ISSN 1883-1818

AAMT Journal

Asia-Pacific Association for Machine Translation

October 2018 No.69

アジア太平洋機械翻訳協会

目　次

CONTENT

巻頭言

機械翻訳の休眠打破

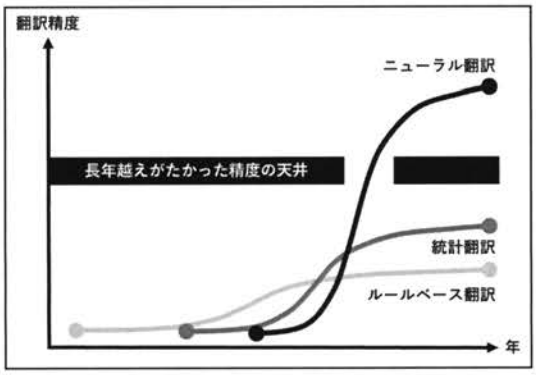
～ 機械翻訳を一挙に普及させ言語の壁を破壊しましょう ～

隅田　英一郎
(国研)情報通信研究機構

2018年6月にAAMT会長の職を拝命した隅田の初心を表明させていただきます。

1. これまでの機械翻訳

AAMTは、2年半後の2021年4月にちょうど30歳となるところですが、その誕生から今日までの間、機械翻訳技術の重点は、ルールベース翻訳にはじまり、統計翻訳を経て、ニューラル翻訳へと入れ替わりました（図参照）。



日本語と英語との対では、残念ながら、ルールベース翻訳、統計翻訳は広くは普及しませんでした。日本語と英語との対は人間にも翻訳が難しい言語対ですから、機械にとっては尚更難しい。機械翻訳の精度は当時不十分で広く導入するには時期尚早でしたが、誤訳をプロの翻訳者が正す仕事（ポストエディット）が多数行われました。当時のポストエディットは効率が悪く、翻訳者と翻訳会社にとってトラウマとなりました。研究者の努力で翻訳精度は上昇し続けましたが、機械翻訳に対する否定的な意見を払拭するには至らず、改善率も鈍り、翻訳精度は天井に近づきつつあるように思われました。

2. これからの機械翻訳

2016年に登場したニューラル翻訳は、天井かと思われていた翻訳精度を軽やかに越え、今までの機械翻訳と180度異なる福音をもたらしています。例えば、貿易の基本である英文ビジネスレター作成や高精度が必須である特許翻訳を大きく効率化し、結果として残業時間削減や有給休暇消化率の改善等、関係者の働き方を改革しつつあります。また、論文・特許の多言語検索サービス、技術情報配信・電子商取引・各種予約の多言語化、インバウンド向け音声翻訳のアプリや専用端末等の多数の活用事例が出ています。

ニューラル翻訳技術には、対訳データの着実な蓄積や世界中の研究者の競争による活発なアルゴリズムの改良という強力な推進力がありますから、翻訳精度の改善は続き、幅広い分野で活用されていくでしょう。

3. 機械翻訳の休眠打破とAAMT

まるで「桜」の休眠打破のように、機械翻訳は、花芽のまま長い厳しい冬を経験し今まさに春が来て開花しようとしています。

「機械翻訳の研究開発者と機械翻訳を活用する人々」が交流し連携する目的で設立された本協会は、機械翻訳普及の加速に最適な組織です。

2018年6月の総会で選出された理事の方々とともに一丸となって一挙に言葉の壁を破壊したいと思います。

よろしくお願いいたします。

Foreword

Breaking out of the Dormancy of Machine Translation

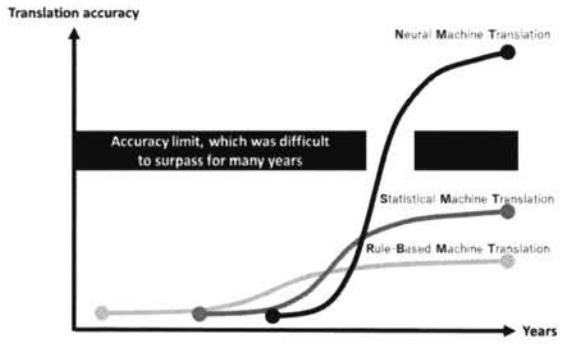
～ Let's spread machine translation in a single stroke to break down language barriers ～

Sumita Eiichiro
National Institute of Information and Communications Technology

The following presents the goals of Eiichiro Sumita, who was appointed to the post of President of the AAMT in June 2018.

1. Previous machine translation

The AAMT will reach the age of 30 in April 2021, which is two and a half years away. Between its founding and the present time, the priority of machine translation technology, which began as rule-based machine translation, passed through statistical machine translation and has now become neural machine translation (see the figure below).



In Japanese/English pairs, rule-based machine translation and statistical machine translation, regrettably, were not widespread. Even humans have difficulty translating between Japanese and English; it is all the more difficult for a machine. At that time, the accuracy of machine translation was inadequate, and it would have been premature to widely introduce it; professional translators often corrected translation errors (post-editing). This post-editing was inefficient and traumatic for translators and translation companies. Through the efforts of researchers, translation accuracy continually improved, but this failed to eradicate negative opinions about machine translation. The improvement rate was sluggish, and it was presumed that translation accuracy was approaching its upper limit.

2. Future machine translation

Neural machine translation, which appeared in 2016, easily surpassed the accuracy of earlier machine translations, which had been considered to be at its upper limit. This is a splendid achievement that has transformed the field of machine translation. Here we can take 1) the composition of English business letters, which is a foundation of trade and 2) patent translation demands high precision, as examples. Neural machine translation is steadily improving the way those involved in these tasks work by sharply boosting the efficiency, resulting in less overtime work and the improvement of the percentage of utilized paid leave. Neural machine translation is now used for many purposes including multilingual thesis/patent search services, the handling of multilingual technical information distribution, electronic transactions, and various kinds of reservations, voice translation applications, specialized terminals, and so on.

Based on two powerful forces, namely the rapid improvement of algorithms brought by worldwide competition among researchers and the steady accumulation of bilingual data makes translation accuracy continue to improve and machine translation will come into use in a wide range of activities.

3. Breaking out of the dormancy of machine translation and the AAMT

Just as "cherry blossoms" break out of their dormancy, machine translation has passed through a long harsh winter as a flower bud; now that spring is here, it is surely about to bloom.

This association, which was established in order to introduce and form links between "people involved in the research and development of machine translation and people using machine translation," is the organization best suited to accelerate the spread of machine translation. I look forward to working closely with the directors that were elected at our general meeting in June 2018 to break down the language barrier in a single stroke.

I look forward to your support and encouragement as we pursue this objective.

招待講演

アジア言語を中心とした機械翻訳評価ワークショップを通じて

中澤　敏明
東京大学

1. はじめに

本招待講演では、Workshop on Asian Translation（WAT）で毎年行われている、アジア言語を中心とした機械翻訳の性能を競う shared task で得られた知見などについて紹介した。WAT は 2014 年に第 1 回が東京で開催され、第 2 回京都、第 3 回大阪（国際会議 Coling2016 のワークショップとして開催）、第 4 回台湾（国際会議 IJCNLP2017 のワークショップとして開催）と毎年開催されており、2018 年は PACLIC32 のワークショップとして 12 月 3 日に香港で開催される。shared task の参加者は毎年 12 チームほどで、海外からも毎年 3, 4 チームほど参加がある（中国、韓国、ドイツ、チェコ、インド）。扱う言語は日本語を中心として、英語、中国語、韓国語、インドネシア語、ヒンディー語、ミャンマー語など多岐にわたる。

機械翻訳結果の評価方法は BLEU などの自動評価はもちろんのこと、人手でも行っている。人手評価はベースラインシステムとの一対評価（Pairwise Evaluation）と、特許庁が提案している「特許文献機械翻訳の品質評価手順」のうち「内容の伝達レベルの評価」[2]（JPO Adequacy Evaluation）の 2 種類の方法で行っている。JPO Adequacy Evaluation は、各サブタスクのテストセットのうちの 200 文を対象に、2 名の評価者が以下の基準での絶対評価を行う。

表 1: JPO Adequacy Evaluation の評価基準

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

講演では日英・英日特許翻訳タスクの JPO Adequacy Evaluation の結果について、特に以下の 2 つの観点で分析を行った結果を報告した：

・成績トップのシステムの評価者間一致度（κ 係数）が小さい理由

・2 名の評価者間で評価が割れている文の検討

以下、この 2 点について説明する。

2. 成績トップのシステムの κ 係数が小さい理由

英日翻訳の評価結果上位 3 システムは順に SYSTEM A（平均 4.75）、SYSTEM B（4.63）、SYSTEM C（4.40）であったが、評価者間の一致度を示す κ 係数（Cohen's Kappa）は順に 0.32、0.42、0.43 であり、トップの SYSTEM A に対する κ 係数が最も小さい値となった（WAT の方針に習って、システム名を匿名化している）。これは日英翻訳においても同様で、トップのシステムの κ 係数が最も小さくなった。この理由を、英日翻訳の SYSTEM A と SYSTEM B を例にとり説明する。

表 2 に SYSTEM A の評価結果の詳細を、表 3 に SYSTEM B の評価結果の詳細を示す。

表 2: SYSTEM A の評価結果詳細

評価	5	4	3	2	1	計
5	154	11	2	1	0	168
4	11	3	3	0	0	17
3	4	3	4	0	0	11
2	0	0	2	1	0	3
1	0	0	0	1	0	1
計	169	17	11	3	0	200

表 3: SYSTEM B の評価結果詳細

評価	5	4	3	2	1	計
5	141	15	1	0	0	157
4	8	7	6	1	0	22
3	4	2	6	2	0	14
2	0	0	4	1	1	6
1	0	0	0	0	1	1
計	153	24	17	4	2	200

この表から純粋に二人の評価者の評価が一致しているものの割合p_oを計算すると、SYSTEM A では対角線上の数字を足した 162(=154+3+4+1+0)を 200 で割って 0.81 であり、SYSTEM B では 0.78 となるため、SYSTEM A の方が一致していることになる。一方でκ係数は二人の評価者の評価が偶然に一致する可能性も考慮している。偶然に一致する可能性p_eは、評価者 X が評価 N をつけた割合をp_{XN}とすると、以下の式で計算される（2 名の評価者を A と B とする）。

$$p_e = \sum_{N=1}^{5} p_{AN} \times p_{BN}$$

この式に則って SYSTEM A と SYSTEM B のp_eを計算すると、それぞれ 0.72 と 0.62 となり、SYSTEM A の方が偶然に一致する可能性が高いことになる。κ係数は上記p_oとp_eを使って、以下のように計算される。

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

ざっくり言うと、偶然に一致する可能性を差し引いて一致率を計算していることになり、偶然に一致する可能性が高い場合ほど、κ係数は小さく見積もられることになる。上記の場合、2 つのシステムにおいてp_oの値はさほど変わらないが、p_eの値に大きな違いがあるため、SYSTEM A の方がκ係数が小さくなっているのである。

SYSTEM A のp_eの値が大きくなる要因の一つは、SYSTEM A の翻訳結果が良く、どちらの評価者も 5 をつける割合が高くなっているからだと考えられる。このように、非常に良いシステム（もしくは非常に悪いシステム）においては、κ係数が不当に低く見積もられる可能性があることに注意が必要である。一般的に、評価結果がなんらかの値に偏る傾向が強い場合には、κ係数の解釈には注意が必要であると考えられる。つまりκ係数が低いからと言って、評価結果がばらついていると短絡的に考えることは危険なのである。

3. 2 名の評価者間で評価が割れている文の検討

次に日英・英日の特許翻訳結果において、2 名の評価者間で評価が 2 以上離れている文について、どのような傾向があるのかを検討した。

3.1 評価者のミス

単純に評価者の評価ミスと思われるものがいくつか見受けられた。以下に例を示す（各翻訳文の後ろにある 1 組の数字が評価値である）。

入力	Accordingly, the present invention relates to a rod-like crystal of CZTS.
正解	それ故、本発明は、CZTS のロッド状結晶体に関する。
出力	したがって、本発明は、<unk>の棒状結晶に関する。(3, 5)

この例では NMT 特有の<unk>がそのまま出力されているにも関わらず、一方の評価者が評価 5 としている。

入力	However no long text and even only pictures may be included in pages for many book images.

正解	しかし、多くの本画像のページには、長いテキストは含まれず、ピクチャしか含まれないことがある。
出力	しかし、多くの書籍画像のページには、長いテキストや絵も含まれなくてもよい。(2, 5)

この例は入力の英語文の解釈が難しいが、正解と比較すると出力は明らかに誤っている。しかしながら一方の評価者が評価 5 としている。

入力	A plug 72 provided on the rear end of the ink pack 7 faces the plug opening 335 .
正解	口栓用開口部 335 には、インクパック 7 の後端に設けられた口栓 7 2 が臨んでいる。
出力	プラグ開口 335 には、インクパック 7 の後端に設けられたプラグ 72 が設けられている。(2, 5)

出力では“faces”が正しく訳出されていないため誤訳であるが、一方の評価者は評価 5 としている。

入力	第 1 熱媒体経路 232 はその内部に熱媒体としての水を通流させる。
正解	Water as the heat medium flows through the inside of the first heat medium passage 232.
出力	The first thermal medium path 232 flows through the water as a heat medium. (1, 4)

この例では入力と正解とで主語が変化しており、それゆえ入力が使役文であるのに対し、正解は能動態の文になっている。一方の出力は主語は入力と一致しているが、能動態で訳してしまっているため翻訳としては誤りとなっている。

入力	同軸芯線 52a は、導線で構成される。
正解	The coaxial core 52 a is constituted by a conducting wire.
出力	The coaxial core 52 a is constituted by a conductor. (2, 5)

conductor だけでも導線という意味があるため、出力も正しいと考えられるが、評価者の一方が評価 2 としている。

3.2 「重要情報」の定義の違いによる評価の差

特許庁の基準では、「重要情報」がどれだけ正確に伝達されているかが評価ポイントになるが、「重要情報」の定義は評価者の主観に委ねられている。以下に示す例ではこの定義の評価者間のズレが影響している可能性がある。

入力	For example, FIG. 1A shows a plan view of a videoconference room with a typical arrangement.
正解	例えば、図 1A は、典型的な構成からなるビデオ会議室の平面図である。
出力 1	例えば、図 1A は、典型的な構成を有する会議室の平面図を示す。(3, 5)
出力 2	例えば、図 1(a)は、一般的な配置の会議室の平面図を示す。(3, 5)
出力 3	例えば、図 1a は、典型的な構成を有する会議室の平面図を示す。(3, 5)

この例ではすべてのシステムにおいて「ビデオ」という部分が抜けているが、一方の評価者は評価 5 としている。これは「ビデオ」の部分がさほど重要ではないと判断し、ほかの重要情報が全て含まれているため評価 5 としている可能性もある。

入力	In this example, assume that jobs have been deleted and A′ =12 is obtained.
正解	今回の例ではジョブが削除され A'=12 となったとする。
出力 1	この例では、ジョブが削除され、A'=12 であると仮定する。(3, 5)
出力 2	この例では、ジョブが削除され、a=12 が得られると仮定する。(3, 5)

この例では情報の不足はないように見えるが、一方の評価者は「仮定する」と訳していることが誤りと判断している可能性がある。

入力	The recording medium 212 is a memory card freely removable from, for example, the camera body 200.
正解	記録媒体 212 は、例えばカメラ本体 200

	に着脱自在になされたメモリカードである。
出力	記録媒体 212 は、例えばカメラ本体 200 から着脱自在に着脱自在のメモリカードである。(3, 5)

この例では出力に重複が見られるが、評価者の一方が評価 5 としている。確かに重複があっても重要情報が正しく伝達されていると考えられなくもない。特許庁基準では過剰に出力された情報をどう扱うかの基準が設けられていないため、このような評価割れが起こる可能性がある。

入力	第 1 熱媒体経路 232 はその内部に熱媒体としての水を通流させる。
正解	Water as the heat medium flows through the inside of the first heat medium passage 232.
出力	The first thermal medium path 232 flows through the water as a heat medium. (1, 4)

この例では入力と正解とで主語が変化しており、それゆえ入力が使役文であるのに対し、正解は能動態の文になっている。一方の出力は主語は入力と一致しているが、能動態で訳してしまっているため、文全体の意味としては誤りとなっている。ここで、文全体の意味が誤っているから評価 1 とするか、個々の名詞は正しく訳出されているから評価 4 とするかで差が生まれていると考えられる。次に示す例も同様である。

