル機械翻訳モデルの提案を行う．提案手法の特徴は，エンコーダ側に構文情報として句構造を取り入れたニューラル機械翻訳モデルであること，並びに, 句へのアテンション機構を有することである．特に，後者により，目的言語側の単語出力と原言語側の句との対応関係を学習することができる. 実験では，英日コーパスを用いて提案モデルの学習を行い，翻訳の自動評価指標である BLEU スコアと RIBES スコアを用いて翻訳性能の評価を行った．その結果，提案手法が，従来の系列に基づくニューラル機械翻訳モデルよりも両自動評価指標において高いスコアを示していることを確認した．

デコーダ側への導入　係り受け解析器から得られた係り受け情報を，正確なデコードを誘導するための情報として利用する新たなニューラル機械翻訳モデルを提案する．複数のタスクを１つのニューラルネットワークで同時学習することで，タスク全体の性能を改善

Introducing Syntactic Structure into Neural Machine Translation
Akiko Eriguchi
Microsoft Corporation

できることが報告されており [8]，構文解析器などから得られる情報の活用や，異なるタスク間の情報の共有は，機械翻訳タスクにおいても有用であることが示唆されている．また，一つのモデルとしてより直接的に記述することで更なる性能改善が期待できる．提案手法では，解析器から得られた係り受け解析情報を正解とみなし，これにより翻訳の出力が誘導される．そして，提案手法内の係り受け解析モデルと翻訳モデルは，各モデルのネットワークがより密接に接続されている．いくつかの翻訳コーパスを用いた実験では，従来のニューラル機械翻訳モデルの性能を改善することを確認した．

2. ニューラル機械翻訳

ニューラル機械翻訳モデルは，ベクトル空間を介した系列変換モデルとして提案された．本節では，系列に基づくニューラル機械翻訳モデルを紹介する．

エンコーダ・デコーダモデルでは，リカレントニューラルネットワークを用いて，原言語の文（単語の系列データ；$\boldsymbol{x} = x_1, x_2, \ldots, x_N$）を固定長ベクトル空間に埋め込み，その固定長ベクトル空間から目的言語の文（単語の系列データ；$\boldsymbol{y} = y_1, y_2, \ldots, y_M$）を出力する．ステップ i でのエンコーダの隠れ層 $\boldsymbol{h}_i$ は，直前の隠れ層 $\boldsymbol{h}_{i-1}$ と i 番目の入力単語 x_i の単語ベクトル $Emb_x(x_i)$ から，$\boldsymbol{h}_i = f_{enc}(\boldsymbol{h}_{i-1}, Emb_x(x_i))$, と算出される．ここで f_{enc} は非線形関数を表し，$\boldsymbol{h}_0$ は 0 ベクトルとする．文末まで再帰的に計算したのち，エンコーダの最終隠れ層 $\boldsymbol{h}_N$ をデコーダの初期値として設定する（$\boldsymbol{s}_1 = \boldsymbol{h}_N$）．

アテンション機構を導入したニューラル機械翻訳モデルでは，目的言語における j 番目の出力単語の予測分布は，$p(y_j \mid \boldsymbol{y}_{<j}, \boldsymbol{x}) = \mathrm{softmax}(\boldsymbol{W}_s \tilde{\boldsymbol{s}}_j + \boldsymbol{b}_s)$, と計算される．ここで，$\mathbf{W}_s$ と $\mathbf{b}_s$ は重み行列とバイアス項である．ステップ j の隠れ層 $\tilde{\boldsymbol{s}}_j$ は，同ステップのデコーダの隠れ層 $\boldsymbol{s}_j$ 並びに，文脈ベクトル $\boldsymbol{d}_j$ を用いて，$\tilde{\boldsymbol{s}}_j = \tanh(\boldsymbol{W}[\boldsymbol{s}_j; \boldsymbol{d}_j] + \boldsymbol{b})$, と算出される．ここで，$\boldsymbol{W}$, $\boldsymbol{b}$ は重み行列とバイアス項を表し，;は行列同士の結合を表す．ステップ j における文脈ベクトル $\boldsymbol{d}_j$ は，エンコーダの各隠れ層の注目度合い（アテンション）を考慮して構成されたベクトルであり，$\boldsymbol{d}_j = \Sigma_{i=1}^{N} \alpha_j(i)\, \boldsymbol{h}_i$, としてアライメントスコア $\alpha_j(i)$ によるエンコーダの各隠れ層 $\boldsymbol{h}_i$ (i=1, …, N) の重み付き和として算出され，$\alpha_j(i) = \exp(\boldsymbol{h}_i \cdot \boldsymbol{s}_j) / \Sigma_{k=1}^{N} \exp(\boldsymbol{h}_k \cdot \boldsymbol{s}_j)$, として全体の和が 1 となるよう正規化された確率分布の形とする．$\boldsymbol{h}_i \cdot \boldsymbol{s}_j$ は，エンコーダ並びにデコーダの隠れ層の類似度を表す．図 1 にアテンションに基づくニューラル機械翻訳モデルの概要図を示す．

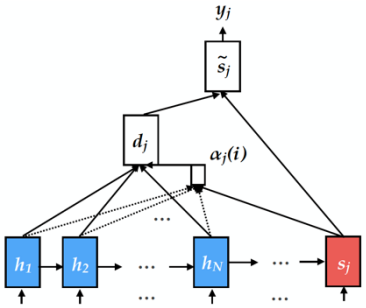


図 1: ニューラル機械翻訳モデル．

ニューラル機械翻訳モデルの目的関数は，学習データ D の入力文と出力文の全ペアに関する対数尤度として，$J(\Theta) = (1/|D|) \Sigma_{(\boldsymbol{x},\boldsymbol{y}) \in D} \log p(\boldsymbol{y}|\boldsymbol{x})$, とする．ここで Θ は，モデルが含む隠れ層，単語ベクトル，アテンションなどの計算に用いられる全てのパラメータを表す．学習時は，パラメータ更新に確率的勾配降下法を用いる．

3. 原言語側に句構造情報を導入したニューラル機械翻訳

3.1 句構造を導入したニューラル機械翻訳モデル

前節で紹介した系列に基づくニューラル機械翻訳モデルは，文を単語の系列として扱い，言語に内在する構文構造を何ら利用していない．ここでは言語の構文構造に着目し，ニューラル機械翻訳モデルに導入するために，句構造に基づくニューラル機械翻訳モデルを

新たに提案する. 図 2 に, 提案モデルの概要図を示す.

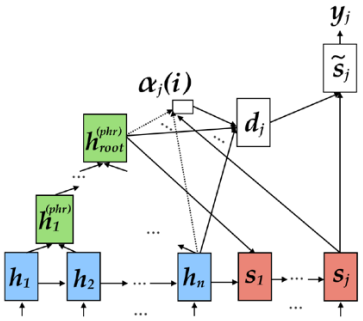


図 2: 句構造を考慮したニューラル機械翻訳モデル.

文はある種の構造情報を有していると見ることができる. HPSG 文法 [9]に従えば, 文は複数の句から構成されており, 2 分木構造で表現される. この 2 分木構造では, 句構造情報が単語から順次ボトムアップ方式で構成されている.

句構造エンコーダは, この句構造情報に従って, 既存の系列エンコーダ上に句ベクトルを再帰的に計算し, 句構造に基づく文ベクトル（ルートノード）を構成するものである. k 番目の句に対する親の隠れ層 $\boldsymbol{h}^{(phr)}_k$ は, 左右の子どもの隠れ層 $\boldsymbol{h}^{left}_k$, $\boldsymbol{h}^{right}_k$から, $\boldsymbol{h}^{(phr)}_k = f_{Tree}(\boldsymbol{h}^{left}_k, \boldsymbol{h}^{right}_k)$, と計算される. ここで, f_{Tree} は Tree-LSTM [10]の関数を表す. そして, 句構造エンコーダにおける葉ノードの隠れ層は, 系列に基づくエンコーダで計算された隠れ層により初期化されている. 葉ノードではないノードには, Tree-LSTM を用いており, 葉ノードの LSTM ユニットを左右の子ノードとして受け取り, 句の隠れ層をルートノードまで再帰的に計算していく. Tree-LSTM の末端葉ノードに, 前からの系列データを受け取って構成されてきた LSTM で初期化することで, 構成されていく句ベクトルが前からの文脈情報をよりよく捉えることができる. これにより, 同じ文内に同じ単語が複数回出現した場合などでも, 位置に応じた異なる句ベクトルを構成することができる. 予備実験から, 葉ノードを単語ベクトルのみで初期化したオリジナルの木構造に基づくエンコーダと比較したところ, 葉ノードを初期化した提案手法の方が, 高い翻訳性能を示すことを確認している.

図 2 で示すように, 系列に基づくエンコーダの最終隠れ層 $\boldsymbol{h}_N$ と句構造エンコーダの最終隠れ層（ルートノード）$\boldsymbol{h}^{(root)}_k$ を子どもの隠れ層として受け取り, 初期デコーダ $\boldsymbol{s}_1$ を, $\boldsymbol{s}_1 = g_{Tree}(\boldsymbol{h}_N, \boldsymbol{h}^{(root)}_k)$, として計算する. 提案手法では, 系列に基づく文ベクトル, 句構造に基づく文ベクトルの両者からデコーダの初期ベクトルを構成するとしたが, 全ての文の構文解析ができるとは限らない. 構文解析器が解析に失敗した場合は, $\boldsymbol{h}^{(root)}_k = \boldsymbol{0}$,としてデコーダの初期化を行う. これにより, 提案手法では構文解析を失敗した文も含め, 任意の文を取り扱うことができる.

この句構造エンコーダに対して, さらにアテンション機構の導入を行う. 提案する句構造エンコーダへのアテンション機構では, 系列に基づくエンコーダに含まれる隠れ層に加えて, 句構造エンコーダに含まれる句の隠れ層の両者を対象とする. これにより, 単語毎のデコードの際に, 原言語の文における各単語, あるいは単語よりも大きな単位である各句を考慮することができる. j 番目の文脈ベクトル $\boldsymbol{d}_j$ は, 系列に基づくエンコーダの隠れ層, 並びに句構造エンコーダの隠れ層をそれぞれアテンションスコア $\alpha_j(i)$で重み付けした総和として, $\boldsymbol{d}_j = \sum_{i=1}^{N} \alpha_j(i) \boldsymbol{h}_i + \sum_{i=N+1}^{2N-1} \alpha_j(i) \boldsymbol{h}^{(phr)}_i$, のように算出される. ここで, 句構造エンコーダは, 葉ノードが N個のとき, N個の節ノードを有することに注意する. 以降, 単語の予測は第 2 節で紹介した方法に従う.

3.2 実験および考察

実験では, ASPEC 英日コーパス 10 万文を使用した. また, 入力文の句構造解析器には, HBSG 文法に基づく解析器 Enju (http://kmcs.nii.ac.jp/enju/) を用いた. 表 1 は, 各言語対におけるモデルの自動評価指標 BLEU 及び RIBES をまとめたものである. どちらの指標においても提案手法により性能が改善していることがわかる. データの規模を 150 万文まで増やした際, NMT と提案手法のそれぞれの性能は(RIBES,

BLEU)=(81.60, 34.64), 及び(81.66, 35.05)となり，提案手法による性能改善が見られる．

表 1：英日翻訳における実験結果．

	RIBES	BLEU
NMT	76.2	24.9
NMT (逆入力)	75.3	24.0
提案手法	76.4	25.6

提案モデルにより句構造情報がどのように利用されているかを分析する．図 3 に入出力間で学習されたアテンションスコアを示す. 翻訳文中の単語”マトリックス"と，もっとも高いアテンションスコアを有していたエンコーダの隠れ層は，英単語”matrix"を入力した際のものであり，その際のアテンションスコアは 0.77 であった．これに対して，単語”液晶"は翻訳語に相当する”liquid crystal"（アテンションスコア 0.07）よりもより高いアテンションスコア（0.41）”liquid crystal for active matrix"と関連付けられている．これは，デコーダの隠れ層がすでに出力済みの単語情報（“活性 マトリックス の") を文脈として受け取っているため，それらの文脈に相当する入力（“for active matrix"）を含めた名詞句”liquid crystal for active matrix"と高く関連付けられたためと考えられる．このように，提案するアテンション機構では，系列に基づくエンコーダの隠れ層に加え，句構造エンコーダの隠れ層も対象とすることで，出力単語と入力単語の関連度と並んで，出力単語と入力句の関連度もまた柔軟に学習していることがわかった．

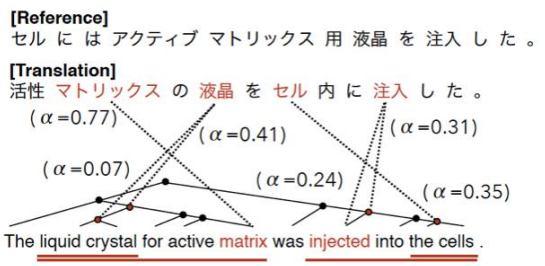


図 3：学習されたアテンションスコア．

ニューラル機械翻訳モデルの翻訳結果では，正解語”アクティブ"の類義語である”活性"という単語が生成されている．この他の翻訳結果においても，”女"と”女性"，”アメリカ航空宇宙局"と”NASA"などといったように，類義語・同義語が生成されている場合があった．今回用いた自動評価指標は，参照訳中の単語が生成結果に含まれているかの一致率に基づいており，単語の類犠牲については考慮されていない．このため，上述のような翻訳生成は不正解となり指標に反映されていない．このようなニューラル機械翻訳特有の翻訳の性質に対して，今後，類犠牲などを考慮した自動評価指標に関する研究の重要性がさらに増すと考える．

4. 目的言語側に係り受け構造を導入したニューラル機械翻訳

4.1 機械翻訳と目的言語における係り受け構文解析の同時学習モデル

Recurrent Neural Network Grammar（RNNG）[11]は，句構造解析と文生成の同時学習モデルであり，単一のニューラルネットワークからなる．ここでは，句構造解析は shift-reduce 操作による遷移型モデルによって記述されており，単語はスタックとバッファへ移動（shift）し，スタック中にある単語に対して適当なアクションを適用（reduce）させていくことで，スタック上に文全体の句構造が構築されていく．RNNG では，さらに，shift アクション時に単語生成を行っており，文生成タスクとの同時学習モデルとなっている．

生成された文に対する全アクションはアクション列 $\boldsymbol{a} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_L)$ からなり，ステップ t におけるアクション $\boldsymbol{a}_t$ の生成確率は，スタック，アクション，バッファの隠れ層 $\boldsymbol{h}^s_t$, $\boldsymbol{h}^a_t$, $\boldsymbol{h}^b_t$ それぞれから，$p(a_t \mid \boldsymbol{a}_{<t-1}) = \mathrm{softmax}(\boldsymbol{W}\tanh[\boldsymbol{h}^s_t; \boldsymbol{h}^a_t; \boldsymbol{h}^b_t] + \boldsymbol{b})$， と計算される．それぞれの隠れ層は，Stack-LSTM [11]を用いて計算され，Stack LSTM では，各ユニットはスタック上で扱われ，TOP の指すユニットと入力を受け取ることで新たな隠れ層の生成が行われる．

提案モデルでは，RNNG モデルにおけるバッファをニューラル機械翻訳モデルのデコーダそのものとみなし，単語の生成もまたニューラル機械翻訳モデルのデコーダに従うものとする．RNNG モデルのスタック，アクションの初期隠れ層 $\boldsymbol{h}^{s}_{0}$, $\boldsymbol{h}^{a}_{0}$には，デコーダの隠れ層の初期化と同様に，両方向エンコーダの各最終隠れ層から Tree-LSTM ユニットを用いて計算する．また，スタックで用いる単語ベクトルのパラメータは，デコーダの単語ベクトルを共有するものとする．予備実験では，スタックとデコーダ間で単語ベクトルのパラメータを共有することで翻訳性能が改善することを確認している．図 4 に提案モデルの概要図を示す．図中では，デコーダにおけるアテンション機構などは省略している．

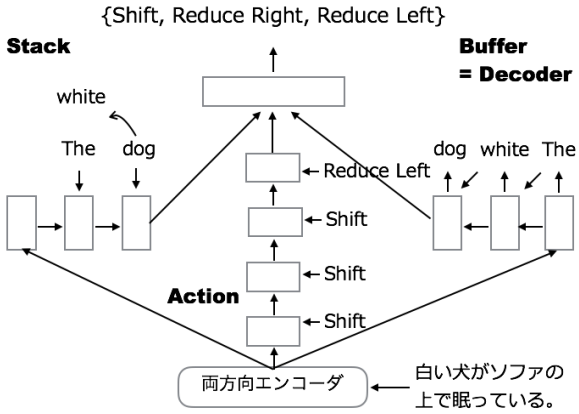


図 4: 目的言語側における係り受け構造を導入したニューラル機械翻訳.