入力	また、その固体撮像装置を用いた電子機器を提供することを目的とする。
正解	Also, it is desirable to provide an electronic system using the solid-state imaging device.
出力	It is also an object of providing an electronic device using the solid state imaging device. (1, 4)

この例でも述部の訳が少しズレているが、必要な部品は全て正しく訳出されている。このような差は、例えば否定の有無といった場合にも起こりえる。

3.3 訳語の不統一

専門用語などが入力文中で複数回出現した際にその訳が統一されていない場合や、より適切な訳がある場合などに、評価が割れる傾向があるようだ。

入力	For information, new items are not limited to added items but include changed items.
正解	なお、新規項目は、追加された項目に限らず、変更された項目を含む。
出力 1	情報のために、新しいアイテムは、追加項目に限定されず、変更項目を含む。(3, 5)
出力 2	情報については、新しいアイテムは、追加項目に限定されず、変更されたアイテムを含む。(3, 5)

“item”という単語の訳が「アイテム」であったり「項目」であったりと訳が揺れている。

入力	In a particular embodiment, the photo sensor is a photo-multiplier tube.
正解	具体的な一実施形態ではそのフォトセンサは光電子増倍管である。
出力 1	特定の実施形態では、光センサは光マルチプライヤー管である。(3, 5)
出力 2	特定の実施形態では、フォトセンサーは photo-multiplier チューブである。(1, 4)

“photo”が「フォト」や「光」と訳されていたり、photo-multiplier tube が様々に訳されている。決まった訳語がなさそうな場合には、一般的には専門用語全体で統一して和名もしくは洋名（カタカナ）で訳すべきであり、和名と洋名が混ざって訳されることは好ましくないと考えられる。

入力	FIG. 15 is a representation of one unit of a carboxymethyl cellulose molecule.
正解	図 15 は、カルボキシメチルセルロース分子の 1 つの単位の模式図である。
出力	carboxymethyl セルロース分子の一単位の表示である。(4, 2)

この例では英単語がそのまま日本語文に訳出されて

いる。頭字語など英単語をそのまま訳出すれば良い場合ももちろん存在するが、日本語訳が存在する場合には日本語にするべきであると考えられる。しかしながらこの例のように、英単語のまま訳出されても意味はわかると考えることもでき、基準を決めないと評価が割れる要因になる。

入力	糖タンパク質 D は、いくつかの宿主受容体のうちの 1 つを認識し、結合し得る。
正解	Glycoprotein D recognizes and can bind to one of several host receptors.
出力	Sugar proteins D may recognize and couple one of several host receptors. (3, 5)

この例では「糖タンパク質」が”Sugar proteins”と訳出されているが、これは誤訳で”Glycoprotein”が正しい訳である。この例からは、専門用語の正しい訳を知っていないと正しく評価できないということと、現在の翻訳システムが専門用語を構成的に翻訳してしまうことの弊害が見て取れる。

3.4 一文だけでは訳が確定しない

稀なケースではあるが、分脈がないと訳が確定できないものもあった。

入力	The encryption method in this case is AES as mentioned above.
正解	この際の暗号化方式は、上述のように AES である。
出力 1	この場合の暗号化方法は、上述のような AES である。(3, 5)
出力 2	この場合の暗号化方法は、上述したような AES である。(3, 5)

この例では、AES というものが 1 つしかないならば（前の文脈でそのような記述がされているならば）正解訳とする必要があるが、AES には何種類かあり、そのうちの一つが前の文脈で記載されているならば、出力が正しい訳となる。このように、文脈情報がないと正しく翻訳できないものも、割合は少ないが存在する。

4. おわりに

本招待講演では WAT について紹介し、WAT2017 の特許日英・英日翻訳タスクの JPO Adequacy Evaluation に対して、2 つの観点からの分析を行った結果を報告した。質の良い翻訳システムの評価において κ 係数が不当に低く算出されることを明らかにし、κ 係数だけを見ても評価の信頼性は測れないことを指摘した。今後は κ 係数だけでなく、評価値の詳細も明らかにする必要があると思われる。

また評価結果が 2 名の評価者間で乖離している例を分析したところ、その要因は評価者の単純なミスによるもの、評価尺度の定義の解釈の違いによるもの、訳語の不統一によるものなどがあった。評価者のミスはある程度は仕方がないが、評価尺度の定義や訳語の不統一の扱いなどは事前に基準を決めておくなどすれば、より一致度の高い評価が行えると思われる。

WAT は今後も毎年開催する予定であり、今後はさらに多くのアジア諸国を巻き込んで成長して行きたいと思っている。そのためには、AAMT との連携を強化することは非常に重要であると考えている。また機械翻訳の研究だけでなく、機械翻訳の利用者との架け橋になるよう、様々な活動を行っていく。

招待講演 & AAMT長尾賞受賞講演

実用化が始まった音声翻訳サービス

安西健
凸版印刷株式会社

1. はじめに

2015年、凸版印刷は、総務省が所管の国立研究開発法人情報通信研究機構（以下NICT）が中心となって発足した「グローバルコミュニケーション計画」に参画し、多言語音声翻訳技術を活用した製品・サービスの社会実装に向けた取り組みを開始した。この年は、訪日外国人旅行者が2,000万人に迫り、観光やショッピングなどの旅行消費の増加が始まった時期で、「言葉の壁」を越えたコミュニケーションとして音声翻訳が注目され始めた年となった。

それから3年が経過し、「訪日外国人」のみならず、日本に住む「在留外国人」の増加により、企業や自治体などにおいて外国人とのコミュニケーションの必要性がさらに高まっている。多言語音声翻訳技術もニューラルネット翻訳技術の採用により翻訳精度が大きく向上、社会実装できる時期を迎えている。

2．音声翻訳システムに関しての最初の取り組み

凸版印刷と株式会社フィートは、NICTの委託研究「自治体向け音声翻訳システムに関する研究開発」（2015年度から2019年度の5年間）を受託し、自治体窓口業務に対応した国内で初めての音声翻訳システムの研究開発を進めている。

この研究では市役所等の窓口業務を分類・分析、業務の改善が見込まれる住民登録や子育ての分野で使用する専用アプリを開発し、実証実験を実施中である。分析した会話の対訳コーパスを用いて翻訳精度の改善を行うとともに、自治体窓口に適したユーザーインターフェイスの開発にも取り組んでいる。

自治体窓口の担当者からのヒアリングでは、長い会話での翻訳精度や専門用語への対応に課題があるが、下記メリットが挙がっている。

・外国人をスムーズに案内できるようになった。
・何も解決できずに帰庁されることがなくなった。
・外国人に対応する職員の不安が減った。
・外国人に対する窓口対応時間が減った。

などの意見があり、社会実装に向けた音声翻訳アプリに対する期待は大きい。

3．機械・音声翻訳を活用したサービス展開

1）音声翻訳サービス「VoiceBiz（ボイスビズ）」（2018年6月リリース）

「VoiceBiz」は、急増する訪日外国人や在留外国人などとの多言語コミュニケーションを支援する音声翻訳サービスで、スマートフォンやタブレットからダウンロードが可能（使用にあたっては契約が必要です）である。

特長

1．音声翻訳11言語、テキスト翻訳30言語に対応！
2．ID/PASS認証と台数課金機能搭載！
3．固有名詞と定型文を登録可能！
4．サーバー構築やアプリ開発不要！

また、近年の変化として見逃せないのが、教育機関で日本語指導が必要な外国人児童生徒が増えていることである。生活指導、学習補助、家庭訪問などの教育現場で保護者も含め、スムーズな意思疎通を図ることに課題をかかえており、現在、コミュニケーションツールとして学校での実証実験を開始している。

2）テキスト翻訳サービス「ジャパリンガル」

（2017年9月リリース）

施設、店舗の翻訳需要が見込まれているインバウンド翻訳市場に対し、テキスト原稿を短納期で翻訳するサービスとしてジャパリンガルのサービスを開始。機械翻訳に人手校正を加えたWEBサービスとして、利活用されている。

特長

1. 安い：1文字7円～。人手翻訳に比べ約半分に。
2. 速い：2,000文字以内なら翌営業日納品。
3. 安心：AI翻訳後に翻訳者が訳文を修正。

今後、様々な分野の翻訳結果を機械翻訳エンジンへ反映することで、ベースとなる機械翻訳精度の向上が見込まれる。

4．今後の展開

音声翻訳が社会実装されたサービスとして、2018年4月から全国の郵便局約2万局（簡易郵便局は除く）に、訪日・在留外国人向け窓口サービスの向上を目的とした多言語翻訳アプリを導入した。郵便局窓口でよく使用される専門用語や定型文を搭載しており、翻訳精度が向上している。

加えて、鉄道、観光、医療、タクシーなど分野でも、一部実用化が始まっている。2019年ラグビーワールドカップ、2020年東京オリンピック・パラリンピックを控え、業種にかかわらず外国人とのコミュニケーションに関しての課題解決が急がれる。

また一方で、労働者不足や高齢化から日本の労働力が不足し、建築、農業、介護分野などでも、外国人労働者の拡大が見込まれている。使用目的やシーンに合わせて使い易い音声翻訳サービスを提供していくことができれば、急速に一般社会に普及すると考えている。

AAMT 長尾賞受賞講演

日中・中日機械翻訳実用化の取り組み

岩城 修*、中澤 敏明*、黒橋 禎夫**
*国立研究開発法人科学技術振興機構、**京都大学大学院情報学研究科

1. はじめに

科学技術振興機構（以下、JST）は、研究開発活動の効率的実施を促し、科学技術の振興を図ることを目的に、国内外から収集した科学技術文献に抄録、索引を付与した文献情報データベース[1]を整備し、インターネット等を活用して研究者・技術者が利用し易い形で提供している。これは、国内の科学技術文献を網羅的に収集している唯一のデータベースとして、日本の科学技術振興に必要不可欠な国家財産ともいえる。

また、近年の中国発の学術論文等の増加に伴い、日本国内に中国の科学技術文献情報を流通させ、中国の科学技術力の認知度向上を図ることにより、日中の科学技術協力関係を良好にし、強いては日本の科学技術力向上に資することが望まれる。そこで、JST では中国国内で発行される科学技術文献 約10,000誌の中から厳選した約1,300誌に掲載される文献の書誌、抄録、索引を中国文献データベース[2]（JSTChina）として提供し、日本語での検索を可能としている。

一方、増え続ける科学技術文献に対し、海外で出版される外国語文献も広くカバーし、即時性を確保しつつ日本語で標題や抄録を提供するためには、これまでのような人手による翻訳では対応困難な状況であり、科学技術文献への高精度な機械翻訳の導入が必要不可欠となっている。

本稿では、JST が京都大学とともに、2013 年度から 2017 年度までの 5 年間の計画で実施した「日中・中日機械翻訳実用化プロジェクト」の取り組みや成果を紹介し、JST における機械翻訳実用化の現状について言及する。

2. 日中・中日機械翻訳実用化プロジェクト

このプロジェクトは、日本と中国が連携し、日本側はJST、京都大学大学院情報学研究科（黒橋・河原研究室）、中国側は中国科学技術信息研究所（以下、ISTIC）、中国科学院自動化研究所*、北京交通大学*、ハルビン工業大学*が参加した。（*の機関は 2014～2015 年のみ参加）

プロジェクトの目標として、

- 科学技術文献の翻訳に特化した機械翻訳エンジンの開発と、日中の言語資源データの整備
- 実用的な日中・中日機械翻訳精度の達成、および国内外の機械翻訳システムの研究開発動向（例えば Google 翻訳）に対する先進性の確保
- 機械翻訳エンジンの他言語対への適用性の確保

を掲げた。

プロジェクト開始時は、それまで京都大学で開発された用例ベース機械翻訳エンジン（Example-Based Machine Translation：以下、Kyoto-EBMT[3]）の改良に取り組み、当初設定した 4 年目の目標である翻訳精度（後述）を 3 年で達成することが出来た。しかし、国内で最高峰の精度達成には至っておらず、また人手による翻訳と比較すると精度向上の余地が大きい状況であった。

一方、2014 年に提案されたニューラル機械翻訳（Neural Machine Translation：NMT）によってそれまでの機械翻訳の精度を大きく上回る可能性が示されたことから、本プロジェクトでもニューラル機械翻訳エンジン（以下、JST・京大 NMT[4]）の独自開発に速やかに着手した。

併せて、日英、中英の専門用語辞書を英語の用語をピボットとして統合することにより、400 万語規模の日英中対訳用語辞書を構築した。

ISTIC との連携については、日本側、中国側でそれぞれ機械翻訳の訓練データとなる対訳コーパスの整備を進め、最終的に全体で約 500 万件以上の中国語・日本語の対訳コーパスを構築した。

3. ニューラル機械翻訳エンジンの精度評価

開発した JST・京大 NMT の翻訳精度を他研究開発事例とまず比較評価するため、機械翻訳の国際評価ワークショップ WAT 2016[5]（The 3rd Workshop on Asian Translation）に投稿した結果、科学技術翻訳タスク（ASPEC[6]-JC データ）において日中および中日の翻訳でともに 1 位を獲得することができた。

次に、プロジェクトの目標設定で定めた翻訳精度評価基準に基づき評価を行った。具体的には、図 1 に示した基準で、重要情報が正しく翻訳されている割合、ならびに文章としての理解のし易さなどを人手により 1〜7 レベルで評価した。

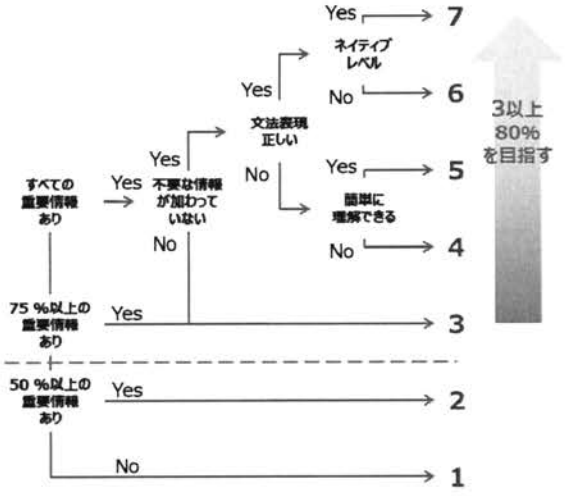


図 1　設定した精度評価基準と目標値

その結果、図 2 に示すように、2016 年 3 月時点で Kyoto-EBMT によりプロジェクトの目標である「7 段階評価で評価 3 以上（75%以上の重要情報が正しく訳せている）が 80%」を達成していたが、同年 10 月時点で JST・京大 NMT でも同目標を達成するとともに、評価 4 以上の割合が大幅に増加した。さらに、その後の改善により、プロジェクト終了時（2018 年 3 月）には評価 3 以上が 97%、評価 5 以上（情報に過不足がなく容易に理解可能）がほぼ 60%となる翻訳精度を達成した。

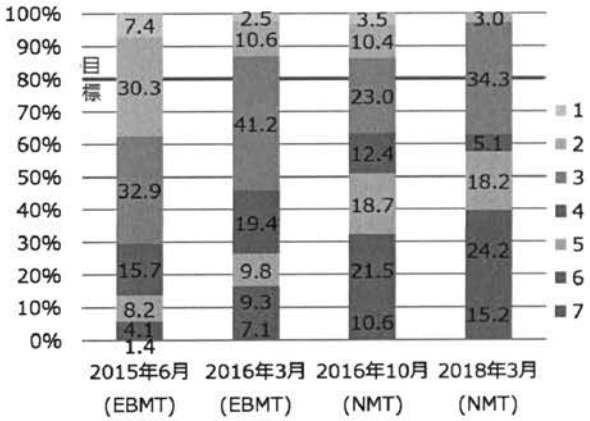


図 2　精度評価の結果

また、図 3 は中国語の科学技術文献（生物、化学、電気工学など）からランダムに選定した 100 文を 4 つのエンジンで翻訳し、各文について、3 名の評価者が最も良い翻訳結果を選択した結果の割合を示したものである。

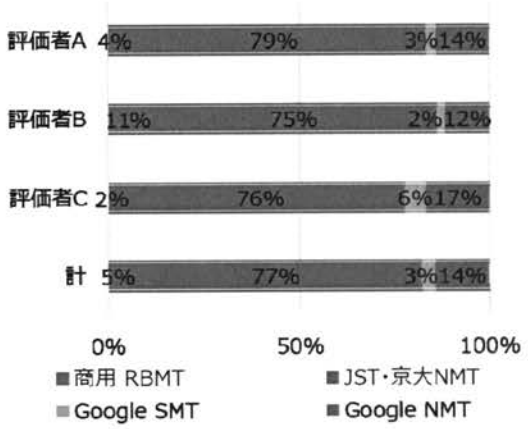


図 3　相対評価

この結果から、どの評価者も 100 文中約 4 分の 3 は

JST・京大 NMT の結果が最も良いと判断したことが分かる。ニューラル機械翻訳は、従来のルールベース機械翻訳や統計的機械翻訳より自然で滑らかな翻訳文を出力することができ、科学技術文献に特化することで専門用語等の翻訳精度も向上した結果と考えられる。

4. 実用化の状況と今後の取り組み

プロジェクトの目的である実用化の取り組みとして、

i. JSTChina のデータ作成のための中日および英日機械翻訳への適用

ii. WEB による日中・中日機械翻訳の公開

を推進した。

実用化に際しては、ニューラル機械翻訳の課題として一般的に知られている内容も含め、以下の対応を行った。

① GPU による処理の高速化

ニューラル機械翻訳では多くの行列演算を伴うことから、訓練は数週間に及ぶことがある。そこで GPU（Graphics Processing Unit）を用いて演算の高速化を図るとともに、翻訳時も複数の GPU を用いて大量データに対する処理時間の短縮を図った。

② 専門用語の訳抜け・誤訳の改善

ニューラル機械翻訳では、専門用語など低頻度語の訳抜けや誤訳が生じることがある。このため、化学物質名などを含む専門用語の対訳辞書を作成し、翻訳時にはこれらの用語を特別なタグに置換え、翻訳後に置換処理を行うことで、訳抜け・誤訳の改善を図った。

③ 翻訳文に生じる繰り返しの改善

ニューラル機械翻訳では、入力文を過不足なく翻訳することができず、翻訳結果に単語・句・節の繰り返しが出現することがある。そこで、翻訳結果に表れた繰り返しパターンの一部を自動削除することにより、不自然な翻訳結果の改善を図った。

1．で紹介した JSTChina は 2007 年より公開しており、2017 年度末までの収録記事件数は 1981 年以降に出版された約 230 万件となっている。近年の中国発の学術論文の増加に伴い、2017 年度は約 35 万件収録し、2018 年度は約 50 万件の収録を予定している。

収録している中国文献の記事のうち、約 9 割は中国語（一部は英語併記）から日本語への翻訳を行うが、残りの約 1 割は英語のみの文献であり、英語から日本語への翻訳を行う。このため、プロジェクトの目標でもあった他言語対への適用例として、JST・京大 NMT の英日機械翻訳への適用を図った。

英日機械翻訳の訓練データには、JST が保有する科学技術文献標題用対訳コーパス約 2,300 万文対、抄録用約 1,400 万文対を用い、訓練を実施した。その結果、機械翻訳の自動評価値である BLEU 値、RIBES 値において、JST がそれまで用いていた統計的機械翻訳エンジン、ならびに評価当時でニューラル機械翻訳に移行したと言われる Google 翻訳を上回る精度を達成できることが分かった。

図 4 に、JDreamⅢ[1) 2)] での JSTChina の検索表示画面を示す。

図 4 JDreamⅢでの JSTChina の検索表示画面

中国語の標題ならびに抄録については中日機械翻訳した結果を表示し、原文に英文標題と英文抄録がある場合は併せて参考表示している。また、原文が英語標題と英文抄録のみの場合は英日機械翻訳した結果を表示している。

次に、開発した日中・中日機械翻訳システム JST・京大 NMT をインターネットを活用して広く利用できるよう 2018 年 5 月より無料公開した。図 5 に日中・中日機械翻訳の WEB 画面[7] を示す。科学技術文献と特許文献を対象とした中国語から日本語、日本語から中国語への翻訳がそれぞれ可能である。

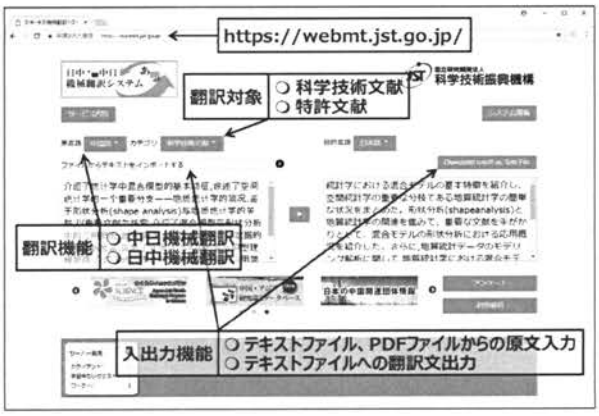


図 5　日中・中日機械翻訳システムの WEB 画面

以下、今後の取り組みについて言及する。

①　新語を含む対訳コーパスの整備

新語や新しい記述・表現などを含む対訳コーパスの拡充は、新規分野の論文等の翻訳精度向上に有効である。このため、人手翻訳、または機械翻訳の結果を修正した記事を毎年一定規模収集し、継続的な訓練により翻訳精度の維持向上を図る。

②　対訳辞書の整備と活用

ニューラル機械翻訳は扱える語彙サイズが限られており、低頻度語の訳抜けや誤訳の大きな要因となっている。これらは対訳コーパスの拡充だけでは対応し切れず、専門用語等は対訳辞書を使う方が安定した訳が得られるため、その整備と活用を継続的に進める。

③　対訳コーパス（訓練データ）の品質向上

JST では過去に人手翻訳した科学技術文献の標題・抄録データを大量に保有しており、これらから対訳コーパスを作成し、訓練に用いた。一方、訓練データに適さない対訳コーパス(直訳になっていない抄録など)も含まれることが分かっており、それらの選別や改善による翻訳精度向上も可能と考えている。

5. おわりに

本稿では、本年 3 月で終了した日中・中日機械翻訳実用化プロジェクトの成果であるニューラル機械翻訳エンジン（JST・京大 NMT）とその実用化の状況について紹介した。本プロジェクトにおいて開発したニューラル機械翻訳エンジンは、実用化の観点で他に先駆けた取り組みと自負しているが、未だ多くの研究開発が行われており、本稿で触れた課題についてもさらにその解決が図られていくものと考えられる。

一方、エンジンの性能向上に加え、訓練用の対訳コーパスの整備は、特に科学技術の分野では新たな専門用語の学習が不可欠であり、継続的な取り組みを推進する計画である。

最後に、本プロジェクトがアジア太平洋機械翻訳協会の名誉ある長尾賞を受賞したことに対し、関係の皆様に感謝します。

【参考】

1)　文献情報データベース
http://jglobal.jst.go.jp/（J-GLOBAL）
http://jdream3.com/（JDreamⅢ）
2)　中国文献データベース（JSTChina）
http://jdream3.com/（JDreamⅢ）
3)　用例ベース機械翻訳エンジン（KyotoEBMT）
http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?KyotoEBMT/
4)　JST・京大ニューラル機械翻訳エンジン（KyotoNMT）
https://github.com/fabiencro/knmt/
5)　WAT 2016
http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2016/
6)　ASPEC（Asian Scientific Paper Excerpt Corpus）
http://orchid.kuee.kyoto-u.ac.jp/ASPEC/
7)　日中・中日機械翻訳システム
https://webmt.jst.go.jp/（無料公開中）

AAMT 長尾賞学生奨励賞受賞講演

ニューラル機械翻訳における二値符号モデル

小田　悠介
グーグル合同会社

1. はじめに

ニューラルネットワークを用いた一貫型の機械翻訳法であるエンコーダ・デコーダ [1, 2, 3] が登場したことにより，これまで入出力間の語の並びと対応関係を陽に扱うことで実現されていた機械翻訳は，連続空間上のベクトル間の演算のみを用いた推論に基づく手法に置き換えられることとなった．翻訳モデルの学習データが十分に用意できる場合，エンコーダ・デコーダ，及びこれを発展させた種々のニューラル翻訳モデルは，従来手法と比較して翻訳性能の面で有効であることが知られている．一方，ニューラル翻訳モデルの動作には通常 GPU 等の高い計算能力を持つハードウェアが必要となる．これは，機械翻訳を実現するために必要なニューラルネットワークとして比較的巨大な構造を持つものを要求されることに起因する．機械翻訳システムを製品として運用する立場からこの事実を眺めると，高性能なニューラル翻訳モデルを問題なく動作させられるだけの十分なハードウェアを用意するか，性能の低下と引き換えに，比較的小規模なハードウェアでも運用可能な軽量なモデルを採用するかの選択を迫られるということに他ならない．ここで，翻訳性能を維持したまま，実際の計算量がより小さく抑えられたニューラル翻訳モデルを採用できるのであれば，これらのハードウェアに対する要求を直接緩和できることとなる．実際の製品システムでもニューラル翻訳モデルが採用され始めた昨今，このようなモデル自体の計算量の削減を目指す研究は今後より重要になるものと考えられる．

図 1 はニューラル翻訳モデルで一般的に用いられる注意機構付きエンコーダ・デコーダの概略図である．

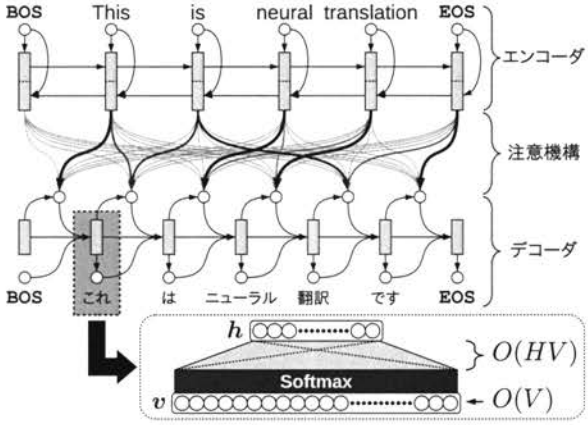


図 1:　注意機構付きエンコーダ・デコーダと出力層の計算量（[8]より）．

図中の横方向の矢印はリカレント・ニューラルネットワークであり，単語間の時間的な前後関係を表現する．縦方向，および曲線による矢印はリカレント結合以外の情報の流れを表す．これら各々の矢印は，実際のネットワークでは全結合層か，その変形として実装されることとなる．単一の全結合層が入出力ベクトル間のアフィン変換と何らかの非線形関数の組合せにより実装されるものとすると，全体の計算量を支配するのはアフィン変換であり，その大きさは $O(H_i H_o)$ となる．ここで H_i, H_o はそれぞれ入出力のノード数である．

ニューラル翻訳モデル中には様々なノード間の結合が存在するが，その中で特に計算量の大きいものは，目的言語側の単語を実際に生成する出力層である．ニューラル翻訳モデルが単語を生成するには，中間層の連続値ベクトルを何らかの形で離散符号に対応付けなければならない．このような単語生成過程で一般的に使用されるソフトマックスモデルでは，事前に決めた語彙（通常数万語程度が選択される）に対して，語彙に含まれる単語のベクトル表現と，中間層から渡され

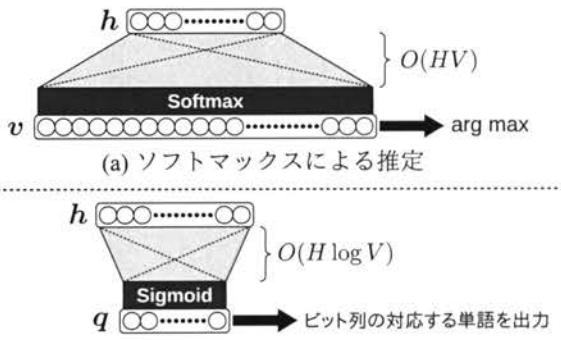
る入力ベクトルとの内積を求める．このようにして計算された内積のうち，最も大きな値を持つ単語を出力として選択するのである．ソフトマックスモデルの計算も全結合層と本質的に同じであるため，その計算量は入力ベクトルのノード数を H ，出力語彙サイズを V とすると $O(HV)$ となる [1]．V が数万程度となり，出力層が目的言語側の単語数分だけ繰り返されることを考えると，この計算量はネットワーク全体でも特に巨大であり，翻訳器全体の計算コストに強く寄与するものと考えられる．実際，ソフトマックスモデルの計算手順を工夫したり，適当な仮定に基づいて計算を省略したりすることで軽量化を図る手法は複数提案されている [4, 5, 6, 7]．

ソフトマックスモデルの計算量が巨大化するのは，個々の単語を直接出力として扱うことに本質的な原因がある．では，単語を直接考慮するのではなく，何か別の構造をネットワークから出力し，ここから間接的に単語を推定することで，ソフトマックスモデルよりも少ない計算量で単語を生成することはできないだろうか．このような考えに基づくひとつのモデルとして，二値符号予測に基づくニューラル翻訳モデルを提案した [8]．次章以降ではこの二値符号モデルに関する解説，及びその効果の説明を行う．なお，本稿ではモデルの概略の説明のみに留め，より具体的な議論は当該論文に譲るものとする．

2. 二値符号モデルによる単語予測

図 2 に示すのは，従来のソフトマックスモデルと提案手法である二値符号モデルの概略図である．二値符号モデルでは単語を直接推定するのではなく，特定の単語に曖昧性なく対応付けられたビット列を代わりに推定することにより，出力層の計算量の削減を行う．

[1] ソフトマックスモデルとしたのは，単語の出現確率を求めるのにソフトマックス関数を用いるためである．この関数は実数の大小関係を保存するため，出力確率が必要なければ実際には省略可能である． なお，ソフトマックス関数を実行する場合の計算量も $O(HV)$ である．



(a) ソフトマックスによる推定

(b) 二値符号推定

図 2: ソフトマックスモデルと二値符号モデル ([8]より)

表 1: ASPEC コーパスから得られる二値符号の例

単語	出現順位	符号語
(未知語)		00000000 00000000
(文頭記号)		00000000 00000001
(文末記号)		00000000 00000010
の	1	00000000 00000011
,	2	00000000 00000100
に	3	00000000 00000101
…	…	…
研究	81	00000000 01010000
得	82	00000000 01010001
解析	83	00000000 01010010
…	…	…

単語にビット列を対応付ける点に関しては，文字コードのようなものを想像すると容易に理解できると考えられる．文字コードに様々な符号化手法が存在するのと同様に，単語にビット列を対応付ける手法も一通りではなく，どのような手法が効果的であるかは議論の余地がある．事前実験でいくつかの符号化手法を試した結果，単語の出現頻度の順位に基づいて二進数を割り当てる手法が比較的良い性能を示すことが分かった．表 1 に実際に ASPEC コーパス [9] から得られる二値符号の例を示す．

二値符号モデルでは，ニューラル翻訳モデルの出力層において，ソフトマックス層の代わりにロジスティ

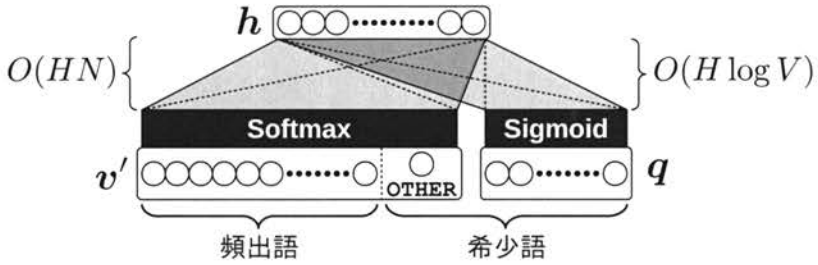


図 3: ソフトマックス・二値符号混合モデル（[8]より）

ック・シグモイド関数を使用し，ノードごとに独立で 0 から 1 の値を生成する．この出力値は符号語の対応する位置のビットが 1 となる「確実さ」[2] を表す数値として解釈され，モデルの学習過程を通して実際に出力すべきビットの値を近似するよう調整される．