4.2 実験および考察

実験では，ASPEC 日英コーパス，News Commentary v8 露英，捷英（チェコ語-英語），独英コーパスを使用した．日英学習データは，10 万文対を用いた．また，出力文の係り受け解析器には，SyntaxNet [12]を用い，英語の単語分割は SyntaxNet の出力に従い，日本語の単語分割には KyTea [13]を用いた．表 2 は，各言語対におけるモデルの自動評価指標 BLEU 及び RIBES をまとめたものである．BLEU スコアでは独英を除いて，また RIBES では全ての言語組において性能が改善していることがわかる．表中の記号※はブートストラップリサンプリングにより優位差が見られた($p<0.05$)ことを示す．

表 2: 実験結果

BLEU

	JA-EN	RU-EN	CS-EN	DE-EN
NMT	17.88	12.03	11.22	16.61
NMT + RNNG	※**18.84**	※**12.46**	※**12.06**	16.41

RIBES

	JA-EN	RU-EN	CS-EN	DE-EN
NMT	71.27	69.56	69.59	73.75
NMT + RNNG	※**72.25**	※**71.04**	※**70.39**	※**75.03**

日英データを用いて，提案モデルのさらなる分析を行った．分析方法としては，3 種類ある RNNG モデルのどのコンポーネントに最も効果があるのかを見るために，各コンポーネントを除去する ablation study を行った．表 3 はその結果をまとめたものである．RNNG のコンポーネントのうち Stack を取り除いた時，最大 1.02 ポイント BLEU スコアが悪化したことから，Stack が最も RNNG の中で性能改善に寄与したことがわかる．Stack は係り受け構造情報をベクトル表現するためのネットワークであることから，提案手法では，係り受け構造情報をうまく利用することでニューラル機械翻訳モデルを改善できたと考えられる．

表 3: 提案モデルにおける ablation study

BLEU

	Ja-EN (Dev.)
NMT + RNNG	**18.60**
w/o Buffer	18.02
w/o Action	17.94
w/o Stack	**17.58**
NMT	17.75

図 5 に提案モデルが予測した翻訳文結果とその係り受け構造を示す．従来のニューラル機械翻訳モデルでは出力文のみが予測されるのに対して，提案モデルでは，翻訳結果に加え，どの単語がどのような係り受け

関係を持つかが示されている．例えば，動詞”has”と名詞”temperature”の間には”has”が”temperature”という語を主語（nsubj） として持つという関係があることがモデルにより予測されている．しかしながら，”The”と”transition”の間のラベルに誤りがあるように，係り受け構造推定の性能にはまだ課題があることがわかる．

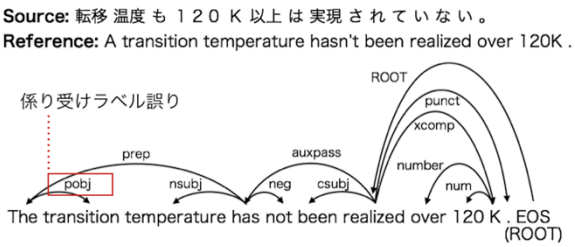


図 5：提案モデルにより予測された翻訳文とその係り受け構造．

5. おわりに

本研究では，文に内在する構造情報として句構造データと係り受け構造データの 2 種類に着目し，ニューラル機械翻訳モデルへの導入を行った．従来のニューラル機械翻訳モデルが原言語側，目的言語側それぞれにおける文を系列データとして扱っていたのに対して，本研究では，原言語側，目的言語側において構文構造データをそれぞれ導入することを目指した．実験結果を通して，提案手法による有効性や提案モデルが期待通りの学習を行えていることを確認した．

構造データは，今回焦点を当てた句構造や係り受け構造と言った構文構造情報のみに止まらない．Web テキストのようなタグ付きテキスト，数式など，様々な場面で構造データを扱うことが考えられる．今後の課題は，そのような個々のケースにおいて，1)どのような形で構造データを考慮あるいは取り入れながら，主タスクである翻訳性能を改善していくのか，また，2)学習されたモデルの分析にこのような構造データの情報が役立てられないか，などを検討していくことである．

謝辞

本稿で紹介した研究を行うにあたって，3年間ご指導くださった東京大学鶴岡慶雅教授に深謝いたします．また，ニューヨーク大学短期滞在中にご指導いただいたニューヨーク大学Kyunghyun Cho准教授にも深くお礼申しあげます．共著者の橋本和真さんからは多くの有益な助言をいただきました．重ねて感謝いたします. 本研究はJSPS 科研費15J12597の助成の支援を受けたものです.

参考文献

[1] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on EMNLP, pp. 1724--1734, 2014.

[2] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the 32nd ICML, 2015.

[3] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. Proceedings of Proceedings of ICLR, 2015.

[4] T. Luong, H. Pham, and C. D. Manning. Effective approaches to attention-based neural machine translation. In Proceedings of the 2015 Conference on EMNLP, 2015.

[5] G. Neubig. Travatar: A forest-to-string machine translation engine based on tree transducers. In Proceedings of the 51st Annual Meeting of the ACL: System Demonstrations, 2013.

[6] 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, J. Dean. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arxiv: arXiv: 1609.08144, 2016.

[7] H. Hassan, A. Aue, C. Chen, V. Chowdhary, J. Clark, C. Federmann, X. Huang, M. Junczys-Dowmunt, W. Lewis, M. Li, S. Liu, T. Liu, R. Luo, A. Menezes, T. Qin, F. Seide, X. Tan, F. Tian, L. Wu, S. Wu, Y. Xia, D. Zhang, Z. Zhang, and M. Zhou. Achieving Human Parity on Automatic Chinese to English News Translation. arxiv: arXiv:1803.05567, 2018.

[8] T. Luong, Q. V. Le, I. Sutskever, O. Vinyals, and L. Kaiser. Multi-task sequence to sequence learning. Proceedings of ICLR, 2015.

[9] I. A. Sag, T. Wasow, and E. Bender. Syntactic Theory: A Formal Introduction. Center for the Study of Language and Information, Stanford, 2nd edition, 2003.

[10] K. S. Tai, R. Socher, and C. D. Manning. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In Proceedings of the 53rd Annual Meeting of ACL and IJCNLP, pp. 1556--1566, 2015.

[11] C. Dyer, A. Kuncoro, M. Ballesteros, and A. N. Smith. Recurrent neural network grammars. In Proceedings of NAACL, 2016.

[12] D. Andor, C. Alberti, D. Weiss, A. Severyn, A. Presta, K. Ganchev, S. Petrov, and M. Collins. Globally normalized transition-based neural networks. In Proceedings of ACL, 2016.

[13] G Neubig, Y. Nakata, and S. Mori. Pointwise pre- diction for robust, adaptable Japanese morphological analysis. In Proceedings of ACL, 2011.

解説記事

最新のニューラル機械翻訳に関する技術紹介

中澤　敏明
東京大学

1. はじめに

本稿では、2019 年に国際会議で発表された論文のうちのいくつかを紹介する。紹介するのは 1. Bridging the Gap between Training and Inference for Neural Machine Translation (ACL2019, best paper) [1]、2. Revisiting Low-Resource Neural Machine Translation: A Case Study (ACL2019) [2]、3. Post-editese: an Exacerbated Translationese (MT-Summit2019, best paper) [3]の 3 つである。

2. Bridging the Gap between Training and Inference for Neural Machine Translation (ACL2019, best paper)

本論文で解決を試みる NMT の問題は 2 つある。Auto-regressive な NMT のデコーダーでは、ある時点での出力を次の入力として利用する。ここで、訓練時は実際の出力ではなく、参照訳を入力することが一般的である (teacher forcing)。しかしテスト時は当然参照訳を持っていないため、デコーダーの出力をそのまま次の入力として用いる。この訓練時とテスト時の環境の違い (Exposure Bias) が翻訳精度に悪影響を及ぼすため、これを解消することがこの論文の目標の一つ目である。もう一つは論文の著者らが overcorrection と呼ぶ問題である。これは例えば参照訳が “We should comply with the rule.” である場合に、デコーダーが “We should abide” と出力したとすると、その次の前置詞は “by” が正しいのだが、文全体の loss を最小化するために “with” と出力してしまうという問題である。さらに、もし “by” を出力できたとしても、今度は “by” に引きずられて “by the law.” という訳を出力してしまう可能性もある。

これらの問題を解消するために、本論文は訓練時のデコーダーへの入力を、参照訳の単語とオラクルの単語を確率的に選択するという方法を取っている。実はこの方法と似た方法は既にいくつか提案されている([4], [5], [6]など)のだが、オラクルの単語の選び方が既存手法とは異なるという主張のようである。オラクル単語は word oracle と sentence oracle の 2 つを提案している。Word oracle は通常の softmax をかける前の出力スコアに Gumbel noise を足すことで効率的なサンプリングを行う。Sentence oracle は beam-search で n-best の出力文を生成し、その中で最も BLEU 値が高い文をオラクルとして用いるというものである。なお参照訳とオラクル文との長さを揃えるために、オラクル分の長さを強制的に参照訳に合わせる(翻訳を打ち切る)という処理をしている(Force Decoding)。

実験の結果、NIST の中英翻訳においてはベースラインの Transformer と比較して word oracle で平均 0.5 ポイント、sentence oracle で平均 1.5 ポイントの BLEU スコアの向上を達成している。また WMT の英独翻訳においても同様に 1.3 ポイントの向上を達成している。

なお本論文は ACL2019 の best paper ではあるのだが、正直そこまですごいかと言われると、そんなことはない気がする（MT に詳しい何名かに聞いたがみな口を揃えて「なぜこれが？」と言った反応であった)。

3. Revisiting Low-Resource Neural Machine Translation: A Case Study (ACL2019)

The Latest Technologies for Neural Machine Translation
Toshiaki Nakazawa
The University of Tokyo

本論文は、これまで多くの論文で報告されてきた「NMT は low-resource な設定ではうまく訓練できず、Phrase-based SMT よりも精度が低い」という結論を再考し、これに異を唱えるものである。これまでの low-resource NMT 研究のほとんどは、ベースラインシステムを構築する際に high-resource な状況でのハイパーパラメーターなどの設定をそのまま流用しており、low-resource な設定に合うようにチューニングしていないことが問題だと著者らは指摘している。著者らはこれまでに提案された様々なテクニックを組み合わせて利用する best-practice を示し、これにより low-resource な設定でも PBSMT の性能を超えることが可能であることを示した。また low-resource NMT の研究では単言語コーパスや他の言語の対訳コーパスをいかに利用するかを主眼にしたものが多いが、本論文ではこのような相補的な言語資源を一切利用せずとも上記の性能を達成できるとしている。つまり、「NMT は data-hungry である」というこれまでの通説を覆したと言える。

著者らが示した best-practice に含まれるのは、BiDeep RNN [7]、label smoothing [8]、dropout [9]、word dropout [10]、layer normalization [11]、tied embeddings [12]、サブワードの語彙サイズの調整、ハイパーパラメーターのチューニング、lexical model [13]である。以下に簡単にこれらについて説明する。

BiDeep RNN は RNN 層を複数重ねる（Deep Stack）構造と、各 RNN 層の内部に複数の RNN を直列につなげた（Deep Transition）構造を併用する手法である（図 6）。

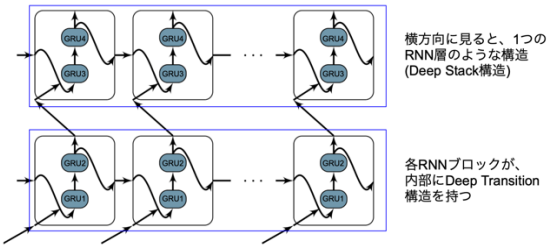


図 6：BiDeep RNN の構造

Label smoothing は分類問題において、正解となるラベルを one-hot ベクトルのような極端な分布で表現するのではなく、正解以外のラベルにも小さな値を与えておくことで過学習を抑制する手法の一つである。Dropout/word dropout も過学習抑制手法の一つであるが、中でも word dropout は入力文のうちのいくつかの単語を mask してしまう（その単語の embedding を全部 0 にする）方法である。

Layer normalization は共変量シフトを解消するための手法の一つで、あるニューロンに一つ前の層から入ってくる値（重みをかけた後の値）を、同一層の他の全ニューロンに入る値の中で正規化する。共変量シフトとは訓練途中で、各層への入力の分布が変化してしまい、うまく学習ができなくなるという現象で、これを抑制する方法は layer normalization の他にも Batch Normalization、Weight Normalization、Instance Normalization、Group Normalization などいくつか存在する。

Tied embeddings は Decoder の入力（1 つ前の出力単語をベクトルに変換したもの）と出力（出力層で目的言語の語彙サイズ次元に変換したもの）の embedding matrix を共通にするという手法で、これによりパラメータを削減することができるし、精度も若干向上することが報告されている。なお原言語と目的言語で語彙が共通（shared sub-word）なら encoder の embedding matrix も共通にすることが可能であるが、この論文では decoder の 2 つだけを共通にしている。

サブワードの語彙サイズについては、high-resource 設定ではあまり精度に影響しないが、low-resource では小さくした方が良いとしている。これは low-resource な状況では、低頻度なサブワードが発生してしまい、その分散表現がうまく学習されないことが要因である。そこでサブワードセットを学習した後に、閾値以下の低頻度なサブワードを高頻度なサブワードでさらに分割する（実験ではコーパス中の頻度 10 以下のサブワードをさらに分割している）。

ハイパーパラメーターの設定については、

low-resource な状況ではネットワークサイズは小さくし（過学習の抑制）、バッチサイズは小さくし（データ数に対してバッチサイズが大きすぎると学習が収束しない）、dropout を積極的に使う（過学習の抑制）方がいいとしている。

最後の lexical model は、入力文の各単語の embedding に対して、attention の重み付き和を取り、residual connection つき feed forward 層を通したものを、次の単語の予測に使うというもので、出力と入力の間での単語アライメント情報のようなものを積極的に利用したいという発想で取り入れられている。

以上のテクニックを組み合わせて用いることで図 7 のように 10 万単語(約 5000 文)という非常に少ない対訳コーパスであっても PBSMT よりも良い翻訳精度が達成できたと報告している。

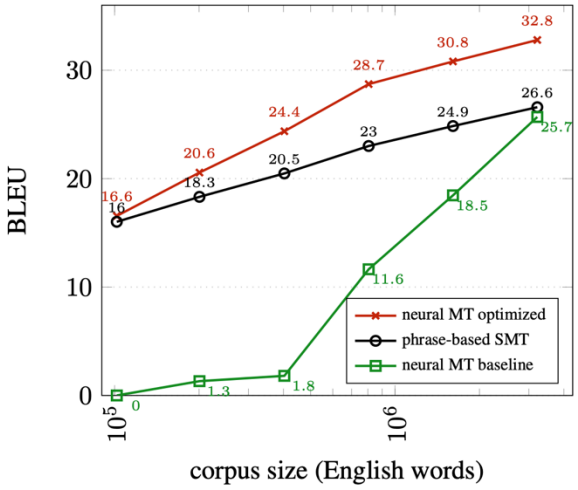


図 7：独英翻訳におけるコーパスサイズと BLEU スコアの関係

4. Post-editese: an Exacerbated Translationese (MT-Summit2019, best paper)

人間によって翻訳された文は、その言語のネイティブが書いた文と比較して simple で normalized で explicit で、元の文の特徴（文の長さなど）の影響を受けているという傾向があることが知られており、これを “Translationese” と呼ぶ。要は一種の方言のような特徴を持っているのである。これは人間の翻訳の質が悪いということではなく、そもそも人間の翻訳は特定のスタイルに従うようにとか、与えられた用語集の訳語を使うようにとかいうような制約の下で行われるものであるため、当然の性質である。

本論文はこの Translationese のような特徴が、post-edit された文にも存在するか（post-editiese は存在するか）、存在するとしたらその性質はどのようなものか、MT の手法の違いにより性質に違いはあるかの 3 点をいくつかの分析によって明らかにした。以下では分析の方法について述べる。

Lexical Variety: 以下の式で計算される type-token ratio (TTR)により、lexical variety を調べている：

$$TTR = \frac{number\ of\ types}{number\ of\ tokens}$$

TTR が大きいほど、様々な種類の単語が使われていることになる。この TTR を人間の翻訳（HT)、機械翻訳(MT)、ポストエディット（PE)の 3 つで比較したところ、HT > PE > MT の順になった。ポストエディットによって MT よりはより多様な表現が足されているが、人間の翻訳には及ばないという結果である。また翻訳システムごとの違いを調べたところ、TTR が最も高いのは SMT で、これと比較して RBMT や NMT は低い値となった。NMT は SMT よりも lexical variety は少ないという結論である。

Lexical Density: 以下の式で計算される lexical density を文の情報密度と定義し、これを比較している：

$$lex\_density = \frac{number\ of\ content\ words}{number\ of\ total\ words}$$

なお HT もネイティブの文章と比較して lexical density が低いことがわかっているという。分析の結果、PE も MT も HT よりは lexical density が低いが、両者の間には明確な差はないことがわかった。また PE の中では NMT を post-edit したものが他よりも lexical density が低く、MT の中では SMT が他よりも

lexical density が高いことがわかった。

Length Ratio: 原言語文と目的言語文の長さの比を以下の式で計算し、これを比較している：

$$length\ ratio = \frac{|length_{ST} - length_{TT}|}{length_{ST}}$$

ST は source text、TT は target text である。結果として MT は入力文を過不足なく翻訳するため、HT に比べて length ratio は小さく、PE も MT の結果をなるべく活かすように行われるため、HT よりも length ratio が小さいことがわかった。しかしこれは PE や HT を行う人の習熟度にもよることもわかった。

Part-of-Speech Sequences:目的言語文（*T*）を品詞列に変換し、原言語（*SL*）と目的言語（*TL*）の品詞列上でそれぞれ構築した言語モデル（LM_{SL}、LM_{TL}）で *T* の perplexity を計算し、この差を調べることで *T* が原言語と目的言語のどちらに近い構造（語順など）をしているかを調査した。結果として、HT > PE > MT の順に perplexity の差が大きいことがわかった。これは機械翻訳の結果が人間の翻訳に比べて語順の並べ替えが少ないという既存の報告とも合致する。またこの順に原言語文の影響が小さいとも言える。また SMT と NMT を比較すると、NMT の方が perplexity の差が大きい (=原言語文の影響が小さい) こともわかった。

以上の分析から、post-editese は存在し、translationese よりも simple で、（文の長さが）normalized で、元の文の影響を受けていることがわかり、さらに post-edit する前の生の機械翻訳は、さらに simple で、より元の文の影響を受けていることがわかった。

筆者らは今後、生産性などの観点から post-edit が増えるであろうが、そうなると目的言語においてpost-editiese の影響がどんどん強くなり、目的言語がどんどん simple で元の文の影響を強く受けたものに徐々に変化して行ってしまうのではないかという警鐘も鳴らしている。

5. おわりに

本稿では 2019 年に国際会議で発表された論文のうちのいくつかを簡単に紹介した。この他にも、例えば NAACL2019 の best paper の一つである “Probing the Need for Visual Context in Multimodal Machine Translation” は「マルチモーダル翻訳は本当に必要か？」という問いに対して、「文の情報と画像の情報が相補的であれば有効」という結論を導いているが、結論自体は当たり前であるし、その結論を導くための実験設定が人工的（非現実的）であり、個人的には正直微妙だと思っている。

最近は似たような内容の論文がたくさん出ており、違いを見出すことが難しくなってきている。またネットワーク構造をいじっただけの論文で、評価も数値(BLEU)でしかしていないようなものも多くあり、面白い論文を見つけること自体が難しくなっている気がしている。SNS 等で情報収集をするなど、効率的に面白い論文を探すことが大切になっているのかもしれない。

2014 年 9 月に最初の NMT が発表され、RNN+attention 機構を採用した NMT が大流行し、その後 2017 年 6 月には Transformer が発表され、今ではどの論文も Transformer を使っている。本稿の著者がオーガナイズしている Workshop on Asian Translation でも、上位のシステムはほぼ間違いなく Transformer を採用している。このように最近では 2,3 年ごとにパラダイムシフトが起きている状況なので、そろそろ次のパラダイムシフトが起きるかもしれない。

参考文献

[1] Wen Zhang, Yang Feng, Fandong Meng, Di You and Qun Liu: Bridging the Gap between Training and Inference for Neural Machine Translation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. (ACL2019), 4334-4343.