このように定義される二値符号モデルでは，出力層では符号化手法が必要とする最大のビット数分だけ数値を生成すれば，単語を推定するのに必要十分であることが分かる．最も効率的な符号化手法，つまり表現可能なビット列に無駄なく単語を割当てる場合を考えると，語彙サイズ V に対し，全ての単語を割当てるのに必要なビット数は $\lceil \log_2 V \rceil$ となり，出力層に要する計算量は $O(H \log V)$ まで削減されることとなる．一例として，ニューラル翻訳モデルで一般的に用いられる語彙サイズである $V = 65536 = 2^{16}$ を考えると，二値符号モデルにおける出力層のサイズは 16 となり，約 4000 分の 1 程度の計算量となることが分かる．

3. 二値符号モデルの頑健性の向上

このように二値符号モデルはソフトマックスモデルと比較して極めて効率的であるが，これは同時に，出力層の表現能力を大幅に制限することと等しい．実際，以降の節で示すように，ニューラル翻訳モデルの出力層を単純な二値符号モデルのみとした場合，ソフトマックスモデルと比較して翻訳精度が大幅に低下してしまう．この問題のいくつかの面に関しては，二値符号モデルの構造を工夫することで緩和することが可能である．本稿では二値符号モデルに対する 2 種類の改良法を示す．

3.1 ソフトマックス・二値符号混合モデル

自然言語のユニグラム出現頻度は Zipf の法則 [10] に従うことが知られている．例えば表 1 に示した単語「の」は日本語では頻繁に用いられるのに対し，「解析」は特定の文脈でしか用いられず，これらの出現頻度には大きな差がある．この現象を二値符号モデルから見ると，実際に出力層で学習されるビット列は，出現頻度の高い単語を表現するものに大きく偏ることとなり，語彙の大部分を占める希少語に関するビット列を学習する機会が相対的に損なわれてしまう．

ここで，頻出語と希少語の推定をネットワーク的に分離することにより，お互いの学習が干渉せず，より正しい推定を行うことができるようになると考えられる．これを実現する最も簡単な方法として，図 3 に示すようなソフトマックス層と二値符号層の両モデルを並列に結合した混合モデルを提案した．混合モデルでは，まずソフトマックス層を使って頻出語か「その他」の単語であるかを推定する．「その他」が推定された場合，二値符号層を使用して改めて単語の推定を行う．出力層で部分的にソフトマックスモデルを用いることとなるため，純粋な二値符号モデルと比較すると計算量は増加してしまうが，頻出語として選択すべき単語集合は語彙全体と比較して小さく抑えられるため，純粋なソフトマックスモデルとの比較では依然として計

[2] 出力層の生成する数値を確率値の一種として解釈することが妥当であるどうかは，二値符号モデルの学習に使用した損失関数の定義に依存する．このため，本稿ではより一般的に「確実さ」とした．

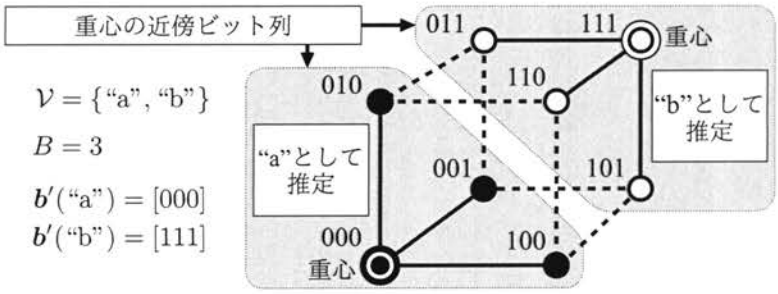


図 4: 冗長な符号表現による二値符号の頑健性向上（[8]より）

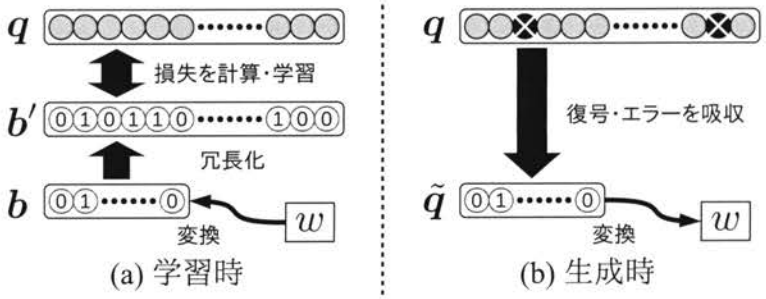


図 5: 誤り訂正符号を導入した場合の学習時・生成時の動作（[8]より）

算量が有利であることが分かる．

3.2 誤り訂正符号の適用

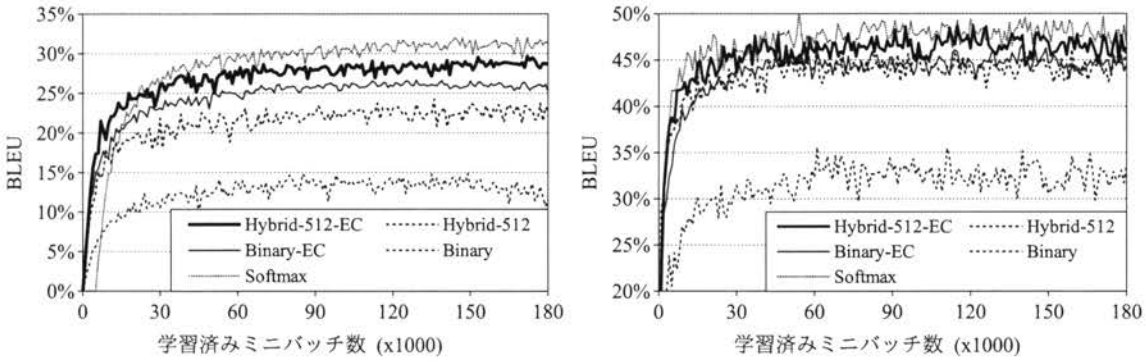
二値符号モデルの別の特徴として，ビット列に対する単語の割当てが密であればあるほど，モデルとしての頑健性が低下するという問題がある．たとえば，n ビットで表現可能な 2^n 種類のビット列全てに対して異なる単語が割当てられている場合，1 ビットの推定誤りが全く異なる単語を出力してしまう原因となる．この問題を回避するためには，割当て先のビット列に余裕を持たせる，つまり単語の割当てられたビット列同士のハミング距離がある程度大きくなるよう疎に配置するのが有効であると考えられる．この考え方を図 4 に示す．図中の「重心」としたビット列に単語が割当てられている場合，実際に得られたビット列と各重心とのハミング距離を調べることで，ある程度の推定誤り（図 4 では 1 ビット以内）を吸収可能であることが分かる．

このような考え方は一般的に誤り訂正符号と呼ばれ，様々な特徴を持つ符号化手法が提案されている．本研究では二値符号モデルとの適合性から，これらのうち畳込み符号 [11] と呼ばれる手法の一種を用いて，表 1 に示した原ビット列をより冗長なビット列へ変換した．誤り訂正符号を用いた場合のモデルの学習方法，及び単語の生成方法を図 5 に示す．学習時には原ビット列ではなく，畳込み符号により冗長化されたビット列をニューラルネットワークの出力層で学習し，生成時にはこの冗長な符号を推定する．このとき出力されるビット列に誤りが含まれていた場合でも，畳込み符号の復号アルゴリズムにより，最終的に正しい単語を推定する可能性を高めることができる．

混合モデルと誤り訂正符号は手法的に独立しており，両者を同時に適用することでより頑健性が向上すると考えられる．

4. 二値符号モデルの性能評価

図 6 に，ASPEC コーパス，および BTEC コーパス [12] により実際に二値符号モデルを学習させた際の学習曲線を示す．ここで用いた翻訳モデルは図 1 に示した単層リカレント・ニューラルネットワークの基本的なものであり，翻訳モデル単体の性能としては



(a) ASPEC　　(b) BTEC

図 6: NMT モデルの学習の推移（[8]より）

最先端のものではないが，本手法を評価するには十分有効であると考えられるものである．従来手法と二値符号モデルの各手法において，モデルの構造が異なるのは出力層のみであり，その他の構造は完全に同一である．このため，学習曲線の異なり具合は純粋に出力層の違いに起因するものと考えてよい．混合モデルに関しては，頻出語 511 種類，および「その他」記号を含めた 512 種類の推定にソフトマックス層を使用した[3]．

図 6 から，純粋な二値符号のみのモデル (Binary) はソフトマックスモデル (Softmax) と比較して大幅に性能が低下していることが観察される．これに対し，混合モデル (Hybrid-512) や誤り訂正符号 (Binary-EC) ではいずれも 10 ポイント程度 BLEU スコア [13] の改善が見られ，これらの手法が二値符号モデルの改良法として有効であることを示している．また両者を同時に適用した場合 (Hybrid-512-EC: 混合モデルの二値符号側を誤り訂正符号化) にはさらなる性能の向上が見られ，ソフトマックスモデルにほぼ匹敵する性能を達成していることが分かる．

このように，二値符号モデルを用いた場合でも，適切にモデルの改良を行うことで，ソフトマックスモデルとほぼ同様の翻訳精度を達成可能であることが分かった．では，二値符号モデルの本来の目的である計算量の削減についてはどうだろうか．理論的な側面に関しては前節までに示したので，ここでは実際の計算機上でモデルを動作させた場合の結果について述べる．図 6 で議論した 5 手法について，その出力層のパラメータ数，および GPU, CPU 上での実際に消費した平均計算時間を表 2 に示す[4]．

まず，従来のソフトマックスモデルでは，出力層だけで 1000 万個以上のパラメータが必要となることが分かる．これは前述したように，ソフトマックスモデルでは各単語それぞれに対して個別にスコアを計算する必要があるためである[5]．これに対し二値符号モデルでは，最も多くのパラメータを必要とする Hybrid-512-EC で 30 万個弱であり，ソフトマックスモデルと比較して大幅にパラメータ数を削減可能であることが分かる．ソフトマックスモデル，および二値符号モデルの各手法とも，パラメータ数は計算に必要なメモリに比例する．このため，表 2 の数値は出力層の実際のメモリ消費を直接反映していると考えてよい．

GPU 上での計算速度に関しては，二値符号モデルはソフトマックスモデルより 2 割程度高速に動作する

[3] 実験に使用したものと同様のコードを下記で公開している．
https://github.com/odashi/nmtkit

[4] 本実験実施時点で，GPU には NVIDIA GeForce GTX TITAN X を実験設定につき 1 デバイス使用した．GPU は世代ごとの性能差が大きいため，ここに示した数値を，読者が本稿を読んだ時点での GPU にそのまま適用するのは恐らく適切ではないことを注記しておく．

[5] 語彙サイズはそれぞれ ASPEC で 65536，BTEC で 25000 とした．

表 2: 各手法の出力層のパラメータ数と平均計算時間（[8]より抜粋）

コーパス	モデル	パラメータ数	平均計算時間		
			学習 (GPU) [ms/64 文]	テスト (GPU) [ms/文]	テスト (CPU) [ms/文]
ASPEC	Softmax	33.6 M	1026.	121.6	2539.
	Binary	8.21 k	711.2	73.08	122.3
	Hybrid-512	271. k	843.6	81.28	127.5
	Binary-EC	22.6 k	712.0	78.75	164.0
	Hybrid-512-EC	285. k	850.3	80.30	180.2
BTEC	Softmax	12.8 M	325.0	34.35	323.3
	Binary	7.70 k	250.7	27.98	54.62
	Hybrid-512	270. k	300.7	28.83	66.13
	Binary-EC	21.5 k	255.6	28.02	69.76
	Hybrid-512-EC	284. k	307.8	28.44	56.98

ことが分かる．これはメモリ消費の改善と比較すると劇的ではないように見えるが，この理由は GPU の高い並列度が関係している．ソフトマックスモデルは全体の計算量は膨大だが，各単語のスコア計算を完全に並列化することが可能であり，GPU のような並列度の高いデバイスを用いた計算に適している．このため，GPU 上での比較では二値符号モデルの利点が相対的に確認しづらいこととなる．

一方 CPU を用いた場合には，二値符号モデルがソフトマックスモデルの 5 倍から 10 倍程度の速度で動作しており，計算量から期待される通りの高速化を達成している．CPU は GPU と比較すると並列度が高くはなく，二値符号モデルのような計算量を直接圧縮する手法による恩恵を受けやすいためである．

誤り訂正符号を用いる場合，テスト時の計算時間には CPU 上での復号アルゴリズムの実行によるオーバーヘッドが含まれており，表 2 でも実際に Binary-EC と Hybrid-512-EC で僅かに計算時間の増加が観察される．復号アルゴリズムはニューラルネットワークのような大量の数に対する並列計算ではなく，少数の値からなる代数計算が主であり，複雑な条件分岐を高速に実行可能な CPU 上で行う方が効率的であると考えられる．このように，実行するアルゴリズムの特徴に応じて実際の計算を行うデバイスを振分けようとした場合，振分けの条件はなるべく単純であるのが望ましい．本手法に関しては，誤り訂正符号のスコア生成までをニューラルネットワーク (GPU)，それ以降の復号アルゴリズムを CPU で行うというように明確に計算を分離可能なため，計算デバイスの振分けという観点でも望ましい手法であるといえる．

5. おわりに

本稿では，ニューラルネットワークによる翻訳モデルに二値符号予測を取り入れることで，翻訳性能を維持したまま，メモリ・計算時間両面での効率化を達成可能であるという例を示した．本稿で紹介した手法の要点は，従来ニューラルネットワーク内部で完結していた単語予測を部分的に外部のアルゴリズムに追い出すことで，ネットワーク自身の計算量を大幅に削減した点にある．近年ニューラルネットワークの分野でよく研究されている一貫型のモデルではなく，本手法のような，問題全体を各アルゴリズムが得意とする部分問題として切り分け，それらを適切に繋ぎ合わせるという考え方も依然として重要であると考えられる．本手法はあくまで一例であり，まだ様々な点で研究の余地が残されている．本手法に続く，より効率的な手法が新たに生み出されることを願っている．

参考文献

[1] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to sequence learning with neural networks." In Advances in neural information processing systems, pp. 3104–3112.

[2] Bahdanau, D., Cho, K., and Bengio, Y. (2014). "Neural machine translation by jointly learning to align and translate." In Proc. ICLR.

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

[4] Morin, F. and Bengio, Y. (2005). "Hierarchical Probabilistic Neural Network Language Model." In Proc. AISTATS, Vol. 5, pp. 246–252.

[5] Serrá, J. and Karatzoglou, A. (2017). "Compact Embedding of Binary-coded Inputs and Outputs using Bloom Filters." In Proc. ICLR.

[6] Chen, W., Grangier, D., and Auli, M. (2016). "Strategies for Training Large Vocabulary Neural Language Models." In Proc. ACL, pp. 1975–1985.

[7] Grave, É., Joulin, A., Cissé, M., Grangier, D., and Jégou, H. (2017). "Efficient softmax approximation for GPUs." In Proc. ICML, pp. 1302–1310.

[8] 小田, Arthur, Neubig, 吉野, 中村 (2018). "二値符号予測と誤り訂正を用いたニューラル翻訳モデル." In 自然言語処理, Vol. 25, No. 2.

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

[10] Zipf, G. K. (1949). Human behavior and the principle of least effort. Addison-Wesley Press.

[11] Viterbi, A. (1967). "Error bounds for convolutional codes and an asymptotically optimum de- coding algorithm." In IEEE transactions on Information Theory, Vol. 13, No. 2, pp. 260–269.

[12] Takezawa, T. (1999). "Building a bilingual travel conversation database for speech translation research." In Proc. of the 2nd international workshop on East-Asian resources and evaluation conference on language resources and evaluation, pp. 17–20.

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

Research Report

Introduction to Toshiba (China)'s CJ-JC Neural Machine Translation

Pengfei Chen, Hui Di, Masanori Hattori
Research & Development Center, Toshiba China Co., Ltd.

Abstract

In Toshiba (China), we are studying about speech and natural language processing technologies and its practical applications, as one of the corporate laboratory of Toshiba group.

This report briefly introduces our recent activities, especially on neural machine translation (NMT) development. Our NMT engine is basically implemented by incorporating state-of-the-art methods, and we are mainly focusing on its practical issues, corpus collection and management.

1. Introduction

Toshiba uses rule-based method for the machine translation products for a long time. Rule-based machine translation (RBMT) is more stable in trends of error and translation quality than other statistical methods. This nature makes translation systems useful for aiding users to understanding or translating foreign language texts roughly, because they can estimate what is happening in the translation process. However, pre-defined translation rules are hard to be maintained and used for translation on new domains.

On the other hand, data-driven method, especially NMT changes the paradigm by its strong ability of distributed semantic representations and exploitation of context, which brings evolutional improvement in translation quality. Nowadays, due to the rapid advances of NMT techniques, given large scale of high quality parallel corpus, NMT would generate human-level translation, which greatly reduces the translation cost.

Toshiba (China) research & development center is now studying about NMT technology towards applying NMT in real applications.

Our research interests include tracking of frontier techniques, detection and correction of NMT characteristic translation failures, construction and refinement of bilingual corpus.

2. Recent NMT R&D trend

End-to-end NMT is based on encoder-decoder framework, which sequentially encodes source sentence to fixed-length vector from which generates a translation.

In this chapter, we briefly summarize recent NMT R&D topics.

Sutskever et al. presented a general end-to-end approach to sequence learning [2], which used multilayered Long Short-Term Memory (LSTM) [15] to map the input sequence and decode the target sequence, demonstrated that LSTM can learn sensible phrase and sentence representation. But using a fixed-length vector was a bottleneck in improving performance of encode-decoder framework, especially for long sentence translation.

Hence, Gehring et al. enhanced NMT by introducing attention mechanism to automatically search for parts of a source sentence that are relevant to predicting a target word[3]. Lin et al. improved alignment from standard attention and alleviated under-translation and over-translation problem with coverage mechanism [16]. The coverage vector was used to keep track of the attention history and fed to the attention model to help adjust future attention, which lets NMT system to consider more about untranslated source words. Sennrich et al. proposed refining the raw result with deliberation networks, which imitate human's behavior like reading news and writing articles [18]. A

deliberation network contains a first-pass decoder to generate a raw sequence and a second-pass decoder to polish and refine the first round result by looking into future words in it.

To reduce sequential computation of RNN, Gehring et al. used entire convolution network to learn sequence to sequence model (ConvS2S) to accelerate training [4]. ConvS2S creates hierarchical representations over input sequence and uses attention in every decoder layer. It is equipped with gated liner units and residual connections.

The most recent innovative NMT model is Transformer [5], which uses only attention based on multi-head mechanisms. It advantages in parallelizable computation capability compared with RNN and in learning dependencies between distant positions compared with ConvS2S. Transformer also adopts Encoder-decoder framework, with layers composed by multi-head self-attention layer and feed forward layer. Both encoder and decoder have several stacked identical layers. Encoder contains two sub-layers: multi-head self-attention mechanism and position-wise fully connected feed-forward network. Decoder includes three sub-layers with inserting a sub-layer to perform multi-head attention over the output of the encoder. Self-attention layer in decoder masks the right positions after current position to guarantee prediction for current position depending only on the already known words. There are residual connections around eachsub-layer followed by layer normalization in both encoder and decoder.

Here are some studies to improve the performance of Transformer model. Dehghani et al. replaced absolute position representation with relative position representation to consider distance between input elements by edges [19]. Graves et al. used multiple self-attention branches with different weight, achieving better performance and faster converging [20]. Papineni et al. improved transformer model with universal computation and employed an adaptive computation time (ACT) mechanism [19]. Universal Transformer is bound by the number of revisions of each symbol's representation while ACT dynamically adjusts the number of computation steps before halting. AAN (Average attention networks) Transformer is a promotion of original Transformer, which addresses decode-time inefficiency without loss of quality [16]. The framework of AAN Transformer is illustrated in Figure 1 [6]. The average layer depends on previous positions and gating layer. During decoding, AAN uses a cumulative uniform averaging operation across previous position, while the original Transformer looks

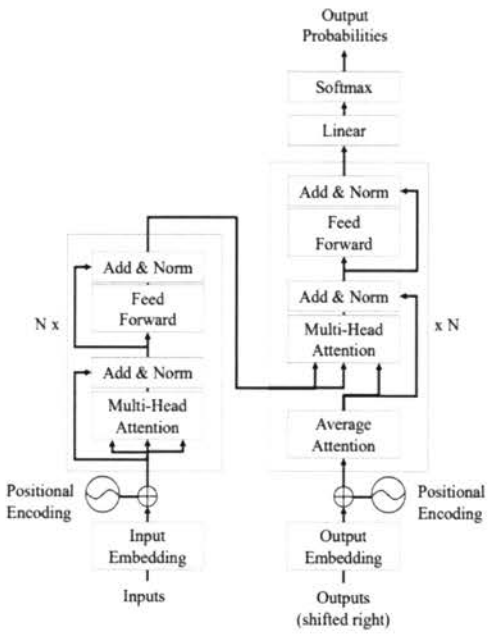


Figure 1. AAN Transformer framework

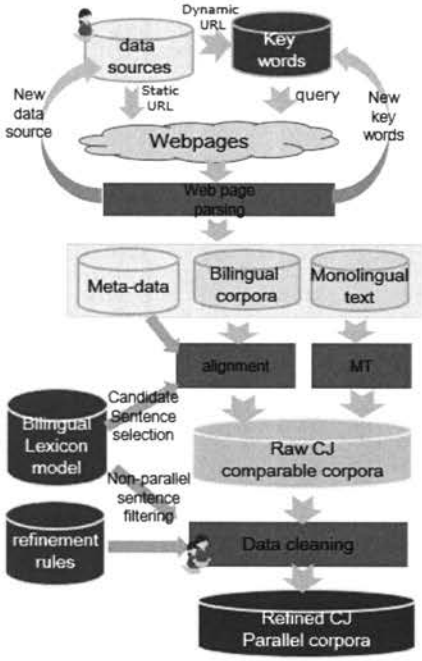


Figure 2. Data collection and Cleaning

back at entire history.

In China, there are many researchers and companies did a lot of works on NMT aiming at improving NMT performance. Because of remarkable performance of Transformer, many researches were conducted based on Transformer. Several companies has deployed online service based on Transformer.

3. Our goal and approach

3.1 Target specification

We aim at building a state-of-art CJ-JC NMT engine based on frontier techniques and large scale of high quality corpus, which could be applicable for CJ-JC translation in various domains.

3.2 Target specification

Transformer model dominated NMT since proposed because of its high performance, our NMT system is also based on Transformer. For fast decoding, our system adopts AAN, decoding time is linear to the output length.

Additionally, our online decoder also process requests in batch mode, which re-packages multiple sentences in different requests to one batch to conduct translation in parallel.

3.3 Data collection and cleaning

For NMT, it's crucial to have enough high-quality bilingual corpora. We accumulated large amount of Chinese-Japanese parallel sentences by web page crawling, semi-automatic data alignment and cleaning. The process is illustrated in Figure 2.

1) Data crawling

   Due to scarcity of parallel sentence in web pages, we crawled not only bilingual comparable sentence pairs, but also monolingual Chinese or Japanese texts, crawled monolingual sentences are translated with high performance MT engine to get sentence pairs.

   The data source for crawling we used covered many domains and genres, like news, blogs, and encyclopedia.

   The pre-prepared keyword dictionary was used to query some data searching web pages and the keywords could be expanded by extracting similar or related words that exist in crawled pages. Similarly, data source was expanded with related web pages. Domain of the seed words affects the domain of the crawled data.

2) Text alignment.

   Regarding non-parallel data, we must align source sentences to target sentences that exist in the same or other web pages. Now, we adopt two approaches to implement the alignment.

   A) Rule-based alignment.

      We exploited the meta-data of the texts for document alignment, paragraph alignment and sentence alignment. The meta-data we used include document meta-data, like language type in the URL, date of the news, sentence numbers of the paragraph, position of paragraph in one document or position of sentence in one paragraph. In addition, considering the specific language feature that Japanese and Chinese texts contain a lot of same Chinese characters, two paragraphs are aligned if enough same Chinese characters exist in them.

   B) Alignment model.

      The alignment score of the candidate sentences was computed by lexicon model that trained by Moses [11] with our former accumulated high quality CJ corpora:

$$\mathrm{lex}(e \mid f, \alpha) = \prod_{i=1}^{\mathrm{length}(e)} \frac{1}{\left|\{j \mid (i,j) \in \alpha\}\right|} \sum_{\forall (i,j) \in \alpha} \omega(e_i \mid f_j)$$

      In which, e is the source sentence and f is the target sentence, e_i denotes the i^{th} word of source sentence and f_j denotes the j^{th} word of target sentence, $w(e_i \mid f_j)$ is their translation probability. Finally, we got 15 million CJ parallel sentences.

3) Filter and refine low quality sentence pairs

   According to our experiment, the quality of

corpus plays important role in outputting high translation performance for NMT. We cleaned the noises for collected corpora with following methods.

A) Filter low quality sentence.

We use the same bilingual lexical model as mentioned before to estimating quality of each sentence pairs, removing the ones that rank last 10%.

B) Noise data filtering rules

- Sentence length comparison
- Mismatch of numerical format, punctuation and English characters.
- Extra space or special characters
- Translation error of specific words

C) Human checking.

Some complex or special noises that could not be matched by rules were revised by human.

After filtering out low quality sentences, the corpus size decreased to 12M sentence pairs.

4) Format and segment sentences.

Before training and decoding, the corpus is pre-processed as follows.

- Process English tokens.
- Convert full characters to half characters.
- Convert upper case characters to lower case characters.
- Convert traditional Chinese characters to simplified Chinese characters.
- Remove redundant spaces and characters.
- Segment the sentences.
- Re-segment to sub-words with BPE, preventing UNK in translation.

5) Post-processing.

Post-process module removes extra space, converts punctuation and repacks the translated sentences based on original sequence.

4. Fundamental experiment

4.1 Experimental settings

We carried out our experiment on Tesla P40 with 20G memory, using CUDA V9.0.0. To accelerate the training speed we installed MKL, SparseHash and tcmalloc.

We trained the model using different hyper parameters on 12 million parallel sentences and limited the vocabulary size to 32k. Out of vocabulary words were mapped to a unique token "UNK". We set the dimensionality d of all input and output layers to 512, and dimensionality of inner feed forward network layer to 2048. We employed 8 parallel attention heads in both encoder and decoder layers. During the training, we used label smoothing with value, attention dropout and residual dropout with a rate of p=0.1.During the decoding, we employed Beam Search algorithm and set the beam size to 4. As different sequence has different lengths, which may result in wasting of computation in the batch, to address this problem, sentences with roughly same length are batched, and the *max-length* of sentence size is set to 100. Adam optimizer with β_1=0.9, β_2=0.98 and ε=10^{-9} was used to tune model parameters, and learning rate was varied under a warm-up strategy with *warmup_steps*=4000.

Our testing data was developed by randomly selected from Web, including 511 sentences in various domains, such as IT, travel, and politics. The reference translation include 5 sentences.

4.2 Performance

Automatic evaluation and human evaluation were conducted in our system. Automatic evaluation metrics include BLEU [7], NIST [8], Rouge [10], TER [9] and Bleu-Ter. We compared our result with two competitors *Online B* and *Online A*. The translation results of *Online B* and *Online A* were both got from their open online translation web site on 30th, July, 2018.

The automatic evaluation result is as Table 1. In all of evaluation metrics, our system outperforms *Online B* and *Online A*. The human evaluation result is consistent with automatic evaluation as Figure 3.

The evaluation result on specific domains is shown as Table 3 and Table 4. Only result of general domain of *Online B* in automatic evaluation is obvious better than

System		BLUE5	TER	BLEU5 -TER	NIST	ROUGE		
						F	R	P
Ours	CJ	**69.72**	**43.72**	**26**	**12.21**	**0.93**	**0.90**	**0.96**
Online A		51.38	60.26	-8.88	10.00	0.89	0.89	0.89
Online B		68.65	46.71	21.94	12.04	0.92	0.89	0.95
Ours	JC	**65.23**	**37.34**	**27.89**	**12.11**	**0.91**	**0.90**	**0.92**
Online A		39.00	55.40	-16.40	9.34	0.87	0.87	0.87
Online B		59.38	40.82	18.56	11.57	0.90	0.90	0.90

Table 1 Translation results compared with other systems

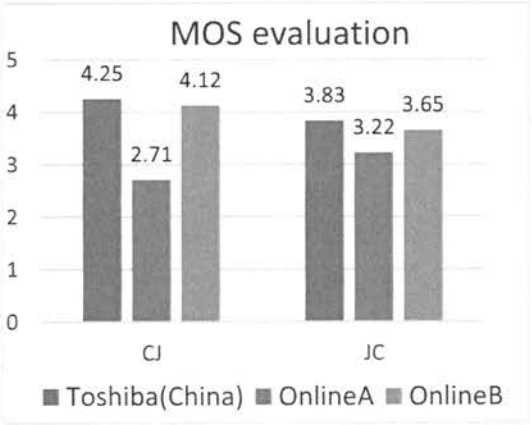


Figure 3 Human evaluation result

	Transformer +BPE	Transformer +BPE+AAN
CJ	43.53s	22.50s
JC	21.09s	18.50s

Table 2 Translation time for 511 sentences

Direction	System	Travel	General	IT	Economics	Politics
CJ	Ours	66.02	70.49	**69.84**	**72.51**	**66.35**
	Online A	44.6	57.75	49.83	54.48	42.52
	Online B	**66.26**	**78.08**	63.33	67.85	65.5
JC	Ours	**68.41**	**56.43**	**72.97**	**68.23**	**58.01**
	Online A	43.43	31.14	45.54	41.08	33.65
	Online B	62.51	54.17	66.42	60.32	51.58

Table 3 comparison of automatic evaluation on different domains

Direction	System	Travel	General	IT	Economics	Politics
CJ	Ours	**3.81**	**3.86**	**4.09**	**4.11**	**4.16**
	Online A	2.95	3.08	3.74	3.83	3.72
	Online B	3.7	3.7	3.97	4.04	3.93
JC	Ours	**4.17**	**3.97**	**4.07**	**4.13**	**3.74**
	Online A	3.43	3.15	3.73	3.57	2.98
	Online B	3.97	3.88	3.87	3.93	3.45

Table 4 comparison of human evaluation on different domains

us. But in human evaluation, our system is better in all domains.

We have the following conclusions:

(1) On Our developed test-sets, our system outperforms other competitors on multiple metrics in both JC and CJ direction.

(2) Our CJ direction is better than JC direction.

(3) The main reasons that our system perform well : we have huge and relatively good quality corpora and good pre-processing and post-processing methods.

(4) AAN accelerates decoding without loss of translation quality.

(5) AAN accelerates more in CJ direction than that in JC direction. The reason is that acceleration rate is related to length of target sentence. In theory, acceleration rate equals "(source sentence length + target sentence length) / source sentence length", the longer of the target sentence, the bigger of the acceleration rate. Generally speaking, Japanese sentence is longer than corresponding Chinese sentence, so in CJ direction acceleration rate is bigger.

5. Discussion

5.1 Practical issues on NMT

Although our NMT has gained high performance, there still exist some problem.

- Incorrect translation of proper noun. e.g., "党 的 十九 大" is translated to "党 の 19", "十九 大" is a proper noun.
- Under translation problem. e.g. "要 羊毛 洗 模式 不 要 快速 洗 模式" is translated to "羊毛 洗い モード で は ない で ください 。"
- Double negative sentence translation error. e.g. "不能 让 我 的 烦恼 没 机会 表白" is translated to "私 の 悩み を 告白 する 機会 が ない".
- Inaccurate translation of polysomic words. e.g. "抄 近路" is translated to "近道 を 写し ます".

5.2 Application scenarios

1) Document translation with human or low performance MT needs high cost, which could be remarkably reduced by exploiting our high performance NMT and few human post-processing like below.

   A) Terminology translation replacement.
   For domain-specific terminologies, to avoid wrong translation from NMT, the corresponding word could be automatically replaced with that from terminology bilingual dictionary, which maybe got from user or be extracted based on user history document.

   B) Reliability estimation.
   NMT outputs reliability score for each sentence, human only needs to modify a small fraction in low reliability sentences. For convenience of user revision, we could provide results from other NMT systems. Considering problem of source sentence that results in bad translations, NMT system may also provide quality estimation score for the source sentence.

   C) Caching correct translation.
   Besides caching correct sentence with traditional translation memory, for a few translation errors like part of word or phrase errors that exist in NMT result, user can also cache them after a few revisions, so that next time same sentence/phrase/word will be translated as expected.