[2] Rico Sennrich and Biao Zhang: Revisiting Low-Resource Neural Machine Translation: A Case Study. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019), 211-221.
[3] Antonio Toral: Post-editese: an Exacerbated Translationese. In Proceedings of Machine Translation Summit XVII Volume 1: Research Track (MT-Summit2019), 273-281.
[4] Samy Bengio, Oriol Vinyals, Navdeep Jaitly and Noam Shazeer: Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1 (NIPS2015), 1171-1179.
[5] Arun Venkatraman, Martial Hebert and J.. Andrew Bagnell: Improving Multi-Step Prediction of Learned Time Series Models. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI2015), 3024-3030.
[6] Alex M Lamb, Anirudh Goyal, Ying Zhang, Saizheng Zhang, Aaron Courville, and Yoshua Bengio: Professor Forcing: A New Algorithm for Training Recurrent Networks. In Proceedings of Advances in Neural Information Processing Systems 29 (NIPS2016), 4601-4609.
[7] Antonio Valerio Miceli Barone, Jind˘rich Helcl, Rico Sennrich, Barry Haddow, and Alexandra Birch. 2017. Deep Architectures for Neural Machine Translation. In Proceedings of the Second Conference on Machine Translation, Volume 1: Research Papers, Copenhagen, Denmark.
[8] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Z. Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2818–2826.
[9] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15:1929–1958.
[10] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Edinburgh Neural Machine Translation Sys- tems for WMT 16. In Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers, pages 368–373, Berlin, Germany.
[11] Lei Jimmy Ba, Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer Normalization. CoRR, abs/1607.06450.
[12] Ofir Press and Lior Wolf. 2017. Using the Output Embedding to Improve Language Models. In Proceed- ings of the 15th Conference of the European Chap- ter of the Association for Computational Linguistics (EACL), Valencia, Spain.
[13] Toan Nguyen and David Chiang. 2018. Improving Lexical Choice in Neural Machine Translation. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 334–343, New Or- leans, Louisiana.

解説記事

人間の翻訳と機械の翻訳（１）：翻訳者は何を翻訳しているか？

影浦　峡
東京大学・大学院情報学環／教育学研究科

1. 少し長い背景

2019 年 7 月から 5 年弱（2024 年 3 月まで）の予定で、国立情報学研究所・NICT・関西大学・名古屋大学とともに新たな研究プロジェクト「翻訳規範とコンピテンスの可操作化を通した翻訳プロセス・モデルと統合環境の構築」(科学研究費基盤 S)を開始しました。機械翻訳が「実用化」されると言われる中で「実用化」には色々なレベルがあってとりわけ日本で約 2000 億円・世界で約 5 兆円の産業規模で世界経済の成長速度を超えて成長している産業翻訳の実際を想定しその中で機械翻訳がどのように「実用化」されているか・されうるかを真面目に考えようとするならば、翻訳プロセスで何がなされているかを高解像度で記述し共有する必要がある、というのが大元の問題意識の一つです。

もちろんこのように言うと、実践の側からはすでにできているのだからそのプロセスを記述するのは後追いにすぎないとの批判が出る可能性はあり、理論の側からはまさにそういうことをやってきたのに今更何をと言われそうではありますが、翻訳ではなく言語とのアナロジーでいささか緩やかに語るならば私たちは母語を使えるけれどそのプロセスは研究対象になっていますし、また例えば英語の冠詞の使い方を冠詞を有さない言語を母語とする人が適切に運用するための具体的な指針となる解像度で説明するものはなく、翻訳に関しても同様の問題はあるので謙虚に考えたいというのがこうした批判への答えになります。

機械翻訳に目を向けると深層学習による how のブレークスルーはしかしながら同時に how のブラックボックス化を進め、end-to-end で従来のステップを省くことが（少なくとも一見したところ）できるがゆえに[1]、そもそも end-to-end でといったときにもともとそこで扱われている（はずの）ことにおいて何がなされていて私たちは何をしてしまっていたのかという what の問いが改めて先鋭化されたとも言えます（同時にこの what の問いを促す状況は how の問いを人間が問うこととはどのようなことなのかという問いも提起しています。名古屋大学の佐藤理史先生が岩波『科学』の座談会で「言葉がわかるってどういうことか、われわれがまったくわかっていない、ということ」が東ロボの研究を通してわかったと述べていることは少しこの点に関係しています[2])。佐藤先生はまた同じ対談の中で「AI の 1 つの本質というのは、今、擬人化してしか説明できないことを、擬人化せずに説明する方法なんじゃないかと思う。たとえば『判断する』とか、『考える』とか、『わかる』というのは、全部擬人化していうわけですね。システムがわかるとか。もしそういうシステムができたならば、擬人化しない用語を使って説明できるようになっているんじゃないかと」[3]とも語っており、それは極めて大切なポイントなのですが end-to-end で何となくできてしまう（ような気になる）技術の時代にそこでは計算機でできたこととは別に改めて何をいかにやっているかを理解する課題が生まれて振り出しに戻ることになりそうです。

自然言語処理の先端で研究をしている Google の工藤拓さんは、しばらく前に次のように言っていました。

> NLP が他の ML 応用と少し異質なのは、データでなんとか学習可能な問題とは別に、知識がないとどうにも解けない問題がいっぱいあること。固有名詞の翻訳・読み推定が例。知識を学習データとして与えても学習される保証ないし、直接知識から出力させたいジレンマに陥る[4]。

What do translators translate?
Kyo Kageura
Interfaculty Initiative in Information Studies/Graduate School of Education, The University of Tokyo

この発言では「知識」を処理の側から遡行して考えているように見受けられますが、その際には処理の側あるいは知識そのものとしてどこかでそれが何かがわかられていることを先取りしており、けれども今、私たちは「それは何か」と言う問いが見え隠れしているところにいるので「知識」とは何かというもう一つ手前の問いに改めて直面していると言うことができます。少し飛躍して言うと、ここでの問題はフッサールが実験心理学は決して実験の対象とはならない意識そのものの分析を前提としていると述べたときに[5]、そしてヤコブソンがソシュールの音素はすでにそれに対応した外在的データを想定していると述べたときに[6]、直面していた問題に近いものです。

そこにいきなり飛び込んで悩むのは一つの手ですが、その前に極めて具体的なところでまだ色々とできることはあるだろう、基本的にあまりわかっていない領域を前にもちろんその領域がそもそもわかられていないのはその領域をわかることがそもそも何を前提としてしまうことになるのかという困難な問題を抱えているからだと言えば確かにその通りではあるけれどそれに気づかぬふりをして粛々と測量できるところを測量していくことが必要だし重要であるような段階はあきらかに存在すると考え、「翻訳とは何か？」という問いを「翻訳とは、何に対し、どのような要素・属性を同定し、それにどのような操作を加え、どのような基準でその操作を終了させる行為か？」と定式化しなおしてそれを具体的な翻訳プロセス・モデルとして書き下すことを目指す研究プロジェクトを立てたわけです。ここで「翻訳プロセス」は目線の動きとかそういったことではなく、外在的に操作可能なものとして捉えます。前置きが長くなりましたが、本稿では「何に対し」の一番大元を考えてみることにします。

2. 翻訳者が翻訳するもの

「翻訳する」が取りうる目的語を考えてみましょう。

1. 「文を翻訳する。」
2. 「文章を翻訳する。」
3. 「文書を翻訳する。」
4. 「小説を翻訳する。」
5. 「修理書を翻訳する。」
6. 「本を翻訳する。」

いずれも特に不自然ではないと(多くの人が思うと)思います。ここで「翻訳」を、多少ぎくしゃくした表現になりますが「言語分析」に変えてみます。

1. 「文を言語分析する。」
2. 「文章を言語分析する。」
3. 「文書を言語分析する。」
4. 「小説を言語分析する。」
5. 「修理書を言語分析する。」
6. 「本を言語分析する。」

1 と 2 は（もともと「言語分析」をサ変動詞化したことからくる不自然さを除けば)問題なさそうですが、3 からは少し奇妙に感じられます。3 からは、「文書の言語を」（あるいは「文書の言語表現を」）等々と言った方が自然な感じです。このことは、「文」と「文章」が言語（表現）のカテゴリ（少し曖昧な表現ですが）を表す一方、「文書」、「小説」、「修理書」、「本」は言語表現のカテゴリではないことを示しています。少し視点を変えて「言語分析する」を単に「分析する」に変えて、例えば「修理書を分析する」や「本を分析する」としたとき、分析の対象／レベルは言語学的な意味での言語表現とは違う何かであることも示唆されます。そして「翻訳」が「文書」「小説」「修理書」「本」を目的語として不自然でなく取れることは、「翻訳」という行為が言語表現（のみ）を対象としたものではないことを、もちろん日常的な言語使用の感覚に依拠しすぎるのはよくないことですが[7]、示唆しています。「X 冊翻訳した」とは言いやすいけれど、「X 冊言語分析した」というのは少し奇妙で、後者は「X 冊分言語分析した」ならば自然に理解できる感じがします。

なお、日本語の「翻訳」と「言語分析」に関してここで述べたことは、英語でも、筆者が多少はわかるフランス語でもスペイン語でも、ほぼ成り立ちます。ま

た、私の研究室に所属する韓国／朝鮮語を母語とする大学院学生は「X を翻訳する」であげた 6 つの例に対応する韓国／朝鮮語について「文を翻訳する」に対応する表現は少し不自然でそれ以外は自然であると述べており、中国語を母語とする大学院学生はほぼ対応する（ただし限定詞等を補足しないと自然さを判断しにくく、「本を翻訳する」は「一冊の本を翻訳する」「この本を翻訳する」に対応する表現でないと不自然といったことはあるようですが）中国語の表現について日本語の場合と同様の判断を下しています。

文（sentence）や文章（sentences/writing/text：この表現は境界に位置している感じですね）が言語表現のカテゴリであるのに対して、文書、小説、修理書、本は言語表現のカテゴリではない。私たちは、「小説を読む」「修理書を書く」「本を読む」などと当たり前に言います。このことは、「読む」「書く」といった行為が言語学的な意味での言語プロセスではない（あるいはそれにとどまらない）ことを示唆しています。

興味深いことに、欧州共通翻訳修士（European Master's in Translation）の 2017 年コンピテンス枠組みは「言語的および文化的コンピテンス」の基本的な部分を修士課程で翻訳を学ぶにあたっての前提（prerequisite）であると述べています[8]。また、翻訳者の高橋さきの氏は、Google 翻訳について、自分が翻訳者になる前にやっていたことをやっていると述べています[9]。翻訳が言語的なプロセスを前提とするけれども言語的なプロセスではないことあるいはそれにとどまらないことを、翻訳関係者は理解しているのです。技術の側から言うと、NMT における目標言語の流暢性向上は SMT やルールベースに比べて顕著に目に付くところですが、それゆえにこそ、翻訳を学んだことはないけれどそれなりに言葉はできる学部学生の訳と似ているといった点も目につくようになりました[10]。SMT より NMT の方が精度は高いはずなのに、MT+PE で PE を学生が担当すると NMT の誤りが見つけにくいために必ずしも効率が上がるわけではないとの報告もあります[11]。肯定的に捉えるならば、このことは、機械翻訳技術が翻訳を始めることができる地点に到達しつつあることを示しています。

では、翻訳者は何を翻訳しているのか。次に少し積極的な特徴づけを試みることにします。

3. 言語表現・文章・文書

前節で観察した例では、「言語分析する」に関して「文章」と「文書」の間に（あるいは「文章」を挟んで「文」と「文書」の間に）、境界が観察されました。ここでは文章は言語学が扱う概念と割り切って、「文章」と「文書」の相違を少し天下り的に整理します。

文章は定義上言語表現からなるけれど文書には図表も含まれるというのは両者の差異としてわかりやすいポイントですが問題はその差異そのものではなくそれでは図表も含むことが可能だけれどやはり言語表現を中心的な要素とする文書というものはそもそもどのようなものであるから図表を含んだ上で同一性を維持する単位として成立しているのかという点です。

「文章」を言語学的な概念と考え「文書」をそうでないものと考えるときに最初の便利な手掛かりになるのはフーコーの次のような言葉です。

> つまり、言語［ラング］とは、無限数個の言語運用［パフォーマンス］をゆるす有限な規則の集合であるのだ。これに対して、言説とは、じっさいに述べられた言語的まとまり［シークエンス］のみからなる、つねに有限で現働的［アクチャル］に限定された集合である。〔・・・〕言語の分析が、ある言説の事実に関して問う問いとはつねに、「どのような規則にもとづいて、このような言表がつくられたのか？　またしたがって、どのような規則によって他の同じような言表をつくることができるのか？」であるのに対して、言説の記述が立てるのはまったく別の問いであって、「このような言表が出現した、しかも、他のいかなる言表もその代わりには出現しなかったのは、どのようなわけなのか？」という問いなのだ[12]。

これを参考にすると、文章を、与えられた言語表現を前に、それを他でもあり得たものと考えたとき、すなわち「言語学的」な視点を採用したときに認識される単位、文書を、そうでしかなかったものと見なす態度において認識される単位と特徴付けることができるでしょう。フーコーは後者を言説の分析と呼んでいますが、もっとベタに言うとそれは図書館学における言語表現の扱いで典型的に採用されている視点です[13]。

この区別は機械翻訳と人間の翻訳との違いに比較的よく対応しています。機械翻訳は「言語」処理技術の応用で、起点言語のどのような文章（今のところ基本的には文）が与えられても対応できる一般的な処理のメカニズムを構築することを課題として発達してきました。「このような言表」から「他の同じような言表」にも対応できる規則を抽出することで実際に「他の同じような言表」に対応しようという課題です。一方、欧州共通翻訳修士のコンピテンス枠組みが示しているように[8]、人間の翻訳においてはそこまでは前提であって、翻訳はそこから始まります。翻訳者は、与えられた言語表現が歴史的に生み出された文書群のどこに位置付けられるどのような存在であるのかを検討し、言語表現としては他でもあり得たけれどそれにもかかわらずこのようなかたちでだけ出現した固有のものとして扱います。

実はこれは、翻訳に特異的なことではありません。前節で「読む」「書く」といった述語に言及しつつ述べたことと関連しますが、実際に、私たちが有意味な言語活動を行なうときには言語表現をここで言う文章としてよりも文書とみなしています。例えば、誰も、自分の研究を進めるにあたって先行研究を無作為に抽出し「別に他でもあり得たものだからそれでよい」とは言いません。査読者が「この研究に必須のはずのこの論文が引用されていない」といったコメントをするとき、議論は文章のレベルでではなく文書のレベルで行われていることが当然の前提とされています。

ところで、ある言語表現を純粋に歴史的に固有の単独的な存在とみなすだけならば、私たちは一般にそれに対して博物学的に列挙し整理し登録する以外の作業はできません。それだけだと私たちの有意味な言語活動は極めて限られるしまたそもそも特定の言語表現を生み出したり翻訳したりすることがどこに位置付けられるかもうまく説明できません。では私たちが文書を扱うとき何を扱っているのか。ここでフーコーが「言表」について述べた言葉が参考になります。

> 言表は、その誕生の時間的かつ空間的な座標と全面的に連動するにはあまりに反復可能であり（言表は､それが出現する日付および場所とは別物である)、一つの純粋形式と同じくらい自由であるには自身を取り囲み自身を支えるものにあまりに縛り付けられている（言表は、諸要素の一つの集合に関わる構築法則とは別物である）ということ。そのようなものとして、言表が備えているのは、変容することの可能なある種の重さであり、それが置かれている領野に関係する一つの重量であり、多様な使用を許す一つの恒常性であり、単なる痕跡のように不活性ではなく自分自身の過去にまどろむことのない一つの時間的永続性である。一つの言表行為が再開ないし再想起されうるのに対し、また､(言語学的ないし論理学的な)一つの形式が再現働化されうるのに対して、言表は反復されうるということ､ただしその反復は、常に厳密な諸条件のもとで可能であるということ。これが、言表に固有の特徴なのである[14]。

いささか乱暴に言うと、この領域は、比較的長い時間にわたる持続性の中で、私たちが言語や文化を超えて「合理的に疑問の余地なき」という理念のもとで殺し合わずに相違を確認したりすり合わせたりすることができる可能性を言語における思考とコミュニケーションにおいて維持するための条件とその中での言語使用からなるものです。言葉の意味は使用からわかるとか言葉の意味は使用を通して変わるという主張は逆説的にもまさにこの領域（それは再びいささか乱暴に言うと言葉の意味はイデア的同一性を有するという理念がそれなりに担保されているということでもあります）が蝕まれることなしに存在していることを前提と

しているからこそ可能なもので、それに甘えつつそのことを忘却してだらしない言語使用を賞賛する先に何が待っているかを私たちはあまり幸せとはいえない世界の現状から知ることができます。技術の側面では、そのことは、Tay が 24 時間にして人種差別主義者のナチス礼賛になったことに露呈しています[15]。

保守的な主張になりますが、私たちはその時々において妥当な有意味性を確保するために存在する条件をそこにノイズが入っていたとしてもイデア的な同一性を担保するものとして前提とせざるを得ません。実際、それを抜きにした言葉の使用はまさに今ここにある社会における差別や力関係を反映したハラスメントになりがちです。言語表現を文書としてみることは、「合理的に疑問の余地なき」に向けた思考を可能にする条件を維持し引き継いでいく行為に参加することで、翻訳はそのような行為の一貫です。これに対して言語表現を文章としてみることは、形式的に許容可能な範囲を考えることで、その範囲は言語学的な表現の適格性に関わりますが有意味性を担保しません（このことは言語学的な意味での意味論にも当てはまります）。