Since data security is quite a big problem, we could consider providing on-premise MT service, which saves user data on their internal servers to avoid data leak risk.

In addition, if user could provide specific domain parallel data, such as accumulated patent translation data, we could fine-tune the parameter of baseline NMT model to output higher accurate translation.

2) Assist personnel to translate oral language.
Since good pre-processing of colloquial expressions and data accumulation of travel domain, using our NMT to assist personnel to remove the CJ-JC communication obstacle is also feasible.

6. Summary

We built a translation engine based on state-of-art Transformer model for CJ and JC translation, which outperforms several competitors on randomly selected test sets and gets practical decoding time by using AAN and batch decoding. Besides the advances of these open techniques, the main reason of the high performance comes from building large corpus, which were crawled from Internet and cleaned with filter algorithm and human defined rules.

There still exists some problems in our translation system, such as terminology translation error, under translating problem etc. We will focus on dealing with these issues with proper techniques. To provide user with better MT service, we will continue to improve the performance in the future.

Reference

[1] Nal Kalchbrenner and Phil Blunsom. Recurrent continuous translation models. In Processing of EMNLP, 2013.

[2] Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. InAdvances in Neural Information Processing Systems (2014), pp. 3104–3112, 2014.

[3] Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. In International Conference on Learning Representations (2015).

[4] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and YAAN N. Dauphin. Convolutional sequence to sequence learning. arXiv preprint arXiv:1705.03122v2, 2017.

[5] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In the AANual Conference on Neural Information Processing Systems (NIPS).

[6] Biao Zhang, Deyi Xiong, and Jinsong Su. Accelerating neural Transformer via an average attention network. In Proceedings of the 56th AANual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany. Association for Computational Linguistics. 2018.

[7] Papineni, K., Roukos, S., Ward, T., jing Zhu, W.: Bleu: a method for automatic valuation of machine translation. pp. 311–318. 2002.

[8] George Doddington. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In Proc. ARPA Workshop on Human Language Technology. 2002.

[9] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul, "A Study of Translation Edit Rate with Targeted Human AANotation," Proceedings of Association for Machine Translation in the Americas, 2006.

[10] Lin, Chin-Yew. 2004. ROUGE: a Package for Automatic Evaluation of Summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain, July 25 - 26, 2004.

[11] Koehn P, Hoang H, Birch A, et al. Moses: open source toolkit for statistical machine translation[C]. Meeting of the ACL on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 177-180. 2007.

[12] Rico Sennrich, Barry Haddow and Alexandra Birch : Neural Machine Translation of Rare Words with Subword Units Proceedings of the 54th AANual Meeting of the Association for Computational Linguistics (ACL 2016). Berlin, Germany. 2016.

[13] Dehghani M, Gouws S, Vinyals O, et al. Universal Transformers [J]. 2018.

[14] Alex Graves. Adaptive computation time for recurrent neural networks. arXiv preprint arXiv:1603.08983, 2016.

[15] Hochreiter, S. and Schmidhuber, J. Long short-term memory. Neural Computation, 9(8), 1735–1780. 1997.

[16] Tu Z, Lu Z, Liu Y, et al. Modeling Coverage for Neural Machine Translation [J]. 2016.

[17] Tu Z, Liu Y, Shang L, et al. Neural Machine Translation with Reconstruction [J]. 2016.

[18] Y. Xia, F. Tian, L. Wu, J. Lin, T. Qin, N. Yu, and T.-Y. Liu, "Deliberation networks: Sequence generation beyond one-pass decoding," in Advances in Neural Information Processing Systems. 2017.

[19] Shaw P, Uszkoreit J, Vaswani A. Self-Attention with Relative Position Representations [J]. 2018.

[20] Ahmed K, Keskar N S, Socher R. Weighted Transformer Network for Machine Translation [J]. 2017.

Appendix

Domain	Dir.	Source Sentence	Target Sentence
Patent	CJ	为了在嵌入式设备等中以高性能和低成本实现这种分层互连神经网络，提出了使用模拟硬件或数字硬件来实现分层互连神经网络的方法。	このような階層型相互接続ニューラルネットワークを組み込み機器等で高性能かつ低コストで実現するために、アナログハードウエアやデジタルハードウエアを用いて階層的相互接続ニューラルネットワークを実現する方法が提案されている。
	JC	遺伝子情報提供者から画像データを含む遺伝子情報を取得し、遺伝子情報利用者にとって利用しやすい状態で提供できる遺伝子情報解析システムを構築する。	从遗传基因信息提供者取得包含图像数据的遗传基因信息，构筑能够在遗传基因信息利用者容易利用的状态下提供的遗传基因信息解析系统。
Law	CJ	经营者提供商品或者服务时造成消费者或者其他受害人人身伤害的，应当赔偿医疗费、护理费、交通费等为治疗和康复支出的合理费用，以及因误工减少的收入。	経営者が商品やサービスを提供する際に消費者やその他の被害者の人身傷害を引き起こしたのは、医療費、介護費、交通費などの治療とリハビリのための合理的な費用と、ミスで減少した収入を賠償すべきである。
	JC	この法律は、共同で営利を目的とする事業を営むための組合契約であって、組合員の責任の限度を出資の価額とするものに関する制度	该法律是为了共同经营以盈利为目的的事业的工会合同，通过确立以工会成员的责任限度为出资的价值的制度，以谋求个人或法人共同进

		を確立することにより、個人又は法人が共同して行う事業の健全な発展を図り、もって我が国の経済活力の向上に資することを目的とする。	行的事业的健全发展，以提高我国经济活力为目的。
Software Design	CJ	根据合同或项目任务书等有关文件，在对用户进行多次调研的基础上，逐项说明该软件各项功能的详细需求，描述为完成各项功能所需要的输入、输出、处理及达到的目标。	契約書やプロジェクトタスクなどの書類に基づいて、ユーザーに対して何度も調査を行った上で、このソフトウエアの各機能の詳細な要件を説明し、各機能を完了するために必要な入力、出力、処理、および達成された目標について説明します。
	JC	大容量のマルチメデイアファイルのストリーミングなど、再度利用しないデータを読み取る場合、そのデータをファイルシステムのキヤッシユに追加しないようにファイルシステムに指示します。	如果读取不希望再次使用的数据，例如大型多媒体文件流，请指示文件系统不要将该数据添加到文件系统缓存中。
Travel	CJ	我急需一个房间要明晚住，请问你们还有空房吗？	明日の夜に泊まりたいのですが、空いている部屋がありますか?
	JC	時計台がランドマークの上海税関の隣にある建物です。ドーム型の屋根が特徴的で、ヨーロッパの建物を見ているかのようでした。重厚感も感じられ、存在感がある建物でした。	钟楼是地标上海海关旁边的建筑物。圆顶型的屋顶很有特点，好像在看欧洲的建筑物一样。 也能感受到厚重感，是一座有存在感的建筑物。

Table 5 Translation examples

シンポジウム参加報告

Women in Localization Japan 第 14 回イベント報告

目次 由美子

XTM International Ltd.

Women in Localization Japan 第 14 回イベント報告

日時：2018 年 06 月 08 日（金）19:00 ～ 21:30

開催場所：ネットアップ株式会社 会議室

セッション I

テーマ：SAP におけるローカリゼーションの現状と挑戦

登壇者：佐野栄司 Eiji Sano, SAP ジャパン株式会社

セッション II

テーマ：2020 年に向けてローカリゼーション業界が目指すべきもの

登壇者：尾立源幸 Motoyuki Odachi, Oyraa

セッション III

テーマ：LocWorld Tokyo #36 の報告

登壇者：上田有佳子 Yukako Ueda, Women in Localization 日本代表

Women in Localization Japan 事務局メンバー：

上田 有佳子、千葉 容子、澤村 雅以、加藤めぐみ、森 みゆき

ローカリゼーションに携わる女性のための非営利団体として 2008 年に米国加州で設立された「Women in Localization」の日本支部は 2015 年に発足し、3 ヶ月ごとに勉強会兼ネットワーキングのイベントを開催している。発足当初は約 10 名であった会員は今や 100 名を越え、クライアント企業のみでなく、翻訳会社や個人翻訳者、翻訳ツールベンダーを含む。ローカリゼーション関連の新しい技術や共通の問題をトピックとしたセッションを行っており、この 6 月のイベントでは日本におけるローカリゼーションの取り組みにスポットライトが当てられた。

セッション I

『SAP におけるローカリゼーションの現状と挑戦』

佐野 栄司氏

「SAP」といえば、グローバルに活躍している企業であり、その翻訳需要の高さからも社名を聞いたことがない業界人は少ないのではないだろうか。SAP ジャパン株式会社にて SAP ランゲージサービス ジャパンのディレクターを務める佐野氏は、「企業向けのソフトウェアを作って販売しているドイツの会社である」と自社を簡潔に説明してくれた。予測分析や機械学習、クラウドを基盤としたマーケティングへ AI（Artificial Intelligence、人工知能）を組み込むことなどにも取り組んでいるマーケット リーダである。急成長を遂げるデータベース ベンダーでもあり、同社のソリューションはクラウドのみでなく、オンプレミスやハイブリッドでも提供されている。1972 年の創設当初は約 5 名であった社員数は 2018 年には 140 ケ国 9 万人を超え、約 10 社であった顧客も 180 ケ国 40 万社近くに昇っている。

同社の翻訳・通訳の需要を満たすSAPランゲージサービス（SAP Language Services、SLS）は、SAP製品、サービスおよび社内コンテンツを標準的に約40の言語に翻訳している。SLSの拠点はドイツ本社のほか日本、カナダなどに点在するが、通訳サービスの取り扱いが最も多いのは日本であるとのこと。SLSは41ケ国の120を越えるサプライヤーと協業し、2800名以上の翻訳者により2017年には約10億ワードを翻訳した実績を有する。ソース言語は英語のみではなく独語のこともある、また、近年のストラテジーとして機械翻訳を標準プロセスとして取り入れ始めているという紹介もあった。同社独自の翻訳ツールを開発し、外部顧客へも提供しているそうだ。

SAP社の現状としては、製品のクラウド化が加速し、新技術への対応が求められる中、イノベーションも促進されている。これに伴って、労働環境の改善や人材育成も進んでいるそうだ。

SLSが直面する課題として、急激な翻訳分量の増加、予算や人材を含む翻訳リソースの限界、クラウドの開発サイクルや翻訳プロセスのオートメーションなどが挙げられた。SAP社は数十社の吸収・合併を年間で実施しており、合併された企業の製品なども翻訳対象となる。単なる翻訳分量の増加だけではなく、既存する用語集やスタイルガイドなどとの統一も当然ながら課題となる。さらにはSAP社への売り上げの貢献も求められている。

特にSLSジャパンとしては、オペレーション拠点が中国へ移転されるなど世界的な経済情勢に単純に飲み込まれることなく、日本チームの重要性や存在意義を訴え、日本チームからの価値提供を継続していくという課題も抱えているそうだ。会社の指針や上位組織のストラテジーを反映した目指すべきゴールを明確に周知し、ビジネスの変化を受け入れ、自己啓発や新しい知識の習得も意識してメンバーと共に邁進しているとのこと。

セッションII

『2020年に向けてローカリゼーション業界が目指すべきもの』

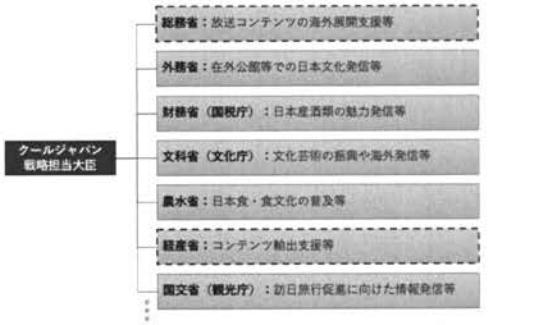
尾立 源幸氏

尾立氏は講演の冒頭で、海外で放映された日本のアニメーションの数コマを見せてくれた。そのうちの1つは、あるキャラクターが視聴者へ向けて棒付きキャンディーを差し出している。次のスライドに示された日本原作の同じコマでキャラクターが手の平に握っているのは、1本のタバコであった。未成年者に喫煙を促してはならないという配慮からこのような変更が発生したのだそうだ。

ローカリゼーションとは、「産業翻訳、ビジネスや技術に関わる文書の翻訳」と定義されているが、海外の地域性や文化、法律や規制、商品やサービスに対するニーズを理解した上で進出することが注意されている。

日本政府の内閣府が推進する「クールジャパン推進戦略」では、コンテンツのローカライズおよびプロモーション支援もその一環として含まれており、コンテンツ産業の海外市場獲得希望を2011年の0.7兆円から約10年で約2〜3兆円に拡大するという目標を掲げている。

クールジャパン推進戦略の体制



日本の魅力を海外に発信する支援政策としては、大きく4つの事業がある。そのうちの1つである「ジャパン・コンテンツ・ローカライズ&プロモーション支援助成金」では、コンテンツを海外へ展開する際に必要な「ローカライズ」（字幕や吹き替えなど）や「プロモーション」（国際見本市への出展、PRイベントの実施など）に関する費用の補助を行っている。海外への発信に対する総合的な支援を実施し、日本ブームを創出することによる関連産業の海外展開の拡大や、観光などの促進につなげることを目的としている。2年間の予算総額約155億円において、3815件が採択されたとのこと。

また、観光立国推進基本法により、観光立国の実現は国家戦略として位置づけられていることも手伝って、訪日外国人の数は2011年頃から毎年約500万人ずつ飛躍的に増加している。政治家としての経歴をもつ尾立氏は、中国を対象として初めての個人観光ビザの発給に2009年に取り組まれたそうだ。その後も、タイやマレーシアに対するビザ免除、インド、フィリピン、ベトナムに対するビザ発給要件の緩和にも取り組まれたとのこと。

ビザ発給の緩和化は訪日外国人数の増加に効果を発揮したと考えられるが、新たな問題がもたらされたことも否めない。たとえば訪日外国人に対する緊急医療の支払いが滞る、医療現場における「言葉の壁」が挙げられる。

尾立氏は現在、「Oyraa」（オイラ）という名前のオンデマンドでプロの通訳者を利用出来るアプリの開発と世界展開にあたってOyraaのローカライズに取り組まれている（https://www.oyraa.com/）。インバウンドとアウトバウンドの両面において言語の障壁をなくすという目的を掲げ、グローバル化に取り組まれている。

セッションIII

『LocWorld Tokyo #36の報告』

上田 有佳子氏

4月3日～5日に舞浜のヒルトン東京ベイにて開催された LocWorld Tokyo #36 では、『Let's Talk About Japanese Again in Japan』というパネルディスカッションが執り行われ、上田氏よりその報告があった。

日本語は世界で最も難しい言語と言われており、Women in Localizationのジャパン チャプターとして取り組んでいるトピックについてのパネル ディスカッションを行った。今後はこのトピックに関するポータルを製作し、ナレッジを蓄積して世界に対して発信していくことを目的としている。

LocWorld Tokyo #36では新しいツールが導入され、会場の参加者を対象としてリアルタイムで統計を取得し、スクリーンに結果を映し出すことができた。以下、結果をいくつか紹介する。

当該セッションの参加者のうち、52%以上が翻訳バイヤー（翻訳会社にとってのソースクライアント）であることが示され、翻訳会社が24%、ツールベンダーと個人翻訳者それぞれが4%、その他が16%であった。

日本語を母国語とする参加者が52%と示された。

さらに、翻訳バイヤーのみを対象とした質問では、67%が機械翻訳を使用しているという数値も示された。

同じ質問を翻訳会社のみを対象としたところ、60%という全体の半分以上の数値が示された。

さらに、翻訳会社を対象に英語から日本語の翻訳プロジェクトで機械翻訳を使用しているかという質問に対しても、55%という数値が示された。

100%マッチに対するレビューを実施しているかという質問に対しては、71%がしているとの回答だった。「読みやすさ」や「品質のため」という理由が挙げられる一方、「コンテキストがなければ意味がない」、「コストや時間を消費する」という意見もあった。

質疑応答においては、ローカリゼーション業界の最新状況が伺える内容が網羅されており、こちらも一部紹介したい。

Q. プロセスの自動化に伴い、プロジェクトマネージャは減少しているか？

A. 「コーディネータ」タイプの PM は減少していると言える。コミュニケーションに長けている、「アカウント マネージャ」と呼ばれるような PM は需要が高いと考えられる。

Q. 翻訳バイヤーとしては翻訳会社が機械翻訳の使用を把握できるか？

A1. 翻訳会社と確認すべきである。

A2. 翻訳会社は合意しないかぎり機械翻訳を使用すべきではない。

Q. 100%マッチのレビューに AI が導入されているか？

A1. 可能性はある。大きなデータでコンテキストを AI によって判別するようなことができれば、100%マッチの信頼性は高まるであろう。

A2. 現在そのような取り組みがなされていると聞いたことはない。

シンポジウム参加報告

Women in Localization Japan 第 15 回イベント報告

目次 由美子

XTM International Ltd.