言語表現を文書として捉える視点は、「合理的疑問の余地なき」に向けた私たちの実際の言語使用が妥当性を有する範囲の存在を前提としそれを確保するものとしてある一方、言語表現を文章として捉える視点はその点は別のところで扱われるものとしてそれよりもはるかに広い範囲を適格なものとする、そして翻訳の視点は前者である、というのが暫定的な結論となります。

4. 文書に接することと解像度

長尾真先生が京大の総長をなさっていたと同じ時期に東大の総長をしていた蓮實重彦氏がプルーストとヴァレリーは同じ年の生まれだけれどプルーストは古びないのにヴァレリーはいかにも古臭いといった趣旨のことを言っていたことに多少は同意しつつも個人的には少なくとも一部のものについて決して嫌いではないポール・ヴァレリーが、風景画をめぐって次のように書いています。

> 森の樹々とか都市の外の土地というものは、わたしたちにとって、動物たちよりずっと親しみの薄いものであるために、恣意的な姿勢が芸術において増大し、たとえ粗雑であれ単純化が習わしとなる。人間の脚や腕を、まるで一本の枝でも描くように描かれたら、わたしたちは不快感を覚えるにちがいない。わたしたちは、鉱物界や植物界のフォルムに関して、可能事と不可能事をひどく区別しにくくしてしまった。こうして風景画には取っ付きやすいところが多くなる。だれもかれもが絵を描きはじめた[16]。

例えば翻訳の先端で活動している高橋さきの氏のようなプロフェッショナルの眼には、言語処理が語っている翻訳は、「人間の脚や腕を、まるで一本の枝でも描くように」描いているようなものと映るのかもしれません。興味深いことに、翻訳と言語処理との間ではなく、一つレベルをずらして言語処理と機械学習との間を見ると、そこでもどうやら同様のずれが意識されているようで、そのことは、Yoav Goldberg 氏が 2017 年に公開したウェブ記事に示されています[17]。

果物やお金と違い思考と知識は他人と共有しても減らないので（その「価値」が減るのは経済のシステムの問題で情報と知識の問題ではありませんし、完全に独自の他の人には誰にもわからない知識が価値を持たないことを考えると共有できること／共有することは思考と知識の基本的な要請でもあります）、縄張り争いをするのではなく「翻訳とは何か」を共有しようとするとき、けれどもその共有可能性はこれまでのところかなりの程度「できる」ことを前提としていて（あるいは佐藤先生の言葉を借りると「擬人化」されていて）、民主的な理解の共有のために必要な解像度を有していなかったことに気づきます。そしてそのための言葉を私たちは十分に有していなかったこと、有していても共有してこなかったことに。本稿では、翻訳者が何を翻訳しているかを一番大元のところで解像度をあげるのが私たちが新たに始めた研究プロジェクトの課題だと言いながらそれを裏切ってしまうような粗いかたち

で言葉にしてみました。

連載の枠をいただいたので、次回は、ここで「文書」と述べたもの―翻訳が対象としているもの―はこの世界においてどのような存在としてあるかを考えていきたいと思います。

注・参考文献

[1] 坪井祐太・海野裕也・鈴木潤 (2017) 『深層学習による自然言語処理』東京：講談社.

[2] 池上高志・石黒浩・梅田聡・佐藤理史・中島秀之・開一夫 (2019) 「[座談会] 人工知能研究は何をめざすか（前編）」『科学』89(4), pp. 371-383.

[3] 池上高志・石黒浩・梅田聡・佐藤理史・中島秀之・開一夫 (2019) 「[座談会] 人工知能研究は何をめざすか（後編）」『科学』89(5), pp. 460-469.

[4] 工藤拓 (2019) https://twitter.com/taku910/status/1100035871183536128

[5] Edmund Husserl (1987) “Philosophie als strenge Wissenschaft,” *Logos: Internationale Zeitschrift für Philosophie der Kultur* 1(3), pp. 289-341. [佐竹哲雄訳 (1969) 『厳密な学としての哲学』東京：岩波書店]

[6] Roman Jakobson (1976) *Six Leçons sur le Son et le Sens* [花輪光訳 (1977) 『音と意味についての六章』東京：みすず書房]

[7] 日常言語の分析が可能なのは、イデア的同一性を何らかのかたちで先取りしているからです。これは「日常言語学派」等の「主張」や「考え」とは関係なく論理的に要請されることです。

[8] European Master’s in Translation (2017) *Competence Framework 2017*. European Commission.

[9] 山田優（モデレータ）・井口耕二・高橋さきの・藤田篤・Chenhui Chu・宮田玲（パネラー）(2018) 公開シンポジウム「翻訳におけるテクノロジーを考える」日本通訳翻訳学会第 19 回大会.

[10] 影浦峡 (2017) 「改めて、翻訳とは何か：Google NMT が使える時代に」言語処理学会第 23 回年次大会, pp. 931-934.

[11] 山田優・大西菜奈美 (2018) 「それでも学生はポストエディターになれるのか？ ニューラル機械翻訳（Google NMT）を用いたポストエディットの検証」言語処理学会第 24 回年次大会, pp. 738-741.

[12] Michel Foucault (1968) Sur l'archéologie des sciences. Réponse au Cercle d'épistémologie. *Cahiers pour l'Analyse* 9, pp. 9-40. [石田英敬訳 (1999)「科学の考古学について―「認識論サークル」への回答」蓮實重彦・渡辺守章監修『ミシェル・フーコー思考集成 3―歴史学 系譜学 考古学』東京：筑摩書房, pp. 100-143]

[13] 著者は図書館情報学研究室に所属しています。

[14] Michel Foucault (1969) *L’Archéologie de Savoir*. Paris: Gallimard. [慎改康之訳 (2012) 『知の考古学』東京：河出文庫]

[15] Peter Bright (2016) “Tay, the neo-Nzai millennial chatbot, gets autopsied,” *Ars Technica*. https://arstechnica.com/information-technology/2016/03/tay-the-neo-nazi-millennial-chatbot-gets-autopsied/

[16] Paul Valéry (1938) *Degas Dance Dessin*. Paris: Gallimard. [清水徹訳 (2006) 『ドガ ダンス デッサン』東京：筑摩書房]

[17] Yoav Goldberg (2017) “An adversarial review of ‘Adversarial Generation of Natural Languag’ - Or, for fucks sake, DL people, leave language alone and stop saying you solve it.” https://medium.com/@yoav.goldberg/an-adversarial-review-of-adversarial-generation-of-natural-language-409ac3378bd7

イベント報告

MT フェア 2019 報告

内山　将夫
情報通信研究機構

1. はじめに

MT フェアは、2019 年 6 月 19 日に、アジア太平洋機械翻訳協会（AAMT）の主催により開催されました。140 名以上の参加があり、会場全体に熱気がありました。以下では、MT フェアについての、報告者の主観によるまとめを報告したいと思います。

2. AAMT 長尾賞・長尾賞学生奨励賞

MT フェア冒頭には、AAMT 長尾賞・長尾賞学生奨励賞の発表がありました。第 6 回 AAMT 長尾賞学生奨励賞は江里口瑛子氏（博士学生時代は東京大学、現在は Microsoft Corporation）、第 14 回 AAMT 長尾賞はソースネクスト株式会社が受賞しました。

江里口氏の講演では、博士論文についての概要が発表されました。従来は、文字列から文字列への翻訳としてニューラル機械翻訳（NMT）が設計・利用されていたところに対して、原文・訳文の構文情報を活用した NMT を提案した先駆的な研究です。講演の中では、研究の難しい点として、たとえば、訳文側の構文情報を活用した場合に、確かに訳文側の構文情報を訳文の生成とともに構築できるのであるが、それにより大幅に精度が向上するわけではないというものがありました。報告者の感想としては、訳文側の構文情報の利用法に関しては、訳文側にタグを埋め込まなければいけないようなときに、タグ情報を構文情報とみなしての処理ができないかというものがありました。

ソースネクスト株式会社からは川竹一氏による講演がありました。本講演によりますと、ポケトークWは 40 万台が累計で販売されているということです。また、ユーザーの声が多数紹介され、機械翻訳が、ポケトーク W により、実際のコミュニケーションを助けていることが分かりました。ポケトーク使用に関する 1 か月間の統計では、101 カ国でポケトークが使われたことが分かったそうです。

報告者の感想としては、半世紀以上にわたる機械翻訳の研究成果が、ついに幅広くコミュニケーションに実用されたことが嬉しかったです。また、機械翻訳の精度が重要なことはもちろんとして、それをどのようにうまく使えるようにしていくかが非常に重要だと思いました。

3. アジア翻訳ワークショップ（WAT）の紹介、黒橋禎夫氏（京都大学・AAMT 副会長）

WAT は、アジア言語を対象とした機械翻訳評価ワークショップです。対象としている言語は、日本語、中国語、韓国語、インドネシア語、ヒンディー語、インド系諸語、ミャンマー語、英語、ロシア語、クメール語です。WAT は、2014 年より毎年開催しています。また、2019 年は EMNLP-IJCNLP2019 併設で行われます。

MT Fair 2019 Report
Masao Utiyama
National Institute of Information and Communications Technology

翻訳タスクの分野は、論文、特許、新聞・ニュース、混合、レシピです。2014 年当初は、 ASPEC という論文コーパスによる論文翻訳タスクを実施しました。このコーパスは現在、論文の機械翻訳に関する標準的なコーパスになっています。

参加チーム数は 2014－2018 年は、12－15 です。2019 年は 20 チームほどとかなり盛況になりました。

翻訳結果評価については、自動評価については、自動評価サーバーを利用して随時できるようになっています。人手評価については、一対比較評価は、ベースラインシステムの機械翻訳結果と比較しています。正確性評価としては、一対比較評価結果の上位数チームについて、特許庁による内容の伝達レベルの評価をしています。

一対比較評価の対象文はテストセット中の 400 文を利用しています。5 人の評価者が、1 文毎に、ベースラインよりも良いか、悪いか、同程度かを評価しています。正確性評価としては、特許庁による特許文献機械翻訳の「内容の伝達レベルの評価」を 5 段階で評価しています。これは、一対比較対象文中の 200 文を文毎に 2 人が評価しています。

歴史的な発展としては、日英・英日論文機械翻訳について、2014－2018 年について、 BLEU 値も向上しているし、正確性評価に関する人手評価値も 3 点台から 4.5 点に近い段階まで上昇しています。つまり、2018 年の時点で 90%程度の内容の伝達度が達成されています。また、日中・中日論文機械翻訳についても 4.5 点程度に達しています。

他のタスクについては、特許機械翻訳については 4.5 点程度を言語対によらずに達成しており、混合については言語対にもよりますが 3－4 点という状況です。日英の新聞機械翻訳についてはコーパスサイズが小さいこともあり 2 点程度です。

2018 年のパネルディスカッションでは、アジア各国での機械翻訳の研究および実務利用の状況などについて話し合いました。また、アジア各国が協力し、機械翻訳技術の発展と普及に取り込むことで合意しました。これにより、今後の AAMT の真の Asia-Pacific 化が期待されます。

WAT のオーガナイザーは各国のメンバーから構成されています。2019 年では、EMNLP-IJCNLP2019 @香港の併設ワークショップとして開催（11 月 3 日か 4 日）し、翻訳タスク開催だけでなく、研究論文も募集します。翻訳タスクとしては、日本語とロシア語、タミル語と英語、クメール語と英語、適時開示文書、マルチモーダル翻訳などが増えました。また、適時開示文書の翻訳タスクでは、重要な数値・固有名詞にフォーカスして、それらが正確に翻訳できているかを評価します。

最後に、WAT に関連する国際会議として WMT の紹介もありました。WMT は基本的には英語を中心とした翻訳タスクを開催しています。BLEU 値としては 30 以上のものがでています。（WAT における日英・英日ニュース翻訳の BLEU は 20 程度）また、Microsoft Asia が多くのタスクで良い成績を収めています。日本からは NICT がチェコ語とドイツ語の機械翻訳について良い成績を収めています。

WAT は、アジアの機械翻訳研究の活性化・データ整備等に一定程度の成果を得ています。この会議・コミュニティを成長させ AAMT の真の Asia-Pacific 化の基盤とすることが期待されます。

4. 招待講演 1：「トランスパーフェクトの機械翻訳ソリューションについて」 Diego Bartolome 氏（TransPerfect）

まず、AI の紹介として、Hype Cycle for Emerging Technologies, 2018 というものがありました。これは、技術の発展の過程とその技術に関する期待度合いを、いったんは速やかに頂点に達した後で、同じくらい急速に下がり、その後また緩やかに上昇しなだらかになるカーブで描く模式図です。それによれば、AI は、2018 年においては、期待がまさに頂点に達している段階ということです。報告者としても、AI に関する世間の期

待はかなり高いと思っていたので、納得のいく図でした。

また、AI は、非常にたくさんの分野に応用されるようになりました。特に、自然言語処理は広く活用されています。かつての機械翻訳は、翻訳精度が低いにもかかわらず、人間翻訳者の仕事を奪うものであると恐れている翻訳者もいましたが、現在では状況が異なっています。現在の機械翻訳の精度は非常に向上しており、翻訳メモリと同様に、翻訳全体の生産性を向上するものと考えられています。

TransPerfect は世界最大の翻訳サービス会社です。TransPerfect では、40％程度の翻訳が AI-based workflow で処理されています。また、40 言語方向以上で NMT が使われています。また、何百億もの単語が機械翻訳で処理されています。

機械翻訳は、1954 年から現在まで様々な手法が研究開発されており、着実に進歩してきましたが、現在の NMT であっても、まだたくさんの問題が残っています。

機械翻訳を使うことで、翻訳サービスの範囲を広げることができます。ポストエディットとしては、Full か Light かなどがあり、また、機械翻訳のみを利用するときでも、カスタマイズした NMT を使うとか、TransPerfect 提供のカスタマイズなしの NMT を使うとかの選択肢があります。これらからどのサービスを使うかは、クライアントの目的や予算によるものです。

機械翻訳品質の比較という点では、TransPerfect のエンジンは、ベンチマークとしているエンジンと比べて、15－23%程度、編集距離（edit distance）の観点から優れています。

ビジネスへの機械翻訳の利用という点では、小売り・ファッション・ライフサイエンス等様々なパートナーがいます。E-Discovery においては、機械翻訳が非常によく使われています。E-Discovery では、非常に大量の複数言語の文書を読む必要があります。また、金融方面でも機械翻訳が非常によく使われています。

TransPerfect の NMT は、他の無料・低コストの NMT エンジンと比べて、セキュアであり、利便性も高いです。NMT の用途としては、出版等は目的としていないで、 casual translations が目的です。

その他の NMT の用途としては、E コマース、B2C などがあります。E コマースでは、クライアントの翻訳に関する予算を最適化してビジネスの目標を達成します。また、クライアントに合わせて、機械翻訳のカスタマイズなどをしています。今後の方向としては、API の提供やポストエディターの育成が重要です。

報告者にとって、当講演で最も印象に残った言葉は「Optimize budgets」です。つまり、クライアントの予算に合わせて最適なソリューションを提供することが大切だと思いました。

5. 招待講演 2：「AI を取り巻くデータの課題と責任～データから見た研究とビジネスの壁～」田丸健三郎氏（AI データ活用コンソーシアム・日本マイクロソフト株式会社）

AI 活用において、海外と日本には大きな差があります。海外では聴覚障碍者が教室にいる場合には、AI 活用により音声認識情報を提供することが増えていますが、日本国内ではそのような例はありません。

AI 活用が進まないのは、日本固有のデータがないことが原因です。データの活用に関して、米国では、法律等で禁止されていなければ問題ないとして機微な情報の収集と活用をしています。一方、日本では、ガイドライン等で指針が示されていなければ心配だということで、データの収集と活用が進んでいません。AI データ活用コンソーシアムの目的は、データの円滑な流通を促進することです。

日本におけるデータの課題としては、日本固有のデータが圧倒的に不足しており、その結果として、海外データによる偏った AI モデルができています。たとえば、赤べこなどの日本文化固有の画像データは海外データには含まれていません。

複数のモダリティによる AI モデル構築の重要性が増しています。また、データ（特に言葉）は、継続的に変化するため、AI モデルの性能を保つためには、テキストデータ・音声データ・行動データなどの様々なマルチモーダルなデータを継続的に更新する必要があります。

様々なデータ＋AI により都市課題が解決します。たとえば、気象・交通・GPS 等のデータを活用して、交通事業者横断的な最適化された運行スケジュールにより混雑を緩和することができます。また、障碍者と健常者の円滑なコミュニケーションの実現に対して、さまざまなデータの提供・共有基盤の構築を通して、AI 研究を促進します。

AI について責任の重要性が増しています。つまり、ソフトウェア・プログラムに関する製造物責任がありますが、データを学習させた AI を用いたアプリケーション開発では、学習の部分で不確定性・説明不能性がでてきます。

信頼できる AI のためには、信頼できるデータが必要です。データの信頼性には、データソースの信頼性に加えてアノテーションの質も重要です。ただし、対象とする問題によって、同一データであっても、アノテーションの仕様は異なります。

データの信頼性と AI 品質の関係としては、たとえば、アノテーションの誤りや悪意のあるアノテーションによりデータの品質が悪化し、それにより、AI の品質も悪化する可能性があります。したがって、アノテーション作業者の信頼性が重要になってきます。

データの信頼性の確保は容易ではありません。たとえば、データソースが何かを想定することは一般には難しいです。たとえば、センサーからのデータであっても、センサー内部の処理により、生データでなくて、加工済みのデータが出力されることがあります。また、データにバイアスがあるときには、AI の品質も影響されます。たとえば、データに人種のバイアスがあれば、AI の品質にも人種のバイアスがでてきます。

価値ある AI データ基盤は、実装・実現が非常に困難です。価値ある AI データのためには、データ自体だけでなく、そのデータの来歴「provenance」も同時に確保する必要があります。つまり、何が、誰が、何時、どのように、生成・クレンジング・アノテーション・評価したデータなのか、を記録する必要があります。

AI データ活用における国内における課題としては、データを持つ企業と研究技術を持つ企業間の連携があります。たとえば、コンプライアンス確認のために、膨大な顧客との会話データがありますが、国内では、これを AI で処理していません。（海外では AI で処理することで効率化ができています）ただ、データ保有企業と AI 研究企業との連携のためには、個人情報の取り扱い等が問題となります。