Women in Localization Japan 第 15 回イベント報告

日時：2018 年 09 月 07 日（金）19:00 ～ 21:30

開催場所：ネットアップ株式会社 会議室

セッション I

テーマ：Expedia's Journey with Machine Translation

登壇者：Claire Pagès and Tommaso Rossi, Expedia Group

セッション II

テーマ：Dell EMC サポートコンテンツ翻訳における NMT 導入事例

登壇者：森山崇範 Takanori Moriyama, Dell EMC

Women in Localization Japan 事務局メンバー：

上田 有佳子、千葉 容子、澤村 雅以、加藤めぐみ、森 みゆき

ローカリゼーションに携わる女性のための非営利団体として 2008 年に米国加州で設立された「Women in Localization」は、本年、記念すべき 10 週年を迎え、そのささやかなお祝いも含めたイベントが開催された。日本支部は 2015 年に発足し、3 ヶ月ごとに勉強会兼ネットワーキングのイベントを開催している。発足当初は約 10 名であった会員は今や 100 名を越え、クライアント企業のみでなく、翻訳会社や個人翻訳者、翻訳ツールベンダーを含む。ローカリゼーション関連の新しい技術や共通の問題をトピックとしたセッションを行っており、この度はグローバル企業での機械翻訳への取り組みが紹介された。

Session I

"Expedia's Journey with Machine Translation"

Mrs. Claire Pagès and Mr. Tommaso Rossi

Women in Localization Japan chapter advances and welcomed speakers online from abroad for the first time. Claire and Tommaso from Expedia Group, the worldwide leader of online booking sites in travel, were willing to carry out a presentation from London, United Kingdom. They talked about their works with Machine Translation (MT).

As benefits of using MT, Claire mentioned about cost and time. The translation vendors provide discount for their post-editing works comparing to "pure" translation tasks. Also, MT generates translation at faster speed, thus it makes it possible for them to have more time to translate more materials.

Currently, Expedia utilizes 2 types of MT engines as they see different benefits. It is possible for them with Statistical MT (SMT) to specify the language variants such as Chinese for Simplified Chinese, Hong Kong and Taiwan. Meanwhile, Neural MT (NMT) engine does not offer multiple language variants but they found that it generates more fluent translations for longer sentences. It was also pointed out that NMT cannot be customized, and less accuracy can be seen for terminology.

Also, Expedia is aware of the necessity of testing MT before its implementation as experience shows that not all content types or languages are suitable for MT. Especially, they are concerned with Asia-Pacific languages. Moreover, they noted that there are currently more and more options for MT engines with varying results at a content type or language level.

By now, Expedia has built up connectors with different MT engines, a process to measure quality and efficiency gains, and partnership with their external suppliers. They also addressed their concerns on security.

During the Q&A session, they pointed out it is necessary to increase MT knowledge for their internal resources also via trainings and learning, so that all translators can consider MT as an aid rather than a threat.

セッション II

『Dell EMC サポートコンテンツ翻訳における NMT 導入事例』

森山 崇範氏

森山氏は、社外に対して Web 上で公開している Support Knowledgebase の機械翻訳（MT、Machine Translation）の利用について紹介してくれた。含まれる情報の一部には、もともと日本語や中国語で執筆されたものもあるが、ほとんどが英語で執筆され、17 言語への翻訳を展開しているとのこと。

スピードが重視される内容については機械翻訳のみ、品質が重視される場合は人手翻訳（HT、Human Translation）、それ以外の大部分は機械翻訳＋事後編集（PE、Post-Edit）を実施するという使い分けをしているそうだ。地域性と言えるのだろうか、英語が広く普及している国では、翻訳するよりも英文のままでの情報公開が好まれることもあれば、品質に関わらず現地の言語に翻訳することがなによりも望まれる場合もある。特に日本の場合は、訳文の正確性が重視されるとのこと。

Expedia 社と同様、コンテンツの種類による MT の使い分けが言及され、言語バリアントについても紹介があった。たとえば製品の売上に直結するマーケティング資料についてはローカル言語への翻訳が必要、サ

ポート コンテンツについてはユーザによる問題解決が重要であり方言を使い分ける必要性は低いと考えられるが、法律関係書類となると特別に現地語での高品質での翻訳が必要とされる。自らに対しても基準を明確にすることの重要性が指摘された。

また「Hot Issue」と呼ばれるような、一時的に数多くのユーザが参照するようなトピックについてはPEを実施する、予算などの都合上でMT訳を公開して後追いでPEを実施することもあるそうだ。

同社ではNMTを使い始めるにあたって、従来のSMTから切り替えて生産性が落ちないかという評価をパートナー企業と推進しているとのこと。トレーニングが可能なNMTも出てきているとのことで、用語管理などに関する弱みが改善させられるのではと期待しているそうだ。品質評価についてはTAUS DQFを活用しているとのことだが、PEを実施していないRaw MTの翻訳品質を自動的に評価することが目下の課題とのこと。

登壇者プロフィール

Claire Pagès
Expedia Group
Director, Localization Programs

In her role as Director of Localization Programs for the Expedia Group, Claire Pagès manages a team of Program and Project Managers to ensure that the 15+ brands and nearly 200 points of sale they provide services to present localized contents that are complete, accurate and fluent in all languages. This means localizing millions of words each month to maintain the sites & products, with varied contents such as property descriptions, user interface, marketing, SEO, geo, travel guides, help, training, etc. Claire comes with about 15 years' experience in localization, having held many different roles and responsibilities in the industry, both on the supplier and client side. She started her career as a French translator, a role she did for a couple of years before moving onto management roles. When she is not at work, Claire loves travelling and discovering new horizons and cultures with her young family.

Tommaso Rossi
Expedia Group
Senior Manager, Vendor Management Global Supply Operations

Tommaso Rossi has worked in the translation and localization industry for almost fifteen years. After earning a degree in interpreting and translation in 2003, he worked as a freelance interpreter for a year, mainly for the UN and other international organizations, then started a career in translation. Tommaso first worked as a project manager for an Italian translation company then as a vendor manager in Brussels. He currently lives in London and works for

Expedia as the head of vendor management inside the Lodging Partner Services department. He has developed solid experience in various localization domains such as freelancing, project management, machine translation, translation management systems and the translation supply chain management. Outside the localization world, Tommaso loves reading, traveling, watching movies and he is determined to improve his cooking skills.

森山 崇範

Dell EMC

Advisor, eServices & SDS Knowledge Management

2006 年入社。日本のサポート部門にて社内用ナレッジベースのコンテンツ管理およびコール削減に従事。2009 年グローバル・ナレッジ・マネージメント部門に異動。次期 KB ツールの APJ への展開とコンテンツ翻訳管理を担当。SMT の導入開始。2013 年、お客様向けサポート・ウェブ・サイト管理部門に異動。現在サポート・サイト管理の APJ リード業務と並行してコンテンツ翻訳管理をグローバルで担当。2017 年から NMT の導入プロジェクト開始。

AAMT会員のひろば

AAMT 会員の新たな交流の場を AAMT Journal 誌面上で提供するべくスタートいたしました「AAMT 会員のひろば」、会員の皆さまのご助力をいただきまして、第一回の No.41 のスタートから第二十四回を迎えることができました。今号では、個人会員一名の皆さまからのご寄稿をいただいております。

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

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

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

個人会員（敬称略・50 音順）

会員名

二宮 崇（愛媛大学 大学院理工学研究科 電子情報工学専攻／NINOMIYA Takashi)

自己紹介

私は現在、愛媛大学で、自然言語処理、特に、機械翻訳の研究に従事しております。東大辻井研の出身で、言語学的文法理論に基づく構文解析をずっと行っておりましたが、2006 年から東大中川研に移り、機械学習と構文解析の研究を行いました。言語に興味があるというよりは、人工知能、コンピュータ科学、哲学に興味があって、言語を解析するためのコンピュータの実現を夢見て、そのために、言語学の理論に基づく構文解析の研究をただひたすらやっていました。言語学的文法に基づく構文解析は、ルールや論理に基づく合理主義的方法論の代表格となりますが、言語には明らかに構造や規則性があるので、この方法論は正しいやり方だと思う一方、矛盾しているかもしれませんが、人間の思考や言語活動は、記号論的表現では表現しきれない、もっとぼやっとしたところがあると思っていたので、いつかどこかで破綻して、もっと別のやり方を考えないといけなくなるのではないだろうかと思っていました。折しも 2000 年代は、80 年代から 90 年代に続くデータの蓄積の結果、合理主義的方法論から、確率モデルや機械学習に代表される経験主義的方法論にシフトしていた時代でした。合理主義が正しいのか、経験主義が正しいのか、どちらをとるのか問われる時代となっていたのですが、この頃、純粋理性批判の解説を読んで、なるほど、しかし、合理主義と経験主義は必ずしも相矛盾するものではなく、経験を元に合理性に還元する合理性を仮定する、そんな考え方もあるのかと、すごく感心した覚えがあります。言語学の理論に基づく構文解析は、基本的に、合理主義的方法論、つまり、文法、辞書が存在し、それらを演繹することで解析を得る方法論が主流でしたが、この方法論に限界を感じていたので、合理的かつ経験的な方法論として、文法(句構造規則)を与えた上で、データから辞書(語彙)を獲得する語彙化文法の獲得の研究を行いました。この方法論は、今まで実用的に供することができな

かった言語学的文法理論のための構文解析を実用的にしたことで、画期的なものだと思っていましたが、合理主義者からは、これは言語学的文法ではない、と批判されたりして、散々でした。機械学習は、経験主義的手法ではありますが、モデルが与えられた上でパラメータを学習しているという見方で考えると、機械学習も、経験から合理性に還元する合理性を説明していると考えられなくもありません。機械学習の枠組みの中で、もっと言語学的な知見を活用したモデルを考えたら、いつかこの２つの方法論は一体化するのかもしれない、と思うようになりました（今から考えると、この予想はあまり当たっておらず、言語学的知見は、どちらかというとアノテーションという形でコーパスに外在化され、数学的なモデルのパラメータの中に還元されるようになったのではないかと思います)。中川研に移ってからは、経験的手法から新しい手法を考えるようにならないといけないと思い、オンライン学習など機械学習の研究も行うようになりました。言語処理においては、組み合わせ素性(特徴)をいかにつめこむか、ということが焦点となっていましたが、いくらたくさんの組み合わせ素性をいれてもそんなに性能があがるわけでもなく、漠然とですが、多様体学習など、特徴の抽象化が必要なのかな、と思っていました。この頃、AAMT/Japio特許翻訳研究会の委員になり、機械翻訳にも関わりを持つようになりました。

2010 年からは、愛媛大の人工知能研究室に移りました。こちらに来てからは PI として研究室を運営することになり、研究の幅をもたせるために、構文解析だけではなく、特徴抽出、機械翻訳、自動要約の研究も行うことにしましたが、大学の業務(教育と管理運営)が思ったより大変で、ここ 2,3 年になってようやく研究のための時間がとれるようになってきたというところです。2012 年以降は、ご存知のとおり、第３次 AI ブームが到来し、深層学習が大きく注目を浴びるようになりました。多様体学習の研究をしたかったので、深層学習はちょうど良いテーマだと思い、人工知能研究室でも積極的に深層学習を用いた自然言語処理の研究を行うようになりました。統計機械翻訳の頃は、開発も既存ツールのハックも容易ではなく、本質的な機械翻訳の研究を行うことが難しかったのですが、ニューラル機械翻訳は、chainer 等の深層学習ツールが充実していることと、Python で実装しているということがあって、学部生でも１週間ほどで動くところまで作り上げることができて、非常に驚きました。現在は、もっと複雑な仕組みになってきていますが、それでも学生個人個人が実装できています。BLEU 等の翻訳精度においても従来手法より高く、人間が翻訳したかのような質の高い翻訳が実現できつつあります。しかし、一方で、ニューラル機械翻訳を使っていても、まだ翻訳がおかしい、と思うことがよくあります。おそらく状況に応じて翻訳することができていないのでしょう。現在の自然言語処理の多くは、テキストやそのアノテーションなど記号だけから学習を行い、記号だけを手がかりに解析を行っているため、現実世界と記号の世界が結びついていないことが問題点として挙げられています（シンボルグラウンディング問題)。ニューラル機械翻訳でも、現実世界がどうなっているのかコンピュータが理解できておらず、そのため状況にあわせた適切な翻訳ができていないのではないかと思います。京大の森先生から声をかけていただいて、最近、シンボルグラウンディングの研究を一緒にすることになりました。上に書いたように、人間の思考や言語活動は、ぼやっとしていて、よくわからないのですが、深層学習とシンボルグラウンディングの問題を同時に扱うことで、新しい人工知能の世界を切り拓くことができるのではないかと考えています。

MT/翻訳とのかかわり

故磯崎先生の研究になりますが、辻井研で開発された構文解析ツールを使い、英語の主辞を日本語の語順に近くなるように後置すると、英日機械翻訳の性能が格段に向上するということを示された研究があります。今まで構文解析の研究を行ってきましたが、構文解析不要論という言葉があるように、構文解析ツールが実際に役に立つ産業界のタスクはあまり存在せず、構文解析研究者としては辛い立場にいました。それが磯崎先生の研究によって、構文解析が機械翻訳に役立てられることがわかって非常に嬉しく、同時に勇気づけられる思いがしました。この研究をきっかけに愛媛大でも機械翻訳の研究を始めたのですが、統計機械翻訳のツールは非常に複雑で、思ったような機械翻訳の研究はなかなかできず、歯がゆい思いをしていました。しかし、近年、深層学習をコア技術とする第3次AIブームが到来し、2014年にはニューラル機械翻訳が提案され、新参者でも比較的容易に機械翻訳を実装することができるようになりました。時間はかかりましたが、2017年には、構文解析技術であるCKYアルゴリズムの計算を模したニューラルネットワークによるニューラル機械翻訳を提案し、愛媛大からもようやく本格的な機械翻訳の論文をだすことができました。

今年度、NICTの委託研究「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、「深層学習によるマルチモーダル文脈理解と機械翻訳の高度化」のテーマが採択され、愛媛大学も分担者として参加することになりました。前述のシンボルグラウンディングに関わりのある研究テーマで、次世代の機械翻訳を実現するために、画像、シーン、文脈など様々な情報を取り入れた高度機械翻訳を実現することを目的とした研究課題です。参加メンバーは、代表者である東工大（岡崎先生)、分担者である東大（中澤先生、鶴岡先生、中山先生)、愛媛大（二宮、田村先生)、NHK（山田様、後藤様、美野様)、NHKエンジニアリングシステム（田中様、金次様、山崎様)、時事通信社（朝賀様、川上様、大嶋様、齋藤様)となっていて、この6者で協力して研究課題を遂行していきます。

ＭＴおよび翻訳業界に期待すること

多くの日本人はやはり英語等の外国語を非常に苦手としています。この状況を改善すべくMT技術が世の中の役に立てれば非常に良いと思っています。そのために、まず、機械翻訳の会社がたくさん誕生し、大きく成長することを期待しています。次に、大学や公的機関が行うべきことかもしれませんが、学習済みのモデルや翻訳エンジンを公開することで、多くの会社における多言語化の手助けができればいいのではないかと思っています。難しいかもしれませんが、特定の翻訳タスクのためのモデルを再学習するための計算資源も提供できるとこの業界が活性化するのではないかと思います。

ＡＡＭＴへの要望

(もうなっているのかもしれませんが…)産官学を結ぶ場となれば良いと思っています。学術界から、最新技術の提供を行ったり、計算資源や言語資源を提供する場になれば良いと思っています。

第 28 回通常総会および関連行事の報告

AAMT 事務局

当協会の第 28 回通常総会が 2018 年 6 月 18 日（月）13 時より・ホテルアジュール竹芝にて開催された。総会後、招待講演会、さらにモデレーテッド セッションが行われた。そして第 13 回 AAMT 長尾賞授与式と第 5 回 AAMT 長尾賞学生奨励賞授与式と、引き続いての AAMT 長尾賞学生奨励賞および長尾賞受賞者による記念講演会が盛況のうちに行われた。各参加者数は下記に示す通りである。

総会の参加者：個人会員：出席 13 名、委任状参加者 32 名、
　　法人会員：出席 25 法人（法人会員参加者 54 名）、委任状参加 4 法人

報告会・展示会・講演会・AAMT 長尾賞関連の式及び講演会：
　　個人会員：　20 名
　　法人会員：　54 名
　　会員外一般：18 名
　　合計：　　92 名

懇親会：
　　個人会員：　15 名
　　法人会員：　31 名
　　会員外一般：12 名
　　合計：　　58 名

第 28 回通常総会

1. 開会の辞
2. 副会長挨拶　　東京工科大学名誉教授・飯田 仁
3. 出席会員の確認
4. 議案
   第 1 号議案　2017 年度事業報告（案）　第 2 号議案　2017 年度決算報告（案）
   第 3 号議案　2018 年度事業計画（案）　第 4 号議案　2018 年度収支予算（案）
   第 5 号議案　理事・監事人事(案)　　その他・会員提案事項
6. 閉会の辞

報告会　2017 年度活動報告

I　機械翻訳課題調査委員会	委員長	長瀬 友樹(富士通研究所)
II AAMT/Japio 特許翻訳研究会	副委員長	須藤 克仁 (奈良先端科学技術大学院大学)
IIIインターネット WG	リーダー	富士 秀(富士通研究所)
IV編集委員会	委員長	宇津呂 武仁(筑波大学)

招待講演会

・「アジア言語を中心とした機械翻訳評価ワークショップを通じて」

中澤 敏明（東京大学 大学院情報理工学系研究科　特任講師）

・「機械・音声翻訳技術を活用した多言語翻訳サービスのご紹介 」

安西 健（凸版印刷株式会社 情報コミュニケーション事業本部）

展示会

【法人会員による展示】

十印の機械翻訳ソリューションのご紹介

- Gen-pak（玄白） & T-tact -

株式会社十印

みんなの自動翻訳@KI（商用版）/Translation Designer

株式会社川村インターナショナル

翻訳支援ツール「ヤラクゼン」

八楽株式会社

Japio の機械翻訳への取り組み

一般財団法人日本特許情報機構

翻訳バンクとみんなの自動翻訳@TexTra による高精度 NMT

国立研究開発法人情報通信研究機構

ID カード型ハンズフリー音声翻訳端末

富士通株式会社

Mirai Translator™企業向け機械翻訳サービスのご紹介

株式会社みらい翻訳

自社製エンジンの無料翻訳サイト「CROSS-Transer」

株式会社クロスランゲージ

【個人会員による展示】

「MTの効果的な活用

吉川 潔

【AAMT各委員会活動報告会】

1. 機械翻訳課題調査委員会　　委員長 長瀬 友樹（富士通研究所）

２０１８年度総会後の組織・人事について

・ 会長、副会長、理事、顧問、委員長、ワーキンググループリーダーは下記の通りとする。

・ （敬称略：50 音順）

会長	隅田　英一郎	（国立研究開発法人　情報通信研究機構）
副会長	安達　久博	（株式会社サン・フレア）
	黒橋　禎夫	（京都大学）
理事	石川　弘美	（株式会社十印）
	宇津呂　武仁	（筑波大学）
	釜谷　聡史	（東芝デジタルソリューションズ株式会社）
	小谷　克則	（関西外国語大学）
	小林　明	（一般財団法人 日本特許情報機構）
	田中　英輝	（元ＮＨＫ放送技術研究所 現ＮＨＫエンジニアリングシステム）
	長瀬　友樹	（株式会社富士通研究所）
	永田　昌明	（日本電信電話株式会社）
	中村　哲	（奈良先端科学技術大学院大学）
	二宮　崇	（愛媛大学）
	森口　功造	（株式会社川村インターナショナル）
	山畑　征四郎	（株式会社インターグループ）
	山本　和英	（長岡技術科学大学）
	Key-Sun Choi	（韓国 KAIST）
	Virach Sornlertlamvanich	（タイ SIIT）
監事	古谷　祐一	（スピード翻訳株式会社）
	福嶋　健一郎	（株式会社クロスランゲージ）
顧問	長尾　真	（京都大学名誉教授）
	辻井　潤一	（国立研究開発法人　産業技術総合研究所）
	井佐原　均	（豊橋技術科学大学）
	中岩　浩巳	（名古屋大学）
	飯田　仁	（東京工科大学名誉教授）
	小谷　泰造	（株式会社インターグループ）

機械翻訳課題調査委員長	長瀬　友樹	（株式会社富士通研究所）
AAMT/Japio 特許翻訳研究会委員長	辻井　潤一	（産業技術総合研究所）
編集委員長	宇津呂　武仁	（筑波大学）
インターネットワーキンググループリーダー	富士　秀	（株式会社富士通研究所）
事務局長	石川　弘美	（株式会社十印）

（事務局長は第１回理事会にて理事より選任された）

第 13 回 AAMT 長尾賞・第 5 回 AAMT 長尾賞学生奨励賞授与式及び記念講演会

第 13 回 AAMT 長尾賞

選考委員長：飯田 仁（東京工科大学名誉教授）

選考委員：　宇津呂 武仁(筑波大学)

黒橋 禎夫(京都大学)

隅田 英一郎（情報通信研究機構）

2018 年度は以下の２件が受賞となった。

1)受賞者：京都大学・科学技術振興機構

日中・中日機械翻訳実用化プロジェクト

受賞理由：

科学技術論文翻訳のためにコーパス・辞書の開発構築を進め、大規模日英中専門用語辞書を公開するとともに、最先端の機械翻訳技術であるニューラル機械翻訳をいちはやく独自に開発し、国際ワークショップ開催を通してその有効性を広く知らしめた功績は顕著である。

その成果を一般に公開し、日中間の科学技術交流の促進に供していることから、機械翻訳システムを使った新しい社会サービス実現という点で長尾賞の趣旨に合致するものであり、長尾賞の受賞にふさわしい。

推薦者：　安達 久博（株式会社サン・フレア）

同意人：　相良 美織（株式会社バオバブ）

授賞式

左から、中澤 敏明氏（東大）、岩城 修氏（科学技術振興機構）、中岩 浩巳氏（授賞式当時ＡAMT 会長）

2)受賞者：凸版印刷株式会社情報コミュニケーション事業本部
ソーシャルイノベーションセンター情報インフラ本部コンテンツ企画部

受賞理由：

企業連携により多様な自動翻訳の事業化を進め、商業接客音声翻訳サービスを現場店舗に大規模に展開している実績、さらに日本郵便向け音声翻訳システムを全国２万の郵便局に設置している実績、そして自治体音声翻訳の常設を目指した取り組みが高く評価できる。このような社会にインパクトを与える多様な事業展開は機械翻訳システムの実用化促進という点で長尾賞の趣旨に合致するものであり、長尾賞の受賞にふさわしい。

推薦者：　安達 久博（株式会社サン・フレア）

同意人：　相良 美織（株式会社バオバブ）

授賞式

左から安西 健氏、紅林 美紀雄氏

第5回 AAMT 長尾賞学生奨励賞

選考委員長：永田 昌明（日本電信電話株式会社）

選考委員：　釜谷 聡史（東芝デジタルソリューションズ）

鄭 育昌（富士通研究所）

山本和英（長岡技術科学大学）

受賞者：小田悠介 奈良先端科学技術大学院大学（現在 グーグル合同会社）

受賞対象論文：

二値符号予測と誤り訂正を用いたニューラル翻訳モデル

小田 悠介・Philip Arthur・Graham Neubig・吉野 幸一郎・中村 哲

自然言語処理 第25巻2号（2018年3月）pp.167-199（ジャーナル論文）

受賞理由：

本論文は、ニューラル翻訳モデルのデコーダの出力層において、単語出力確率の計算量を従来法の対数程度まで削減する方法を提案している。提案法は、従来のソフトマックスによる単語出力確率の計算の代わりに、各単語に対応付けられたビット列を予測することにより単語出力確率を求める。提案法は、単体で使用した場合には翻訳精度の低下を招くので、本論文ではさらに、提案法と従来法との混合モデルや、二値符号に対する誤り訂正手法を提案法に適用する方法を提案し、実験により、CPUを用いたデコーディングにおいて、精度の低下を防ぎつつ、最大10分の1程度までの計算量を削減できることを示した。

本論文は、非常に独創的かつ実用性が高い研究成果の報告であり、AAMT長尾学生奨励賞にふさわしいと考える。

推薦者：　中村哲教授（奈良先端科学技術大学院大学）

受賞者：　小田悠介氏

協会活動報告

(2018 年 5 月～2018 年 8 月)

機械翻訳課題調査委員会

2018 年月 5 月 25 日 (2018 年度　第 2 回)

① 各 WG の活動について (各 WG に分かれて議論)

(WG1、WG2)

・AAMT/Japio 特許翻訳研究会　拡大評価部会からの依頼事項について

・18 年度の予算案について

・機械翻訳フェア

プログラム (案)

招待講演の講演者

展示会の申し込み状況

委員会報告 (今年はプレゼン形式で報告)

・アンケート

アンケートはプッシュ型の方法で、もっと多くの人にメール出す

申込と同時にアンケート依頼を送る

プライバシーポリシー

アンケートの内容を決める

(WG3)

・この数か月の動向の説明

・今後の活動について

・機械翻訳の評価方法の標準化

・評価基準

・Memsource の仕様調査と UTX との対応検討

2018 年 6 月 13 日 (2018 年度　第 3 回)

① 各 WG の活動について (各 WG に分かれて議論)

(WG1、WG2)

・機械翻訳フェア

展示準備

AAMT 展示

アンケート

(WG3)

・アジェンダ

・ISO 多言語用語集 国内委員会

・品質評価

・宿題

2018年月7月　20日（2018年度　第4回）

① 各WGの活動について（各WGに分かれて議論）

(WG1、WG2)

・機械翻訳フェアの振り返り

プログラム全体

AAMT展示

アンケート

・新体制について

・各WGの計画詳細化

・次回の予定

インターネットWG

① AAMTホームページの更新

編集委員会

No.69（2018年10月発行）の原稿依頼、編集作業を行った。

AAMT/Japio 特許翻訳研究会

2018 年 5 月 11 日（金）(2018 年度　第 1 回)

議事次第：

1. 議事録の確認
2. 発表　愛媛大学　大学院理工学研究科　田村　晃裕先生
3. 29年度報告書の内容紹介
4. 次回の開催について
   - 今後の研究会・拡大評価部会開催の日時予定（場所）
   - 次回研究会予定（平成30年6月開催予定）
   - 主な議題
   - 新年度の活動計画書

2018年6月22日（金）(2018年度　第2回)

議事次第：

1. 議事録の確認
2. 発表　豊田工業大学　三輪　誠　先生
3. 各委員　今年度研究計画のご報告
4. 次回の開催について
   - 今後の研究会・拡大評価部会開催の日時予定（場所）
   - 次回研究会について（平成30年7月27日開催確定）
   - 主な議題

2028年7月27日（金）(2018年度　第3回)

議事次第：

1. 議事録の確認
2. 発表　奈良先端科学技術大学院大学 須藤　克仁先生
   「大規模特許対訳データに関する検討報告」
3. シンポジウム招待講演者について
4. 次回の開催について
   - 今後の研究会・拡大評価部会開催の日時予定（場所）
   - 次回研究会について
   - 主な議題

AAMT/Japio 特許翻訳研究会　拡大評価部会

2018 年 5 月 11 日（金）(2018 年度　第 1 回)

議事概要：

1. 各グループの活動発表
   - 自動評価グループ
   - 人手評価グループ
   - テストセットグループ
2. その他
3. 次回開催日予定　　平成 30 年 9 月もしくは 10 月開催予定

編集後記

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

AAMT ジャーナル 69 号をお送りします。

今号の巻頭言は、AAMT 新会長の情報通信研究機構 隅田英一郎様より御寄稿を頂きました。

AAMT の活動として、今号に先立ちまして、2018 年 6 月に開催されました総会におきまして、東京大学の中澤敏明先生、凸版印刷株式会社の安西健様より貴重な御講演を賜りましたが、今号におきましては、御講演内容についての貴重な御寄稿を頂きました。

あわせて、AAMT 長尾賞の選考結果を受けまして、受賞者の方々から受賞内容についての貴重な御寄稿を頂きました。本年の AAMT 長尾賞の受賞は 2 件で，内 1 件についての受賞内容の御寄稿は、凸版印刷株式会社の安西健様よりの御講演内容の御寄稿と兼ねておりました。もう 1 件の御受賞につきまして、受賞者である科学技術振興機構 岩城修様、中澤敏明様(長尾賞御受賞の御寄稿での肩書は科学技術振興機構)、京都大学の黒橋禎夫先生より、受賞内容についての貴重な御寄稿を頂きました。また、AAMT 長尾賞学生奨励賞の選考結果を受けまして、受賞者である奈良先端科学技術大学院大学(現グーグル合同会社)小田悠介様より、受賞理由となった「ニューラル機械翻訳における二値符号モデル」についての紹介記事を御寄稿頂きました。

研究報告として、Toshiba China の Pengfei Chen 様，Hui Di 様，Masanori Hattori 様より” Toshiba (China)'s CJ-JC Neural Machine Translation”についての紹介記事を御寄稿頂きました。

また、イベント報告として、XTM International の目次由美子様より、本年 6 月と 9 月に開催されました”Women in Localization Japan”の参加報告を御寄稿頂きました。

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

AAMT

Asia-Pacific Association for Machine Translation

AAMT 入会のご案内

AAMT は、機械翻訳の発展を目的として、機械翻訳の研究者、開発者、製造者、利用者が集まった任意の組織です。委員会による定期的な調査研究をはじめ、機関誌の発行、シンポジウムの開催など活動を行っています。
機械翻訳にご関心のあるすべての方にご入会をお勧めします。

＊＊AAMT 会員の特典＊＊

1.AAMT Journal の購読ができます。

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

2.機械翻訳関連の最新情報をメールでお届け

会員専用メーリングリストで、最新の機会翻訳関連の情報をお届けします。
MT 新製品、新サービスの紹介、国際会議、シンポジウムのお知らせ、WEB での MT 関連記事の紹介など盛りだくさんです。

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

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

4.関連機関の主催する国際会議に参加できます。

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

AMTA* ：Association for Machine Translation in the Americas
EAMT** ：European Association for Machine Translation

年会費は以下の通りです。

法人会員：入会金　1 口　10,000 円　　年会費 1 口 50,000 円
個人会員：入会金　　1,000 円　　年会費　　5,000 円（学生は学生会費 1,000 円）

ご関心のある方は、事務局までお問い合わせください。

アジア太平洋機械翻訳協会（AAMT）
ホームページ：http://www.aamt.info
電子メール：aamt-info@aamt.info

AAMT

Asia-Pacific Association for Machine Translation

入会申込書

以下の通り、アジア太平洋機械翻訳協会の会員申し込みを致します。

申込日　２０　　年　　月　　日

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

コメント

AAMT

AAMTジャーナル No.69

発行：アジア太平洋機械翻訳協会（AAMT）
ホームページ：http://www.aamt.info
住所：〒171-0014　東京都豊島区池袋2−55−2鈴木ビル3階
(株)日本システムアプリケーション内
phone：03-5951-3961　fax：03-5951-3966
編集委員会：宇津呂 武仁　小谷 克則　大倉 清司
阿部 さつき　釜谷 聡史　河野 弘毅
表紙(図部分)デザイン：阿部 さつき
事　務　局：石川 弘美　荻野 孝野
印　刷　所：株式会社ユリクリエイト

Asia-Pacific Association for Machine Translation (AAMT)
c/o Japan System Application Co., Ltd.
Suzuki Building 3F 2-55-2, Ikebukuro, Toshima-ku Tokyo 171-0014, JAPAN
aamt-info@aamt.info
phone:+81(0)3-5951-3961 fax:+81(0)3-5951-3966