データに関して、取り扱いが変わってきています。旧来の分析では、データから統計量を抽出して分析用のグラフなどを作ったら、その段階で分析が終わりました。一方、AI では、データからモデルが作成された後は、モデルが永続的に利用されます。そのため、いわば、データが永続的に利用されるとも言えます。したがって、データを販売する側としては、データ利用により作られたモデルの利用方法までわからないと、データを販売できません。また、AI に外部データを活用するときには、その外部データについて、権利・個人情報・説明責任などの統制が重要になってきます。

AI データ活用コンソーシアムでは、知的財産・法令・データ共有基盤という 3 つの大きな課題に取り組みます。コンソーシアムの 7−8 割が研究者です。コンソーシアムにより、データ活用契約手続きの簡素化、会員間連携による AI 課題解決、組織内データ基盤構築の実現、ニーズに応じた AI データ収集提供のモデルが達成できます。

コンソーシアムの理事発起人としては、長尾先生が入っています。その他にも著名な先生方が関わっています。みんなでデータを流通させていきたいと思います。これまでにスタートアップと関わってきたことから感じる課題としては、データのライセンスが研究・

教育に限るというのが多いです。そもそも商用ライセンスがありません。そのため、スタートアップでのデータ活用が難しいです。これを解決したいと思います。

イベント報告

2019 年度 JTF 関西セミナー『翻訳をアップデートせよ～AI 時代の翻訳力を理論と実践で考える～』プリエディットの可能性を模索する

平岡　裕資
関西大学 外国語教育学研究科 博士前期課程在学

2019 年度第 1 回 JTF 関西セミナーが 2019 年 7 月 11 日（木）14 : 00 ～ 17:00 で開催された。本セミナーは、『翻訳をアップデートせよ　～AI 時代の翻訳力を理論と実践で考える～』というタイトルで、関西大学の山田優教授と彼の研究室に所属する大学院生が発表を行った。この発表では、機械翻訳が使用できる時代において、これから求められる翻訳のあり方を理論と実践の両方から紐解くことに主眼が置かれた。山田氏の発表の詳細は、JTF 関西セミナー報告のウェブページ等[1]でまとめてあるため、本紙では、発表の一部を担った平岡裕資のプリエディットに関連したトピックに焦点を当て、その詳細を報告する。

1. 非母語への翻訳（L2 翻訳）とテクノロジー

これまで、翻訳は第二言語から母語への言語方向が原則とされてきた（ISO 2015; 日本翻訳連盟 2012 など）。しかし、英語の非母語話者による英語への翻訳（ここでは L2 翻訳と呼ぶ）も視野に入れる必要性が見直されている。例えば、EMT（2017）では、ネイティブスピーカー主義ではなく、英語をリンガフランカとして扱われることによって、従来の「外国語から母語へ」という翻訳の原則に逆らう場合があることを明記している。実際に、L2 翻訳が実務としてなされる場合は多くある（Pietzak 2013）。

そして、この L2 翻訳をテクノロジーで支援するという考え方がある（Austermuhl 2016, Yamada 2019）。例えば、検索エンジンを用いたワイルドカード検索や画像検索などを用いることによって非母語話者には獲得しづらい自然な言語表現の選択を助けることができる。同様に、機械翻訳も L2 翻訳の支援に利用できる可能性がある。例えば、MT に入力する前の原文を編集するプリエディットでは、母語である原文を書き換えることによって、英語で書かれた複数の言語表現を選択肢として入手することができる。

2. プリエディットの種類と利点

プリエディットとは、前述の通り、翻訳しやすいように原文を編集する作業である。プリエディットの目的は、MT 結果の品質を向上させることによる後工程の全体的な効率化である。また、プリエディットを介するメリットが顕著に見られるのは一つの言語から複数の言語に翻訳する場合である。ポストエディットではそれぞれの目標言語で複数の修正が必要なるのに対して、プリエディットでは原文を一度のみ編集することで複数の言語における機械翻訳結果の向上を目指すことができる（宮田 2016）。実際にも、日本語から英語への翻訳を想定してプリエディットされた日本語が、中国語と韓国語を目標言語とした場合でも各々の MT 結果が向上する傾向にあるという研究結果がある（宮田・藤田 2017）。

プリエディットは、その作業環境によって二つに分類できる。一つはバイリンガルプリエディットと呼ばれるもので、これは MT 結果を参照しながら原文を書き換えるものである。これはアドホックな（試行錯誤

Report: “Update Translation: Practice and Theory Revealing Translation Competence in the Era of Neural Machine Translation” JTF Seminar in Kansai 2019
Yusuke Hiraoka, Kansai University

な）プリエディットとも呼ばれており（宮田・藤田 2017）、翻訳の品質が向上するまで可能な限り繰り返し編集を行うため、非効率である。このため、バイリンガルプリエディットは一つの工程というよりも、訳出時の部分的な利用が現実的だろう。例えば、原文の表現を変更することで、MT が複数の言語表現を出力してくれるため、その中から適切なものを選択するという使い方ができる。作業者の目標言語の運用能力（主にライティング能力）が低いほど、この利点は大きくなる。

一方でモノリンガルプリエディットと呼ばれるものもある。これは機械翻訳結果を参照せず、あらかじめ決められた編集方法・ルールに従って原文を書き換える作業である。これは、これまで研究の対象とされてきた制限言語（Controlled Language）と類似した概念であるが、既存の言語表現に修正を加えるという点で異なる。モノリンガルプリエディットは基本的には一度のみの書き換えであることから、時間とコストの面で効率的である。また、多言語化を想定した場合、一つの原文に対して修正を加えることで、複数の目標テクストの品質向上を目指すこともできる（宮田 2016）。また、機械翻訳結果を参照しないため目標言語の知識を必要としないことも一つの利点である。モノリンガルポストエディットのように、専門知識がある一方で英語が苦手な人がモノリンガルプリエディットを前行程で行うことで品質向上・効率化を図ることができる。

3. 誰がプリエディットするのか

では具体的に誰がプリエディットの作業者として適切なのだろうか。上で説明したプリエディットの種類と言語運用能力にわけて考える。

まず、バイリンガルプリエディットの場合、機械翻訳の結果を参照する必要があることから、おのずと起点言語と目標言語の両方の高い言語運用能力を要する。また一方で、起点言語の編集能力も重要となる。宮田・藤田（2017）では、編集能力の差によってプリエディットの類型と事例の数が顕著に異なること、そして編集能力の高い作業者の方が十分な品質の機械翻訳結果を産出する傾向にあることがわかった。また、同様の研究で、MT の挙動に対する理解も重要であることが示唆された。これらのことから、普段から翻訳に従事している翻訳者が適していると考える。また、前述したとおり、言語運用能力を補強すると言う意味で L2 翻訳に利用する方がより効果的だろう。

対してモノリンガルプリエディットの場合は、目標言語の知識は必須でない一方で、ルールに従って起点言語表現を書き換えるテクニカルライティングに近い技術が必要となるが、これは訓練を積むことによって習得できる（宮田・藤田 2017）。このことから、原文の理解と編集能力さえあれば、誰でも作業が可能となる。

4. プリエディットを含むワークフロー

では、プリエディットを実際のワークフローに落とし込むとすればどのようになるか。これを人間翻訳を規定する ISO 17100 とポストエディットを規定する 18587 の翻訳ワークフローと比較して考える。

基本的にプリエディットは全体のフローの効率化を図るものであり、MT 結果の品質を保証することは難しい（宮田 2016）。そのため、ISO/DIS 18587.2 のポストエディットより前の段階に入れ込む形が現実的だろう（図 1 を参照）。しかし、プリエディット後にポストエディットを介するかどうかはクライアントとの合意（最終訳にどれくらいの品質を求めるか）によって決定されるか、プリエディットを行なった段階でどれくらいの品質に達するかによって異なる。また、プリエディットされた訳文のチェック作業に関しては明記されたものがないことから、その確認事項およびチェッカーの資格などは、ポストエディットと同様に翻訳会社に依存することになる。

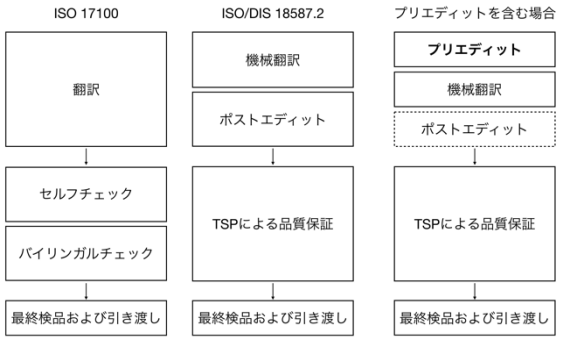


図 1. プリエディットを含む場合のワークフロー: ISO 17100 と ISO/DIS 18587.2 の図は森口（2017）から引用。TSP は Translation Service Provider を意味する。

5. 日英字幕翻訳のためのルールセットの構築

このような背景を踏まえて、筆者らは日英翻訳において字幕を対象としたプリエディット手法の構築を目的とした研究を行なっている。上に見た通り、モノリンガルプリエディットは単一言語話者による機械翻訳の使用を助けること、そして全体のワークフローを効率化できることから、大きなメリットがあると考える。これを考慮して、同研究では、a) 単一言語話者にも簡単に使用できる、b) 高い編集能力を必要としない、c) 字幕翻訳に適した要素を含む、の 3 つの方針に合わせたルールセットを構築することを目的に掲げた。

書き換え事例の収集と類型化

平岡・山田（2019）では、ニューラル機械翻訳（Google 翻訳）を用いた日英翻訳のバイリンガルプリエディットの実験を行った。この実験では、低品質の MT 結果が向上するまで対応する日本語文にプリエディットを繰り替えし、その結果得られた書き換え事例を収集・類型化を行った。MT の品質評価は Goto et al（2013）を改変した 3 段階評価を使用した（表 1 を参照）。

スコア	詳細
Good (3)	原文の情報が完全に翻訳されている。訳文は文法誤りを含まない。母語話者から見ても、語句の選択が自然である。
	語句の選択がやや不自然であるが、原文の情報が完全に翻訳され、訳文は文法誤りを含まない。
Acceptable (2)	原文のあまり重要でない情報の翻訳に些細な誤りがあるが、原文の内容は容易に理解できる。
	重要な情報の一部が欠落したり、誤訳されたりしているが、原文の中核となる情報は何とか理解できる。
Nonsense (1)	原文の意味が全く理解できない。

表 1. 正確性 の 3 段階評価

結果的に、プリエディットを行なったセグメントのほとんどの正確性のスコアが上がった。また、文字数の違反率に大きな差は見られなかった。類型化には、宮田・藤田（2017）の類型をもとに行なった。同研究は、複数の文書ドメインに対して網羅的に書き換え事例を収集しており、今回の研究で対象とする文書でも類似した類型が見られると考えた。

対象とした原文は日本の登録者ランキング上位のユーチューバーが作成する映像である。彼らの発話を書き起こしした日本語字幕に対してプリエディットを行なった。宮田・藤田（2017）に該当する 8 種の類型が得られた（表 1 を参照）。

	分類	類型名
1	構造	語順変更
2		主語の追加・削除
3		読点の追加・削除
4	内容語	特定表現の使用

5	機能語	時制の変更
6	ターミノロジー	固有表現
7	その他:エラー	表記ミス、文法ミス
8		必要要素の削除・復元、原文にない情報の追加

表 2. 書き換えの類型一覧

ルールセットの構築と検証

上記の書き換え手法により翻訳の品質向上が見られたことから、これらの書き換えが品質向上に寄与すると考え、試験的に 3 つの書き換え手法を含むルールセットを構築した。これらは、上記で挙げた書き換え例の中から、使用の頻度が高い、ルール化が可能、操作が容易などの理由で選出した。以下ではルールの詳細をまとめる。例文に含まれる MT 訳は「みんなの自動翻訳@TexTra」の汎用 NMT（2019 年 1〜9 月に使用）から収集した。

ルール 1. 読点の補完

省略された読点を補完する。文章の切れ目や語句の係り受けを明確にさせるとき、語句が列挙されている場合などに挿入する。字幕は読点の代わりに半角スペースが使用されることが多いため、これらを読点に置換する。

a) 語句が列挙されている場合

今 日本は 少子化だけじゃなく うつ病の問題 ダイバーシティ 大介護の問題 財政難 問題山積の国です。Japan is not only a declining birthrate, but it is also a problem of the problem of the problem diversity in the problem diversity and the problem of the problem of the problem.（MT 訳）

→今、日本は、少子化だけじゃなく、うつ病の問題、ダイバーシティ、大介護の問題、財政難、問題山積の国です。Japan is not only a declining birthrate, but also a problem of depression, a problem of diversity, a large care, a financial difficulty, and a problem pile.

b) 係り受けを明確にする場合

私は研究を続けながら、子育てに四苦八苦する母親たちと積極的に交流しました。I continued to interact with my mothers who were struggling with raising children while continuing my research.

→私は、研究を続けながら子育てに四苦八苦する母親たちと積極的に交流しました。I actively interacted with mothers who were struggling to raise their children while continuing to work.

→私は研究を続けながら、子育てに四苦八苦する母親たちと積極的に交流しました。As I continued my research, I actively interacted with my mothers who struggled to raise their children.

ルール 2. 主語と目的語の補完

曖昧な主語と目的語を明示する。曖昧とは、これらが省略されている場合、動詞との繋がりが語順的に不明瞭な場合などを指す。前者では文脈から予想できる主語・目的語を追加し、後者では語順を動詞より前に変更する。

a) 省略されている場合

磨きながら演奏出来るおもちゃ。A toy that can be played while brushing.

→子供が歯を磨きながら演奏出来るおもちゃ。A toy that children can play while brushing their teeth.

b) 動詞との繋がりが語順的に不明瞭な場合

誰かがその帽子を被った瞬間に、頭を突き破って凶暴なエイリアンが「ギャアァッ」と出てくる。The moment someone wears the hat, he pierces his head and the violent aliens come out with the giaceae.

→誰かがその帽子を被った瞬間に、凶暴なエイリアンが頭を突き破って「ギャアァッ」と出てくる。The moment someone wears the hat, a violent alien pierces his head and comes out with "giaceae.".

ルール 3. 固有表現の置き換え

人・土地・商品などの名称、文化的要素、専門用語など、一般的にあまり使用されない単語をあらかじめ目標言語の表記（代用語もしくはローマ字表記）に置き換える。

a) 代用語の場合

そして今代わりにしりとりというゲームを使ってアイディアを考えています。And I am thinking about ideas by using a game called Toritori instead.
→そして今代わりに Shiritori というゲームを使ってアイディアを考えています。And I'm thinking about ideas using a game called Shiritori instead.

このルールセットをもとに、TED Talks の日本語字幕に対してモノリンガルプリエディットを行なった。使用した MT システムは「みんなの自動翻訳@TexTra」の汎用 NMT で、筆者一名で作業を行なった。
機械翻訳結果の評価は平岡・山田（2019）と基本的には同じ手法で行なった。字幕の読みやすさに関しては、TED が規定する文字数制限のルール（字幕に表示できる文字数は表示時間 1 秒につき 21 文字まで）に準拠し、各セグメントの超過率を見た。超過率は以下のようにして求めた。超過率が 1 より大きい場合、そのセグメントは文字数制限に違反していることを意味する。

$$超過率 = \frac{実際の文字数}{表示時間(秒) \times 21}$$

以上のデザインで実験を行なった結果、プリエディット前後でおよそ 3 割程度の MT 結果で正確性の向上が見られ、平均スコアも有意に上昇した。一方で、スコアが低下したセグメントはおよそ 1 割だった。また、プリエディットの段階で解決したい Nonsense の数は MT と MT+PE で差はなく、依然として残っていることから、ルールセットの更なる改良が必要である。

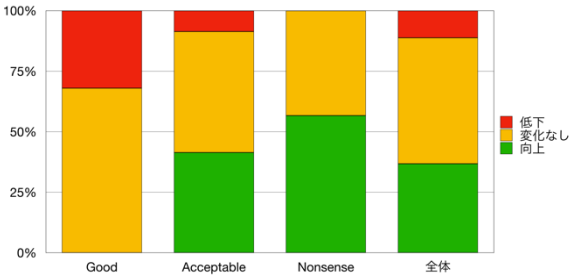
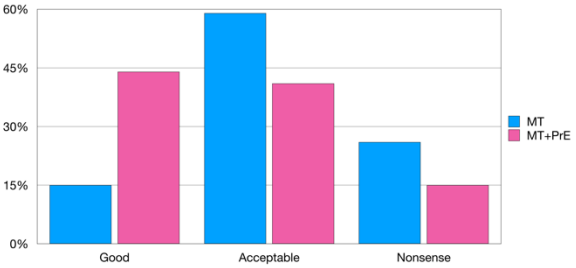


図 2. 各スコアにおける正確性の変化

図 3. 各スコアの割合



	MT	MT+PE
N	25	125
Mean	1.85	2.29**
SD	0.72	0.71
Nonsense	26%	15%

表 3. 正確性の平均スコア：**はウィルコックスの順位和検定で $p<0.01$ を示す。

文字数制限の超過率に関しては、プリエディット前後で超過率がおよそ 0.1 上昇したが、超過率が 1 を超える（1 秒 21 文字を違反する）割合に差はなかった。また、人間翻訳と MT+PE との差は 10%未満だった。つまり、文字数としての読みやすさに大きな影響はないと考えられる。

	MT	MT+PE	HT
N	25	125	25
Mean	0.58	0.69	0.56
SD	0.26	0.32	0.25
違反率	8%	12%	4%

表 4. 各データの超過率

以上の結果から、構築したルールセットは全体的な

品質の向上に寄与する一方で、まだ改善の余地があることが量的に検証された。特に、十分な品質（good enough quality）を実現するために、まずは低品質の数を減らすような改良が望まれる。また、今回の作業者は一名であったため、他の作業者でも同様の結果が得られるのかについても再度検証を行う必要がある。

6. まとめ

この報告ではプリエディットに焦点を当て、概要と筆者らの研究内容についてまとめた。プリエディットはL2 翻訳を支援する有効な手法であり、全体のフローを効率化させるポテンシャルを持つ。筆者らの研究では字幕に特化したルールセット（読点の補完、主語・目的語の補完、固有名詞の英語表記の３つ）を構築し、品質向上に寄与することが検証された。しかし、まだ改善の余地があるため、引き続き調査を行う予定だ。

[1] https://journal.jtf.jp/eventreport/id=770

[2] https://mt-auto-minhon-mlt.ucri.jgn-x.jp

参考文献

- Austermuhl, F. (2016). 'Don't mind the Gap: Using Technology to Empower English L2 Translators', invited talk for an open seminar, e-LINC (English instruction Network Center), Faculty of Foreign Language Studies, Kansai University, 18 May 2016.
- Goto, Isao, Ka Po Chow, Bin Lu, Eiichiro Sumita, and Benjamin K. Tsou. (2013). Overview of the Patent Machine Translation Task at the NTCIR-10 Workshop. In Proceedings of the 10th NII Testbeds and Community for Information access Research Conference, pp. 260-286.
- ISO. (2015). ISO 17100 Translation services - Requirements for translation services. Geneva: International Organization for Standardization.
- ISO. (2017). ISO 18587:2017(en) Translation services - Post-editing of machine translation output - Requirements. ISO Online Browsing Platform. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:en
- Pietrzak, P. (2013). Divergent Goals: Teaching Language for General and Translation Purposes in Contrast. *Second Language Learning and Teaching Correspondences and Contrasts in Foreign Language Pedagogy and Translation Studies*, pp. 233–240.
- Yamada, M. (2019). Language learners and non-professional translators as users. *The Routledge Handbook of Translation and Technology*, pp. 183–199.
- 日本翻訳連盟. (2012). 翻訳で失敗しないために翻訳発注の手引き [lit. For not getting it wrong with translation: a guide to ordering translation]. Retrieved from http://www.jtf.jp/pdf/translation_order.pdf
- 平岡裕資, 山田優.（2019）. 機械翻訳(MT)は字幕翻訳できるのか YouTube 字幕の記述および字幕におけるプリエディットの有効性の検証. 言語処理学会 第 23 回年次大会 発表論文集, pp. 934-937.
- 宮田玲. (2016). 翻訳テクノロジーを学ぶ プリエディット編[pdf]. Retrieved from http://www.apple-eye.com/ttedu/content/pg148.html
- 宮田玲, 藤田篤. (2017). 機械翻訳向けプリエディットの有効性と多様性の調査. *通訳翻訳研究への招待*, 18, pp. 53-72.
- 森口功造. (2017). ISO17100 と ISO/DIS18587.2 の要求事項の確認とポストエディット現場への影響. 言語処理学会 第 23 回年次大会 発表論文集, pp. 1157-1159.

イベント報告

MT Summit XVII（第 17 回）参加報告

小野　淳也
国立研究開発法人　情報通信研究機構

1. はじめに

MT Summit は，アジア，欧州，アメリカの順に隔年にて開催される機械翻訳(MT)に関わる会議であり，今年は 17 回目の開催であった．

研究者に加え，開発者，翻訳者，利用者やプロジェクト立案者が集う MT Summit の特色は，機械翻訳の技術発展だけでなく，社会実装に向けて議論する観点からも重要性を増している．

本会議に加えて，9 つのワークショップと 4 つのチュートリアルが開催された．本会議では，研究分野の Research トラックと，利用者・翻訳者目線の Users/Translators トラックに大きく分かれる．

参加者は 300 名強（290 名以上が事前登録）であった．主催者の Andy Way 氏によると，日本からの参加者は 25 名ほどであった．女性の参加者が多かった印象であり，本会議の招待講演者は全 3 件とも全員女性であった．Best paper award は，プレゼンテーションの質も考慮して，Antonio Toral 氏の”Post-editese: an Exacerbated Translationese”[1] が選ばれた．

ニューラル機械翻訳（NMT）のパラダイムシフトにより，機械翻訳の精度が劇的に向上し，高品質な出力が得られるようになったことで，実際の翻訳のワークフローの中に MT をベースとして，翻訳者が必要に応じて関わるというスタイルが広まりつつある．そこで，本会議の Users/Translators トラックを中心に聴講した．Workshop on Human-aided Translation では，翻訳のワークフローで用いられる各種技術のうち，主に品質推定（Quality Estimation; QE）および自動後編集（Automatic Post-Editing; APE）の導入事例についての紹介がなされた．また，QE や人間による後編集（Post-Editing; PE）の効果に関するものが多かった．

図 1: 会場の外観

- 開催地と会期

Dublin City University, ダブリン，アイルランド
会場：The Helix（図 1）
会期：2019 年 8 月 19～8 月 23 日
　　19～20 日：ワークショップ，チュートリアル
　　21～23 日：本会議
次回（第 18 回）開催地：2021 年 シアトル，アメリカ

2. 翻訳業界における世界の競争力

Vasiljevs 氏の発表[2]では，翻訳業界における世界の競争力をアジア，欧州，北アメリカの 3 地域ごとに比較した結果を示している．比較項目は，図 2 のように，Research, Innovation, Investments, Market Dominance, Industry, Infrastructure, Data の 7 項目ごとに，0 から 3（優位性あり）のポイントで示している．Research に関しては，2010 年から 2018 年 6 月までの機械翻訳研究における論文数に対して，著者

MT Summit XVII (17th) Report
Junya Ono
National Institute of Information and Communications Technology

別にランキングしており，図 3 のように TOP10 には日本人は本機構の隅田英一郎氏が 2 位，内山将夫氏が 7 位の二人であった．元データは scopus データベース (https://www.sco-pus.com/)から取得したとのこと．

図 2: 世界の競争力 [2]

AUTHOR NAME	NUMBER OF PUB.	COUNTRY	REGION
1.Way, A.	75	Ireland	Europe
2.Sumita, E.	67	Japan	Asia
3.Liu, Q.	55	Ireland	Europe
4.Casacuberta, F.	45	Spain	Europe
5.Specia, L.	44	UK	Europe
6.Zhao, T.	40	China	Asia
7.Utiyama, M.	35	Japan	Asia
8.Xiong, D.	35	China	Asia
9.Zhang, M.	34	China	Asia
10.Zhou, M.	34	US	America
11.Ney, H.	31	Germany	Europe
12.Yvon, F.	31	France	Europe
13.Neubig, G.	29	Japan	Asia
14.Zong, C.	29	China	Asia
15.Liu, Y.	28	China	Asia
16.Turchi, M.	28	Italy	Europe
17.Van Genabith, J	28	Germany	Europe
18.Costa-Jussà, M.R	27	Spain	Europe
19.Finch, A.	26	Japan	Asia
20.Toral, A.	26	Netherlands	Europe

図 3: MT 研究における論文数の著者ランキング [2]

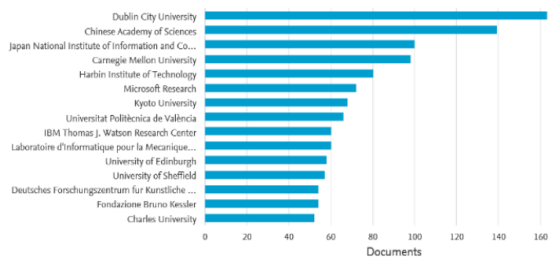


図 4: MT 研究における論文数の組織ランキング [2]

本発表[2]では，欧州における翻訳業界の優位性と足りない部分を調査した結果を示しており，欧州において Research, Innovation においては強みではあるが，それ以外においてはアメリカに遅れを取っていると示している．図 2 のアジアの競争力は，全体的にポイントが低く，Research は 1 ポイントとなっている．しかしながら，図 3 のように Top10 の名前にはアジア 5 人がランクされており，また図 4 の組織別の Top15 でもアジアがランクされているため，この結果は慎重に見るべきではある．

3. 全体的な傾向

全発表および，ワークショップ末尾のパネルディスカッションも含めた議論を通して，Post-Editing (PE), その次に Low-Resource のキーワードが印象的であった．以下，PE の要点を記載する．

- MT と PE の状況

翻訳業界全体が翻訳のワークフローの一つとして，MT を前段に置く可能性を評価・検証する動きになっている．MT を前段に置く PE は翻訳のワークフローにおいて標準的な工程と認識されている．

PE の結果は MT のバイアスを受けた Translationese あるいは Post-editese ではあるが，正しい訳文でもある．PE の結果を MT の訓練にフィードバックすることで，より高品質の訳文になることで，PE の頻度や PE に要する時間が短くなるなど，効率を上げる効果が確認されている．

また，MT の出力が流暢であるため，訳文の適否や修正箇所の検出に要する認知的負荷が大きく，PE を嫌う作業者は多い．

- 翻訳者と PE の問題

MT を利用しない生粋の翻訳者が，MT+PE に容易に移行できない理由として，自分自身の翻訳手法による高い生産性を既に実現していること，現在使用している慣れたツール・技術を使うこと（例えば，CAT

Tools, regular expressions, dictaphones…）が理由として挙げられる．今の MT+PE のままでは，移行するきっかけとして十分ではない．一方，翻訳者は正しい技術や具体的な内容を説明・支援すれば，非常に高いスループット(生産性)を実現する．

- Automatic Post-Editing (APE) の状況

APE の効果があるのは，そもそも機械翻訳の性能が低い場合であり，例えば，RBMT + Phrase-based APE や，PBSMT + Neural APE のケースであり，NMT の出力を APE によって改善することは容易ではない．そのため，現時点では，翻訳業界における利用者・企業が，ワークフローの中に APE を導入することに懐疑的な傾向である．

- Quality Estimation (QE) の状況

MT の結果を QE に通し，精度の高いものは直接ユーザへ返し，低いものは PE で修正していくシステムのように，QE の役割は明確ではあるが，翻訳のワークフローへの導入は限定的である．翻訳のワークフローにおいては，文レベル QE よりも語レベル QE の方が有用であるが，語レベル QE の方が難しい．文レベル QE は，文書分野を限定した場合は，PE を行うべきか，そのまま出版に耐えるかの分類に使い得るが，分野を限らない場合は性能が出ない．

QE は，翻訳のワークフローだけでなく，MT の改良の指針になり得る．例えば，MT の出力で精度が低いものを検出し，人手で訳を与えて MT の訓練にフィードバックすること，訓練データの誤りを検出して修正または除外することが挙げられる．

- 翻訳のワークフローにおける技術評価

MT や APE の性能は，BLEU や TER で自動評価されることが多い．しかし，翻訳のワークフローでは，そもそも BLEU が既に十分高いことが前提であるため，例えば，BLEU が 70 以上または， TER が 20 以下であり，自動評価結果は人間の評価と相関せず，自動評価が十分に機能していない．これは，参照訳が Translationese の場合に特に顕著である．翻訳のワークフローへの導入や企業のサービスローンチの段階では，MT や QE の効果を適切に判断するために人間評価のみが信用され，徹底的かつ継続的に行われる．

- MT Summit の参加者

参加者の中には MT の研究者や技術者だけでなく翻訳者も多数見受けられ，PE や QE の話題に追従しており，建設的な質疑が行われていた．WMT の QE タスクへの参加や LSP と同様の翻訳業務を行うなど，社内で様々な役割の人間が有機的に協業できているとのことである．

4. おわりに

ニューラル機械翻訳（NMT）によるパラダイムシフトにより，翻訳精度が劇的に向上し，翻訳業界全体がワークフローの一つとして，機械翻訳（MT）を導入し，後編集（PE）により，出版物にも使用可能な訳文を生成している．MT＋PE の傾向は今後も進むと思われる．しかしながら，NMT の訳文が流暢であることで，PE による作業負荷が高い傾向にあることや，生粋の翻訳者にとっては，まだ MT+PE への移行は容易ではない等の問題がある．一方で，ローカリゼーション業界における主なビジネスモデルは，世界レベルの速度で展開され，より良いモデル・技術のために時間を要しても導入させていくことが重要である．

また，翻訳業界が NMT 導入への流れがあるため，Low-Resource の言語やタスクに対しても NMT の可能性を検証している発表が印象的であった．NMT の導入の前にデータ集め・データのクリーニングを優先的に行う必要がある．

これまでは，人手をかける対象が評価や PE であったが，より高品質で多様なデータの作成にも今以上に注力していく必要がある．

参考文献

[1] Antonio Toral. (2019). Post-editese: an Exacerbated Translationese. In *Proceedings of Machine Translation Summit XVII Volume 1: Research Track*.

[2] Vasiļjevs, A., Skadiņa, I., Sāmīte, I., Kauliņš, K., Ajausks, Ē., Meļņika, J., Bērziņš, A. (2019). Competitiveness Analysis of the European Machine Translation Market. In *Proceedings of Machine Translation Summit XVII Volume 2: Translator, Project and User Tracks*.

イベント報告

PSLT 2019 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、2005 年の第 1 回から数えて 8 回目となるワークショップを、機械翻訳サミットの本会議前日にあたる 8 月 20 日（火）に開催した。オーガナイザは須藤が代表となり，綱川隆司先生（静岡大），宇津呂武仁先生（筑波大）のご両名とともに務めた。プログラム委員として研究会委員及び 6 名の国内外の研究者にご協力いただき、技術講演論文の査読を依頼した。

招待講演者の選定は昨年度末から研究会での議論を通じて行い、特許庁、機械翻訳研究者、技術文書翻訳に関連して医療言語処理研究者に依頼した。

技術講演に関しては、数日投稿期限を延長した結果 4 件の投稿があり、プログラム委員による査読の結果そのすべてを採択した。

従来本ワークショップは 1 日開催としていたが、本年は技術論文の投稿件数が伸び悩んでいることを考慮して半日の開催とした。しかしながら、結果的に 4 件の技術講演があったことと、本会議主催者から途中休憩の時間指定を受けたことにより、プログラム編成に苦慮することとなった。ワークショップ当日の参加者は 20 名強（事前登録 17 名）であり、およそ半数が日本からの参加者であった。

2. 招待講演

1 件目は、University of Amsterdam の Christof Monz 准教授による、ニューラル機械翻訳におけるドメイン適応についての講演であった。ニューラル機械翻訳が広く使われるようになるにつれて、実用において対象分野の対訳コーパスの不足という問題が重要となっており、現在は世界的に様々な手法が検討されている状況である。講演では、大量の分野外コーパスによる事前学習とごく少量の対象分野コーパスによる fine-tuning （微調整）を行う方式において、繰り返し学習の中で徐々に分野外コーパスの重みを下げていくための手法について紹介された。

2 件目は、特許庁の松谷洋平氏による、特許庁における機械翻訳の導入についての講演であった。特許庁では NICT の技術を活用し、日本語と英語／中国語／韓国語の間の機械翻訳技術を導入しており、J-platpat のような情報公開サイトや審査業務への利用が進んでいることが示された。

3 件目は、LIMSI-CNRS の Aurélie Névéol 氏による、医療言語処理と医療分野における機械翻訳についての講演であった。医療分野における言語処理の研究は英語を中心に行われており、実際の医療現場における様々な言語への対応が難しいという課題があり、フランス語等欧州言語について様々なコーパスの整備を進めていること、また日本でも日本語の医療言語処理データが整備され利用可能であることが示された。

3. 技術講演

技術講演 1 件目は NICT の小野氏らによる、多層 RNN に基づくニューラル機械翻訳における計算の効率化についての発表であった。複数の GPU を利用可能な場合に、各層の RNN を異なる GPU に担当させる（モデル並列）とともに、注視層や出力層のデータ

Workshop Report of PSLT 2019
Katsuhito Sudoh
Nara Institute of Science and Technology (Vice-chair of AAMT/Japio Special Interest Group on Patent Translation)

も分散させる（データ並列）というアイデアに基づく技術である。

技術講演 2 件目は Dublin City University（アイルランド）の Poncelas 氏らによる、ニューラル機械翻訳の適応についての発表であった。通常分野適応の問題はモデル学習時に大量の分野外データと少量の対象分野データを利用して対処するが、提案手法では翻訳対象であるテストセットの原言語側の情報に基づき学習データ中の類似する対訳文を選択的に利用してモデル適応を行った。

技術講演 3 件目は筑波大の飯田氏らによる（発表は共著者の Hung 氏が行った）、RNN に基づくニューラル機械翻訳における多段注視機構についての発表であった。複数ヘッドによる注視機構が Transformer をはじめよく使われているが、提案手法ではさらに異なるヘッドによる注視を混合して注視機構を多段化することで、特に長文における翻訳において精度改善を果たした。

技術講演 4 件目は Tampere University （フィンランド）の Nurminen 氏による、企業等の知財・特許担当者の視点から機械翻訳が業務における意思決定にどう貢献するかを、実際のユーザへのヒアリングによって調査した研究であった。本ワークショップではユーザサイドの話題がこれまで非常に少なかったことから、実際に機械翻訳がどう使われるかという意味で興味深いものであった。

4. 所感

前回と比較してニューラル機械翻訳が広く浸透し、基礎研究に加えて、実用的な研究・検討が進んできている印象を受けた。その一方で特許や技術文書の翻訳に関する多言語のリソースは医療関係の EMEA、論文抄録の ASPEC 等存在はするものの、機械翻訳研究のメインストリームとなっている WMT の News タスクのように広く用いられているとは言い難いのが実情である。特許等技術文書の翻訳は比較的大規模な対訳

招待講演時の会場の様子

コーパスが入手しやすく、文脈に依存する代名詞や省略等の問題も比較的小さく、また長文の翻訳等の問題もあって機械翻訳研究の対象として有益であると考えられる。しかしながら、実際に共通タスクとして利用できるリソースが限定されていること、それと関連し世界的な認知度がまだ低いのが実情と言えよう。他方、日本の特許庁における NICT 技術の利用や、欧州特許庁と Google との連携等実応用も進みつつあり、もはや基礎研究の対象ではないという見方もあるのかもしれない。ただ、（筆者の私見ではあるが）専門用語や長文の問題のように統計的機械翻訳の時代からその困難さが取り沙汰されてきた問題についての顕著な改善はまだ得られておらず、今後のさらなる発展が期待されるところである。

本ワークショップについても、技術的な内容のみならず、ユーザ視点での分析・提言や研究開発のためのリソースの提案等、これまで以上に幅広い展開を検討してもよいのではないかと考えられる。AAMT/Japio 特許翻訳研究会としてこの分野の貢献にどのような役割を果たすことができるか、継続して議論をする必要があろう。

イベント報告

テクニカルコミュニケーションシンポジウム 2019　発表報告

新田　順也
エヌ・アイ・ティー株式会社

1. 開催概要

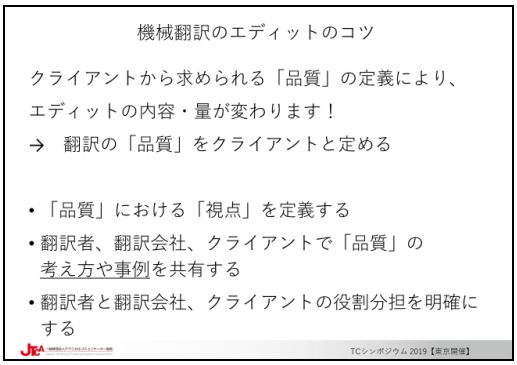
2019 年 8 月 27 日～28 日に東京学芸大学･小金井キャンパスにて開催されたテクニカルコミュニケーター協会主催のテクニカルコミュニケーションシンポジウム 2019 にて、株式会社川村インターナショナルの森口功造氏とともにミニセッションに登壇しました。このセッションでは､｢機械翻訳を活用したエディティング～プリエディットとポストエディットによる最適化～」のタイトルの元、日英翻訳におけるエディティングの手法や考え方を紹介しました。

最初に森口氏が、自社におけるニューラル機械翻訳（NMT）の導入事例に基づいた NMT の課題を説明しました。NMT 訳文を人手で修正する必要性を説明し、課題の 1 つである「原文に起因する問題」の対策として原文修正（プリエディット）のメリットを説明しました。続いて、私はプリエディットの狙いとプリエディット方法を説明し、最後に参加者を交えたディスカッションをしました。

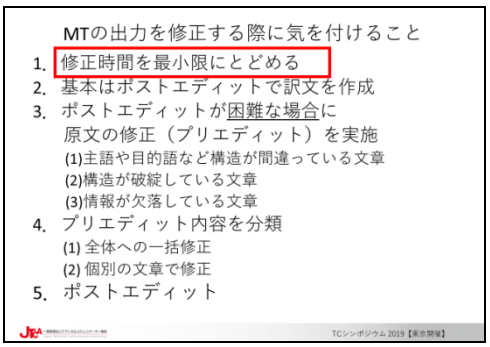
本稿では、プリエディットの具体例とディスカッション内容を報告します。森口氏の発表内容は、2018 年の同シンポジウムで発表されたことに近いため、本報告では割愛します。

2. プリエディットの狙い

機械翻訳を活用して翻訳する際には、翻訳者は翻訳文の品質に関してクライアントと合意する必要があります。エディティングの対象箇所やその修正内容は、合意した品質に応じて変わります。そのため、品質定義をしない限り修正箇所を特定できません。



エディティングとは、あくまでも NMT による訳文とクライアントの求める品質とのギャップを埋める手段です。さらに、プリエディットはポストエディットの作業時間を最小化する手段にすぎません。そのため、「ポストエディットをしすぎない」ことが重要です。私は、ポストエディットの作業が増えるような場合（例：文章構造が破綻している場合）にのみプリエディットをすればよいと思っています。



3. プリエディットの具体例

プリエディットを「文書全体で行う一括処理」と「文章ごとに行う個別処理」に区分します。

Report on Presentation at Technical Communication Symposium 2019
Junya Nitta
NIT Inc.

3-1. 文章全体で行う一括処理の例

原文ファイルで文字列を一括置換します。丸数字（①、②、③など）は誤訳の原因になるので、丸括弧で囲んだ数字に置換します。「⇒」（矢印）も誤訳の原因になるので、英文で使われる--> などの別の文字に変換します。さらに、明示的に括弧で区切ると出力結果がよくなることを紹介しました。たとえば、「⇒」を「[→]」や「(-->)」のように置き換えると出力が安定します。

この際に森口氏は、丸括弧で専門用語を囲むと出力が向上するという類似の事例を紹介しました。

3-2. 文章ごとに行う個別処理の例

文章ごとに修正する例として、無生物主語への変換方法を紹介しました。和文技術文書中の「●では、」や「●においては、」における「●」が文章の主語に該当するケースがあります。このような場合、これを無生物主語とする英文に訳したほうが自然な表現になることが知られています。

「この論文においては、消費増税の是非について議論されている。」という文章の機械翻訳は、「In this paper, the pros and cons of the consumption tax increase are discussed.」となります。これを「この論文は、消費増税の是非について議論されている。」とすると、無生物主語の「This paper discusses the pros and cons of the consumption tax hike.」という訳文を得られます。

このように「においては、」を「は、」に修正するだけで出力が変化するのです。主語と述語が正しく対応するように文章全体を修正する必要はないのです。

森口氏はこれに関連して、文章を区切るために文中に句点を挿入する手法を紹介しました。入れ子構造等の複雑な文章は機械翻訳で誤訳の可能性が高まります。このような場合には、文中の読点の直後に句点を挿入するだけで出力を変えられます。翻訳支援ツール（CAT ツール）を用いる場合、原文のセグメントを分割する必要がないため便利な方法です。この修正においても、原文は文法的に正しい日本語でなくてもよいのです。

4. 聴講者との質疑応答

4-1. CAT ツールと機械翻訳との併用方法は？

積極的に併用します。CAT ツール上で原文修正が可能な場合があるのでプリエディットのノウハウを活かせます。機械翻訳よりも翻訳メモリのほうが使いやすい場合があるので、柔軟に訳文を取捨選択します。

4-2. 機械翻訳を用いる場合の生産性の向上は？

クライアントの要求する品質や分野、言語方向により効率が変わります。一概に言えません。

4-3. プリエディットの内容の二次的な活用方法は？

プリエディット内容で AI をトレーニングするには多くの文対が必要となり現実的ではありません。翻訳メモリのように検索をする手法も試しましたがあまり役立たないように思います。

4-4. エンジン性能の進化は？

日英翻訳における無生物主語構文への変換が自動的になされることが以前よりも増えてきました。一般に性能が向上しているようですが退化する場合もあります。たとえば、「彼女は多趣味である。」を数か月前は「She has many hobbies.」と訳せましたが、現時点では「She is a hobby.」となってしまいます。

5. まとめ

今後も実務者目線で機械翻訳の活用方法を研究し、研究成果を発表していきたいと思います。

イベント報告

次世代音声言語研究シンポジウム 2019 の報告

山田　優
関西大学

1. 概要

2019 年 9 月 2 日（月）に奈良先端科学技術大学院大学ミレニアムホールにて『次世代音声言語研究シンポジウム 2019』が開催された。本シンポジウムは、奈良先端科学技術大学院大学の中村哲教授が代表を務める科研費プロジェクト「次世代音声翻訳の研究」の一環として開かれた。公式ページ [1] にもあるように、そのプロジェクトメンバーを中心に、音声翻訳研究に関連する各分野の第一人者を招聘して研究発表および討論を行った。

これまで機械翻訳といえば、書き言葉を訳す自動翻訳がよく知られているが、旅行会話などを中心にスマートフォンで動作するような音声翻訳の技術も一定の成果を収めている。しかし、話者の発話が終わる前に文末を待たずに内容を理解しながら翻訳を行う自動同時通訳に関しては、まだ技術開発の余地が多い。本プロジェクトの究極の目標は、この自動同時通訳機を開発することである。このために、音声認識技術、機械翻訳・通訳技術、音声合成技術を融合して、リアルタイムかつ漸進的な訳出処理を行える通訳機を開発することである。また開発の機械学習・評価用の通訳コーパスの構築も視野に入れている。

本稿ではその『次世代音声言語研究シンポジウム 2019』の報告をまとめる。特に研究プロジェクトの中核となる「同時翻訳」≒「同時通訳」技術に関する説明、デモ、発表をおこなった須藤克仁氏（奈良先端大准教授）の「講演音声の同時翻訳：翻訳技術と同時通訳コーパスの収集」を中心にレポートする。発表は①通訳と翻訳の違い、②訳出遅延時間を乗り越える技術と実際のデモ、③通訳コーパス構築の 3 部構成になっていた。以下、この発表順序に従って報告する。

2. 通訳と翻訳の違い

通訳と翻訳の違い、より具体的に、音声の同時通訳と書記言語の翻訳との違いはなんだろうか。一言でいえばそれは、原文（原発話）が提示されてから、訳出までの時間（遅延時間 latency）の違いである。翻訳は、原文テクストが静的で、時にスタイルガイドなどのルールや制約事項などにも従いながら、ゆっくりと腰を据えて訳していくものである。一方、同時通訳は原文が動的（通常は音声）なので、訳出に時間をかけすぎると、原文がどんどんと進んでいってしまう。例えば以下の原文を訳出することを考えてみる。

(1) The relief workers (2) say (3) they don't have (4) enough food, water, shelter, and medical supplies (5) to deal with (6) the gigantic wave of refugees (7) who are ransacking the countryside (8) in search of the basics (9) to stay alive[2].

翻訳では、文末を待ってから、(1) The relief workers 救援担当者は→(9) to stay alive 生きるための…という順序で訳出することができる。しかし、同時通訳で、この順序で訳してしまうと、遅延時間が長くなってしまい、また次の文が聞こえだしてしまうので、訳出処理がどんどんと遅れていってしまう。できるかぎり、原文の順序に近い形で訳したい。

そのために、普通、人間の通訳者がどうしているかというと、大まかに以下の 4 つの方略を使っていると考えられる。

Report: Symposium on Next Generation Spoken Language Research 2019
Masaru Yamada
Kansai University

a) 訳せそうなら訳し始める
b) 原文の順序にできるだけ近づけて訳す
c) 訳出の言葉を完結に圧縮したりする
d) 先を予測して待ち構える

これら４つのうち、現時点では a と b の２つしか技術的に達成できていないので、須藤氏は自らの技術を自動通訳機ではなく自動翻訳機のレベルであると述べている。

3. 訳出遅延を乗り越える技術

さて、同時翻訳を達成するために、本プロジェクトもニューラル技術を採用している。従来の統計システムに比べて、どこまで訳したのかを把握しにくい、訳抜けがあるなどの欠点が指摘される一方で、流暢性の良さ、そして開発コードのシンプルさなどの利点からニューラル・システムを採用している。

技術的関心の一つは訳出品質の「流暢性」を保ちつつ、「遅延（latency)」を最小限に抑えられるようにするための最善トレードオフを達成することだ。下記のように、英語の原文の文末を待たずに、適当なチャンクで日本語の訳出ができるようにするイメージである。

I am going to take you	皆様をお連れします。
on a journey	旅へ
in the next 18 minutes	今から 18 分後です。

このための提案手法は、既存の「Wait-K」[3] を援用し、それに独自の手法を組み合わせるというものであった。Wait-K では、単純に固定数の単語（トークン）を待ってから訳出を開始するというものである。擬似的な同時通訳がシミュレートできる利点があるのだが、翻訳の原文単位が短すぎると訳出精度が下がってしまうリスクや、短い入力からニューラルが勝手な予測をして訳出を進めてしまうリスクが伴う。例えば、入力に「ブッシュ大統領がプーチンと」とだけしか入力されていないのにもかかわらず、すでに訳文側に「President Bush meets with Putin」と勝手な「meets　会う」を予測して訳文が出てきてしまう。この場合は、そのあとに続く原文がたまたま「会合する」となっており，正しい訳文っぽくなったから良いものの、これを、本当の同時通訳の予測のプロセスと考えるには不十分である。

そこで、今回、次に出現する原文の単語の確率が高くるようなアルゴリズムを wait-k に組み、ベストな訳出単位を予測して＜wait 訳出を待て＞という指示を込もうというのが狙いだ。例えば、「This」のあとに、「is」が来るのが正解の時に（確率を max 1.0 とすると)、これ以外の単語（例：「are」「a」「this」など文法的に正しくないもの）が出現する確率をできるだけ低く見積もれるように学習させる。この機械学習に音声認識でよく使われる CTC（Connectionist Temporal Classification）を採用したというのが、今回の提案手法[4]になる。

結果は、概ね固定の単語数を待って訳出を行う Wait-k に比して、短文で約４単語（± 3)、長文で 20 単語超くらい待ってから訳出できるようになり、翻訳品質（BLEU、RIBES）においてもほぼ劣化なく、場合により従来手法を上回ることもできた。Wait-k で固定された単語数だけを待って訳出する場合、間違った予測をして誤訳してしまうケースも、本提案手法では、適切に待ってから訳出を行うことで誤訳が避けられるケースが観察された。一方で、本手法では、必要以上に訳出を待ってしまう場合もあったので、この点に関しては学習データを変えてみるなど、改良の余地も残った。

尚、講演では試作品のデモも披露された。TED Talk の英日通訳と、須藤氏自身が実際に読み上げる英語をその場でリアルタイムに通訳してみせた。翻訳品質の精度に関しては、お世辞にも素晴らしいとは言えなかったが、訳出されるタイミングに関しては、非常にポテンシャルを感じた。これからの向上が楽しみである。

4. 通訳コーパス構築とまとめ

最後に開発・テスト用の通訳コーパス構築について

の話があった。本科研プロジェクトは、合計 400 時間程度の日英・英日通訳コーパスデータの構築を目標にしている。音源のメインは TED Talks、ほかにもニュースカンファレンスやシンポジウムの通訳で、実際に人間の通訳者に通訳を行ってもらっている。テストセットのデータについては、経験 10 年以上、3 年以上、3 年以下の人のパフォーマンスを用意している。これまでに約 200 時間弱のデータ収集ができている。ほとんどは、TED Talks の通訳であるが、シンポジウムの通訳や、プレスカンファレンス（立教大学　松下佳世准教授の科研プロジェクトから提供を受けたもの）なども含まれる。

通訳コーパス収録で難しいのは、まず TED Talks という素材の特徴である。原発話者のスピーチ原稿は、非常に入念に練られており、おそらく何度も練習をしてからプレゼンテーションを行っているので、ポーズや淀みが少なく、スムーズ過ぎて、ある意味で不自然な話し方であるということ。これを同時通訳する通訳者にとっては、とても負担になるというのだ。

また、通訳収録を依頼する前に渡しておくべき事前資料も、悩みのタネであるという。通常、通訳者には、仕事の前に、スライドなどの資料を渡しておき、ある程度の準備をしてもらってから本番に臨んでもらうのが、オーセンティックなやり方である。しかし、TED Talks の場合、スピーカーの略歴情報などしか無いのが実情で、スピーチのトピックに関する周辺情報の入手が非常に困難である。とはいえ、公開されている映像なので、プレゼンテーションのスクリプトの全原稿は存在するわけだが、それを事前に渡して見てもらうと、今度は、通訳本番前に全文をわかってしまった状態になるのも、これまたリアリティに欠けてしまう。この塩梅を見極めながら、コーパス構築を進めているということである。

講演の最後に、須藤氏は、このような困難が伴うけれども、同時通訳は、非常にチャレンジングなタスクであり、それだからこそ、とても面白いと語っていた。それを計算機にやらせるということを考えても、また人間が言語的な処理をどのように行っているのかを考えるにも、通訳はとても興味深いということ。

さらに、評価という観点からも興味深いと付け加える。そもそも「良い通訳」とは、どのように評価されるものなのか。これは翻訳された訳出プロダクトの品質だけではない部分も、通訳パフォーマンス評価に関係してくることを示唆している。最後に、研究の過程で副次的に人間通訳者を支援できる技術開発も可能となるかもしれないと述べられた。この発表を聞いて、筆者は、本プロジェクトが幅広い野望と射程をもっていることをあらためて痛感した。

5. 発表内容

このように本プロジェクトは大きな目標に向かって進んでおり、次世代音声言語研究シンポジウム 2019 は、そのための関連研究の結集であった。以下に、その内容を列挙しておく。ここから垣間見られるように、複雑な通訳の解明とその技術的実現のスコープの広さを感じていただけることだろう。

- Special Talk: Intelligent Systems and a Language Transparent World Alexander Waibel (Carnegie Mellon University, USA/ Karlsruhe Institute of Technology, Germany)。
- 同時通訳に特有の難しさを乗り越えるには : 通訳コーパス構築プロジェクトの経験から
  松下佳世（立教大学）・ 山田優（関西大学）
- 同時に聞いて話すニューラル音声通訳に向けて
  Sakriani Sakti (NAIST)
- パラ言語音声翻訳と音声から音声への End-to-end 音声翻訳の試み
  中村哲（奈良先端大）
- 「賢い」同時通訳支援ツールに向けて
  Graham Neubig (CMU)
  講演音声の同時翻訳 : 翻訳技術と同時通訳コーパスの収集
  須藤克仁（奈良先端大）

- 音声翻訳のための柔軟な音声合成の進展
  戸田智基（名古屋大学）・高道慎之介（東京大学）
- 独立深層学習行列分析に基づく教師有り/半教師有り信号強調
  猿渡洋（東京大学）
- EEG からの精神状態および認知負荷の計測
  田中宏季（奈良先端大）
- 特定個人の音声対話アバターを 1 枚の写真からインスタントに生成する技術
  森島繁生（早稲田大学）

以上、自然言語処理の今後の可能性と興味を再認識できる、非常に有意義なシンポジウムであった。領域横断研究などの可能性が十分考えられ、まさに研究の学際性が求められる分野であると思った。筆者は人間の通訳翻訳研究を専門にしているので、これから、自然言語処理研究者らと協同研究できる機会が増えるのを楽しみにしたい。

【参考文献】

[1] https://ahcweb01.naist.jp/s2s-symposium-2019/
[2]水野的（2016）『同時通訳の理論』朝日出版
[3]Mingbo Ma, Liang Huang, Hao Xiong, Kaibo Liu, Chuanqiang Zhang, Zhongjun He, Hairong Liu, Xing Li, and Haifeng Wang. Stacl: Simultaneous translation with integrated anticipation and controllable latency. arXiv preprint arXiv:1810.08398, 2018.
[4]帖佐 克己・須藤 克仁・中村 哲英, 日同時通訳におけるニューラル機械翻訳の検討, 言語処理学会 第 25 回年次大会 発表論文集 (2019 年 3 月)、534-537

法人会員 PR

ユビキタス環境に対応した多言語音声翻訳システムの応用展開

株式会社フィート

1. はじめに

訪日外国人の増加に伴い、空港、鉄道、観光、宿泊、及び商業等の施設では、日本語の文字情報や音声情報の案内は、多言語化が必要とされています。また、労働者人口が限られるなか対人業務には効率化が求められ、音声翻訳機能を利活用した機械化が進んでいます。

株式会社フィート（2005 年設立、代表取締役 奥山美雪）[1] は、スマートフォン向けアプリとしての音声翻訳のみならず、IoT デバイスをはじめとしたあらゆる機器がネットワークに繋がる、ユビキタス環境に対応した多言語音声翻訳システムを展開しています。様々なビジネスの場面で業務の効率化やおもてなしを実現する多言語音声自動翻訳サービスの開発環境やシステム・インテグレーションを提供しています。

2. 多言語音声翻訳システムの開発環境

音声翻訳機能（音声認識・機械翻訳・音声合成）の利活用が様々な場面で求められ、固定のサービスに留まらないプラットフォームとしての音声翻訳システムの提供の流れが進んでいます。

多言語音声自動翻訳サービスの開発環境は、次に示す特徴的な多言語音声翻訳 API 及び SDK です。なお、音声翻訳エンジンは（国研）情報通信研究機構（NICT）の多言語音声翻訳技術をベースにしています。

・多言語音声翻訳 API（図 1）

－ クライアントの端末の環境に制限されず、HTTP 及び WebSocket 通信が可能な既存デバイスに音声翻訳機能のアドオンが可能です。

・多言語音声翻訳 SDK（for iOS / Android）

－ シンプルで利用しやすい I/F で、新規アプリの構築のほか、既存アプリへの音声翻訳機能追加が容易です。

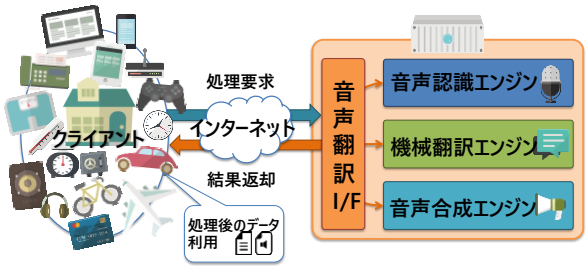


図 1 多言語音声翻訳 API

機械翻訳と音声認識や音声合成とを組み合わせた利用が可能です。現在、機械翻訳は 30 言語に、音声入出力は 11 か国語（日本語、英語、中国語（簡体字）、韓国語、インドネシア語、タイ語ほか）に対応しています。専門用語の固有名詞登録により翻訳の精度を高めることができ、サービスの付加価値向上に繋がります。

3. 多言語音声翻訳システムの適用事例

自治体窓口向けには、外国人との会話を支援する翻訳アプリ（図 2）を手掛け、実証実験を進めて自治体固有の単語を蓄積し、業務の外国語対応に貢献しています[2]。また、ホテルではフロントと客室との内線電話、駅構内では、案内係のアバターが利用者に多言語で施設を案内するデジタルサイネージを利用した実証実験[3] が行われるなど、システム適用先は多岐に亘っています。

図 2 翻訳アプリ例

さらに、建設現場では外国人労働者の業務を支援す

Developing and Deploying Multilingual Speech Translation Systems for the Ubiquitous Environment
Feat Limited

る眼鏡型ウエアラブル端末、及び、スポーツでは聴覚障害者へのバリアフリー支援として審判の指示を表示する装置などで展示協力し、今後、提供するシステムの利活用により音声翻訳の応用の世界が広がります。

4. 参考文献

1) 株式会社フィート：http://www.feat-ltd.jp，’19.10.1
E-mail：contact@feat-ltd.jp
2) 株式会社日本経済新聞社：自治体へ音声翻訳アプリ，日本経済新聞，2019.7.30
3) 東日本旅客鉄道株式会社：「案内ＡＩみんなで育てようプロジェクト」共同実証実験について，’19.6.18

法人会員 PR

みんなの自動翻訳@TexTra と翻訳バンクのご紹介

内山　将夫

国立研究開発法人　情報通信研究機構

1. みんなの自動翻訳@TexTra の特徴

みんなの自動翻訳@TexTra（以下、みんなの自動翻訳）は、情報通信研究機構（NICT）で研究開発している自動翻訳エンジンの Web サイトです：

https://mt-auto-minhon-mlt.ucri.jgn-x.jp/

当サイトは、無料で非商用に利用可能です。また、NICT の技術移転先企業では、当サイトと同等の商用サービスを提供しています。

当サイトは、日本語・英語・中国語・韓国語の 4 言語を中心に多言語に対応しており、一般的な Web 翻訳と同様に、誰でも簡単にお使いいただけます。

さらに、翻訳支援エディタをインストールなしでご利用いただけるほか、ワードやエクセルなどのオフィスソフトウェアのアドインが利用可能です。また、各種翻訳支援ツール（Trados 等）から自動翻訳エンジンを呼び出すことができますし、自分のプログラムから直接 WebAPI を呼ぶこともできます。

このように、みんなの自動翻訳は、自動翻訳に関する一式を便利にご提供しています。

2. みんなの自動翻訳の歴史

みんなの自動翻訳は、2014 年 6 月 19 日にオープンしました。オープン当初は、統計的機械翻訳技術に基づく自動翻訳技術を提供していました。その後、2017 年 6 月にニューラル機械翻訳（NMT）を導入しました。以下に、みんなの自動翻訳に関する NICT からの発表をリストします。これらは http://www.nict.go.jp サイトで確認できます。

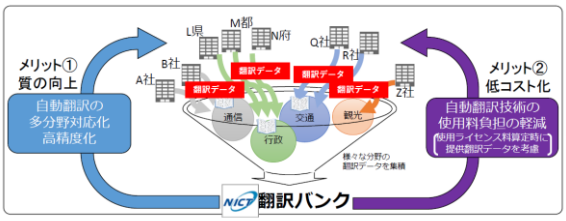
2014/6/19	第 9 回 AAMT 長尾賞を受賞
2014/6/19	***「みんなの自動翻訳@TexTra®」を一般公開***
2014/7/28	NICT と特許庁が多言語特許文献の高精度自動翻訳の実現に向けて協力合意
2015/3/30	科学技術文献データベースの作成に「高精度自動翻訳システム」を導入
2016/4/1	NICT と特許庁の特許文献の機械翻訳に関する協力の継続について
2017/1/18	高精度でセキュアな英文特許自動翻訳の提供開始
2017/6/28	***ニューラル機械翻訳で音声翻訳アプリ VoiceTra が更なる高精度化を実現***
2017/9/8	***『翻訳バンク』の運用開始－自動翻訳システムのさらなる高精度化に向けて、様々な分野の翻訳データを集積－***
2018/6/28	国立研究開発法人情報通信研究機構と MSD 株式会社 AI 多言語音声翻訳アプリ「VoiceTra」（ボイストラ）に、医学事典「MSD マニュアル」の 10 言語翻訳データを活用することで合意
2018/7/26	“VoiceTra” の音声翻訳技術が“POCKETALK® W”に採用
2018/7/31	スタンドアローンでの高精度ニューラル機械翻訳を実現 ～株式会社ログバーの新型オフライン翻訳機「ili® PRO」に採用へ～
2019/4/23	自動車法規文の自動翻訳をニューラル技術で高精度化 ～トヨタとの共同研究を通じ、英日・中日翻訳の実用度が向上～
2019/9/30	IR・金融分野向け自動翻訳エンジンの性能向上を確認 ～日本財務翻訳との共同研究により実用化へ～
2019/10/7	大規模翻訳データによる製薬業界向け AI 自動翻訳システムの最適化 ～情報通信研究機構（NICT）と R&D Head Club の共同開発～

Introduction of みんなの自動翻訳@TexTra
Masao Utiyama
National Institute of Information and Communications Technology (NICT)

3. 翻訳バンクによる NMT の高精度化

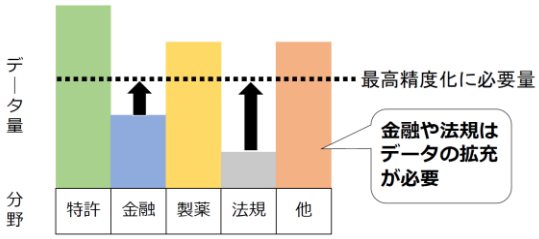
NICT 発表を時系列で確認すると、NMT の導入により、自動翻訳が一気に実用化するようすがわかると思います。

この実用化には、NMT だけでなく、「翻訳バンク」（https://h-bank.nict.go.jp/）が重要でした。「翻訳バンク」とは、NICT に翻訳データを集積し、自動翻訳の多分野化・高精度化を進める取組です。翻訳バンクでは、下図のように、NICT の自動翻訳技術の使用ライセンス料の算定の際に、提供が見込まれる翻訳データを勘案して負担を軽減する仕組みを導入しています。



NMT は優れた自動翻訳アルゴリズムではありますが、学習（訓練）データである対訳データがなければ自動翻訳エンジンはできません。NICT では、みんなの自動翻訳をオープンする前から、特許庁などと協力して、対訳データの収集に努めてきました。翻訳バンクはこの取り組みを加速するものです。

まず、高精度 NMT には、次図に示すように、対訳データが重要です。



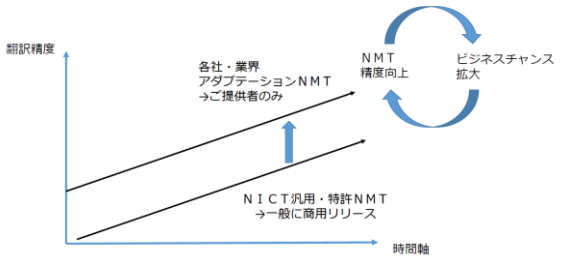
次に、集積された対訳データを活用して、NICT の汎用 NMT を構築することにより、高精度な共通基盤としての NMT が実現可能です。更に、「アダプテーション」という技術を適用することにより、さらに高精度な NMT を構築できると考えています。

「アダプテーション」というのは、NMT においては、既に訓練された NMT を、翻訳対象の分野に応じて追加で訓練することです。

まず、NMT の訓練自体は、非常に大まかにいえば、「現時点の NMT で原文を翻訳してみて、その結果が参照訳文と異なる度合いに応じて、NMT のパラメタを調整すること」で実施されます。この過程を大量の対訳文に対して何回か繰り返すことにより、共通基盤となる NICT 汎用 NMT が構築されます。

次に、上記パラメタ調整が完了した汎用 NMT に対して、翻訳対象分野の対訳文を利用して、同様なパラメタ調整を追加実施するのがアダプテーションです。このように、アダプテーションでは、すでに学習済みの汎用 NMT パラメタを翻訳分野に応じて追加調整するので、比較的少量の対訳データで当該分野に対する高精度な翻訳エンジンを構築できます。

また、翻訳バンクに対訳データをご提供いただくことで、NICT 汎用 NMT の底上げが期待できますが、それに加えて、アダプテーション NMT については、対訳データご提供者のみが利用できます。そのため、NICT 汎用 NMT とアダプテーション NMT には、次図の関係が成り立ちます。



このように、翻訳バンクでは、提供者が NMT の精度向上の恩恵を一番受ける枠組みとなっています。

4. 今後の展開

NICT では、みんなの自動翻訳に最新の NMT 研究成果を展開すると同時に、翻訳バンクにより集積された対訳データを活用することにより、実用的で高精度な NMT を社会還元することを目標としています。皆様のご協力をよろしくお願いいたします。

法人会員 PR

株式会社十印の AI 翻訳ソリューション

石川　弘美
株式会社十印 マーケティング部

1. 十印と機械翻訳

十印は 30 年以上、機械翻訳に取り組んできました。日本で開発が始まった 1980 年代前半には、機械翻訳の第一人者である長尾真教授（元京都大学総長）の依頼により、Mu-プロジェクトに参画し、言語研究を主とする部署を創設、翻訳システムメーカー様へ辞書を提供してきました。

2016 年のニューラル機械翻訳の登場後も、いち早く NMT の機械翻訳エンジンを使用したサービスに取り組み、お客様への機械翻訳導入と翻訳サービス提供しています。

また、エンジンを提供するだけではなく、お客さまの使用環境、使用されるコンテンツに最適なエンジンを選択し、翻訳支援ツールとの組み合わせ、プリエディット・ポストエディットなど導入から運用までをトータルで支援しています。

特に現在は、Google 機械翻訳エンジンを手軽に使用できる「Gen-pak」と、国立研究開発法人情報通信研究機構（以下、NICT）の開発したエンジンを使用した「T-tact AN-ZIN」というサービスを提供しています。

2. Gen-pak

Google NMT エンジンを API 経由でセキュアに利用できるようにしたソリューションです。用語統一が難しいと言われている NMT エンジンですが、独自のテクノロジーT-term により用語集の適用を実現しました。機能を翻訳に絞っており、訳出後の編集環境がありませんが、非常にわかりやすいインターフェイスで、訳した内容を簡単に・すぐに知りたい方に適しています。

テキスト翻訳はもちろん、ドラッグ＆ドロップするだけで Word や Excel、PowerPoint 形式のファイルを、書式を保持しながらそのまま翻訳できます。

なお、Gen-pak の名称は、西洋語から日本語への本格的な翻訳を初めて行ったとされる杉田玄白に由来しています。

3 .T-tact AN-ZIN

NICT が開発したニューラル自動翻訳エンジン「みんなの自動翻訳＠TexTra®」を民間企業や個人が商用目的で使用することができるようにしたサービスで、2019 年 7 月にリリースしました。簡単な手続、かつ低価格で NICT の高性能機械翻訳エンジンをビジネスに利用できます。日本語・英語・中国語・韓国語の 4 言語をはじめとした十数言語ペアの翻訳が可能な、高性能で安全なクラウド型自動翻訳システムです。

従来の機械翻訳サービスでは、翻訳対象の原文もしくは原文/訳文のペアを機械翻訳サービスの提供会社に取得されるセキュリティ上のリスクがありましたが、「AN-ZIN」では翻訳対象文書がユーザーの合意なしに取得されることはありません。また、通信のすべてが暗号化されますので、盗み見といったセキュリティ上のリスクから解放されます。

さらに、API を利用して、ユーザー専用の翻訳環境（プライベートクラウド/オンプレミス）を構築することも可能です。翻訳サービスに半世紀以上携わっている十印だからこそできる、導入から運用まで、導入されるお客様の環境に合わせたサポートをいたします。

なお、AN-ZIN の名称は、徳川家康に重用され、日本で最初に英語を翻訳したとされるイギリス人、三浦按針（ウィリアム・アダムズ）に由来しています。

TOIN's AI Translation Solution
Hiromi Ishikawa
TOIN Corporation

編集後記

内山　将夫

AAMT

本年から AAMT ジャーナル「機械翻訳」の編集長となった内山将夫です。どうぞよろしくお願いいたします。前編集長の宇津呂さん（現在 AAMT 副会長）が６年間編集長をした後の引継ぎです。

当ジャーナルは次のような方針で雑誌の編集をしていきたいです。

- 編集委員のメンバーがまずは自分で読みたい雑誌をつくる。
- ご寄稿いただいた方・法人のキャリア形成・ビジネス成功に資する。
- 機械翻訳専門誌として、機械翻訳に関する読者の適切な理解を促進する。

発行回数は６月号と１２月号の年２回で、AAMT の Web サイトから電子的に発行・一般公開されます。また、AAMT 主催のイベント等で印刷冊子が配布される場合があります。

当ジャーナル記事の著作権は著者に帰属しますが、機械翻訳の産官学における発展のために、自由な情報の流通を推進するという観点から、“Creative Commons Attribution-ShareAlike 4.0 International Public License” https://creativecommons.org/licenses/by-sa/4.0/ をお願いしています。また、当ジャーナルでは法人会員 PR として、AAMT 法人会員からの PR 記事も掲載します。

本号では、長尾真先生に「標準語」というタイトルの巻頭言をいただきました。機械翻訳研究者の私の感想としては、もし、機械翻訳エンジンへの入力テキストが規範的な日本語だけであったなら、機械翻訳は、だいぶ高精度だろうな、というものでした。ぜひ、規範的な日本語が世の中に広まってほしいと思いました。

長尾先生には、巻頭言だけでなく、表紙の「機械翻訳」という題字もいただきました。書については素人ですが、とても良い、伸びのある字だと思いました。本ジャーナルも、いただいた書のように、大きく伸びていきたいです。

本号では、第 6 回 AAMT 長尾賞学生奨励賞受賞記念記事を江里口瑛子様、解説記事を、中澤敏明様、影浦峡様、イベント報告を、内山、平岡裕資様、小野淳也様、須藤克仁様、新田順也様、山田優様に執筆いただきました。皆様お忙しい中、大変ありがとうございました。また、法人会員ＰＲとしては、株式会社フィート様、株式会社十印様、国立研究開発法人情報通信研究機構様からのご寄稿がありました。

どの記事も面白くまた全体として機械翻訳分野の広がりを感じるものだと思います。

編集委員のメンバー・事務局については、奥付けをご参照いただければと思いますが、解説記事・イベント報告をご執筆いただいたり、他組織との連携を担当いただいたり、印刷・デザインの手配をしていただいたり等、当ジャーナルを発行するにあたって、本質的にご参加いただけました。おかげさまでなんとかジャーナルの発行ができました。

今後ともよろしくお願いいたします。

Editor’s note
Masao Utiyama
AAMT

AAMTジャーナル「機械翻訳」No. 71

【　発　行　】アジア太平洋機械翻訳協会（AAMT）
　　ホームページ：https://aamt.info/
【　住　所　】〒619-0289
　　京都府相楽郡精華町光台3-5
　　国立研究開発法人情報通信研究機構　先進的翻訳技術研究室内
【Ｐｈｏｎｅ】0774-98-6332
【編集委員会】内山将夫　後藤功雄　中澤敏明　新田順也　園尾聡　森口功造　隅田英一郎
　　仲山裕子　小谷克則　宇津呂武仁　山田優　石川弘美
【表紙デザイン】泉谷東十郎
【　題　字　】長尾真
【事務局】石川弘美　奥麻里
【印刷所】株式会社 プリントパック

Asia-Pacific Association for Machine Translation (AAMT)
c/o National Institute of Information and Communications Technology
3-5 Hikaridai Seika-cho Soraku-gun Kyoto 619-0289, Japan
Phone: +81(0)774-98-6332


AAMT
Asia-Pacific Association for Machine Translation
アジア太平洋機械翻訳協会